

# Содержание

<b>I</b>	<b>Теория вероятностей.</b>	<b>6</b>
<b>1</b>	<b>Введение.</b>	<b>6</b>
<b>2</b>	<b>Дискретное пространство элементарных исходов.</b>	<b>8</b>
2.1	Случайные события. Действия над событиями. . . . .	8
2.2	Дискретное вероятностное пространство. . . . .	11
2.3	Классическое определение вероятности. . . . .	12
2.4	Геометрическая вероятность. . . . .	12
<b>3</b>	<b>Произвольное пространство элементарных исходов.</b>	<b>13</b>
3.1	Общее определение вероятности. Аксиомы теории вероятностей. . . .	13
3.2	Условные вероятности. Умножение вероятностей. Независимость со- бытий. . . . .	16
3.3	Формула полной вероятности и формула Байеса . . . . .	20
3.4	Последовательные испытания. Схема и формула Бернулли. . . . .	22
3.5	Предельные теоремы для схемы Бернулли. . . . .	23
<b>4</b>	<b>Случайные величины и их законы распределения.</b>	<b>25</b>
4.1	Понятие случайной величины. . . . .	25
4.2	Числовые характеристики случайной величины. . . . .	31
4.3	Примеры распределений случайных величин. . . . .	35
<b>5</b>	<b>Многомерные случайные величины или случайные векторы.</b>	<b>41</b>
5.1	Случайный вектор и его распределение вероятностей. . . . .	41
5.2	Независимые случайные величины. . . . .	47
5.3	Числовые характеристики случайного вектора. . . . .	49
5.4	Регрессия и линейная регрессия. . . . .	54
<b>6</b>	<b>Функции случайной величины.</b>	<b>57</b>
6.1	Закон распределения монотонной функции случайного аргумента. . .	57
6.2	Числовые характеристики функции случайной величины. . . . .	59
6.3	Закон распределения функции двух случайных величин. . . . .	60
<b>7</b>	<b>Предельные теоремы. Закон больших чисел</b>	<b>62</b>
7.1	Неравенство Чебышева . . . . .	62
7.2	Закон больших чисел. Теорема Чебышева . . . . .	64
7.3	Центральная предельная теорема. . . . .	66
<b>8</b>	<b>Элементарная теория процессов.</b>	<b>66</b>
8.1	Понятие о случайной функции. . . . .	66
8.2	Характеристики случайных функций. . . . .	68
8.3	Числовые характеристики случайных функций. . . . .	70
8.4	Важнейшие классы случайных процессов . . . . .	72

8.5	Линейные операторы. . . . .	72
8.6	Спектральное разложение стационарной случайной функции. . . . .	74
<b>II</b>	<b>Математическая статистика</b>	<b>80</b>
<b>9</b>	<b>Введение</b>	<b>80</b>
<b>10</b>	<b>Статистические модели</b>	<b>80</b>
10.1	Вероятностные и статистические модели . . . . .	80
10.2	Модель независимой однородной выборки с неизвестной функцией распределения . . . . .	82
10.3	Важнейшие параметрические семейства . . . . .	83
10.3.1	Семейство бернуллиевских распределений . . . . .	83
10.3.2	Семейство биномиальных распределений . . . . .	84
10.3.3	Семейство показательных распределений . . . . .	84
10.3.4	Семейство нормальных (гауссовских) распределений . . . . .	84
<b>11</b>	<b>Задачи математической статистики</b>	<b>85</b>
11.1	Задачи оценивания . . . . .	85
11.1.1	Задача оценивания параметра . . . . .	85
11.1.2	Задача оценивания числовых характеристик . . . . .	85
11.1.3	Функция потерь и функция риска . . . . .	86
11.1.4	Задача доверительного оценивания . . . . .	87
11.2	Задача проверки статистических гипотез . . . . .	88
11.2.1	Гипотеза и альтернатива . . . . .	88
11.2.2	Тесты проверки гипотез . . . . .	88
11.2.3	Ошибки I и II рода и их вероятности . . . . .	89
11.2.4	Подход Неймана – Пирсона . . . . .	90
11.2.5	Асимптотические задачи проверки гипотез . . . . .	90
11.3	Оценивание и проверка гипотез . . . . .	91
<b>12</b>	<b>Статистические задачи, связанные с неизвестной функцией рас- пределения</b>	<b>92</b>
12.1	Эмпирическая функция распределения . . . . .	92
12.2	Статистические свойства ЭФР . . . . .	93
12.2.1	Свойства ЭФР при фиксированном значении аргумента . . . . .	93
12.2.2	Свойства ЭФР ”в целом” . . . . .	94
12.3	Доверительные области для функции распределения . . . . .	96
12.4	Критерии согласия Колмогорова . . . . .	96
<b>13</b>	<b>Задачи статистического оценивания</b>	<b>97</b>
13.1	Одномерная характеристика . . . . .	97
13.2	Асимптотически нормальные оценки . . . . .	99
13.3	Выборочные характеристики и их свойства . . . . .	100
13.3.1	Выборочное математическое ожидание . . . . .	100

13.3.2	Выборочные моменты . . . . .	102
13.3.3	Выборочная дисперсия . . . . .	103
13.3.4	Выборочные медиана и квантили . . . . .	104
13.3.5	Выборочная ковариация и корреляция . . . . .	105
13.4	Метод подстановки в задаче построения оценок . . . . .	106
13.4.1	Идея метода подстановки . . . . .	106
13.4.2	Метод моментов . . . . .	107
13.4.3	Метод максимального правдоподобия . . . . .	108
13.5	Задача доверительного оценивания . . . . .	111
13.5.1	Принцип построения доверительных областей . . . . .	111
13.5.2	Асимптотические доверительные интервалы . . . . .	113
13.5.3	Доверительное оценивание на основе ОМП . . . . .	116
<b>14</b>	<b>Оптимальное и асимптотически оптимальное оценивание</b>	<b>116</b>
14.1	О сравнении качества оценок . . . . .	116
14.1.1	Минимаксный подход . . . . .	117
14.1.2	Асимптотически минимаксные оценки . . . . .	118
14.2	Свойства функции правдоподобия (одномерный параметр) . . . . .	118
14.3	Неравенство Рао–Крамера и эффективные оценки . . . . .	119
14.4	Асимптотические свойства ОМП . . . . .	121
14.4.1	Асимптотическая нормальность ОМП . . . . .	121
14.4.2	Асимптотическая минимаксность ОМП . . . . .	122
<b>15</b>	<b>Оценка плотности распределения</b>	<b>123</b>
15.1	Гистограмма как оценка плотности распределения . . . . .	123
15.1.1	Построение гистограммы . . . . .	123
15.1.2	Статистические свойства гистограммы . . . . .	124
15.1.3	Интегральный квадратичный риск и наилучший выбор длин интервалов группировки . . . . .	125
<b>16</b>	<b>Критерий хи-квадрат.</b>	<b>127</b>
16.1	Дискретная случайная величина. . . . .	127
16.2	Проверка параметрической гипотезы. . . . .	129
	<b>Список литературы.</b> . . . . .	<b>130</b>

# Теория вероятностей.

## 1 Введение.

Теория вероятностей - это математическая наука, изучающая закономерности случайных явлений. В данном учебном пособии будет дано краткое изложение основ теории вероятностей. При изучении курса «Теория вероятностей», как и при изучении курса высшей математики, наибольшую трудность вызывает применение теории при решении задач. Изложение сопровождается большим количеством примеров и задач, при решении которых часто используются комбинаторные методы. В связи с этим, во введении приведем некоторые основные определения и формулы комбинаторики.

Комбинаторикой называется область математики, в которой изучаются вопросы о том, сколько различных комбинаций, подчиненных тем или иным условиям, можно составить из заданных объектов.

**Определение 1.1** *Множество называется упорядоченным, если каждому элементу этого множества сопоставлено натуральное число (номер) от 1 до  $n$ , где  $n$  - число элементов множества. Если  $n = \infty$ , то множество называется счетным.*

**Определение 1.2** *Отличающиеся друг от друга порядком, наборы, составленные из элементов данного конечного множества, называются перестановками этого множества.*

**Теорема 1.1 (О числе перестановок)**  $P_n$  - число перестановок из  $n$  элементов определяется по формуле:

$$P_n = n!, \text{ где } n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n.$$

*Доказательство.* Отведем для размещения элементов данного набора  $n$  занумерованных мест. Первое место может занимать элемент с любым из номеров  $1, \dots, n$ . Пусть, например, это элемент с номером 1. На остальных  $(n - 1)$  местах могут стоять элементы с различными наборами номеров из чисел  $2, \dots, n$ , отличающиеся друг от друга только порядком. Число таких наборов равно числу перестановок  $P_{n-1}$ .

Но на первом месте может стоять элемент с любым номером. Следовательно, число различных наборов с фиксированным первым элементом равно  $P_{n-1}$ . Поскольку на первое место можно поместить  $n$  различных элементов с номерами от 1 до  $n$ , то число всех отличающихся друг от друга наборов множества из  $n$  элементов будет  $nP_{n-1}$ , то есть

$$P_n = nP_{n-1}.$$

Это равенство справедливо для любого  $n$ , поэтому можно написать цепочку равенств:

$$P_n = nP_{n-1} = n(n-1)P_{n-2} = n(n-1)P_{n-3} = \dots = n(n-1)(n-2)\dots 2 \cdot P_1.$$

Но  $P_1 = 1$ . Следовательно,  $P_n = n(n-1)\dots 2 \cdot 1 = n!$ .

**Определение 1.3** Упорядоченные наборы, состоящие из  $k$  элементов, взятых из данных  $n$  элементов, называются размещениями из  $n$  элементов по  $k$ .

**Теорема 1.2 (о числе размещений)**  $A_n^k$  – число размещений из  $n$  элементов по  $k$  определяется по формуле:

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

*Доказательство.* Отведем для элементов размещения  $k$  занумерованных мест. Поскольку размещения упорядочены, то у каждого элемента есть номер места, на котором он расположен: 1-й, 2-й, ...,  $k$ -ый. На первое место поставим, например, элемент с номером 1. Тогда из оставшихся  $n-1$  элементов нужно составить всевозможные размещения по  $k-1$  элементу и поставить их на свободные  $k-1$  мест. Следовательно, число различных размещений с единицей на первом месте равно  $A_{n-1}^{k-1}$ . Так как на первом месте можно поместить  $n$  различных элементов, то число всех размещений из  $n$  по  $k$  равно  $nA_{n-1}^{k-1}$ , то есть

$$A_n^k = nA_{n-1}^{k-1}.$$

Это верно для любых  $n$  и  $k$  таких, что  $k \leq n$ . Получаем

$$A_n^k = n(n-1)A_{n-2}^{k-2} = n(n-1)(n-2)A_{n-3}^{k-3} = \dots = n(n-1)\dots A_{n-k+1}^1 = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}, \text{ так как } A_{n-k+1}^1 = n-k+1.$$

**Определение 1.4** Неупорядоченные наборы, состоящие из  $k$  элементов, взятых из данных  $n$  элементов, называются сочетаниями из  $n$  элементов по  $k$ .

**Теорема 1.3 (о числе сочетаний)**  $C_n^k$  – число сочетаний из  $n$  элементов по  $k$  определяется по формуле:

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

*Доказательство.* Формулу для числа сочетаний проще всего вывести, основываясь на формулах для числа размещений и перестановок. Нетрудно заметить, что если сначала составить различные сочетания из  $n$  элементов по  $k$ , а потом в каждом из сочетаний различными способами поменять порядок, то получатся различные размещения из  $n$  элементов по  $k$ . Следовательно,  $A_n^k = C_n^k \cdot k!$ . Отсюда получаем

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}.$$

Коэффициенты  $C_n^k$  называются *биномиальными коэффициентами*, так как они входят в формулу бинома Ньютона

$$(a+b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}.$$

Свойства коэффициентов  $C_n^k$ :

- 1)  $\sum_{k=0}^n C_n^k = 2^n$ .
- 2)  $C_n^k = C_n^{n-k}$ , где  $k = 0, \dots, n$ ;  $C_n^0 = C_n^n = 1$ .
- 3)  $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$ , где  $k = 1, \dots, n-1$ .

Формула Стирлинга (без доказательства).

Для любых натуральных  $n$  справедливо равенство

$$n! = \sqrt{2n\pi} (n/e)^n e^{\alpha(n)},$$

где  $1/(12n+1) < \alpha(n) < 1/(12n)$ .

**Пример 1.1** Сколькими способами можно расставить на полке десять различных книг?

Используем теорему 1.1:  $P_{10} = 10! = 3628800$ .

**Пример 1.2** В турнире принимают участие 8 команд. Сколько различных предсказаний о распределении первых трех призовых мест можно сделать?

Используем теорему 1.2:  $A_8^3 = 8 \cdot 7 \cdot 6 = 336$ .

**Пример 1.3** Партия состоит из 8 изделий. Из партии выбирается для контроля 3 изделия. Сколькими способами это можно сделать?

Используем теорему 1.3:  $C_8^3 = 8!/(3!5!) = 56$ .

## 2 Дискретное пространство элементарных исходов.

### 2.1 Случайные события. Действия над событиями.

В основе теории вероятностей лежит понятие случайного эксперимента.

**Определение 2.1** Случайным экспериментом или опытом называется процесс, результат которого нельзя предсказать заранее. Опыт характеризуется тем, что его можно повторить неограниченное число раз.

**Определение 2.2** Взаимоисключающие друг друга исходы опыта называются элементарными исходами, или элементарными событиями.

**Замечание 2.1** Взаимоисключающие исходы - это исходы, которые не могут наступить одновременно.

**Определение 2.3** Множество всех возможных взаимоисключающих элементарных исходов опыта называется пространством элементарных исходов или пространством элементарных событий. Обозначим пространство элементарных исходов буквой  $\Omega$ , а элементарные исходы буквой  $\omega$ .

**Пример 2.1** Опыт – подбрасывание игральной кости. Пространство  $\Omega = \{1; 2; 3; 4; 5; 6\}$ , то есть возможными взаимоисключающими исходами являются числа 1, 2, 3, 4, 5, 6.

**Пример 2.2** Опыт – бросание монеты. Исходами являются два события: "выпал герб"( $\Gamma$ ), "выпала решка"( $P$ ), тогда  $\Omega = \{\Gamma, P\}$ .

Будем обозначать события большими латинскими буквами  $A, B, C, \dots$

**Определение 2.4** Будем говорить, что в результате опыта событие  $A$  наступило или произошло, если опыт заканчивается одним из исходов, входящих в событие  $A$ .

**Пример 2.3** Опыт – подбрасывание игральной кости, то есть кубика, грани которого занумерованы цифрами 1, 2, 3, 4, 5, 6. Событие  $A$  – выпадение числа очков, кратного трем. Тогда  $A = \{3; 6\}$ .

**Замечание 2.2** Пространство может содержать несчетное число элементарных событий.

**Пример 2.4** Пусть есть проволока длиной 1 м. Мы растягиваем её за концы, в результате чего происходит разрыв в какой-то точке. Множество исходов – это все точки на проволоке, которые математически можно задать отрезком  $[0; 1]$ , то есть  $\Omega = [0, 1]$ , а каждому исходу  $\omega$  соответствует координата точки разрыва.

**Определение 2.5** Событие  $A = \Omega$  состоящее из всех элементарных исходов, называется достоверным событием. Событие  $\Omega$  обязательно происходит, так как эксперимент всегда заканчивается каким-нибудь исходом.

**Определение 2.6** Событие  $A = \emptyset$  (пустое множество) называется невозможным событием. Невозможное событие никогда не происходит, так как нет ни одного элементарного исхода, благоприятствующего этому событию.

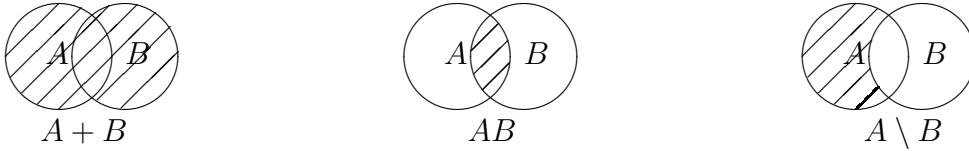
**Определение 2.7** Суммой  $A + B$  событий  $A$  и  $B$ , где  $A, B \subseteq \Omega$ , называется событие, состоящее из всех элементарных исходов, входящих либо в  $A$ , либо в  $B$ .

**Определение 2.8** Произведением  $AB$  событий  $A$  и  $B$ , где  $A, B \subseteq \Omega$  называется событие состоящее только из тех элементарных исходов, которые входят и в  $A$  и в  $B$ .

**Определение 2.9** Разностью  $A \setminus B$  двух событий  $A$  и  $B$ , где  $A, B \subseteq \Omega$  называется событие состоящее из элементарных исходов входящих в  $A$  и не входящих в  $B$ .

**Пример 2.5** Опыт — извлечение одной карты из колоды в 52 карты. Событие  $A$  — извлечена карта красной масти.  $A$  состоит из 26 исходов (имеется 26 различных карт двух красных мастей). Событие  $B$  — извлечение короля, содержит 4 исхода. Событие  $A + B$  (извлечение карты красной масти или короля) состоит из 28 исходов (все карты красной масти + короли черной масти). Событие  $AB$  (извлечение короля красной масти) содержит 2 исхода. Событие  $A \setminus B$  (извлечение карты красной масти, но не короля) содержит 24 исхода (все карты красной масти кроме королей).

**Замечание 2.3** Если события изображать множествами на плоскости, то получаем следующие рисунки, иллюстрирующие действия над событиями



**Определение 2.10** Противоположным (дополнительным) для события  $A$  называется событие  $\bar{A} = \Omega \setminus A$ .

**Определение 2.11** События  $A$  и  $B$  называются несовместными, если их произведение есть невозможное событие  $AB = \emptyset$ .

**Определение 2.12** Событие  $A \subseteq \Omega$  влечет событие  $B \subseteq \Omega$ , если  $A \subseteq B$ , то есть все исходы события  $A$  входят в событие  $B$ .

**Замечание 2.4** Равенство двух событий  $A = B$  означает, что  $A \subset B$  и  $B \subset A$ .

Для событий справедливы стандартные свойства теоретико - множественных операций:

- |  |   |
|--|---|
| 1. $A + A = A$ ;                                 | 2. $A + B = B + A$ ;                      |
| 3. $A \cdot A = A$ ;                             | 4. $AB = BA$ ;                            |
| 5. $A + \Omega = \Omega$ ;                       | 6. $A \cdot \Omega = A$ ;                 |
| 7. $A + \emptyset = A$ ;                         | 8. $A \cdot \emptyset = \emptyset$ ;      |
| 9. $A + \bar{A} = \Omega$ ;                      | 10. $A \cdot \bar{A} = \emptyset$ ;       |
| 11. $(A + B) + C = A + (B + C)$ ;                |   |
| 12. $(AB)C = A(BC)$ ;                            | 13. $A(B + C) = AB + AC$ ;                |
| 14. $\overline{A + B} = \bar{A} \cdot \bar{B}$ ; | 15. $\overline{AB} = \bar{A} + \bar{B}$ . |

**Замечание 2.5** Операции сложения и умножения переносятся на счетное множество событий, при этом свойства действий над событиями сохраняются. Например,  $A \cdot \sum_i B_i = \sum_i AB_i$ .



## 2.2 Дискретное вероятностное пространство.

Пусть задано конечное или счетное пространство элементарных событий  $\Omega = \{\omega_1, \omega_2, \dots\}$ . Будем говорить, что на пространстве  $\Omega$  заданы вероятности элементарных исходов, если каждому исходу  $\omega_l$ ,  $l = 1, 2, \dots$  сопоставлено неотрицательное число  $p(\omega_l)$  и при этом выполнено условие нормировки  $\sum_l p(\omega_l) = 1$ . В этом случае будем говорить, что неотрицательная функция  $p = p(\omega)$  задает распределение вероятностей на  $\Omega$ .

**Определение 2.13** Вероятностью события  $A = \{\omega_{l_1}, \omega_{l_2}, \dots\}$  называется число  $P(A)$ , равное сумме вероятностей элементарных исходов, составляющих событие  $A$ , то есть  $P(A) = \sum_{\omega_{l_k} \in A} p(\omega_{l_k})$ .

**Утверждение 2.1** (Свойства вероятностей.)

1.  $P(\Omega) = 1$ , в силу условия нормировки.
2.  $P(\emptyset) = 0$ , из определения (сумма не имеет слагаемых).
3. Если  $A \subseteq B$ , то  $P(A) \leq P(B)$ , так как сумма равная вероятности события  $B$  содержит дополнительные положительные слагаемые или полностью совпадает с суммой равной вероятности события  $A$ .
4.  $0 \leq P(A) \leq 1$ . Следствие включения  $\emptyset \subseteq A \subseteq \Omega$  и свойств 1 и 2.
5.  $P(\bar{A}) = 1 - P(A)$ .

Доказательство.  $P(A) + P(\bar{A}) = \sum_{\omega_l \in A} p(\omega_l) + \sum_{\omega_l \in \bar{A}} p(\omega_l) = \sum_{\omega_l \in \Omega} p(\omega_l) = 1$ , в силу условия нормировки.

6. Для любых событий  $A$  и  $B$  справедлива формула сложения вероятностей:  $P(A + B) = P(A) + P(B) - P(AB)$ .

Доказательство.  $P(A + B) = \sum_{\omega_l \in A+B} p(\omega_l) = \sum_{\omega_l \in A} p(\omega_l) + \sum_{\omega_l \in B} p(\omega_l) - \sum_{\omega_l \in AB} p(\omega_l) = P(A) + P(B) - P(AB)$ .

**Замечание 2.6** Утверждение 2.1 справедливо не только в дискретном, но и в общем случае.

**Следствие 2.1** Если  $AB = \emptyset$ , то есть события  $A$  и  $B$  несовместны, то  $P(A + B) = P(A) + P(B)$ .

**Следствие 2.2** Если события  $A_l$ , где  $l = 1, 2, \dots, n$  попарно несовместны, то есть  $A_i A_j = \emptyset$  для любых  $i \neq j$ , то  $P(\sum_{l=1}^n A_l) = \sum_{l=1}^n P(A_l)$ .

**Пример 2.6** По мишени стреляют два стрелка. Вероятность события  $A$  - попадания первого стрелка - 0,9; вероятность события  $B$  - попадания второго стрелка - 0,8; а вероятность события  $AB$  - двух попаданий в мишень 0,72. Какова вероятность события  $C$  - хотя бы одного попадания в мишень?

Решение.  $C = A + B$ ,  $P(C) = 0,9 + 0,8 - 0,72 = 0,98$ .

## 2.3 Классическое определение вероятности.

Рассмотрим простейшую модель теории вероятностей, называемую *классической схемой*. В этой модели все исходы предполагаются равновероятными, то есть  $p(\omega_i) = p$ ,  $i = 1, \dots, n$ . По условию нормировки  $np = 1$  следовательно  $p = 1/n$ , тогда  $P(A) = m(1/n) = m/n$ , где  $m = m(A)$  - число исходов благоприятных для события  $A$ ,  $n$  - общее число исходов.

**Пример 2.7** При бросании правильной монеты  $P(\Gamma) = P(P) = 1/2$ .

**Пример 2.8** При бросании игральной кости вероятность того, что выпадает число очков кратное трем  $P(A) = 2/6 = 1/3$ .

**Пример 2.9** Бросаются две монеты. Найти вероятность появления хотя бы одного герба.

*Решение.* Данный опыт имеет 4 равновероятных исхода. Рассмотрим два события:

$A$  - выпадение герба на первой монете, ему благоприятны исходы  $\Gamma P$  и  $\Gamma \Gamma$ ,

$B$  - выпадение герба на второй монете, ему благоприятны исходы  $P \Gamma$  и  $\Gamma \Gamma$ .

Найдём вероятность события  $C = A + B$ ,  $P(C) = P(A) + P(B) - P(AB) = 1/2 + 1/2 - 1/4 = 3/4$ .

Можно решать данный пример и через противоположное событие.

$$P(C) = 1 - P(\overline{C}) = 1 - P(\{PP\}) = 1 - 1/4 = 3/4.$$

## 2.4 Геометрическая вероятность.

Пусть на плоскости задано множество  $\Omega$ . Рассмотрим некоторое подмножество  $A \subseteq \Omega$ . Предположим, что множества  $\Omega$  и  $A$  измеримы, то есть имеют площадь. Пусть результатом эксперимента является случайный выбор точки из множества  $\Omega$ . Будем предполагать, что выбор любой точки множества равновозможен и попадание точки в множество  $A$  зависит только от площади  $A$  и не зависит от формы множества  $A$  и от его расположения в  $\Omega$ . Считаем, что событие  $A$  наступило, если случайно выбранная точка попала в множество  $A \subseteq \Omega$ . Тогда по аналогии с определением 2.13 положим

$$P(A) = S_A/S_\Omega, \quad (2.1)$$

где  $S_A$  - площадь множества  $A$ , а  $S_\Omega$  - площадь множества  $\Omega$ .

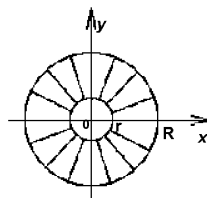
Поясним соответствующую аналогию. В общем случае символом  $mes$  обозначим меру Лебега в пространстве: для прямой - это длина, для плоскости - площадь, для 3-х мерного пространства - объем. Для любого  $n \in \mathcal{N}$ , где  $\mathcal{N}$  - множество натуральных чисел, разобьём  $\Omega$  на  $n$  частей одинаковой меры Лебега. Пусть событие  $A$  целиком состоит из  $k$  таких частей. Тогда, используя классическую схему будем иметь

$$P(A) = \frac{k}{n} = \frac{mes(A)}{mes(\Omega)}.$$

Это утверждение, в силу произвольности  $n$  (с использованием при необходимости предельного перехода  $n \rightarrow \infty$ ), имеет место для любого события  $A$ , имеющего меру Лебега.

Геометрические вероятности обладают всеми свойствами, доказанными в предложении 2.1 и в следствиях к нему.

**Пример 2.10** Имеются две концентрические окружности радиуса  $r$  и  $R$ . Случайная точка оказалась внутри большого круга. Какова вероятность того, что она в заштрихованном кольце?



*Решение.*

Вычислим площади кольца и круга и применим формулу (2.1)

$$S_A = S_{\text{кольца}} = \pi R^2 - \pi r^2 = \pi(R^2 - r^2),$$

$$S_{\Omega} = S_{\text{круга}} = \pi R^2,$$

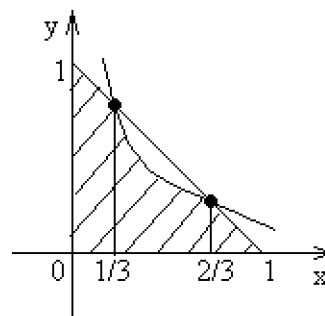
$$P(A) = S_A / S_{\Omega} = \pi(R^2 - r^2) / (\pi R^2) = 1 - r^2 / R^2.$$

**Пример 2.11** Какова вероятность того, что сумма двух наугад взятых положительных чисел, каждое из которых не больше единицы не превосходит 1, а их произведение будет не больше  $2/9$ ?

*Решение.* Пусть  $x, y$  - взятые числа. Имеем  $0 \leq x \leq 1, 0 \leq y \leq 1$ , то есть множество элементарных исходов имеет вид  $\Omega = [0, 1] \times [0, 1]$  и  $S_{\Omega} = S_{\text{квадрата}} = 1$ . Благоприятствующие интересующему нас событию  $A$  значения удовлетворяют условиям:  $x + y \leq 1, xy \leq 2/9$ . Следовательно,

$$S_A = \frac{1}{3} + \frac{2}{9} \int_{1/3}^{2/3} \frac{dx}{x} = 1/3 + (2 \ln 2)/9,$$

$$P(A) = \frac{S_A}{S_{\Omega}} = \left( \frac{1}{3} + \frac{2}{9} \ln 2 \right) / 1 \simeq 0,4874.$$



### 3 Произвольное пространство элементарных исходов.

#### 3.1 Общее определение вероятности. Аксиомы теории вероятностей.

До сих пор мы рассматривали такие случайные эксперименты, в которых количество исходов было конечным или счетным. Пусть  $\Omega$  - произвольное множество. Оно может быть как счетным, так и несчетным. Предположим, что для какого-нибудь случайного эксперимента  $\Omega$  является пространством всех элементарных событий, то есть все точки множества  $\Omega$  являются исходами этого эксперимента, при этом в результате эксперимента появляется только одна точка из  $\Omega$ .

Рассмотрим некоторую систему подмножеств  $\Omega$  и обозначим ее  $S$ .

**Определение 3.1** Совокупность множеств  $S$  называется алгеброй, если выполнены следующие условия:

1.  $\Omega \in S$ .
2. Если  $A \in S$  и  $B \in S$ , то  $A + B \in S$ .
3. Если  $A \in S$ , то  $\overline{A} \in S$ .

Если  $\Omega$  конечно, то в качестве алгебры  $S$  рассматривают систему всех подмножеств  $\Omega$ .

**Пример 3.1** Пусть  $\Omega = \{\omega_1, \omega_2\}$ , тогда

$$S = \{\emptyset, [\omega_1], [\omega_2], [\omega_1; \omega_2]\}.$$

Пусть  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ , тогда

$$S = \{\emptyset, [\omega_1], [\omega_2], [\omega_3], [\omega_1, \omega_2], [\omega_1, \omega_3], [\omega_2, \omega_3], [\omega_1; \omega_2; \omega_3]\}.$$

В общем случае: если  $\Omega$  содержит  $n$  элементов -  $\text{card}(\Omega) = n$ , то  $S$  - алгебра всех подмножеств  $\Omega$  содержит  $2^n$  элементов (ср. со св. 1 на стр. 8).

**Пример 3.2** Пусть  $\Omega = [0, 1)$ , тогда на отрезке  $[0, 1)$  все множества, состоящие из конечного числа непересекающихся интервалов вида  $[a, b)$ ,  $a, b \in [0, 1)$  образуют алгебру.

Используя свойства операций над событиями (подразд. 2.1) и определение алгебры  $S$ , нетрудно доказать следующие утверждения:

1.  $\emptyset \in S$ , так как  $\emptyset = \overline{\Omega} \in S$ .
2. Если  $A \in S, B \in S$ , то  $AB \in S$ , так как  $\overline{AB} = \overline{A} + \overline{B}$ , откуда  $\overline{AB} \in S$  и  $AB \in S$ .
3. Если  $A \in S, B \in S$ , то  $A \setminus B \in S$ , так как  $A \setminus B = A \setminus AB = A\overline{B}$ .
4. Если  $A_l \in S, l = 1, 2, \dots, n$ , то  $\sum_{l=1}^n A_l \in S$  и  $\prod_{l=1}^n A_l \in S$  (без доказательства).

То есть любое конечное число операций над событиями из  $S$  не выводит за пределы алгебры  $S$ . Это свойство называется замкнутостью относительно конечного числа действий над событиями. Если мы хотим чтобы вероятность обладала свойством непрерывности, то необходимо свойство замкнутости распространить на счетное число действий.

**Определение 3.2** Алгебра  $\mathcal{A}$  называется  $\sigma$ -алгеброй, если она замкнута относительно объединения любого счетного набора своих подмножеств. То есть требование 2 в определении алгебры заменяется требованием  $2'$ : если  $A_i \in \mathcal{A}, i = 1, 2, \dots$ , то  $\sum_{i=1}^{\infty} A_i \in \mathcal{A}$ .

**Замечание 3.1** Если пространство элементарных исходов  $\Omega$  конечно -  $\text{card}(\Omega) = n$ , то алгебра всех подмножеств, порожденная исходами  $\omega_i, i = 1, \dots, n$  является  $\sigma$ -алгеброй. Если  $\Omega$  счетно, то обычно рассматривают  $\sigma$ -алгебру всех подмножеств, которая содержит алгебру, порожденную исходами  $\omega_i, i = 1, 2, \dots$ , но не равна ей, так как содержит подмножества  $A$ , такие, что количество элементов в  $A$  и в  $\overline{A}$  - счетно. Например,  $A$  - подмножество всех элементов, занумерованных нечетными номерами.

**Замечание 3.2** Наименьшая  $\sigma$ -алгеброй на прямой, содержащая все интервалы называется борелевской  $\sigma$ -алгеброй и обозначается  $\mathcal{B}$  (здесь "наименьшая" означает, что все остальные  $\sigma$ -алгебры, содержащие интервалы, включают в себя все множества из борелевской  $\sigma$ -алгебры  $\mathcal{B}$ ).

Сопоставим событиям  $\sigma$ -алгебры  $\mathcal{A}$  вероятности, то есть дадим аксиоматическое определение вероятностной меры.

**Определение 3.3** Вероятностной мерой  $P(A)$  называется числовая функция, заданная на  $\sigma$ -алгебре  $\mathcal{A}$  подмножеств  $\Omega : A \subseteq \Omega, A \in \mathcal{A}$  и удовлетворяющая следующим условиям (аксиомам):

- 1)  $P(A) \geq 0$ , для любого  $A \in \mathcal{A}$ .
- 2)  $P(\Omega) = 1$  – условие нормировки.
- 3) Если события  $A_1, A_2, \dots, A_i \in \mathcal{A}, i = 1, 2, \dots$  попарно несовместны, то есть  $A_i A_j = \emptyset$  для любых  $i \neq j$ , то  $P(A)$  обладает свойством счетной аддитивности

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \quad \text{в частном случае } P(A_1 + A_2) = P(A_1) + P(A_2).$$

**Определение 3.4** Пара  $(\Omega, \mathcal{A})$  называется измеримым пространством, множества  $A \in \mathcal{A}$  называются измеримыми множествами или событиями, другие множества не рассматриваются.

**Замечание 3.3** Совокупность всех множеств из  $\Omega$  является алгеброй и  $\sigma$ -алгеброй, но если множество  $\Omega$  не счетно, то нетривиальная вероятностная мера  $P(A)$ , удовлетворяющая аксиомам определения 3.3, определенная на всех множествах из  $\Omega$  может не существовать. В рассмотренном выше примере 3.2 нельзя задать вероятностную меру на всех подмножествах  $\Omega = [0, 1)$  (см., например, [11, стр. 87, 93]). Поэтому обычно рассматривают  $\mathcal{B}[0, 1)$  – борелевскую алгебру на отрезке  $[0, 1)$ , которая содержит все возможные пересечения множеств из борелевской алгебры  $\mathcal{B}$  на прямой и отрезка  $[0, 1)$ .

**Замечание 3.4** Вероятностная мера обладает свойством непрерывности: пусть последовательность событий  $\{A_n \in \mathcal{A}\}$  такова, что  $A_{n+1} \subseteq A_n$  (убывающая последовательность множеств) и  $\bigcap_{n=1}^{\infty} A_n = A$ , тогда  $P(A_n) \rightarrow P(A)$  при  $n \rightarrow \infty$  (первая аксиома непрерывности) или  $A_{n+1} \supseteq A_n$  (возрастающая последовательность множеств) и  $\bigcup_{n=1}^{\infty} A_n = A$ , тогда  $P(A_n) \rightarrow P(A)$  при  $n \rightarrow \infty$  (вторая аксиома непрерывности).

Более того, можно показать, что аксиомы непрерывности эквивалентны, а аксиома счетной аддитивности эквивалентна конечной аддитивности + любая аксиома непрерывности.

**Замечание 3.5** Если  $\Omega \subseteq R_1$ , где  $R_1$  – вещественная прямая, то достаточно определить  $P(A)$  на интервалах, на остальные множества борелевской  $\sigma$ -алгебры она продолжится по непрерывности (подробнее см. [11, ч. II, гл. 1, §1, 2] или [1, гл. 2, §1]).

**Определение 3.5** Пусть  $\Omega$ -пространство элементарных исходов,  $\mathcal{A}$  -  $\sigma$ -алгебра подмножеств  $\Omega$ ,  $P$  - вероятностная мера заданная на  $\sigma$ -алгебре  $\mathcal{A}$ . Тройка  $(\Omega, \mathcal{A}, P)$  называется вероятностным пространством или вероятностной моделью.

Условия в определении  $\sigma$ -алгебры  $\mathcal{A}$  и условия 1, 2, 3 в определении вероятностной меры составляют аксиомы теории вероятностей.

*Свойства вероятностной меры.* В подразд. 2.2 мы вывели свойства вероятностей, для случая конечного или счетного числа исходов эксперимента (такой эксперимент называется дискретным). В общем случае для вероятностной меры также справедливо утверждение 2.1, кроме того она обладает свойством счетной аддитивности и непрерывности.

В качестве примера приведем доказательство формулы сложения вероятностей для любых событий  $A$  и  $B$  в общем случае.

Пусть  $A$  и  $B$  - любые события. Тогда верна следующая формула

$$P(A + B) = P(A) + P(B) - P(AB)$$

*Доказательство.* Справедливы следующие представления:

$$A + B = A + (B - AB), \quad B = AB + (B - AB)$$

В этих представлениях событие  $A$  и  $(B - AB)$ , а также  $AB$  и  $(B - AB)$  несовместны. Действительно

$$A \cdot (B - AB) = A \cdot B - A \cdot AB = AB - AB = \emptyset.$$

$$AB \cdot (B - AB) = AB \cdot B - AB \cdot AB = AB - AB = \emptyset.$$

Отсюда в силу аддитивности вероятностной меры:

$$P(A + B) = P(A) + P(B - AB) \text{ и } P(B) = P(AB) + P(B - AB).$$

Подставив

$$P(B - AB) = P(B) - P(AB) \tag{3.1}$$

в первое из равенств получим формулу сложения вероятностей.

Заметим, что при доказательстве формулы сложения вероятностей мы вывели формулу (3.1), которая представляет самостоятельный интрес.

### 3.2 Условные вероятности. Умножение вероятностей. Независимость событий.

Рассмотрим классическую схему, в рамках которой все исходы равновозможны. Пусть  $\Omega$  - пространство элементарных исходов. Предположим, что некоторое событие  $B$ ,  $P(B) \neq 0$  произошло. Будем рассматривать только те элементарные исходы, которые составляют событие  $B$ . Тогда  $\Omega_1 = B$ , новое пространство элементарных исходов. Выберем исходы из  $A$ , которые входят в  $B$  и обозначим их  $A_1$ . Тогда

$A_1 = AB$ . В рамках классической схемы вероятность события  $A$  при условии, что событие  $B$  произошло есть число  $P(A|B)$ :

$$P(A|B) = \frac{m(A_1)}{\text{card}(\Omega_1)} = \frac{m(A_1)/\text{card}(\Omega)}{\text{card}(\Omega_1)/\text{card}(\Omega)} = \frac{P(AB)}{P(B)},$$

так как  $\text{card}(\Omega_1) = m(B)$ .

Дадим общее определение условной вероятности  $P(A|B)$ ,  $P(B) \neq 0$ .

**Определение 3.6** Рассмотрим событие  $B$ , такое что  $P(B) \neq 0$ . Тогда условной вероятностью события  $A$  при условии, что событие  $B$  произошло, называется число, которое обозначается  $P(A|B)$  и вычисляется по формуле:

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

Поскольку условная вероятность - это обычная вероятность, но лишь на более узком пространстве элементарных исходов, то для нее справедливы все свойства обычной вероятностной меры.

*Свойства условных вероятностей.*

Пусть  $P(B) \neq 0$ , тогда

- 1)  $P(\Omega | B) = 1$ .
- 2)  $P(\emptyset | B) = 0$ .
- 3)  $0 \leq P(A | B) \leq 1$ .
- 4) Если  $A \subseteq C$ , то  $P(A | B) \leq P(C | B)$ .
- 5)  $P(\bar{A} | B) = 1 - P(A | B)$ .
- 6) Формула сложения условных вероятностей.

Для любых событий  $A$  и  $C$

$$P((A + C) | B) = P(A | B) + P(C | B) - P(AC | B).$$

- 7) Формула умножения вероятностей.

Для любых событий  $A$  и  $B$ ,  $P(B) \neq 0$

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

Эта формула является просто другой формой записи определения условной вероятности.

- 8) Общая формула умножения вероятностей.

Для любых событий  $A_1, A_2, \dots, A_n$ ,  $P(A_i) \neq 0$ ,  $i = 1, \dots, n - 1$

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \dots P(A_n|A_1 A_2 \dots A_{n-1}).$$

*Независимость событий.*

**Определение 3.7** Событие  $A$  называется независимым от события  $B$ , если  $P(A|B) = P(A)$ , то есть вероятность наступления события  $A$  не зависит от того, произошло событие  $B$  или нет.

Удобнее использовать определение взаимной независимости событий  $A$  и  $B$ .

**Определение 3.8** События  $A$  и  $B$  называются независимыми, если  $P(AB) = P(A)P(B)$ .

При  $P(B) \neq 0$  эти определения равносильны в силу формулы умножения вероятностей  $P(AB) = P(A|B)P(B) = P(A)P(B)$ , но определение 3.8 не теряет смысла и при  $P(B) = 0$ . Заметим, что при  $P(A) \neq 0$ ,  $P(B) \neq 0$  из условия  $A$  не зависит от  $B$  следует, что  $B$  не зависит от  $A$ .

*Свойства независимых событий.*

1. Пусть  $A$  и  $B$  - независимые события, тогда события  $\bar{A}$  и  $B$  также независимы.

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

При доказательстве использовали формулу (3.1) для нахождения вероятности  $P(B - AB)$ .

**Следствие 3.1** Из независимости событий  $A$  и  $B$  следует в силу симметрии независимость  $\bar{B}$  и  $A$  и независимость  $\bar{A}$  и  $\bar{B}$ .

2. Пусть события  $A$  и  $C$ , а также события  $B$  и  $C$  попарно независимы. Если события  $A$  и  $B$  несовместны, то события  $A + B$  и  $C$  независимы.

Доказательство. События  $AC$  и  $BC$  несовместны, следовательно,  $P((A+B)C) = P(AC + BC) = P(AC) + P(BC) = P(A)P(C) + P(B)P(C) = (P(A) + P(B))P(C) = P(A + B)P(C)$ .

Последнее равенство справедливо в силу аддитивности вероятностной меры.

Можно распространить определение независимости на совокупность событий.

**Определение 3.9** События  $A_1, A_2, \dots, A_n$  называются независимыми в совокупности, если для любого набора индексов  $1 \leq l_1 \leq l_2 \leq \dots \leq l_k \leq n$  выполняется равенство:

$$P(A_{l_1} \cdot A_{l_2} \cdot \dots \cdot A_{l_k}) = P(A_{l_1})P(A_{l_2})\dots P(A_{l_k}).$$

**Замечание 3.6** Из попарной независимости независимость в совокупности не следует (см. [11, стр. 25], [1, гл. 2, §3, пр. Бернштейна] или см. [13], [3]). Независимыми в совокупности являются, например, испытания в схеме Бернулли, изучаемой в подразд. 3.3 - 3.4.

Также вводят понятие независимости испытаний или экспериментов. Для независимых испытаний  $P(A_1A_2) = P(A_1)P(A_2)$ , если  $A_1$  и  $A_2$  любые события происходящие в результате разных испытаний (подробнее см. [1, гл. 2, §3]). В частности, в примере 2.6 можно было вместо того, чтобы задавать вероятность  $P(AB)$ , добавить условие независимости испытаний - стрелки стреляют независимо друг от друга.



**Пример 3.3** Урна содержит 20 шаров, из которых 8 шаров черного цвета, 12 - белого. Найти вероятность того, что при последовательном извлечении трех каких - либо шаров все они окажутся белыми (извлеченные шары в урну не возвращаются. )

Пусть  $A$  - первый шар белый,  $B$  - второй шар белый,  $C$  - третий шар белый. Тогда

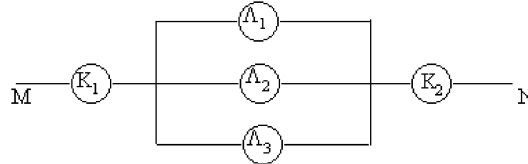
$$P(ABC) = P(A)P(B/A)P(C/AB) = \frac{12}{20} \cdot \frac{11}{19} \cdot \frac{10}{18} = \frac{11}{57}.$$

**Пример 3.4** 33 буквы русского алфавита написаны на карточках разрезной азбуки. Шесть карточек выбираются наугад, одна за другой и укладываются на стол в порядке появления. Найти вероятность того, что получено слово "теория".

$$P(A) = \frac{1}{33} \cdot \frac{1}{32} \cdot \frac{1}{31} \cdot \frac{1}{30} \cdot \frac{1}{29} \cdot \frac{1}{28} = \frac{27!}{33!}.$$

Заметим, что в двух последних примерах можно считать, что предметы извлекаются не последовательно один за другим, а одновременно три шара или шесть карточек. Такая постановка задачи эквивалентна приведенной выше и приводит к классической схеме. Например в примере 3.4 в качестве элементарных исходов можно рассмотреть множество упорядоченных наборов по шесть карточек, случайно выбранных из 33 данных. Тогда количество элементарных исходов  $card(\Omega) = A_{33}^6$  и эти исходы можно считать равновероятными. Событию  $A$  - в результате получено слово "теория", благоприятен один исход. Следовательно,  $P(A) = 1/A_{33}^6 = 27!/33!$ .

**Пример 3.5** ( независимые события ) Электронная цепь между точками  $M$  и  $N$  составлена по схеме:



Выход из строя за время  $T$  различных элементов цепи - независимые события  $P(\overline{K_1}) = 0,6$ ;  $P(\overline{K_2}) = 0,5$ ;  $P(\overline{\Lambda_1}) = 0,4$ ;  $P(\overline{\Lambda_2}) = 0,7$ ;  $P(\overline{\Lambda_3}) = 0,9$ . Здесь черта означает выход из строя соответствующего элемента цепи. Определить вероятность разрыва цепи за время  $T$ .

Решение. Пусть событие  $C$  - разрыв цепи за время  $T$ , событие  $A$  - выход из строя элементов  $K_1$  или  $K_2$ , событие  $B$  - одновременный выход из строя элементов  $\Lambda_1, \Lambda_2, \Lambda_3$ . Тогда  $C = A + B$ ,  $A = \overline{K_1} + \overline{K_2}$ ,  $B = \overline{\Lambda_1}\overline{\Lambda_2}\overline{\Lambda_3}$ . Отсюда, используя независимость событий и формулу сложения вероятностей, имеем

$$P(A) = P(\overline{K_1} + \overline{K_2}) = P(\overline{K_1}) + P(\overline{K_2}) - P(\overline{K_1}\overline{K_2}) = 0,8$$

$$P(B) = P(\overline{\Lambda_1}\overline{\Lambda_2}\overline{\Lambda_3}) = P(\overline{\Lambda_1})P(\overline{\Lambda_2})P(\overline{\Lambda_3}) = 0,252.$$

$$P(A + B) = P(A) + P(B) - P(AB) = 0,8 + 0,252 - 0,8 \cdot 0,252 = 0,8504.$$

### 3.3 Формула полной вероятности и формула Байеса

Пусть требуется определить вероятность некоторого события  $A$ , которое может произойти вместе с одним из событий  $H_1, H_2, \dots, H_n$ .

**Определение 3.10** Набор  $H_1, H_2, \dots, H_n$  называется *полной группой событий*, если они попарно несовместны, то есть  $H_i H_j = \emptyset$  для любых  $i \neq j$ , и их сумма является достоверным событием, то есть  $\sum_{i=1}^n H_i = \Omega$ . В этом случае, события  $H_1, H_2, \dots, H_n$  называются *гипотезами*.

В дальнейшем без потери общности (так как  $P(AH_i) = 0$  при  $P(H_i) = 0$ ) будем предполагать, что  $P(H_i) \neq 0$ ,  $i = 1, \dots, n$ .

**Теорема 3.1 (формула полной вероятности)** Пусть  $H_1, H_2, \dots, H_n$  - полная группа событий. Тогда для любого события  $A$

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

*Доказательство.*  $A = A\Omega = A(H_1 + H_2 + \dots + H_n) = AH_1 + AH_2 + \dots + AH_n$ . Так как  $H_i$  попарно несовместны, то и события  $AH_i$  также попарно несовместны. Имеем  $(AH_i)(AH_j) = A(H_i \cdot H_j) = \emptyset$  для любых  $i \neq j$ . Тогда

$$\begin{aligned} P(A) &= P(AH_1 + AH_2 + \dots + AH_n) = P(AH_1) + P(AH_2) + \dots + \\ &+ P(AH_n) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \dots + P(H_n)P(A|H_n) \end{aligned}$$

последнее утверждение следует из формулы "умножения вероятностей" (свойства 7 условных вероятностей).

Теорема доказана

**Пример 3.6** Имеются две урны: в первой  $a$  белых шаров и  $b$  черных; во второй  $c$  белых и  $d$  черных. Из первой урны во вторую перекладывают, не глядя, один шар. После этого из второй урны берут один шар. Найти вероятность того, что этот шар будет белым.

*Решение.* Пусть событие  $A$  - появления белого шара из второй урны. Рассмотрим следующие гипотезы:

$H_1$  - во вторую урну переложен белый шар,

$H_2$  - во вторую урну переложен черный шар.

В силу того, что шары перекладывались случайно, возникает классическая схема.

$$\text{Тогда имеем } P(H_1) = \frac{a}{a+b}, \quad P(H_2) = \frac{b}{a+b}, \text{ так как в первой урне все-}$$

го находится  $a+b$  шаров, из них  $a$  белых и  $b$  черных. После перекладывания во второй урне  $c+d+1$  шар и либо  $c+1$ , либо  $c$  белых шаров.

$$\text{Следовательно, } P(A|H_1) = \frac{c+1}{c+d+1}; \quad P(A|H_2) = \frac{c}{c+d+1}. \text{ Откуда}$$

по теореме 3.1 имеем 
$$P(A) = \frac{a}{a+b} \cdot \frac{c+1}{c+d+1} + \frac{b}{a+b} \cdot \frac{c}{c+d+1}.$$

Следствием формулы полной вероятности является формула Байеса. А именно, справедлива теорема:

**Теорема 3.2 (формула Байеса)** Пусть заданы  $H_1, H_2, \dots, H_n$  - полная группа событий и событие  $A$ ,  $P(A) \neq 0$ . Тогда для любого  $i = 1, 2, \dots, n$  условная вероятность события  $H_i$  при условии, что событие  $A$  произошло, задается формулой:

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{k=1}^n P(H_k)P(A|H_k)}$$

*Доказательство.* В силу определения условной вероятности имеем  $P(H_i|A) = P(H_i A) / P(A)$ . Используя формулу умножения вероятностей и формулу полной вероятности для нахождения  $P(A)$  получаем формулу Байеса.

**Пример 3.7** В группе из 10 студентов, пришедших на экзамен, 3 подготовленных отлично, 4 - хорошо, 2 - удовлетворительно и 1 - неудовлетворительно. В экзаменационных билетах 20 вопросов. Отлично подготовленный студент может ответить на все вопросы, хорошо подготовленный на 16, удовлетворительно - на 10, неудовлетворительно - на 5. Вызванный наугад студент ответил на три произвольно заданных вопроса. Найти вероятность того, что этот студент подготовлен:

- а) отлично,
- б) неудовлетворительно.

*Решение.* Пусть  $A$  событие, состоящее в том, что студент ответил на 3 вопроса. Рассмотрим четыре гипотезы:

- $H_1$  - студент подготовлен отлично;
- $H_2$  - хорошо;
- $H_3$  - удовлетворительно;
- $H_4$  - неудовлетворительно.

До опыта вероятности гипотез были  $P(H_1) = 0,3$ ;  $P(H_2) = 0,4$ ;  $P(H_3) = 0,2$ ;  $P(H_4) = 0,1$ . Найдем условные вероятности  $P(A|H_i)$ ,  $i = 1, 2, 3, 4$ .

Имеем для отлично подготовленного студента  $P(A/H_1) = 1$ .

Для хорошо подготовленного студента вероятность ответить на первый вопрос  $16/20$ , при ответе на 2-ой вопрос студент знает 15 вопросов из 19, на 3-й - 14 из 18, откуда

$$P(A|H_2) = (16/20)(15/19)(14/18) \simeq 0,4912.$$

Аналогично

$$\begin{aligned} P(A|H_3) &= (10/20)(9/19)(8/18) \simeq 0,1053; \\ P(A|H_4) &= (5/20)(4/19)(3/18) \simeq 0,0088. \end{aligned}$$

По формуле полной вероятности:  $P(A) \simeq 0,5184$ . По формуле Байеса находим условные вероятности гипотез  $H_1$  и  $H_4$ :

$$\begin{aligned} P(H_1|A) &\simeq 0,3 \cdot 1/0,5184 \simeq 0,5787; \\ P(H_4|A) &\simeq 0,1 \cdot 0,0088/0,5184 \simeq 0,0017. \end{aligned}$$

Заметим, что в соответствии с естественным представлением при условии правильного ответа на три вопроса, вероятность гипотезы  $H_1$  - студент подготовлен отлично увеличилась, а гипотезы  $H_4$  - студент подготовлен неудовлетворительно уменьшилась.

### 3.4 Последовательные испытания. Схема и формула Бернулли.

Схема Бернулли заключается в следующем. Производится  $n$  последовательных независимых испытаний с двумя исходами -  $A$  и  $\bar{A}$ . Вероятность появления события  $A$  в каждом испытании одна и та же  $P(A) = p$ . Нас будет интересовать вероятность того, что в серии из  $n$  таких испытаний, событие  $A$  произойдет ровно  $m$  раз. Обозначим буквой  $q$  вероятность дополнительного к  $A$  события  $\bar{A}$ , то есть  $P(\bar{A}) = q$ , где  $q = 1 - p$ .

**Теорема 3.3 (формула Бернулли)** . Справедлива формула:

$$P_n(m) = C_n^m p^m q^{n-m}, \quad (3.2)$$

где  $P_n(m)$  - вероятность того, что при  $n$  испытаниях событие  $A$  произойдет ровно  $m$  раз.

*Доказательство.* Рассмотрим сначала частный случай  $n = 4$ ,  $m = 2$ . Событие  $A$  в четырех испытаниях произошло 2 раза. Тогда возможны следующие случаи:  $AA\bar{A}\bar{A}$ ,  $A\bar{A}A\bar{A}$ ,  $A\bar{A}\bar{A}A$ ,  $\bar{A}A\bar{A}A$ ,  $\bar{A}\bar{A}AA$  и  $\bar{A}AA\bar{A}$ . Может реализоваться любая из этих цепочек. Вероятность любой из этих цепочек одна и та же и равна  $p^2q^2$ . Таких цепочек  $C_4^2 = 6$  то есть  $P_4(2) = C_4^2 p^2 q^2$ .

В общем случае будем рассматривать цепочки вида  $\overbrace{A \dots A}^m \overbrace{\bar{A} \dots \bar{A}}^{n-m}$  вероятность каждой из таких цепочек -  $p^m q^{n-m}$ . Число таких цепочек -  $C_n^m$ , так как число способов выбрать для события  $A$  " $m$ " мест из общего числа " $n$ " мест есть число сочетаний из " $n$ " по " $m$ ", откуда

$$P_n(m) = C_n^m p^m q^{n-m}.$$

*Частные случаи формулы Бернулли.* Будем называть появление события  $A$  успехом, появление события  $\bar{A}$  неудачей.

1. Вероятность одних успехов:  $P_n(n) = C_n^n p^n q^0 = p^n$ .
2. Вероятность одних неудач:  $P_n(0) = q^n$ .
3. Вероятность хотя бы одного успеха:  $P_n(m \geq 1) = 1 - q^n$ .
4. Вероятность хотя бы одной неудачи:  $P_n(m < n) = 1 - p^n$ .

5. Вероятность того, что число успехов заключено в пределах от  $m_1$  до  $m_2$  равна:

$$P_n(m_1 \leq m \leq m_2) = \sum_{m=m_1}^{m_2} C_n^m p^m q^{n-m}. \quad (3.3)$$

**Определение 3.11** Число  $m_0$  наступлений события  $A$  называется *наивероятнейшим*, если оно имеет наибольшую вероятность по сравнению с вероятностями наступлений  $A$  любое другое количество раз:  $P_n(m_0) \geq P_n(m)$ ,  $m = 0, 1, \dots, n$ .

**Теорема 3.4** (без доказательства). *Наивероятнейшее число наступлений события  $A$  в  $n$  испытаниях заключено между числами  $(np - q)$  и  $(np + p)$ , то есть  $np - q \leq m_0 \leq np + p$ . При этом, если число  $(np + p)$  является натуральным числом, то существуют два наивероятнейших числа наступлений события  $A$ :  $m_{01} = np - q$  и  $m_{02} = np + p$ . Если же  $(np + p)$  не является натуральным числом, то  $m_0$  - наивероятнейшее число наступлений события  $A$  единственно.*

Также на практике часто решается следующая задача. Пусть вероятность события  $A$  равна  $p$ . Найти количество опытов, которые необходимы произвести, чтобы с вероятностью  $P$  можно было утверждать, что событие  $A$  произошло хотя бы один раз. Можно показать, что  $n \geq \ln(1 - P)/\ln(1 - p)$ .

**Пример 3.8** *Прибор состоит из 8 однородных элементов, но может работать при наличии в исправном состоянии не менее 6 из них. Каждый из элементов за время работы прибора  $t$  выходит из строя независимо от других с вероятностью 0,2. Найти вероятность того, что прибор откажет за время  $t$ .*

*Решение.* Для отказа прибора требуется выход из строя не менее 2-х из 8 элементов. Используя формулу (3.3) получим

$$\begin{aligned} P_8(2 \leq m \leq 8) &= 1 - (P_8(0) + P_8(1)) = \\ &= 1 - (0,8^8 + C_8^1 \cdot 0,2 \cdot 0,8^7) = 1 - 0,8^7(0,8 + 8 \cdot 0,2) \approx 0,497 \end{aligned}$$

**Пример 3.9** *При установившемся технологическом процессе 80% всей произведенной продукции оказывается продукцией высшего сорта. Найти наивероятнейшее число изделий высшего сорта в партии из 250 изделий.*

*Решение.* Используя теорему 3.4 имеем  $np - q \leq m_0 \leq np + p$ , где  $n = 250$ ,  $p = 0,8$ ,  $q = 0,2$ . Отсюда  $199,8 \leq m_0 \leq 200,8$ . Следовательно,  $m_0 = 200$ .

### 3.5 Предельные теоремы для схемы Бернулли.

В случае, когда число испытаний велико, формулу Бернулли применять не удобно. Рассмотрим несколько приближенных формул для вероятности  $P_n(m)$ , используемых при условии, что  $n$  велико (без доказательства).

**Теорема 3.5 (Пуассона)** *Предположим, что для последовательности схем Бернулли  $n \rightarrow \infty$ ,  $p = p(n) \rightarrow 0$  так, что существует  $\lim_{n \rightarrow \infty} np = \lambda$ . Тогда для любого фиксированного  $m$*

$$\lim_{n \rightarrow \infty} P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda} \quad (3.4)$$

**Замечание 3.7** Возникающее в теореме Пуассона предельное распределение на пространстве элементарных исходов:  $\Omega = 0, 1, 2 \dots$

$$P(\omega_m = m) = \frac{\lambda^m}{m!} e^{-\lambda}, \quad \lambda > 0, \quad m = 0, 1, 2 \dots$$

называется распределением Пуассона.

**Замечание 3.8** Применение теоремы Пуассона для приближенных вычислений дает хорошие результаты при больших  $n$ , малых  $p$ , таких что  $np < a$ ,  $a \approx 9; 10$  и сравнительно с  $n$  небольших  $m$ .

**Пример 3.10** Радиоаппаратура состоит из 2000 элементов. Вероятность отказа одного элемента в течении года равна 0,001. Какова вероятность отказа двух элементов за год?

*Решение.* Используя формулу (3.4) для приближенного вычисления этой вероятности будем иметь  $p = P(A) = 0,001$ ,  $n = 2000$ . Следовательно,  $\lambda = np = 2000 \cdot 0,001 = 2$  и  $P_{2000}(2) \approx (2^2/2!) \cdot e^{-2} = 2/e^2 \approx 0,2707$ .

Непосредственное вычисление по формуле Бернулли дает в этом случае  $P_{2000}(2) \simeq 0,2708$ .

**Теорема 3.6 (локальная теорема Муавра - Лапласа)** Рассмотрим последовательность схем Бернулли при  $n \rightarrow \infty$ ,  $p(n) = p = \text{const} \in (0, 1)$ . Положим  $x_n = (m - np)/\sqrt{npq}$ . Пусть при  $n \rightarrow \infty$  величина  $m = m(n)$  изменяется так, что величины  $x_n \in \Delta$ , где  $\Delta = [a, b]$  - произвольный конечный интервал. Тогда

$$\sqrt{npq}P_n(m) = \frac{e^{-x_n^2/2}}{\sqrt{2\pi}}(1 + \alpha_n),$$

где  $\alpha_n \rightarrow 0$  при  $n \rightarrow \infty$  равномерно по всем  $m$  таким, что  $x_n \in \Delta$ .

В частности, если  $x_n \rightarrow x$  при  $x \rightarrow \infty$ , то  $\sqrt{npq}P_n(m) \rightarrow e^{-x^2/2}/\sqrt{2\pi}$ .

**Замечание 3.9** Функция  $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$  называется функцией Гаусса или плотностью распределения стандартного нормального закона.

**Теорема 3.7 (интегральная теорема Муавра-Лапласа)** Пусть  $m = m(n)$  - число успехов в последовательности схем Бернулли с числом испытаний  $n \rightarrow \infty$  и вероятностью успеха в каждом испытании  $p(n) = p \in (0, 1)$ , тогда для любых вещественных чисел  $x_1 < x_2$

$$P\left(x_1 < \frac{m(n) - np}{\sqrt{np(1-p)}} < x_2\right) \rightarrow_{n \rightarrow \infty} \Phi(x_2) - \Phi(x_1),$$

где  $\Phi(x) = \left(\int_{-\infty}^x e^{-t^2/2} dt\right)/\sqrt{2\pi}$  - функция распределения стандартного нормального закона.

Значения функций  $\varphi(x)$  и  $\Phi(x)$  для  $x \geq 0$  приведены в таблицах, имеющих, например в [2, 3]. Отметим, что справедливы соотношения  $\varphi(-x) = \varphi(x)$ ,  $\Phi(-x) = 1 - \Phi(x)$ . Используя функции  $\varphi(x)$  и  $\Phi(x)$ , сформулированные выше результаты можно преобразовать к виду:

**Теорема 3.8 (локальная теорема Муавра - Лапласа)**

$$P_n(m) \sim \frac{1}{\sqrt{npq}} \varphi(x_n), \quad n \rightarrow \infty, \quad \text{где } x_n = \frac{m - np}{\sqrt{npq}}$$

**Теорема 3.9 (интегральная теорема Муавра - Лапласа)**

$$P_n(m_1 \leq m \leq m_2) \approx \Phi(b_n) - \Phi(a_n), \quad (3.5)$$

где  $a_n = (m_1 - np)/\sqrt{npq}$ ,  $b_n = (m_2 - np)/\sqrt{npq}$ .

**Пример 3.11** В партии из 768 арбузов каждый арбуз оказывается незрелым с вероятностью  $1/4$ . Найти вероятность того, что количество зрелых арбузов будет в пределах от 564 до 600.

*Решение.* Используя формулу (3.5), имеем  $p = 3/4$ ,  $q = 1/4$ ,  $\sqrt{npq} = \sqrt{768 \cdot 3/4 \cdot 1/4} = 12$ ,  $np = 768 \cdot 3/4 = 576$ ,  $a_n = (564 - 576)/12 = -1$ ,  $b_n = (600 - 576)/12 = 2$ .

Отсюда искомая вероятность

$$P(564 \leq m \leq 600) \approx (\Phi(2) - \Phi(-1)) \approx (0,97725 - (1 - 0,84135)) = 0,8186.$$

## 4 Случайные величины и их законы распределения.

### 4.1 Понятие случайной величины.

Вместе с понятием события и его вероятности третьим основным понятием теории вероятностей является понятие случайной величины.

**Определение 4.1** Случайной величиной называется вещественная функция  $X = X(\omega)$ , заданная на вероятностном пространстве  $(\Omega, \mathcal{A}, P)$  и обладающая свойством измеримости: прообраз любого интервала измерим. То есть  $X^{-1}(B) \in \mathcal{A}$ , где  $B \subseteq R_1$  - любой (конечный или бесконечный) интервал. Такая функция называется измеримой относительно  $\sigma$ -алгебры  $\mathcal{A}$ .

**Замечание 4.1** Если  $\Omega$  - пространство элементарных исходов дискретно, то всякая вещественная функция, заданная на  $\Omega$  измерима. Пусть  $(\Omega, \mathcal{A}) = (R_1, \mathcal{B})$ , тогда все непрерывные вещественные функции вещественного аргумента (и многие другие) являются измеримыми относительно борелевской  $\sigma$ -алгебры  $\mathcal{B}$ .

**Пример 4.1** Обозначим через  $X$  - число очков, выпадающее на игральной кости. Здесь  $\Omega$  является дискретным и, следовательно,  $X = X(\omega)$  - случайная величина, возможные значения которой 1, 2, 3, 4, 5, 6 принимаются ею в зависимости от результата испытания - бросания игральной кости.

**Пример 4.2** Пусть  $\omega$  - обозначает наименьший угол между большой стрелкой случайно остановившихся часов и горизонтальным направлением. Тогда рассмотрим в качестве  $\Omega = [0, \pi/2]$  с борелевской  $\sigma$ -алгеброй. Функция  $X = \text{tg}(\omega)$  будет случайной величиной, которая принимает все значения от 0 до  $+\infty$ . В этом случае  $X$  - непрерывная вещественная функция вещественного аргумента и потому измерима.

Случайные величины обозначаются заглавными буквами  $X, Y, Z, \dots$ , а конкретные их значения соответственно малыми буквами  $x, y, z, \dots$ . Наиболее полной вероятностной характеристикой случайной величины служит закон распределения этой величины.

**Определение 4.2** Пусть  $(\Omega, \mathcal{A}, P)$  - вероятностное пространство и  $X = X(\omega)$  заданная на нем случайная величина. Распределением случайной величины  $X$  называется функция множеств  $P_X(B) = P(X^{-1}(B)) = P(\omega \in \Omega : X(\omega) \in B)$ , где  $B \in \mathcal{B}$  - борелевской  $\sigma$ -алгебре на прямой, то есть  $P_X(B)$  - вероятностная мера (аналог длины), заданная на прямой.

**Замечание 4.2** Достаточно функцию  $P_X(B)$  определить для произвольных интервалов  $B$ . Нужно найти прообраз интервала  $B$  - то есть все исходы из  $\Omega$ , для которых значения случайной величины  $X = X(\omega)$  принадлежат интервалу  $B$ . Тогда  $P_X(B)$  - есть вероятность прообраза  $B$ . Иначе,  $P_X(B)$  есть вероятность того, что случайная величина  $X$  приняла значение из интервала  $B$ .

**Пример 4.3** Пусть опыт заключается в двукратном подбрасывании правильной монеты. Исходы данного эксперимента можно записать в виде  $\{ГГ, РГ, ГР, РР\}$ , где  $Г$  - появление герба, а  $Р$  - появление обратной стороны (решки). Слова "правильная монета" указывают на то, что мы имеем классическую схему, то есть все исходы равновозможны и имеют вероятность  $1/4$ . Рассмотрим случайную величину  $X$  - число появившихся гербов.  $X$  принимает 3 значения:  $X(ГГ) = 2$ ,  $X(РГ) = X(ГР) = 1$ ,  $X(РР) = 0$ . В данном случае можно говорить о вероятностях самих этих значений:  $P_X(2) = P(X^{-1}(2)) = P(ГГ) = 1/4$ ,  $P_X(1) = P(X^{-1}(1)) = P(ГР + РГ) = 1/2$ ,  $P_X(0) = P(X^{-1}(0)) = P(РР) = 1/4$ . Кроме того, определена вероятностная мера на всей прямой. Для определения новой "длины" всякого интервала достаточно найти значения  $X$ , лежащие на этом интервале и сложить их вероятности. Так, например,  $P_X(0, 2) = P_X(1) = 1/2$ , так как на открытом интервале  $(0, 2)$  из значений случайной величины  $X$  лежит только 1;  $P_X[0, 2] = P_X(0) + P_X(1) + P_X(2) = 1$ ,  $P_X(-0, 5, 0, 3] = P_X(0) = 1/4$ ,  $P_X[4, 6] = P_X(\emptyset) = 0$ .

Случайные величины могут иметь дискретное, абсолютно непрерывное, сингулярное или смешанное распределение. Мы остановимся на случайных величинах с



дискретным распределением, будем называть их дискретными случайными величинами и на случайных величинах с абсолютно непрерывным распределением, они называются обычно непрерывными случайными величинами.

**Определение 4.3** Случайная величина  $X$  называется дискретной случайной величиной, если множество всех ее значений можно перенумеровать, то есть величина  $X$  принимает не более чем счетное число значений  $x_1, x_2, \dots, x_n, \dots$

Распределение (или закон распределения) дискретной случайной величины  $X$  удобно задавать в виде таблицы. Пусть  $P_X(x_i) = p_i$  - вероятность принять значение  $x_i$ , тогда следующая таблица называется рядом распределения дискретной случайной величины

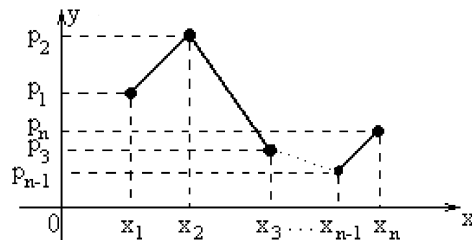
$x_i$	$x_1$	$x_2$	$\dots$	$x_n$	$\dots$
$p_i$	$p_1$	$p_2$	$\dots$	$p_n$	$\dots$

События  $(X = x_1), (X = x_2), \dots$  являются несовместными и образуют полную группу событий, следовательно,  $\sum_i p_i = 1$ . То есть сумма вероятностей всех возможных значений дискретной случайной величины равна единице (закон нормировки).

**Пример 4.4** Пусть  $X$  - число гербов минус число решек выпавших при двух бросаниях правильной монеты. Величина  $X$  может принимать значения  $\{-2, 0, 2\}$ . Приведем ее ряд распределения (см. пример 4.3).

$x_i$	$-2$	$0$	$2$
$p_i$	$1/4$	$2/4$	$1/4$

Чтобы придать ряду распределения более наглядный вид, часто прибегают к его графическому изображению: по оси абсцисс откладывают возможные значения случайной величины, а по оси ординат - вероятности этих значений. Для наглядности полученные точки соединяют отрезками прямых. Такая фигура называется многоугольником распределения:



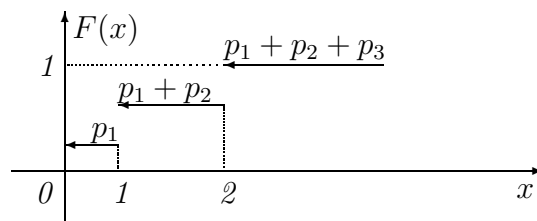
**Определение 4.4** Функция  $F_X(x) = P(X < x) = P_X(-\infty, x)$ , определенная для любого вещественного  $x$  и равная вероятности попадания случайной величины  $X$  на промежуток  $(-\infty, x)$ , называется функцией распределения случайной величины  $X$ .

**Замечание 4.3** Для дискретной случайной величины

$$F_X(x) = P(X < x) = P\left(\sum_{x_i < x} (X = x_i)\right) = \sum_{x_i < x} P(X = x_i) = \sum_{x_i < x} p_i.$$

**Пример 4.5** Найти функцию распределения случайной величины  $X$ , равной числу гербов при двух бросаниях правильной монеты.

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ 1/4, & 0 < x \leq 1, \\ 3/4, & 1 < x \leq 2, \\ 1, & x > 2. \end{cases} \quad \text{График этой функции имеет вид}$$



При помощи функции распределения  $F_X(x)$  находится вероятность того, что значения случайной величины  $X$  попадают на любой промежуток  $[x_1, x_2)$  числовой оси.

**Утверждение 4.1** Вероятность попадания случайной величины на полукрестный промежуток  $[x_1, x_2)$  определяется по формуле

$$P(x_1 \leq X < x_2) = F_X(x_2) - F_X(x_1).$$

*Доказательство.* Событие  $\{X < x_2\}$  представляет собой объединение двух несовместных событий  $\{X < x_1\}$  и  $\{x_1 \leq X < x_2\}$ . Следовательно, в силу аддитивности  $P$  имеем

$$P(X < x_2) = P(X < x_1) + P(x_1 \leq X < x_2),$$

откуда получаем

$$P(x_1 \leq X < x_2) = P_X(X < x_2) - P(X < x_1) = F_X(x_2) - F_X(x_1). \quad (4.1)$$

В дальнейшем мы будем опускать индекс  $X$  в обозначении функций распределения и других характеристик случайной величины, если из контекста ясно о какой случайной величине идет речь.

*Свойства функции распределения.*

F1. Функция  $F(x)$  является неубывающей функцией. То есть если  $x_1 < x_2$ , то  $F(x_1) \leq F(x_2)$ .

*Доказательство.* Пусть  $A = \{X < x_1\}$ ,  $B = \{X < x_2\}$ . При  $x_1 < x_2$  выполняется включение  $A \subset B$ , так как если  $X(\omega) < x_1$ , то  $X(\omega) < x_2$ . По п. 3, утв. 2.1  $P(A) \leq P(B)$ , а это по определению функции распределения и означает, что  $F(x_1) \leq F(x_2)$ .  
F2.  $0 \leq F(x) \leq 1$ ,  $F(-\infty) = 0$ ,  $F(+\infty) = 1$ .

*Доказательство.* Функция  $F(x)$  при любом  $x$  является вероятностью некоторого события и по п. 4, утв. 2.1 принимает значения от 0 до 1. Далее, поскольку  $\{X < -\infty\} = \emptyset$ , а  $\{X < +\infty\} = \Omega$ , то есть эти события являются соответственно невозможным и достоверным, по п.п. 1, 2 утв. 2.1 имеем  $F(-\infty) = P(X < -\infty) = 0$ , а  $F(+\infty) = P(X < +\infty) = 1$ .

Более того из свойства непрерывности вероятностной меры (см. замечание 3.4) и монотонности функции распределения следует, что

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

F3. Функция  $F(x)$  непрерывна слева, то есть  $\lim_{x \rightarrow x_0, x \leq x_0} F(x) = F(x_0)$ .

*Доказательство* этого факта, аналогично следует из свойства непрерывности вероятностной меры (из второй аксиомы непрерывности) и монотонности функции распределения.

**Теорема 4.1** *Всякая вещественная функция  $F(x)$  вещественного аргумента, обладающая свойствами F1, F2, F3 является функцией распределения случайной величины  $X(\omega) = \omega$ , заданной на пространстве  $\{\Omega = R_1, \mathcal{B}, P : P([a, b)) = F(b) - F(a)\}$ .*

*Схема доказательства.* Доказательство заключается в проверке выполнения аксиом теории вероятностей о вероятностной мере  $P = P(B)$ , для  $B \in \mathcal{B}$  -  $\sigma$ -алгебре Бореля. Как уже отмечалось достаточно эти аксиомы проверить для интервалов  $[a, b)$ . Ясно,  $P(\Omega) = F(+\infty) - F(-\infty) = 1$  - условие нормировки выполнено. Условие  $P(B) \geq 0$ ,  $B = [a, b)$  следует из монотонности  $F(x) : F(b) \geq F(a)$  при  $b > a$ . Счетная аддитивность следует из непрерывности слева, но это доказывается достаточно сложно. Функция  $X(\omega) = \omega$  непрерывна и потому является случайной величиной с функцией распределения равной  $F(x)$ .

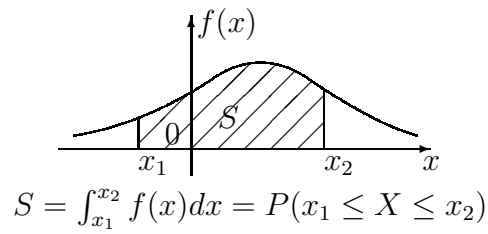
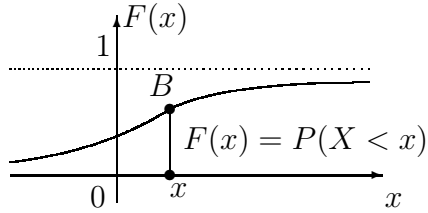
*Из теоремы следует, что достаточно проверить условия F1, F2, F3, чтобы убедиться в корректности вероятностной постановки задачи.*

**Определение 4.5** *Если функция распределения  $F(x)$  для любого  $x \in (-\infty, \infty)$  представима в виде  $F(x) = \int_{-\infty}^x f(y)dy$ , где  $f(y) \geq 0$ , то функция  $f(y)$  называется плотностью распределения случайной величины  $X$ . В этом случае случайная величина  $X$  называется непрерывной случайной величиной.*

**Замечание 4.4** *Функция распределения непрерывной случайной величины, являясь интегралом с переменным верхним пределом, непрерывна.*

- f1.  $\int_{-\infty}^{\infty} f(x)dx = 1$  - свойство нормировки. Это свойство следует из определения плотности и свойства функции распределения  $F(+\infty) = 1$ .
- f2. Обычно полагают  $f(x) = F'(x)$ , в тех точках  $x$ , в которых существует производная.
- f3.  $P(a \leq x \leq b) = \int_a^b f(x)dx$ . Это свойство следует из (4.1), аддитивности интеграла относительно промежутка интегрирования и того, что для непрерывной случайной величины  $P(X = x) = 0$  для любого вещественного  $x$ .

Наглядно геометрический смысл функций  $F(x)$  и  $f(x)$  можно показать на следующих рисунках.



В первом случае величина ординаты точки  $B$  для данной абсциссы  $x$  равна  $F(x)$  - вероятности того, что  $X < x$ . Во втором случае,  $S$  - площадь заштрихованной фигуры равна вероятности того, что  $x_1 \leq X \leq x_2$

**Пример 4.6** Функция распределения непрерывной случайной величины  $X$  имеет вид

$$F(x) = \begin{cases} 0, & x \leq 0, \\ ax^2, & 0 < x \leq 1, \\ 1, & x > 1. \end{cases}$$

Найти коэффициент  $a$ , функцию  $f(x)$  - плотность распределения случайной величины  $X$  и  $P(1/4 \leq X < 1/2)$  - вероятность попадания случайной величины  $X$  на интервал  $[1/4, 1/2)$ .

Решение. В силу непрерывности функции распределения имеем:  $F(1) = ax^2|_{x=1} = 1$ , следовательно,  $a = 1$ .

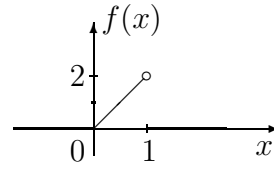
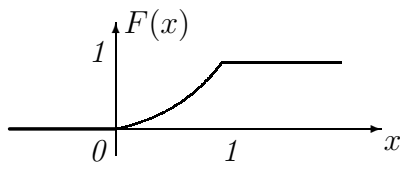
$$f(x) = F'(x) = \begin{cases} 0, & x < 0, \\ 2x, & 0 < x < 1, \\ 0, & x > 1. \end{cases}$$

В точках  $x = 0$ ,  $x = 1$  можно положить, например,  $f(0) = f(1) = 0$ .

$$P(1/4 \leq x < 1/2) = F(1/2) - F(1/4) = (0,5)^2 - (0,25)^2 = 0,1875.$$

Заметим, что в этом случае  $P(1/4 \leq x < 1/2) = P(1/4 \leq x \leq 1/2) = P(1/4 < x < 1/2)$ .

Приведем графики функции распределения и плотности для этого примера.



## 4.2 Числовые характеристики случайной величины.

Важнейшими числовыми характеристиками случайных величин являются моменты различных порядков, к числу которых, в частности, принадлежат математическое ожидание и дисперсия случайной величины.

**Определение 4.6** Математическим ожиданием случайной величины  $X$  называется число  $m_x$ , которое определяется соотношениями

1) для дискретной случайной величины:

$m_x = E(X) = \sum_i x_i p_i$ , где  $x_i$  - все возможные значения случайной величины  $X$ , а

$p_i$  - вероятности с которыми эти значения принимаются:  $p_i = P(X = x_i)$ ;

2) для непрерывной случайной величины:

$m_x = E(X) = \int_{-\infty}^{\infty} x f(x) dx$ , где  $f(x)$  - плотность распределения случайной величины  $X$ .

При этом, если дискретная величина  $X$  имеет счетное число значений, то говорят, что математическое ожидание существует, если ряд  $\sum_{i=1}^{\infty} |x_i| p_i$  сходится, в противном случае говорят, что математического ожидания не существует. Аналогично, для непрерывной случайной величины  $X$ , существование математического ожидания эквивалентно сходимости несобственного интеграла  $\int_{-\infty}^{\infty} |x| f(x) dx$ .

В дальнейшем, говоря о свойствах и связях моментов, будем предполагать, что все соответствующие моменты существуют.

Укажем на механическую аналогию математического ожидания. Пусть в точках  $x_i$  на числовой оси сосредоточены массы  $m_i$ , тогда сумма  $\sum_{i=1}^n m_i x_i$  представляет собой статический момент рассматриваемой системы материальных точек, а сумма  $\sum_{i=1}^n x_i (m_i/m)$  будет равна координате центра тяжести распределения масс этой

системы, если  $m = \sum_{i=1}^n m_i$  - масса всей системы. Но  $\sum_{i=1}^n x_i p_i$  будет аналогична сумме

$\sum_{i=1}^n x_i (m_i/m)$ , так как  $m_i/m \geq 0$ ,  $\sum_{i=1}^n m_i/m = 1$  (то есть можно положить  $p_i = m_i/m$ ).

Следовательно,  $E(X)$  можно рассматривать, как координату центра тяжести распределения масс системы материальных точек с абсциссой  $x_i$  и массой  $m_i$ .

Математическое ожидание называется начальным моментом первого порядка или средним значением случайной величины  $X$ .

### Свойства математического ожидания:

Все свойства математического ожидания являются следствием свойств сумм, рядов или интегралов и справедливы в предположении, что соответствующие математические ожидания существуют. Говоря о суммах или произведениях случайных величин мы будем предполагать, что они заданы на одном и том же вероятностном пространстве.

Е1. Если  $C$ - постоянная величина, то  $E(C) = C$ .

Е2. Постоянный множитель можно выносить за знак математического ожидания:  
 $E(CX) = CE(X)$ .

Е3.  $E(X + Y) = E(X) + E(Y)$ ,  $E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i)$ .

Е4. Если случайная величина  $X \geq 0$ , то  $E(X) \geq 0$ .

Математическое ожидание обладает еще одним свойством, которое будет сформулировано и доказано в подразд. 5.2 ( свойство Е5 математического ожидания ).

Свойство Е3 будет доказано в подразд. 5.1 (см. соотношение (5.2)).

**Определение 4.7** Случайная величина  $\check{X} = X - E(X)$  называется центрированной.

**Определение 4.8** Начальным моментом порядка  $k$  случайной величины  $X$  называется число  $E(X)^k$ . Центральным моментом порядка  $X$  называется число  $E(\check{X})^k$ .

**Упражнение 4.1** Доказать, что

1)  $E(\check{X}) = 0$ ,

2) если  $E(XY) = E(X)E(Y)$  , то  $E(\check{X} \check{Y}) = 0$  (это свойство эквивалентно свойству некоррелированности случайных величин  $X$  и  $Y$ , которое будет введено в подразделе ??).

Указание. Перемножить  $(X - E(X))(Y - E(Y))$  и воспользоваться свойством Е3 математического ожидания.

**Определение 4.9** Дисперсией случайной величины  $X$  называется математическое ожидание случайной величины  $(\check{X})^2$  :

$$D(X) = E(\check{X})^2 = E((X - E(X))^2).$$

В частности, если  $X$  - дискретная случайная величина, то

$$D(X) = \sum_{i=1}^n (x_i - E(X))^2 p_i,$$

если  $X$  - непрерывная случайная величина, то

$$D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx.$$

Квадратный корень из дисперсии называется среднеквадратичным отклонением  $\sqrt{D(X)} = \sigma(X) \geq 0$ .

Дисперсия  $D(X)$  является центральным моментом второго порядка. Величина  $\sigma(X)$  характеризует разброс случайной величины  $X$  вокруг ее математического ожидания.

*Свойства дисперсии:*

D1.  $D(X) \geq 0$ , по свойству E4 математического ожидания и  $D(X) = 0$  тогда и только тогда когда  $P(X = C) = 1$  - где  $C$  постоянная величина, то есть когда случайная величина  $X$  вырождена.

D2. Постоянный множитель выносится из под знака дисперсии в квадрате:  $D(CX) = C^2 D(X)$  для любых постоянных  $C$ .

D3. Дисперсия инвариантна относительно сдвига  $D(X + B) = D(X)$  для любых постоянных  $B$ .

Докажем сразу D2 и D3.

*Доказательство.* Так как  $E(CX + B) = CE(X) + B$ , то

$$\begin{aligned} D(CX + B) &= E((CX + B - E(CX + B))^2) = E((CX - CE(X))^2) = \\ &= E(C^2(X - E(X))^2) = C^2 E((X - E(X))^2) = C^2 D(X). \end{aligned}$$

**Следствие 4.1** Дисперсии случайной величины  $X$  и соответствующей центрированной случайной величины  $\mathring{X}$  равны:

$$D(\mathring{X}) = D(X - E(X)) = D(X).$$

D4. Если  $E(XY) = E(X)E(Y)$ , то  $D(X + Y) = D(X) + D(Y)$ .

*Доказательство.* Используя второе свойство из упражнения 4.1 получим

$$\begin{aligned} D(X + Y) &= D(\mathring{X} + \mathring{Y}) = E((\mathring{X} + \mathring{Y})^2) = \\ &= D(X) + D(Y) + 2E(\mathring{X}\mathring{Y}) = D(X) + D(Y). \end{aligned}$$

D5. Справедливо следующее выражение дисперсии через начальные моменты

$$D(X) = E(X^2) - E^2(X).$$

*Доказательство.* Для доказательства воспользуемся определением дисперсии:

$$\begin{aligned} D(X) &= E((X - E(X))^2) = E(X^2 - 2XE(X) + E^2(X)) = \\ &= E(X^2) - 2E(X) \cdot E(X) + E(E^2(X)) = E(X^2) - (E(X))^2, \end{aligned}$$

так как  $E(X)$  и  $E^2(X)$  - постоянные величины.

**Пример 4.7** Пусть  $X$  - число гербов, выпавших при двух бросаниях правильной монеты. Укажем ряд распределения случайной величины  $X$  (см. пример 4.3).

$x_l$	0	1	2
$p_l$	1/4	2/4	1/4

Следовательно,  $E(X) = 0 \cdot 1/4 + 1 \cdot 2/4 + 2 \cdot 1/4 = 1$ ,  $D(X) = E(X^2) - (E(X))^2 = 0^2 \cdot 1/4 + 1^2 \cdot 2/4 + 2^2 \cdot 1/4 - 1 = 1/2$ .

**Пример 4.8** Пусть случайная величина  $X$  имеет плотность  $f(x) = 1/(\pi(1+x^2))$ ,  $-\infty < x < \infty$ . Распределение с указанной плотностью носит название распределение Коши. Найти  $E(X)$ .

Решение:

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\pi(1+x^2)} dx = \left( \frac{1}{2\pi} \ln(1+x^2) \right) \Big|_{-\infty}^{\infty} = \frac{1}{2\pi} \left[ \lim_{x \rightarrow +\infty} \ln(1+x^2) - \lim_{x \rightarrow -\infty} \ln(1+x^2) \right].$$

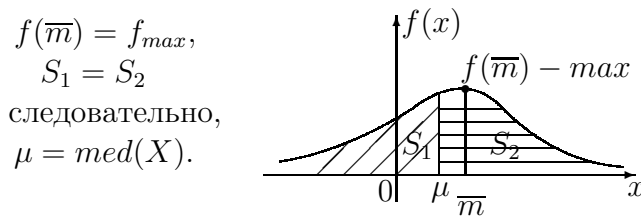
В данном случае несобственный интеграл, задающий математическое ожидание, расходится и случайная величина не имеет математического ожидания.

Мода и медиана случайной величины  $X$ .

**Определение 4.10** Для дискретной случайной величины  $X$  модой называется её наиболее вероятное значение. Обозначается  $\text{mod}(X)$  или  $\bar{m}$ . Для непрерывной случайной величины - модой называются те значения переменной  $x$ , при которых плотность вероятности  $f(x)$  достигает максимума.

**Определение 4.11** Медианой случайной величины  $X$  называется такое число  $\text{med}(X)$ , которое удовлетворяет условиям  $P(X < \text{med}(X)) = P(X > \text{med}(X)) = 1/2$ .

Геометрически:



**Пример 4.9** Дана плотность случайной величины  $X$ :  $f(x) = ae^{2x-x^2}$ ,  $a = 1/(e\sqrt{2\pi})$ . Найти моду этой случайной величины  $X$ .

Решение. Находим  $f'_x(x) = 2a(1-x)e^{2x-x^2}$ ,  $f'(x) = 0$  при  $x = 1$ , причем в точке  $x = 1$   $f'(x)$  меняет знак с “+” на “-”, следовательно, функция  $f(x)$  имеет при  $x = 1$  максимальное значение, то есть  $\bar{m} = 1$ .

**Пример 4.10** Задана плотность распределения

$$f(x) = \begin{cases} 0, & x < 0, \\ x - x^3/4, & 0 \leq x \leq 2, \\ 0, & x > 2. \end{cases}$$



Найдем медиану случайной величины  $X$ .

$$P(X < \mu) = \int_0^{\mu} (x - x^3/4) dx = (x^2/2 - x^4/16)|_0^{\mu} = \mu^2/2 - \mu^4/16 = 1/2.$$

Следовательно,  $med(X) = \pm\sqrt{4 \pm \sqrt{8}}$ , но  $0 < med(X) < 2$  значит,  $med(X) = \sqrt{4 - \sqrt{8}} \approx 1,09$ .

### 4.3 Примеры распределений случайных величин.

Остановимся на примерах распределений случайных величин.

#### 1. Равномерное распределение.

Пусть случайная величина  $X$  принимает любое значение из конечного промежутка  $[a, b]$ , причем существует  $f(x)$  - плотность распределения случайной величины  $X$ :

$$f(x) = \begin{cases} 0, & x < a, \\ 1/(b-a), & a \leq x \leq b, \\ 0, & x > b. \end{cases}$$

Такую случайную величину называют равномерно распределенной на промежутке  $[a, b]$ .

Функцию распределения вероятностей  $F(x)$  найдем с помощью интегрирования:  $F(x) = \int_a^x dx/(b-a) = (x-a)/(b-a)$ ,  $a \leq x \leq b$ .

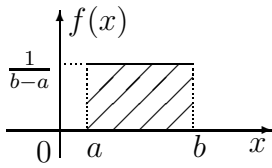
$$\text{Следовательно, } F(x) = \begin{cases} 0, & x < a, \\ (x-a)/(b-a), & a \leq x \leq b, \\ 1, & x > b. \end{cases}$$

Заметим, что функция  $F(x)$  непрерывна и монотонно возрастает от 0 до 1, следовательно условия  $F1$ ,  $F2$ ,  $F3$  (см. теор. 4.1) выполнены, такое распределение вероятностей существует. Аналогично ведут себя и функции распределения остальных, рассмотренных в следующих примерах непрерывных случайных величин. Для дискретных величин функции распределения удовлетворяют  $F1$ ,  $F2$ ,  $F3$  по построению, если  $p_i \geq 0$  и выполнено условие нормировки  $\sum_i p_i = 1$ .

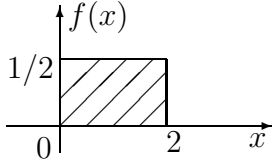
Равномерно распределенная величина имеет моменты всех порядков. В частности,

$$\begin{aligned} E(X) &= \int_a^b \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}, \quad D(X) = \int_a^b \left(x - \frac{b+a}{2}\right)^2 \frac{dx}{b-a} = \\ &= \frac{(b-a)^2}{12}, \quad \sigma(X) = \sqrt{D(X)} = \frac{b-a}{2\sqrt{3}}. \end{aligned}$$

График плотности равномерного распределения приведен на следующем рисунке



**Пример 4.11** Случайная величина  $X$  равномерно распределена на  $[0, 2]$ . Найти  $E(X)$ ,  $D(X)$ .



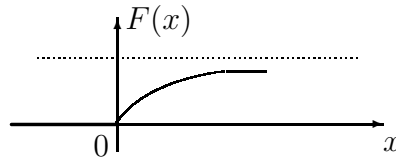
Решение.

Так как  $f(x) = \begin{cases} 1/2, & x \in [0, 2], \\ 0, & x \notin [0, 2]. \end{cases}$   $E(X) = \int_0^2 (x/2) dx = 1$ ,  $D(X) = \int_0^2 ((x - 1)^2/2) dx = (x - 1)^3/6 \Big|_0^2 = 1/3$ .

## 2. Показательный закон распределения.

Пусть  $X$  принимает любые неотрицательные значения  $x \geq 0$ , причем плотность вероятности равна  $f(x) = \lambda e^{-\lambda x}$ , где  $\lambda > 0$  – произвольное положительное число. Говорят, что  $X$  имеет показательное распределение. В этом случае  $F(x) = 0$  при  $x \leq 0$ , при  $x > 0$ :  $F(x) = \int_{-\infty}^x f(x) dx = \int_0^x \lambda e^{-\lambda x} dx = \lambda e^{-\lambda x} / (-\lambda) \Big|_0^x = -e^{-\lambda x} \Big|_0^x = 1 - e^{-\lambda x}$ . Следовательно,

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



Эта функция непрерывна, возрастает от 0 до 1 и  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$ , то есть свойства F1, F2, F3 выполнены.

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = 1/\lambda.$$

$$D(X) = E(X^2) - (E(X))^2 = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx - 1/\lambda^2 = 1/\lambda^2 \quad (\text{интегралы вычисляются по частям}), \quad \sigma(X) = \sqrt{D(X)} = 1/\lambda.$$

**Замечание 4.5** Для показательного распределения математическое ожидание равно среднеквадратичному отклонению.

**Пример 4.12 ( Задача о надежности системы.)** При работе некоторой системы в случайные моменты времени возникают неисправности. Время  $T$  работы системы от момента её включения до возникновения неисправности распределено по показательному закону с параметром  $\lambda$ . Найти вероятность безотказной работы системы в течение времени  $t$ .

*Решение.* В данном случае  $F_T(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$

Если система время " $t$ " работала безотказно, а " $T$ " - время до момента наступления неисправности, то  $T \geq t$ . Тогда искомая вероятность есть  $P(T \geq t) = 1 - P(T < t) = 1 - F_T(t) = e^{-\lambda t}$ .

Заметим, что показательное распределение и только оно обладает свойством отсутствия последействия - при известном настоящем будущее не зависит от прошлого, что позволяет использовать его при построении простейших моделей массового обслуживания.

Точнее, случайная величина  $X$  обладает свойством отсутствия последействия, если выполнены два условия:

$$P(X > x) > 0, \quad \text{для любого } x > 0, \\ P(X > x + y | X > x) = P(X > y), \quad \text{для любых положительных } x, y.$$

3. *Биномиальное распределение.* Пусть  $X$  - случайная величина, равная числу успехов в  $n$  испытаниях Бернулли. И пусть вероятность успеха в каждом испытании равна  $p \in (0, 1)$ , тогда величина  $X$  имеет распределение

$$p_m = P(X = m) = C_n^m p^m q^{n-m} > 0, \quad \text{где } q = 1 - p, \quad m = 0, 1, 2, \dots, n.$$

Как и для схемы Бернулли  $\sum_{m=0}^n p_i = \sum_{m=0}^n C_n^m p^m q^{n-m} = (p + q)^n = 1$ . Следовательно, такое вероятностное распределение существует.

Говорят, что случайная величина  $X$  имеет биномиальное распределение.

$$E(X) = np, \quad D(X) = npq, \quad \sigma(X) = \sqrt{npq}.$$

При  $n = 1$ , то есть, если испытание только одно, будем называть такую случайную величину бернуллиевской.

**Пример 4.13** Вероятность попадания стрелком в мишень равно  $2/3$ . Стрелком сделано 15 выстрелов. Случайная величина  $X$  - число попаданий в мишень. Найти  $E(X)$ ,  $D(X)$ .

*Решение.*  $X$  - случайная величина, распределенная по биномиальному закону;  $p = 2/3$ ;  $q = 1 - p = 1/3$ ;  $n = 15$ . Тогда  $E(X) = np = 15 \cdot 2/3 = 10$ ;  $D(X) = npq = 15 \cdot 2/3 \cdot 1/3 = 10/3$ .

**Пример 4.14** Урна содержит 6 белых шаров и 4 черных. Пять раз достают шар из урны (с возвращением).  $X$  - число вынутых белых шаров. Найти  $E(X)$ ,  $D(X)$ ,  $\sigma(X)$

*Решение.* Имеем  $p = 0,6$   $q = 0,4$   $n = 5$ ,  $X$  - случайная величина, распределенная по биномиальному закону. Тогда  $E(X) = np = 0,6 \cdot 5 = 3$ ,  $D(X) = npq = 0,6 \cdot 0,4 \cdot 5 = 1,2$ ,  $\sigma(X) = \sqrt{1,2}$ .

#### 4. Распределение Пуассона.

Это распределение является предельным для биномиального распределения, при  $n \rightarrow \infty$ , а  $np \rightarrow \lambda$ .

Пусть случайная величина  $X$  имеет распределение

$$P(X = m) = \lambda^m \cdot e^{-\lambda} / m!, \quad m = 0, 1, 2, \dots, \quad \lambda > 0.$$

Проверим неотрицательность и нормировку. Очевидно, что  $p_m = P(X = m) > 0$  при любом  $m = 0, 1, 2, \dots$ . Найдем  $\sum_{m=0}^{\infty} \lambda^m \cdot e^{-\lambda} / m! = e^{-\lambda} \sum_{m=0}^{\infty} \lambda^m / m! = e^{-\lambda} \cdot e^{\lambda} = 1$ ,

так как  $\sum_{m=0}^{\infty} \lambda^m / m! = e^{\lambda}$ .

Такая случайная величина  $X$  называется распределенной по закону Пуассона с параметром  $\lambda$ .

$$E(X) = \sum_{m=1}^{\infty} m \lambda^m e^{-\lambda} / m! = \lambda e^{-\lambda} \sum_{m=1}^{\infty} \lambda^{m-1} / (m-1)! = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

$$E(X(X-1)) = \sum_{m=2}^{\infty} m(m-1) \lambda^m e^{-\lambda} / m! = \lambda^2 \sum_{m=2}^{\infty} \lambda^{m-2} e^{-\lambda} / (m-2)! = \lambda^2.$$

Следовательно,  $D(X) = E(X(X-1)) + E(X) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ .

**Пример 4.15** Автоматическая станция получает в среднем за час 300 вызовов. Какова вероятность того, что за данную минуту она получит точно два вызова? *Решение.* За минуту АТС получает в среднем  $a = 300/60 = 5$  (вызовов), следовательно,  $\lambda = 5$ . Отсюда  $P(X = 2) = e^{-\lambda} \lambda^2 / 2! = e^{-5} 5^2 / 2! = 25/2e^5 \approx 0,0842$ .

#### 5. Нормальное распределение.

Среди случайных величин важное место занимают случайные величины, функция распределения вероятностей которых имеет вид:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(t-a)^2/2\sigma^2} dt. \quad (4.2)$$

Заметим, что  $\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}$  - это так называемый интеграл Пуассона. Отсюда с помощью замены переменных  $u = (x - a)/\sigma$  получим, что

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(t-a)^2/2\sigma^2} dt = 1. \quad (4.3)$$

Следовательно, интеграл (4.2) сходится при любом  $x$  (абсолютно сходится, так как под интегралом стоит положительная функция). Отсюда, функция  $F(x)$  из соотношения (4.2) определена при любом  $x$ , непрерывна как интеграл с переменным верхним пределом, монотонно возрастает в силу аддитивности интеграла и неотрицательности подынтегральной функции. Кроме того,  $\lim_{x \rightarrow -\infty} F(x) = 0$  - в пределе интеграл по пустому множеству и в силу соотношения (4.3)  $\lim_{x \rightarrow \infty} F(x) = 1$ . Соотношения F1, F2, F3 выполнены.

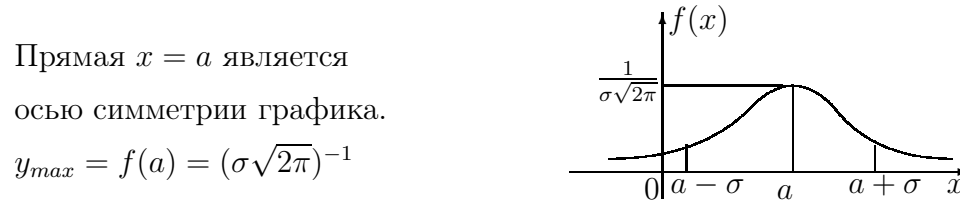
Такие случайные величины называются распределенными по нормальному закону. Часто их называют гауссовыми случайными величинами, а распределение вероятностей - распределением Гаусса.

Пусть случайная величина  $X$ , распределена по нормальному закону (будем это кратко обозначать  $X \sim N(a, \sigma^2)$ ), тогда ее плотность имеет вид

$$f(x) = e^{-(x-a)^2/2\sigma^2} / (\sigma\sqrt{2\pi}), \quad -\infty < x < \infty,$$

Величины  $\sigma > 0$  и  $-\infty < a < \infty$  - параметры распределения.

График функции  $f(x)$  называют кривой Гаусса. Он имеет следующий вид:



Точки  $x = a - \sigma$  и  $x = a + \sigma$  - являются точками перегиба графика.

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-(x-a)^2/2\sigma^2} dx = a,$$

Интеграл вычисляется с помощью замены переменных  $u = (x - a)/\sigma$  с использованием интеграла Пуассона, а также нечетности и интегрируемости функции  $ue^{-u^2/2}$  (из чего следует, что  $\int_{-\infty}^{\infty} ue^{-u^2/2} dt = 0$ .)

$$D(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - a)^2 \cdot e^{-(x-a)^2/2\sigma^2} dx = \sigma^2.$$

Интеграл вычисляется также как предыдущий с использованием дополнительно интегрирования по частям.

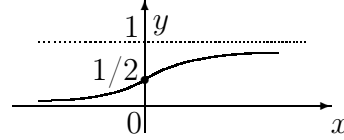
Таким образом, мы выяснили вероятностный смысл параметров  $a$  и  $\sigma$ . Параметр  $a$  является математическим ожиданием случайной величины  $X$ , а  $\sigma^2$  - дисперсией этой случайной величины.

Если  $a = 0$ ,  $\sigma = 1$ , то случайная величина  $X$  называется стандартной нормальной величиной, ее распределение называют стандартным нормальным распределением и обозначают  $N(0, 1)$ ,  $X \sim N(0, 1)$ .

Функция распределения случайной величины  $X \sim N(a, \sigma)$ , определенная в (4.2) выражается через стандартную табулированную функцию

$$F(x) = \int_{-\infty}^x e^{-(t-a)^2/2\sigma^2} dt / (\sqrt{2\pi}\sigma) = \int_{-\infty}^{(x-a)/\sigma} e^{-u^2/2} du / \sqrt{2\pi} = \Phi((x-a)/\sigma),$$
 где  $\Phi(x)$  - функция распределения стандартного нормального закона, уже рассмотренная в подразд. 3.5.

График функции  $y = \Phi(x)$  имеет вид



Для нормально распределенной случайной величины  $P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) = \Phi((x_2 - a)/\sigma) - \Phi((x_1 - a)/\sigma)$ .

**Замечание 4.6** Найдём вероятность того, что отклонение нормальной случайной величины  $X$  от её математического ожидания по модулю меньше некоторого  $\varepsilon > 0$

$$P(|X - a| < \varepsilon) = P(a - \varepsilon < X < a + \varepsilon) = \Phi(\varepsilon/\sigma) - \Phi(-\varepsilon/\sigma) = 2\Phi(\varepsilon/\sigma) - 1.$$

Мы использовали соотношение  $\Phi(-\varepsilon/\sigma) = 1 - \Phi(\varepsilon/\sigma)$ .

**Следствие 4.2** (правило "3σ".) Пусть  $\varepsilon = 3\sigma$ . Для нормально распределенной случайной величины  $X$  со средним  $a$  и дисперсией  $\sigma^2$  справедливо соотношение

$$P(|X - a| < 3\sigma) = 2\Phi(3\sigma/\sigma) - 1 = 2\Phi(3) - 1 \approx 0,9973,$$

то есть в среднем из 10000 значений случайной величины  $X$  только 27 не попадут на интервал  $(a - 3\sigma, a + 3\sigma)$ .

В этом и состоит правило "3σ", которое иногда применяют экспериментаторы при решении некоторых прикладных задач с использованием теории вероятностей и математической статистики, считая событие  $\{|X - a| > 3\sigma\}$  практически невозможным.

**Пример 4.16** Случайная величина  $X$  имеет плотность распределения  $f(x) = (8\pi)^{-1/2} \cdot e^{-(x-3)^2/8}$ . Найдите  $P(1 \leq X < 7)$

*Решение.* Так как  $a = 3$ ,  $\sigma = 2$ , то  $P(1 \leq X < 7) = \Phi((7 - 3)/2) - \Phi((1 - 3)/2) = \Phi(2) - \Phi(-1) = 0,97725 + 0,84135 - 1 = 0,8186$ .

Вычисления значений  $\Phi(2)$  и  $\Phi(-1) = 1 - \Phi(1)$  производятся по таблице.

## 5 Многомерные случайные величины или случайные векторы.

### 5.1 Случайный вектор и его распределение вероятностей.

**Определение 5.1** Рассмотрим упорядоченный набор случайных величин  $X_1, X_2, \dots, X_n$ , заданных на одном и том же вероятностном пространстве  $(\Omega, \mathcal{A}, P)$ . Каждому  $\omega \in \Omega$  поставим в соответствие  $n$ -мерный вектор  $\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$ , называемый  $n$ -мерным случайным вектором или  $n$ -мерной случайной величиной. Таким образом,  $n$ -мерная случайная величина это измеримое отображение  $\mathbf{X}: \Omega \rightarrow R_n$ .

**Пример 5.1** Пусть опыт состоит в двукратном бросании монеты.  $X_1$  - число выпавших гербов,  $X_2$  - число выпавших решек, а  $X_3$  - число гербов выпавших при первом бросании. Здесь  $\Omega = \{\Gamma\Gamma, \Gamma P, P\Gamma, PP\}$ , 3-х мерный вектор  $\mathbf{X} = \mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), X_3(\omega))$  принимает значения

$$\mathbf{X}(\Gamma\Gamma) = (2, 0, 1), \quad \mathbf{X}(\Gamma P) = (1, 1, 1), \quad \mathbf{X}(P\Gamma) = (1, 1, 0), \quad \mathbf{X}(PP) = (0, 2, 0).$$

Также как и в одномерном случае  $n$ -мерный случайный вектор порождает (переносит) на  $R_n$  вероятностную меру:  $P_X(A_n) = P\{\omega \in \Omega : \mathbf{X}(\omega) \in A_n\}$ . Здесь  $A_n$  - параллелепипед в  $R_n$ ,  $P$  - вероятностная мера на пространстве исходов  $\Omega$ , а  $P_X$  - вероятностная мера в  $R_n$ . Для того чтобы определить на  $R_n$  эту новую вероятностную меру (аналог площади) достаточно задать функцию распределения случайного вектора  $\mathbf{X}$ .

**Определение 5.2** Функцией распределения случайного вектора  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  называется функция  $F(x_1, x_2, \dots, x_n)$ , зависящая от  $n$  вещественных аргументов  $x_1, x_2, \dots, x_n$ , равная вероятности тех  $\omega$ , для которых одновременно выполнены неравенства  $X_1 < x_1, \dots, X_n < x_n$ :

$$F(x_1, x_2, \dots, x_n) = P\{\omega \in \Omega : X_1(\omega) < x_1, X_2(\omega) < x_2, \dots, X_n(\omega) < x_n\},$$

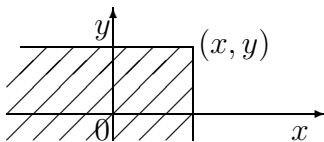
где  $x_i \in (-\infty, \infty)$ , при любых  $i = 1, 2, \dots, n$ .

Далее для простоты будем рассматривать случайные векторы с двумя компонентами  $\mathbf{X} = (X, Y)$ . Функция распределения в этом случае будет иметь вид:

$$F(x, y) = P(X < x, Y < y) = P\{\omega \in \Omega : X(\omega) < x, Y(\omega) < y\}.$$

Геометрический смысл  $F(x, y)$ :

Функция  $F(x, y)$  задает вероятность попадания случайной точки  $(X, Y)$  в бесконечный квадрат с вершиной в точке  $(x, y)$ , лежащий левее и ниже её:



$F_v1$ . Функция распределения  $F(x, y)$  является неубывающей функцией каждого из своих аргументов при условии, что второй аргумент фиксирован (то есть  $F(x, y_0)$  - неубывающая функция аргумента  $x$  при любом  $y = y_0$ , аналогично  $F(x_0, y)$  - неубывающая функция аргумента  $y$  при любом  $x = x_0$ ).

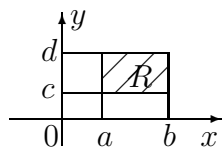
$F_v2$ .  $F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0$ ,  $F(+\infty, +\infty) = 1$

$F_v3$ .  $0 \leq F(x, y) \leq 1$ , так как по определению это вероятность некоторого события.

$F_v4$ .  $F(x, +\infty) = F_X(x)$ ,  $F(+\infty, y) = F_Y(y)$ , где  $F_X(x)$  и  $F_Y(y)$  соответственно функции распределения случайных величин  $X$  и  $Y$ .

$F_v5$ . Функция  $F(x, y)$  непрерывна слева по каждому из аргументов.

$F_v6$ . Пусть  $D = \{(x, y) \in R_2 : a \leq x < b, c \leq y < d\}$  - прямоугольник, тогда  $P((X, Y) \in D) = F(b, d) - F(b, c) - F(a, d) + F(a, c)$ .



Действительно, по теореме сложения вероятностей

$$P(X < a, c \leq Y < d) = F(a, d) - F(a, c),$$

так как мы имеем дело с несовместными событиями.

$$F(b, d) = F(b, c) + F(X < a, c \leq Y < d) + \underbrace{P(a \leq X < b, c \leq Y < d)}_{P((X, Y) \in D)}, \text{ откуда}$$

$$P((X, Y) \in D) = F(b, d) - F(b, c) - F(a, d) + F(a, c).$$

Доказательство пяти первых свойств такое же, как для функции распределения одной случайной величины (см. подразд. 4.1).

Как и для случайных величин будем рассматриваются дискретные и непрерывные случайные векторы.

Рассмотрим сначала две дискретные случайные величины  $X$  и  $Y$ . Пусть случайная величина  $X$  принимает значения  $x_1, x_2, \dots$ , а случайная величина  $Y$  - значения  $y_1, y_2, \dots$ . Одновременное наступление событий  $\{X = x_i\}$  и  $\{Y = y_j\}$  будем обозначать  $\{X = x_i, Y = y_j\}$ , а вероятность их наступления  $p_{ij} = P(X = x_i, Y = y_j)$ .

**Определение 5.3** Соответствие, которое каждой паре значений  $(x_i, y_j)$  дискретных случайных величин  $X$  и  $Y$  сопоставляет её вероятность  $p_{ij}$  называется совместным законом распределения случайных величин  $X$  и  $Y$  или распределением дискретного случайного вектора  $\mathbf{X} = (X, Y)$ .

Совместный закон распределения случайных величин  $X$  и  $Y$  можно задать таблицей:

$X \setminus Y$	$y_1$	$y_2$	$\dots$	$y_n$	$\dots$
$x_1$	$p_{11}$	$p_{12}$	$\dots$	$p_{1n}$	$\dots$
$x_2$	$p_{21}$	$p_{22}$	$\dots$	$p_{2n}$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_m$	$p_{m1}$	$p_{m2}$	$\dots$	$p_{mn}$	$\dots$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$



Свойства вероятностей  $p_{ij}$ :

$p_v 1. \sum_j p_{ij} = p_{X,i}$ , где  $p_{X,i} = P(X = x_i)$ ,  $i = 1, 2, \dots$  - фиксировано.

$p_v 2. \sum_i p_{ij} = p_{Y,j}$ , где  $p_{Y,j} = P(Y = y_j)$ ,  $j = 1, 2, \dots$  - фиксировано.

$p_v 3.$  Свойство нормировки  $\sum_i \sum_j p_{ij} = 1$ .

$p_v 4.$  Для любого  $D$  - борелевского множества на плоскости справедливо соотношение  $P(\mathbf{X} = (X, Y) \in B) = \sum_{i,j: (x_i, y_j) \in B} p_{ij}$ .

Докажем, например, первое свойство. Используя свойства действий над событиями (см. замечание 2.5) будем иметь  $\sum_j \{X = x_i, Y = y_j\} = \sum_j (\{X = x_i\} \cdot \{Y = y_j\}) = \{X = x_i\} \sum_j \{Y = y_j\} = \{X = x_i\} \cdot \Omega = \{X = x_i\}$ , так как  $\sum_j \{Y = y_j\} = \Omega$ .

События  $\{X = x_i, Y = y_j\}$  при разных  $j$  несовместны. Применим аксиому счетной аддитивности

$$p_{X,i} = P(X = x_i) = P(\sum_j \{X = x_i, Y = y_j\}) = \sum_j P\{X = x_i, Y = y_j\} = \sum_j p_{ij}.$$

В дискретном случае функция распределения случайного вектора  $(X, Y)$ , вычисляется по формуле  $F(x, y) = \sum_{i: x_i < x} \sum_{j: y_j < y} p_{ij}$ .

**Пример 5.2** В двух ящиках находится по 6 пронумерованных шаров.

В первом ящике: 1 шар - с номером 1; 2 шара - с номером 2; 3 шара - с номером 3.

Во втором ящике: 2 - с номером 1; 3 - с номером 2; 1 - с номером 3.

Из каждого ящика случайным образом вынули по шару. Пусть  $X$  - номер шара, вынутого из первого ящика и  $Y$  - номер шара вынутого из второго ящика. Составить таблицу распределения случайного вектора  $(X, Y)$ .

Решение. События  $\{X = i\}$  и  $\{Y = j\}$  независимы при любых  $i, j$  в силу независимости экспериментов (разные ящики). Следовательно,  $P\{X = 1, Y = 1\} = P\{X = 1\}P\{Y = 1\} = 1/6 \cdot 2/6 = 1/18$ . Поступая аналогично для вычисления других вероятностей, получим таблицу.

$X \setminus Y$	1	2	3
1	1/18	1/12	1/36
2	1/9	1/6	1/18
3	1/6	1/4	1/12

Рассмотрим теперь случайный вектор  $\mathbf{X} = (X, Y)$ , где  $X$  и  $Y$  - непрерывные случайные величины.

**Определение 5.4** Пусть существует такая функция  $f(x, y) \geq 0$ , что для всех вещественных  $x, y$  справедливо представление

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv, \quad (5.1)$$

тогда случайный вектор  $\mathbf{X} = (X, Y)$  называется непрерывным, а функция  $f(x, y)$  называется плотностью распределения случайного вектора  $\mathbf{X} = (X, Y)$  или совместной плотностью распределения случайных величин  $X$  и  $Y$ .

*Основные свойства плотности распределения случайного вектора:*

$$f_v1. \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = F(+\infty, +\infty) = 1 \text{ - свойство нормировки.}$$

$$f_v2. \text{ Положим } f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}, \text{ в тех точках, где эта производная существует.}$$

$$f_v3. P((x, y) \in D) = \int_D \int f(x, y) dx dy, \text{ для любого } D \in \mathcal{B}(R_2), \text{ где } \mathcal{B}(R_2) \text{ - боре-}$$

левская  $\sigma$ -алгебра на плоскости. Заметим, что это свойство эквивалентно соотношению (5.1).

$f_v4. \int_{-\infty}^{\infty} f(x, y) dy = f_X(x)$ , где  $f_X(x)$  - плотность распределения случайной величины  $X$ . Это свойство справедливо в силу того, что  $F_X(x) = F(x, +\infty) = \int_{-\infty}^x (\int_{-\infty}^{\infty} f(u, v) dv) du$  (см. определение 4.5) и означает, что компоненты непрерывного случайного вектора являются непрерывными случайными величинами.

$f_v5.$  Аналогично,  $\int_{-\infty}^{\infty} f(x, y) dx = f_Y(y)$ , где  $f_Y(y)$  плотность распределения случайной величины  $Y$ .

*Доказательство свойства E3 математического ожидания.*

Проведем доказательство для непрерывного случайного вектора  $\mathbf{X} = (X, Y)$ , в дискретном случае доказательство аналогично. В силу абсолютной сходимости интегралов (по определению математического ожидания) имеем

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y) dx dy = \int_{-\infty}^{\infty} x dx \int_{-\infty}^{\infty} f(x, y) dy + \\ &+ \int_{-\infty}^{\infty} y dy \int_{-\infty}^{\infty} f(x, y) dx = \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy = E(X) + E(Y). \end{aligned} \quad (5.2)$$

**Замечание 5.1** Заметим, что в только что приведенном доказательстве нельзя ограничиться требованием непрерывности случайных величин  $X$  и  $Y$ , так как из этого требования не следует непрерывность случайного вектора  $\mathbf{X} = (X, Y)$ . Например, при  $Y = X$  значения случайного вектора  $\mathbf{X} = (X, X)$  сосредоточены на прямой  $y = x$  и, если  $X$  - непрерывная случайная величина, то его распределение будет сингулярным. То есть вероятность  $P(X = x, Y = y) = 0$ , для любой точки плоскости  $(x, y)$ , множество значений случайного вектора  $V = \{(x, y) = (X(\omega), Y(\omega)), \omega \in \Omega\}$  несчетно, а мера множества  $V$  (площадь) равна нулю, следовательно,  $f(x, y)$  - плотность распределения, не существует.

С использованием интеграла Лебега-Стилтьеса по мере  $P = P(d\omega)$  приведенное доказательство обобщается на общий случай.

*Условные законы распределения вероятностей компонент случайного вектора.*

**Определение 5.5** Условным законом распределения случайной величины  $X$ , входящей в систему  $(X, Y)$  называется её закон распределения, вычисленный при условии, что другая случайная величина  $Y$  приняла определенное значение  $y_{j_0}$  или  $y_0$ . Обозначается  $\{p_{i|j_0}\}$ ,  $i = 1, 2, \dots$ ;  $F(x|y_0)$ ;  $f(x|y_0)$ .

По определению условной вероятности имеем при  $i = 1, 2, \dots$

$$p_{i|j_0} = P(X = x_i | Y = y_{j_0}) = \frac{P(X = x_i, Y = y_{j_0})}{P(Y = y_{j_0})} = \frac{p_{ij_0}}{p_{Y,j_0}} \text{ при } P(Y = y_{j_0}) \neq 0, \quad (5.3)$$

Так как при  $P(Y = y_{j_0}) = 0$  вероятность  $P(\{X = x_i, Y = y_{j_0}\} \subseteq \{Y = y_{j_0}\}) = 0$  положим  $p_{i|j_0} = 0$  при  $P(Y = y_{j_0}) = 0$ ,

Аналогично, можно показать, что плотности условного распределения имеют вид

$$f(x|y_0) = \frac{f(x, y_0)}{f_Y(y_0)} \text{ при } f_Y(y_0) \neq 0 \text{ или } f(y|x_0) = \frac{f(x_0, y)}{f_X(x_0)} \text{ при } f_X(x_0) \neq 0, \quad (5.4)$$

и положим  $f(x|y_0) = 0$ , для  $y_0 : f_Y(y_0) = 0$  или  $f(y|x_0) = 0$ , для  $x_0 : f_X(x_0) = 0$ .

**Пример 5.3** Пусть задано распределение дискретного случайного вектора  $(X, Y)$ :

X	Y		
	-1	0	1
-1	0.5	0.1	0
1	0	0.2	0.2

Найти распределения вероятностей случайных величин  $X$  и  $Y$ , условное распределение  $X$  при условии  $Y = 1$  и условное распределение  $Y$  при условии  $X = -1$ .

*Решение.* Используем для нахождения распределений компонент случайного вектора  $(X, Y)$  свойства  $p_{v1}$  и  $p_{v2}$  вероятностей совместного распределения  $p_{ij}$ , тогда

$$\begin{aligned} p_{X1} = P\{X = -1\} &= \sum_j p_{1j} = P\{X = -1, Y = -1\} + P\{X = -1, Y = 0\} + \\ &+ P\{X = -1, Y = 1\} = 0,5 + 0,1 + 0 = 0,6; \end{aligned}$$

аналогично  $p_{X2} = P\{X = 1\} = 0,4$ . Такие же вычисления проводятся и для  $Y$ . Таким образом,

или

$x_i$	-1	1
$p_{X,i}$	0.6	0.4

или

$y_i$	-1	0	1
$p_{Y,i}$	0.5	0.3	0.2

Для нахождения условного распределения, используя формулу (5.3) имеем:

$$P\{X = -1 | Y = 1\} = \frac{P\{X = -1, Y = 1\}}{P\{Y = 1\}} = 0, \quad P\{X = 1 | Y = 1\} = \frac{P\{X = 1, Y = 1\}}{P\{Y = 1\}} = 1 \text{ или}$$

$x_i$	-1	1
$p_{i 3} = P(X = x_i   Y = 1)$	0	1

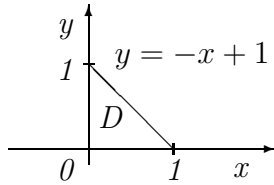
Заметим, что в сечении  $Y = 1$  возникло вырожденное распределение, сосредоточенное в точке  $X = 1$ .

Аналогично, находится условное распределение случайной величины  $Y$  при  $X = -1$

$y_j$	-1	0	1
$p_{j 1} = P(Y = y_j   X = -1)$	5/6	1/6	0

(5.5)

**Пример 5.4** Система случайных величин  $(X, Y)$  равномерно распределена в треугольнике  $D$  с вершинами в точках  $(0; 0)$ ,  $(0; 1)$ ,  $(1; 0)$ . Найти:



1) плотность распределения  $f(x, y)$  случайного вектора  $\mathbf{X} = (X, Y)$ ;

2) плотности компонент  $f_X(x)$   $f_Y(y)$ ;

3) плотности условных распределений  $f(y|x_0)$ .

*Решение.* 1) Плотность равномерного распределения постоянна в треугольнике  $D$  и равна нулю вне треугольника:  $f(x, y) = \begin{cases} c = \text{const}, & (x, y) \in D, \\ 0, & \text{вне } D \end{cases}$

Используя свойство  $f_v1$  совместной плотности  $f(x, y)$ , найдем  $c$ :

$$1 = \int_0^1 \int_0^{-x+1} c \, dx \, dy = c \int_0^1 (-x + 1) \, dx = c \left( -\frac{1}{2} + 1 \right) = \frac{1}{2}c.$$

Таким образом,  $c = 2$ , следовательно,  $f(x, y) = 2$  для всех  $(x, y) \in D$ .

2) Для вычисления плотностей компонент используем свойства  $f_v4$ ,  $f_v5$ , тогда:

$$f_X(x) = \begin{cases} \int_0^{-x+1} 2 \, dy = 2 \cdot (-x + 1 + 0) = 2 - 2x, & x \in (0, 1); \\ 0, & x \notin (0, 1) \end{cases}$$

$$f_Y(y) = \begin{cases} \int_0^{-y+1} 2 \, dx = 2 \cdot (-y + 1 + 0) = 2 - 2y, & y \in (0, 1); \\ 0, & y \notin (0, 1) \end{cases}.$$

3) Используя 5.4 для вычисления плотностей условных распределений, имеем:

$$f(x|y_0) = \begin{cases} 2/(2 - 2y_0) = 1/(1 - y_0), & x \in (0, 1 - y_0), y_0 \in (0, 1); \\ 0, & (x, y) \notin D \end{cases}$$

$$f(y|x_0) = \begin{cases} 2/(2 - 2x_0) = 1/(1 - x_0), & y \in (0, 1 - x_0), x_0 \in (0, 1); \\ 0, & (x, y) \notin D \end{cases}. \quad (5.6)$$

В этом примере распределение в любом сечении является равномерным.

## 5.2 Независимые случайные величины.

**Определение 5.6** Случайные величины  $X$  и  $Y$  называются независимыми, если для любых боревских множеств  $A$  и  $B$  выполнено соотношение

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \quad (5.7)$$

**Замечание 5.2** Событие  $(X \in A, Y \in B)$  на языке теории множеств записывается как пересечение множеств:  $\{\omega : X(\omega) \in A\} \cap \{\omega : Y(\omega) \in B\}$

**Замечание 5.3** Свойство независимости на геометрическом языке площадей фигур звучит очень привычно: "площадь" прямоугольника равна произведению "длин" его сторон. Вероятностная мера не всегда обладает этим свойством, если же для вероятностной меры оно справедливо, то породившие эту меру случайные величины называются независимыми, то есть "поведение" одной из них на вторую "не влияет".

**Теорема 5.1** Необходимым и достаточным условием независимости случайных величин  $X$  и  $Y$  является представимость функции их совместного распределения в виде

$$F_{\mathbf{X}}(x, y) = F_X(x)F_Y(y), \text{ при всех } x, y \in R. \quad (5.8)$$

*Доказательство. Необходимость.* Если случайные величины  $X$  и  $Y$  независимы, то представимость функции их совместного распределения в виде (5.8) следует из определения функции распределения  $F(x, y) = P(X < x, Y < y) = P(X \in (-\infty, x), Y \in (-\infty, y))$  и того, что множества  $(-\infty, x)$ ,  $(-\infty, y) \in \mathcal{B}$  - борелевской  $\sigma$ -алгебре на  $R_1$  при любых  $x, y$ .

*Достаточность. Схема доказательства.* Доказательство того, что из (5.8) следует независимость случайных величин  $X$  и  $Y$  основано на свойстве  $F_v$  функции распределения - из этого свойства следует, что соотношение (5.7) выполнено для любого прямоугольника: пусть  $D = \{(x, y) \in R_2 : a \leq x < b, c \leq y < d\}$  - прямоугольник, тогда

$$\begin{aligned} P(X \in [a, b), Y \in [c, d)) &= P((X, Y) \in D) = F(b, d) - F(b, c) - F(a, d) + \\ &+ F(a, c) = F_X(b)F_Y(d) - F_X(b)F_Y(c) - F_X(a)F_Y(d) + F_X(a)F_Y(c) = \\ &= F_X(b)(F_Y(d) - F_Y(c)) - F_X(a)(F_Y(d) - F_Y(c)) = \\ &= (F_X(b) - F_X(a))(F_Y(d) - F_Y(c)) = P(X \in [a, b)) P(Y \in [c, d)). \end{aligned}$$

Отсюда легко следует, что соотношение (5.7) справедливо для любого множества из алгебры, построенной на базе прямоугольников. Затем доказывается теорема аппроксимации о вероятностной связи алгебры и порожденной ею  $\sigma$ -алгебры (то есть пересечения всех  $\sigma$ -алгебр, содержащих данную алгебру), откуда следует свойство (5.7) для всех боревских множеств  $A$  и  $B$ .

**Теорема 5.2** Для дискретных случайных величин необходимым и достаточным условием независимости случайных величин  $X$  и  $Y$  является представимость вероятности их совместного распределения в виде

$$p_{ij} = p_{X,i}p_{Y,j}, \text{ где } p_{X,i} = P(X = x_i), \quad p_{Y,j} = P(Y = y_j) \text{ при всех } i, j. \quad (5.9)$$

*Доказательство. Необходимость.* Как и в теореме 5.1 из условия (5.9) независимость случайных величин  $X$  и  $Y$  следует в силу того, что  $p_{ij} = P(X = x_i, Y = y_j)$ , а одноточечные множества  $\{x_i\}$  и  $\{y_j\}$  являются борелевскими.

*Достаточность* следует из свойства вероятностей  $p_{ij}$ : для любых борелевских  $A$  и  $B$  имеем

$$\begin{aligned} P(X \in A, Y \in B) &= P((X, Y) \in \{(x, y) : x \in A, y \in B\}) = \\ &= P((X, Y) \in A \times B) = \sum_{i,j: (x_i, y_j) \in A \times B} p_{ij} = \sum_{i,j: (x_i, y_j) \in A \times B} p_{X,i} p_{Y,j} = \\ &= \left( \sum_{i: x_i \in A} p_{X,i} \right) \left( \sum_{j: y_j \in B} p_{Y,j} \right) = P(X \in A) P(Y \in B). \end{aligned}$$

**Теорема 5.3** Для непрерывных случайных величин справедливы два утверждения.

А) Из независимости непрерывных случайных величин  $X$  и  $Y$  следует непрерывность случайного вектора  $\mathbf{X} = (X, Y)$ .

В) Необходимым и достаточным условием независимости непрерывных случайных величин  $X$  и  $Y$  является представимость плотности их совместного распределения в виде

$$f(x, y) = f_X(x) f_Y(y), \quad \text{при всех } x, y \in R. \quad (5.10)$$

*Доказательство пункта А и необходимости пункта В.* Пусть  $X$  и  $Y$  независимые непрерывные случайные величины. Из теоремы 5.1 имеем  $F(x, y) = F_X(x) F_Y(y)$ , откуда по теореме Фубини для любых вещественных  $x, y$

$$F(x, y) = F_X(x) F_Y(y) = \int_{-\infty}^x f_X(u) du \int_{-\infty}^y f_Y(v) dv = \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) du dv$$

По определению 5.4, при  $f(x, y) = f_X(x) f_Y(y)$ , отсюда следует непрерывность случайного вектора  $\mathbf{X} = (X, Y)$  и необходимость пункта В.

*Доказательство достаточности пункта В.* Пусть справедливо равенство (5.10), отсюда следует непрерывность случайного вектора  $\mathbf{X} = (X, Y)$ . Из (5.1), используя теорему Фубини и определение 4.5 получим

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv = \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) du dv \\ &= \int_{-\infty}^x f_X(u) du \int_{-\infty}^y f_Y(v) dv = F_X(x) F_Y(y). \end{aligned} \quad (5.11)$$

По теореме 5.1 из (5.11) следует независимость случайных величин  $X$  и  $Y$ .

**Утверждение 5.1** (Свойство Е5 математического ожидания) Пусть случайные величины  $X$  и  $Y$  независимы и существуют математические ожидания  $E(X)$ ,  $E(Y)$ , тогда  $E(XY) = E(X)E(Y)$ .

*Доказательство.* Докажем это утверждение для дискретных случайных величин, в случае непрерывных случайных величин доказательство проводится аналогично с заменой сумм на интегралы. Используя свойства абсолютно сходящихся рядов (теорему Фубини для интегралов) будем иметь:

$$\begin{aligned} E(XY) &= \sum_i \sum_j x_i y_j p_{ij} = \sum_i \sum_j x_i y_j p_{X,i} p_{Y,j} = \\ &= \left( \sum_i x_i p_{X,i} \right) \left( \sum_j y_j p_{Y,j} \right) = E(X)E(Y). \end{aligned}$$

Из этого утверждения и свойства D4 дисперсий вытекает

**Следствие 5.1** Пусть случайные величины  $X$  и  $Y$  независимы и существуют дисперсии  $D(X)$ ,  $D(Y)$ , тогда  $D(X + Y) = D(X) + D(Y)$ .

**Пример 5.5** Пусть задана плотность распределения случайного вектора:

$$f(x, y) = \begin{cases} Ae^{-x-2y}, & x \geq 0 \text{ и } y \geq 0, \\ 0, & x < 0 \text{ или } y < 0. \end{cases}$$

Найти  $A$ ,  $F(x, y)$ ,  $f_X(x)$ ,  $f_Y(y)$ .

1).  $A \int_0^\infty e^{-x} dx \int_0^\infty e^{-2y} dy = A/2 = 1$  - свойство нормировки, откуда  $A = 2$ .

2).  $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy = 2 \int_0^x e^{-u} du \int_0^y e^{-2v} dv = (1 - e^{-x})(1 - e^{-2y})$ , при  $x \geq 0$  и  $y \geq 0$ ;  $F(x, y) = 0$  при  $x < 0$  или  $y < 0$ .

3).  $f_X(x) = \int_0^\infty 2e^{-x-2y} dy = 2e^{-x} \int_0^\infty e^{-2y} dy = e^{-x}$ , при  $x > 0$  и  $f_X(x) = 0$ , при  $x \leq 0$ .

Аналогично,  $f_Y(y) = 2e^{-2y}$ ,  $y > 0$  и  $f_Y(y) = 0$ ,  $y \leq 0$ . Следовательно,  $f(x, y) = f_X(x)f_Y(y)$  и случайные величины  $X$  и  $Y$  независимы.

Можно найти функции распределения по определению  $F_X(x) = \int_{-\infty}^x \int_{-\infty}^\infty f(u, y) du dy = \int_0^x du \int_0^\infty 2e^{-u-2y} dy = 1 - e^{-x}$ ,  $x \geq 0$   $F_X(x) = 0$ ,  $x < 0$ .

Аналогично,  $F_Y(y) = 1 - e^{-2y}$ ,  $y \geq 0$  и  $F_Y(y) = 0$ ,  $y < 0$ . С другой стороны, в силу независимости  $F(x, y) = F_X(x)F_Y(y)$ , и этот результат следует из пункта 2.

## 5.3 Числовые характеристики случайного вектора.

**Математические ожидания компонент случайного вектора.** Начальные и центральные моменты компонент случайного вектора определяются как и ранее - в подразделе 4.2. Совокупность математических ожиданий  $m_x = E(X)$  и  $m_y = E(Y)$  представляет собой характеристику положения системы:  $(m_x, m_y)$  - координаты центра рассеивания системы. Можно привести формулы для вычисления математических ожиданий и дисперсий компонент случайного вектора  $\mathbf{X}$ . Для дискретного случайного вектора:

$$E(X) = \sum_i \sum_j x_i p_{ij}; \quad E(Y) = \sum_i \sum_j y_i p_{ij}$$

$$D(X) = \sum_i \sum_j (x_i - E(X))^2 p_{ij}; \quad D(Y) = \sum_i \sum_j (y_i - E(Y))^2 p_{ij},$$

Для непрерывного случайного вектора:

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy; \quad E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy$$

$$D(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))^2 f(x, y) dx dy; \quad D(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - E(Y))^2 f(x, y) dx dy.$$

Дисперсии  $D(X)$ ,  $D(Y)$  характеризуют рассеивание случайной точки  $(X, Y)$  соответственно вдоль осей  $OX$  и  $OY$

### **Ковариация, коэффициент корреляции системы случайных величин.**

Для многомерного случайного вектора кроме начальных и центральных моментов рассматривают смешанные начальные и центральные моменты.

**Определение 5.7** Начальными смешанными моментами порядка  $k$  случайного вектора  $\mathbf{X}=(X, Y)$  называются числа  $E(X^i Y^j)$ ,  $i, j = 1, 2, \dots, i+j = k$ . Центральными смешанными моментами порядка  $k$  случайного вектора  $\mathbf{X}=(X, Y)$  называются числа  $E((X - E(X))^i (Y - E(Y))^j)$ ,  $i, j = 1, 2, \dots, i+j = k$ .

Рассмотрим более подробно второй центральный смешанный момент.

**Определение 5.8** Ковариацией случайных величин  $X$  и  $Y$  называется математическое ожидание произведения отклонений этих величин от своих математических ожиданий, обозначается  $K(X, Y)$ ,  $cov(X, Y)$ :

$K(X, Y) = E((X - E(X))(Y - E(Y)))$  или  $K(X, Y) = E(\dot{X}\dot{Y})$ , где  $\dot{X} = X - E(X)$ ,  $\dot{Y} = Y - E(Y)$  - центрированные случайные величины. То есть, ковариация есть математическое ожидание произведения центрированных величин (второй центральный смешанный момент). Ковариация является качественной характеристикой зависимости случайных величин. Если распределение дискретное, то

$$K(X, Y) = \sum_i \sum_j (x_i - E(X))(y_j - E(Y)) p_{ij}.$$

В непрерывном случае,

$$K(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))(y - E(Y)) f(x, y) dx dy.$$

*Свойства ковариации:*

K1.  $K(X, Y) = K(Y, X)$  - следует из определения.



К2. Для любых вещественных чисел  $a$  и  $b$  справедливо равенство  $K(aX + b, Y) = aK(X, Y)$  - следует из того, что  $E(aX + b) = aE(X) + b$  и определения ковариации.

К3.  $K(X, X) = D(X)$ ,  $K(Y, Y) = D(Y)$  - следует из определения.

К4.  $K(X, Y) = E(XY) - E(X)E(Y)$ .

Действительно,  $K(X, Y) = E(\tilde{X}\tilde{Y}) = E((X - E(X))(Y - E(Y))) = E(XY - E(X)Y - XE(Y) + E(X)E(Y)) = E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y)$ .

**Следствие 5.2** В силу свойства E5 математического ожидания (см. утверждение 5.1) для независимых  $X$  и  $Y$  ковариация равна нулю -  $K(X, Y) = 0$ .

**Определение 5.9** Если  $K(X, Y) = 0$ , то случайные величины  $X$  и  $Y$  называются некоррелированными.

**Замечание 5.4** Из следствия 5.2 получаем, что независимые случайные величины являются некоррелированными. Но достаточно просто построить пример (см. пример 5.6) некоррелированных случайных величин, являющихся зависимыми. Для случайных величин  $X$  и  $Y$ , имеющих нормальное распределение, независимость и некоррелированность эквивалентны.

Если  $K(X, Y) = 0$ , то случайные величины  $X$  и  $Y$  зависимы.

**Утверждение 5.2** Если случайные величины  $X$  и  $Y$  некоррелированы, то  $D(X + Y) = D(X) + D(Y)$ .

*Доказательство.* Равенство  $E(XY) = E(X)E(Y)$  следует из свойства К4 ковариации, тогда равенство  $D(X + Y) = D(X) + D(Y)$  следует из свойства D4 дисперсии.

Это свойство можно распространить на любой набор попарно некоррелированных случайных величин. Пусть  $X_1, X_2, \dots, X_n$  набор состоящий из  $n$  попарно некоррелированных случайных величин, тогда  $D(\sum_{i=1}^n X_i) = \sum_{i=1}^n D(X_i)$ .

**Замечание 5.5** В общем случае, используя ковариацию, можно записать

$$D\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n D(X_i) + 2 \sum_{i>j} K(X_i, X_j).$$

В частности

$$D(X + Y) = D(X) + D(Y) + 2K(X, Y).$$

*Коэффициент корреляции.*

**Определение 5.10** Коэффициентом корреляции случайных величин  $X$  и  $Y$  называется величина  $r(X, Y) = \frac{K(X, Y)}{\sigma(X)\sigma(Y)}$ , где, как и ранее,  $\sigma(X) = \sqrt{D(X)}$ ,  $\sigma(Y) = \sqrt{D(Y)}$ .

*Свойства коэффициента корреляции.*

г1. Коэффициент корреляции по модулю не превышает единицу:  $|r(X, Y)| \leq 1$ .

*Доказательство.* Обозначим  $\check{X} = (X - E(X))/\sigma(X)$ ,  $\check{Y} = (Y - E(Y))/\sigma(Y)$  - соответствующие случайным величинам  $X$  и  $Y$  центрированные нормированные случайные величины. Тогда

$$E(\check{X}) = E(\check{Y}) = 0, \quad D(\check{X}) = D(\check{Y}) = 1, \quad K(\check{X}, \check{Y}) = r(X, Y).$$

Отсюда

$$0 \leq D(\check{X} \pm \check{Y}) = D(\check{X}) + D(\check{Y}) \pm 2K(\check{X}, \check{Y}) = 2 \pm 2r(X, Y).$$

Следовательно,  $r(X, Y) \leq 1$ ,  $r(X, Y) \geq -1$ .

г2. Из следствия 5.2 получаем, что если случайные величины  $X$  и  $Y$  независимы, то  $r(X, Y) = 0$ .

г3. Равенство  $|r(X, Y)| = 1$  достигается тогда и только тогда, когда величины  $X$  и  $Y$  связаны линейной зависимостью  $Y = aX + b$ , где  $a \neq 0$  и  $b$  постоянны. При этом  $ar(X, Y) > 0$ , то есть  $a$  и  $r(X, Y)$  одного знака. Доказательство см. например, [1].

**Пример 5.6** Рассмотрим систему случайных величин  $(X, Y)$  с плотностью

$$f(x, y) = \begin{cases} 2(a^2 - x^2 - y^2)/(\pi a^4), & x^2 + y^2 \leq a^2, \quad (a > 0), \\ 0, & x^2 + y^2 > a^2. \end{cases}$$

Найти коэффициент корреляции  $r(X, Y)$ .

Вычисления будем производить в полярной системе координат. Обозначим  $A = 2/(\pi a^4)$ , тогда

$$E(X) = E(Y) = A \int_0^{2\pi} d\varphi \int_0^a r \cos \varphi (a^2 - r^2) r dr = A \int_0^{2\pi} \cos \varphi d\varphi \int_0^a r^2 (a^2 - r^2) dr = 0.$$

Следовательно,  $K(X, Y) = E(XY)$  и

$$K(X, Y) = A \int_0^{2\pi} d\varphi \int_0^a r \sin \varphi r \cos \varphi (a^2 - r^2) r dr = A \int_0^{2\pi} \sin \varphi \cos \varphi d\varphi \int_0^a r^3 (a^2 - r^2) dr = 0, \\ r(X, Y) = 0.$$

В данном случае, случайные величины  $X$  и  $Y$  являются некоррелированными, но зависимыми, так как при  $x \in (-a, a)$ ,  $y \in (-a, a)$

$$f_X(x) = A \int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} (a^2 - x^2 - y^2) dy = \frac{8}{3\pi a^4} (a^2 - x^2)^{3/2}, \quad f_Y(y) = \frac{8}{3\pi a^4} (a^2 - y^2)^{3/2}.$$

Вне указанных промежутков обе плотности равны нулю. Следовательно,  $f(x, y) \neq f_X(x)f_Y(y)$ .

*Ковариационная матрица системы случайных величин.*

Пусть теперь  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ . Моменты  $E(X_i)$ ,  $i = 1, 2, \dots, n$ ;  $D(X_i)$ ,  $i = 1, 2, \dots, n$  вычисляются аналогично двумерному случаю для каждой случайной величины  $X_1, X_2, \dots, X_n$ . Ковариации  $K_{ij} = K(X_i, X_j)$  вычисляются по формулам:

$$K_{ij} = E(\dot{X}_i \dot{X}_j), \quad \text{где } \dot{X}_i = X_i - E(X_i), \quad \dot{X}_j = X_j - E(X_j).$$

**Определение 5.11** Ковариационной матрицей системы случайных величин  $(X_1, X_2, \dots, X_n)$  называется матрица вида  $K = \|K_{ij}\|$ , где  $K_{ij} = E(\dot{X}_i \dot{X}_j)$ .

Данная матрица является неотрицательно определенной симметричной матрицей, имеет на главной диагонали дисперсии  $K_{ii} = D(X_i)$ , при  $i = 1, 2, \dots, n$ . Если  $\det(K) = 0$ , то случайный вектор  $\mathbf{X}$  имеет вырожденное распределение, то есть множество значений случайного вектора  $\mathbf{X}$  имеет в  $R_n$  нулевую меру (см. замечание 5.1), при  $\det(K) \neq 0$  - распределение невырожденное.

Например, пусть для  $\mathbf{X} = (X, Y)$  ковариационная матрица имеет вид  $K = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$ , тогда  $\det(K) = 0$  и  $r(X, Y) = -1$ . То есть  $X$  и  $Y$  линейно зависимы:  $Y = -X + b$ , следовательно, множество значений  $\mathbf{X}$  сосредоточено на прямой  $y = -x + b$  и имеет в  $R_2$  нулевую меру (площадь).

Ковариационная матрица характеризует сосредоточенность распределения вероятностей случайного вектора (рассеивание). Эллипсоид  $(\vec{x} - \vec{a})^T K^{-1} (\vec{x} - \vec{a}) = t$ ,  $t > 0$ , где  $\vec{x} = (x_1, x_2, \dots, x_n)$ , а  $\vec{a}$  - произвольный фиксированный вектор в  $R_n$  называется эллипсоидом рассеивания распределения вероятностей случайного вектора  $\mathbf{X}$  относительно вектора  $\vec{a}$ . Если  $\vec{a} = E\mathbf{X} = (E(X_1), E(X_2), \dots, E(X_n))$ , то эллипсоид рассеивания является геометрической характеристикой концентрации распределения вероятностей случайного вектора  $\mathbf{X}$  около вектора его математических ожиданий  $E\mathbf{X} = (E(X_1), E(X_2), \dots, E(X_n))$ . Сравним со следствием 4.2 - в одномерном случае эллипсоид рассеивания при  $a = E(X)$  будет иметь вид  $(x - E(X))D(X)^{-1}((x - E(X))) = t$  или при  $t = 9$ , будем иметь  $|x - E(X)| = 3\sigma(X)$ . То есть для  $X \sim N(a, \sigma^2)$  эллипсоид рассеивания при  $t = 9$  является границей промежутка  $(a - 3\sigma, a + 3\sigma)$ , в который с вероятностью 0,9973 попадают значения случайной величины  $X$ .

Часто вместо ковариационной матрицы  $\|K_{ij}\|$  в целях наглядности степени зависимости случайных величин  $X_1, X_2, \dots, X_n$  пользуются корреляционной матрицей  $\|r_{ij}\|$ .

**Определение 5.12** Корреляционной матрицей системы случайных величин  $(X_1, X_2, \dots, X_n)$  называется матрица

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 \end{pmatrix}$$

Данная матрица является неотрицательно определенной симметричной матрицей, имеет на главной диагонали единицы, причем

$$r_{ij} = \frac{K_{ij}}{\sigma(X_i)\sigma(X_j)}, \quad \text{где } K_{ij} = E(\dot{X}_i \dot{X}_j) \text{ для любых } i \text{ и } j.$$

Ранее (см. определение 5.5) было дано определение условного закона распределения случайной величины  $Y$  при условии  $X = x$ . Для условного распределения рассматривают условные моменты. Остановимся на вычислении условного математического ожидания и условной дисперсии. В дискретном случае имеем

$$E(Y|X = x_i) = \sum_j y_j P(Y = y_j|X = x_i) = \sum_j y_j p_{ij}/p_{X,i}; \quad (5.12)$$

$$\begin{aligned} D(Y|X = x_i) &= E((Y - E(Y|X = x_i))^2|X = x_i) = \\ &= \sum_j (y_j - E(Y|X = x_i))^2 p_{ij}/p_{X,i}. \end{aligned} \quad (5.13)$$

Для непрерывного случайного вектора

$$E(Y|X = x_0) = \int_{-\infty}^{\infty} y f(y|x_0) dy = \int_{-\infty}^{\infty} y f(x_0, y)/f_X(x_0); \quad (5.14)$$

$$\begin{aligned} D(Y|X = x_0) &= E((Y - E(Y|X = x_0))^2|X = x_0) = \\ &= \int_{-\infty}^{\infty} (y - E(Y|X = x_0))^2 f(x_0, y)/f_X(x_0) dy. \end{aligned} \quad (5.15)$$

**Пример 5.7** В условиях примера 5.3 вычислить  $E(Y|X = 1)$ .

*Решение.* Используя найденное в примере 5.3 условное распределение  $p_{j|1} = P(Y = y_j|X = 1)$  (см. ряд распределения (5.5)) имеем

$$E(Y|X = 1) = (-1) \cdot (5/6) + 0 \cdot (1/6) + 1 \cdot 0 = 5/6.$$

**Пример 5.8** В условиях примера 5.4 вычислить  $E(Y|X = 1/3)$ .

*Решение.* Используя найденную в примере 5.4 условную плотность распределения  $f(y|x_0)$  (см. выражение (5.6)) имеем

$$E(Y|X = 1/3) = \int_0^{1-1/3} (y/(1-1/3)) dy = 3 \left( \int_0^{2/3} y dy \right) / 2 = 1/3.$$

## 5.4 Регрессия и линейная регрессия.

Рассмотрим следующую задачу: пусть задано совместное распределение случайных величин  $X$  и  $Y$ , как по значению случайной величины  $X = x$  предсказать значение случайной величины  $Y$ ? Например, известно совместное распределение веса и роста группы людей. Как, зная рост человека "наилучшим образом" прогнозировать его вес? Для решения такого рода задач служат условное математическое ожидание и регрессия.

**Определение 5.13** Предположим, что задано распределение случайного вектора  $\mathbf{X}=(X, Y)$ . Функции  $\varphi(x) = E(Y|X = x)$  и  $\psi(y) = E(X|Y = y)$  называются функциями регрессии  $Y$  на  $X$  и  $X$  на  $Y$  соответственно. Графики этих функций называются кривыми регрессии.

Можно показать, что если совместное распределение случайных величин  $X$  и  $Y$  является нормальным, то функции регрессии являются линейными функциями:

$$\varphi(x) = E(Y) + r(X, Y) \frac{\sigma(Y)}{\sigma(X)} (x - E(X)) \quad (5.16)$$

$$\psi(y) = E(X) + r(X, Y) \frac{\sigma(X)}{\sigma(Y)} (y - E(Y)). \quad (5.17)$$

Функция регрессии обладает следующим свойством оптимальности. Рассмотрим следующую задачу: найти такую борелевскую функцию  $g(x) = \varphi_b(x)$ , что случайная величина  $g(X) = \varphi_b(X)$  наилучшим образом приближала бы случайную величину  $Y$  в смысле среднего квадратичного отклонения, то есть нас интересует решение экстремальной задачи (то есть функция на которой минимум достигается)

$$\inf_{\varphi_b(x)} E (Y - \varphi_b(X))^2. \quad (5.18)$$

Решением такой задачи является функция регрессии

$$\varphi_b(x) = E(Y|X = x).$$

**Замечание 5.6** Можно показать, что значение экстремальной задачи 5.18 - величина  $E (Y - E(Y|X = x))^2$  является средней условной дисперсией, то есть

$$E (Y - E(Y|X = x))^2 = E(D(Y|X)),$$

где условная дисперсия  $D(Y|X)$  является случайной величиной определяемой как функция от случайной величины  $X$  (см. следующий раздел 6 и (5.13), (5.15)):

$$d(x) = D(Y|X = x), \quad D(Y|X) = d(X).$$

Пусть теперь требуется найти *линейную* функцию  $g(x) = ax + b$ , для которой случайная величина  $g(X) = aX + b$  являлась бы наилучшим приближением  $Y$  в смысле среднего квадратичного отклонения, то есть в задаче (5.18) минимум рассматривается только в классе линейных функций:

$$\inf_{a,b} E (Y - (aX + b))^2. \quad (5.19)$$

Решением задачи (5.19) является функция вида, аналогичного (5.16),

$$y = g(x) = E(Y) + r(X, Y) \frac{\sigma(Y)}{\sigma(X)} (x - E(X)), \quad (5.20)$$

а определяющее ее уравнение (5.20) называется *уравнением линейной среднеквадратичной регрессии  $Y$  на  $X$* .

**Пример 5.9** В условиях примера 5.3 найти уравнение линейной среднеекватичной регрессии  $Y$  на  $X$ .

*Решение.* Используя найденные в примере 5.3 распределения случайных величин  $X$  и  $Y$  найдем математические ожидания и дисперсии

$$E(X) = -0,6 + 0,4 = -0,2; \quad E(Y) = -0,5 + 0,2 = -0,3;$$

$$D(X) = E(X^2) - (E(X))^2 = 0,6 + 0,4 - (-0,2)^2 = 0,96;$$

$$D(Y) = E(Y^2) - (E(Y))^2 = 0,5 + 0,2 - (-0,3)^2 = 0,61.$$

Для нахождения коэффициента корреляции  $r(X, Y)$  найдем сначала

$$\begin{aligned} E(XY) &= (-1) \cdot P(XY = -1) + 0 \cdot P(XY = 0) + 1 \cdot P(XY = 1) = \\ &= -P(X = -1, Y = 1) - P(X = 1, Y = -1) + \\ &+ P(X = -1, Y = -1) + P(X = 1, Y = 1) = \\ &= -0 - 0 + 0,5 + 0,2 = 0,7. \end{aligned}$$

Тогда

$$K(X, Y) = E(XY) - E(X)E(Y) = 0,7 - 0,2 \cdot 0,3 = 0,64.$$

Следовательно,

$$r(X, Y) = \frac{K(X, Y)}{\sigma(X)\sigma(Y)} = \frac{0,64}{\sqrt{0,61 \cdot 0,96}} = \frac{64}{\sqrt{61 \cdot 96}}.$$

Подставляя полученные результаты в уравнение (5.20) получим

$$y = -0,3 + \frac{64}{\sqrt{61 \cdot 96}} \sqrt{\frac{61}{96}} (x + 0,2),$$

то есть уравнение линейной среднеекватичной регрессии  $Y$  на  $X$  имеет вид

$$y = g(x) = -\frac{3}{10} + \frac{2}{3} \left( x + \frac{1}{5} \right).$$

**Пример 5.10** В условиях примера 5.4 найти уравнение среднеекватичной регрессии  $Y$  на  $X$ .

*Решение.* Аналогично примеру 5.8, используя найденную в примере 5.4 условную плотность распределения  $f(y|x_0)$  (см. выражение (5.6)) имеем

$$\varphi(x) = E(Y|X = x) = \begin{cases} \int_0^{1-x} \frac{y}{1-x} dy = \frac{\int_0^{1-x} y dy}{1-x} = \frac{1-x}{2}, & x \in (0, 1), \\ 0, & x \notin (0, 1) \end{cases},$$

то есть кривая регрессии  $Y$  на  $X$  является графиком функции

$$\varphi(x) = \begin{cases} \frac{1-x}{2}, & x \in (0, 1) \\ 0, & x \notin (0, 1) \end{cases}$$

и совпадает с графиком линейной среднеекватичной регрессии  $Y$  на  $X$  на отрезке  $(0, 1)$ .

## 6 Функции случайной величины.

### 6.1 Закон распределения монотонной функции случайного аргумента.

**Определение 6.1** Пусть функция  $y = \varphi(x)$  является борелевской функцией, то есть  $\varphi^{-1}(B) \in \mathcal{B}$  для любого  $B \in \mathcal{B}$  и на измеримом пространстве  $(\Omega, \mathcal{A})$  задана случайная величина  $X$ . Тогда функция  $Y = \varphi(X)$  называется функцией случайного аргумента  $X$ . Функция  $Y(\omega) = \varphi(X(\omega))$  является, в силу определения, случайной величиной, заданной на том же измеримом пространстве  $(\Omega, \mathcal{A})$ .

*Закон распределения функции случайной величины.*

Рассмотрим сначала дискретный случай. Пусть  $X$  - дискретная случайная величина, закон распределения которой задан рядом распределения

$$\begin{array}{c|c|c|c|c|c} x_i & x_1 & x_2 & \dots & x_n & \dots \\ \hline p_i & p_1 & p_2 & \dots & p_n & \dots \end{array}, \quad \text{где } p_i = P(X = x_i), \quad i = 1, 2, \dots, n, \dots$$

Тогда  $Y = \varphi(X)$  также будет дискретной случайной величиной и её закон распределения задается таблицей

$$\begin{array}{c|c|c|c|c|c} Y = \varphi(x_i) & \varphi(x_1) & \varphi(x_2) & \dots & \varphi(x_n) & \dots \\ \hline p_i & p_1 & p_2 & \dots & p_n & \dots \end{array}$$

**Замечание 6.1** Для получения ряда распределения случайной величины  $Y$  нужно, объединить все значения  $x_{ij}$ , для которых  $\varphi(x_{ij})$  принимает одно и то же значение:  $\varphi(x_{ij}) = y_i$ , при этом  $P(Y = y_i) = \sum_j P(X = x_{ij}) = \sum_j p_{ij}$ .

**Пример 6.1** Закон распределения случайной величины  $X$  задан следующим рядом распределения

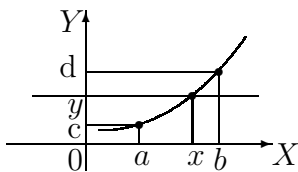
$$\begin{array}{c|c|c|c|c|c} x_i & -2 & -1 & 0 & 1 & 2 \\ \hline p_i & 0,1 & 0,1 & 0,3 & 0,3 & 0,2 \end{array}.$$

Найти закон распределения случайных величин  $Y = X^2$  и  $Z = X^3$ .

$$\text{Имеем } \begin{array}{c|c|c|c|c|c} x_i^2 & 4 & 1 & 0 & 1 & 4 \\ \hline p_i & 0,1 & 0,1 & 0,3 & 0,3 & 0,2 \end{array}, \text{ тогда}$$

$$\begin{array}{c|c|c|c} y_i & 0 & 1 & 4 \\ \hline p_i & 0,3 & 0,4 & 0,3 \end{array}, \quad \begin{array}{c|c|c|c|c|c} z_i = x_i^3 & -8 & -1 & 0 & 1 & 8 \\ \hline p_i & 0,1 & 0,1 & 0,3 & 0,3 & 0,2 \end{array}.$$

Пусть теперь  $X$  - непрерывная случайная величина с плотностью  $f(x)$ . Рассмотрим  $(a, b)$ , такой что  $P(a < X < b) = 1$  и дифференцируемую функцию  $y = \varphi(x)$ , такую что  $\varphi(a) = c$ ,  $\varphi(b) = d$ . Здесь  $a, b, c, d$  могут быть как вещественными числами, так и  $\mp\infty$  соответственно.



Пусть  $y = \varphi(x)$  на  $(a, b)$  монотонно возрастает.

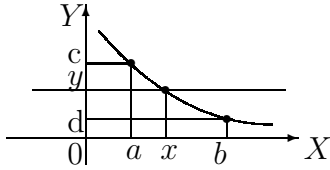
Обозначим  $g(y)$  - плотность распределения случайной величины  $Y$ . Тогда

$$G(y) = P(Y < y) = P(a \leq X < x) = \int_a^x f(x)dx = \int_a^{\psi(y)} f(x)dx,$$

где верхний предел  $x = \psi(y)$ ,  $\psi(y)$  - функция обратная  $\varphi(x)$ . Дифференцируя интеграл по  $y$  (от  $y$  зависит верхний предел), получим

$$g(y) = G'(y) = \left( \int_a^{\psi(y)} f(x)dx \right)'_y = f(\psi(y))\psi'(y).$$

Рассмотрим теперь случай, когда  $y = \varphi(x)$  на  $(a, b)$  монотонно убывает.



В этом случае

$$G(Y) = P(Y < y) = P(x < X \leq b) = \int_b^x f(x)dx = \int_{\psi(y)}^b f(x)dx \Rightarrow g(y) = G'(y) = -f(\psi(y))\psi'(y).$$

Можно объединить эти два случая: пусть  $y = \varphi(x)$  дифференцируемая монотонная на  $(a, b)$ ,  $P(a < X < b) = 1$  функция, а  $X$  - непрерывная случайная величина с плотностью распределения  $f(x)$ , тогда  $g(y)$  - плотность распределения случайной величины  $Y = \varphi(X)$  находится по формуле:  $g(y) = f(\psi(y))|\psi'(y)|$ , где  $\psi(y) = \varphi^{-1}(y)$  - функция обратная для  $\varphi(x)$ .

**Пример 6.2** Пусть случайная величина  $X$  подчинена закону распределения Коши с плотностью  $f(x) = (\pi(1 + x^2))^{-1/2}$ . Найти плотность распределения  $g(y)$  случайной величины  $Y = 1 - X^3$ . Функция  $y = 1 - x^3$  является монотонной на всей числовой оси, поэтому

$$g(y) = f(\psi(y))|\psi'(y)| = \left( \pi(1 + \sqrt[3]{(1-y)^2}) \right)^{-1/2} / (3\sqrt[3]{(1-y)^2}),$$

так как  $x = \psi(y) = \sqrt[3]{1-y}$ .

В общем случае для непрерывной случайной величины  $X$  с функцией распределения  $F(x)$ , плотностью  $f(x)$  и случайной величины  $Y = \varphi(X)$ , где  $y = \varphi(x)$  - борелевская функция можно написать следующую формулу для нахождения  $G(y)$  - функции распределения случайной величины  $Y = \varphi(X)$ . Обозначим через  $D(y) = \{x : \varphi(x) \in (-\infty, y)\}$  - полный прообраз промежутка  $(-\infty, y)$ , тогда

$$G(y) = P(Y < y) = P(\varphi(X) < y) = P(X \in D(y)) = \int_{D(y)} f(x)dx.$$

**Пример 6.3** Пусть  $X \sim N(-1, 1)$ . Найти функцию и плотность распределения случайной величины  $Y = |X|$ .



Найдем  $G(y)$  - функцию распределения  $Y$ .

$$\begin{aligned} G(y) &= P(Y < y) = P(|X| < y) = P(-y < X < y) = F(y) - F(-y) = \\ &= \Phi(y+1) - \Phi(-y+1) = \frac{1}{\sqrt{2\pi}} \left( \int_{-\infty}^{y+1} e^{-x^2/2} dx - \int_{-\infty}^{-y+1} e^{-x^2/2} dx \right) = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-y+1}^{y+1} e^{-x^2/2} dx. \end{aligned}$$

Отсюда плотность  $g(y) = G'(y) = \Phi'(y+1) - \Phi'(-y+1) = (e^{-(y+1)^2/2} + e^{-(y-1)^2/2})/\sqrt{2\pi}$ .

Сформулируем без доказательства важное утверждение о независимости функций аргументы которых являются случайными

**Утверждение 6.1** Пусть  $X_1, X_2, \dots, X_n$  - независимые случайные величины, а функции  $y = g_1(x), y = g_2(x), \dots, y = g_n(x)$  являются борелевскими функциями (например, кусочно-непрерывными), тогда случайные величины  $g_1(X_1), g_2(X_2), \dots, g_n(X_n)$  независимы. При этом функции  $g_1(x), g_2(x), \dots, g_n(x)$  могут быть как различными, так частично или полностью совпадающими:  $g_1(x) = x, g_2(x) = x^2, \dots, g_n(x) = x^n$ , или  $g_1(x) = x, g_2(x) = x, g_3(x) = x^2, \dots, g_n(x) = x^2$ .

## 6.2 Числовые характеристики функции случайной величины.

Рассмотрим непрерывную случайную величину  $X$  с плотностью распределения  $f(x)$ . Пусть задана борелевская функция  $y = \varphi(x)$ . Тогда математическое ожидание случайной величины  $Y = \varphi(X)$  находится по формуле.

$$E(Y) = \int_{-\infty}^{\infty} \varphi(x)f(x)dx,$$

если интеграл в правой части формулы сходится абсолютно.

В случае дискретной случайной величины  $X$

$$E(Y) = E(\varphi(X)) = \sum_i \varphi(x_i)p_i,$$

если ряд в правой части формулы сходится абсолютно.

*Дополнительные свойства моментов.*

В этом разделе мы дополним свойства математического ожидания и дисперсии, перечисленные ранее.

1.  $E(XY) = E(X) \cdot E(Y) + K(X, Y)$ .
2. Пусть  $Y = \sum_{i=1}^n (a_i X_i + b_i)$ , тогда

$$E(Y) = \sum_{i=1}^n (a_i E(X_i) + b_i), \quad D(Y) = \sum_{i=1}^n a_i^2 D(X_i) + 2 \sum_{i < j} a_i a_j K(X_i, X_j)$$

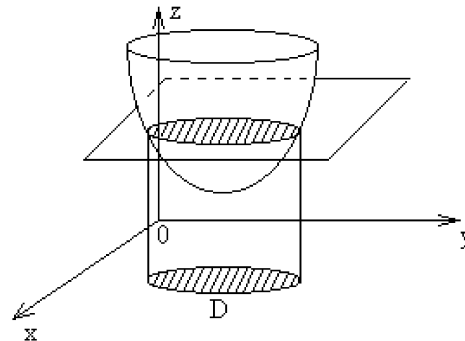
**Пример 6.4** Пусть имеются две случайные величины  $X$  и  $Y$ , связанные соотношением  $Y = -3X + 2$ , известно, что  $E(X) = -1$ ,  $D(X) = 4$ . Определить  $E(Y)$ ,  $D(Y)$ ,  $K(X, Y)$ ,  $r(X, Y)$ .

*Решение.*  $E(Y) = -3E(X) + 2 = 5$ ;  $D(Y) = (-3)^2 D(X) = 36$ ,  $\sigma(X) = 2$ ,  $\sigma(Y) = 6$ . В силу линейной зависимости  $X$  и  $Y$  ( $a = -3 < 0$ ) имеем  $r(X, Y) = -1$ , следовательно,  $K(X, Y) = -\sigma(X)\sigma(Y) = -12$ .

### 6.3 Закон распределения функции двух случайных величин.

Рассмотрим систему непрерывных случайных величин  $(X, Y)$  с плотностью распределения  $f(x, y)$ . Пусть  $\varphi(x, y)$  такая функция, что  $Z = \varphi(X, Y)$  является непрерывной случайной величиной. Найдем закон  $Z$ . Пусть  $G(z)$  и  $g(z)$  соответственно функция  $Z$  и ее плотность. Геометрически функция  $z$

$$\begin{aligned} G(z) &= P(Z < z) = P(\varphi(x, y) < z) = \\ &= P((X, Y) \in D) = \int \int_D f(x, y) dx dy \end{aligned}$$



В данное выражение величина  $z$  входит неявно, через пределы интегрирования. Дифференцируя  $G(z)$  по  $z$ , получим плотность распределения случайной величины  $Z$ :  $g(z) = G'(z)$ .

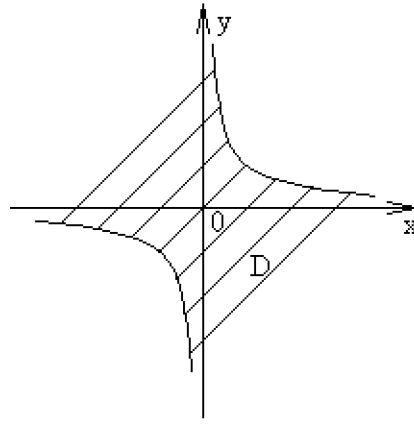
**Пример 6.5** Пусть  $f(x, y)$  - плотность распределения случайного вектора  $\mathbf{X} = (X, Y)$ . Найти  $G(z)$  - функцию распределения случайной величины  $Z = XY$ .

*Решение.* Поскольку  $G(z) = P(Z < z) = P(XY < z)$ , используя свойство  $f_v 3$ , с. 44 находим

$$G(z) = \int \int_D f(x, y) dx dy = \int_{-\infty}^0 dx \int_{z/x}^{\infty} f(x, y) dy + \int_0^{\infty} dx \int_{-\infty}^{z/x} f(x, y) dy.$$

Дифференцируя это выражение по  $z$ , имеем

$$g(z) = G'(z) = - \int_{-\infty}^0 x^{-1} f(x, z/x) dx + \\ + \int_0^{\infty} x^{-1} f(x, z/x) dx.$$



**Пример 6.6** Закон распределения суммы двух случайных величин. Композиция законов распределения.

Рассмотрим систему непрерывных случайных величин  $(X, Y)$  с плотностью  $f(x, y)$ . Найдем функцию распределения случайной величины  $Z = X + Y$ .

$$G(z) = \int_D \int f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) dx dy = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{z-x} f(x, y) dy \right) dx.$$

$$g(z) = G'(z) = \int_{-\infty}^{\infty} f(x, z-x) dx.$$

Это общая формула для плотности распределения суммы двух случайных величин. Аналогично,

$$g(z) = \int_{-\infty}^{\infty} f(z-y, y) dy.$$

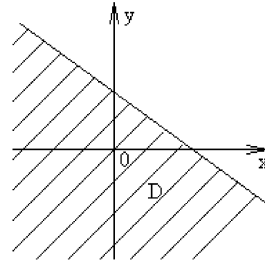
Для независимых случайных величин  $X$  и  $Y$  отсюда получаем

$$g(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \quad \text{и} \quad g(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy.$$

Таким образом, в этом случае  $g(z)$  является сверткой (композицией) исходных плотностей  $f_X(x)$  и  $f_Y(y)$ .

**Пример 6.7** Рассмотрим две независимые случайные величины  $X$  и  $Y$ , подчиненные нормальным законам

$$f_X(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-(x-m_x)^2/(2\sigma_x^2)}; \quad f_Y(y) = \frac{1}{\sigma_y \sqrt{2\pi}} e^{-(y-m_y)^2/(2\sigma_y^2)}.$$



Пусть  $Z = X + Y$ , тогда после преобразований получим:

$$g(z) = \frac{1}{\sqrt{\sigma_x^2 + \sigma_y^2} \cdot \sqrt{2\pi}} e^{-(z-(m_x+m_y))^2/(2(\sigma_x^2+\sigma_y^2))}.$$

Таким образом, сумма (композиция) двух независимых случайных величин имеющих нормальное распределение снова имеет нормальное распределение с параметрами  $m_z = m_x + m_y$  и  $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$ .

Правило композиции нормальных законов может быть обобщено на случай произвольного числа независимых гаусовских случайных величин. Пусть  $X_1, \dots, X_n$  - независимые случайные величины, подчиненные нормальному закону распределения;  $m_{x_1}, \dots, m_{x_n}$  - математические ожидания величин  $X_1, \dots, X_n$ ,  $\sigma_{x_1}^2, \dots, \sigma_{x_n}^2$  - дисперсии  $X_1, \dots, X_n$ . Тогда,  $Z = \sum_{i=1}^n X_i$  - случайная величина подчиненная нормальному закону распределения, причем  $m_z = \sum_{i=1}^n m_{x_i}$ ,  $\sigma_z^2 = \sum_{i=1}^n \sigma_{x_i}^2$ .

Приведем формулы для вычисления моментов функции двух непрерывных случайных величин  $Z = \varphi(X, Y)$ .

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) f(x, y) dx dy, \quad D(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\varphi(x, y) - E(Z))^2 f(x, y) dx dy.$$

## 7 Предельные теоремы. Закон больших чисел

### 7.1 Неравенство Чебышева

**Теорема 7.1** Пусть  $X$  - неотрицательная случайная величина, то есть  $P(X \geq 0) = 1$ , и существует математическое ожидание  $E(X)$ . Тогда для любого  $\varepsilon > 0$  выполнено

$$P(X \geq \varepsilon) \leq \frac{EX}{\varepsilon}. \quad (7.1)$$

*Доказательство.* Приведем доказательство для непрерывной случайной величины  $X$  с плотностью распределения  $f(x)$ . Тогда  $P(X < 0) = 1 - P(X \geq 0) = \int_{x < 0} f(x) dx = 0$ , но  $f(x) \geq 0$ , откуда  $f(x) = 0$  при  $x < 0$  и  $xf(x) \geq 0$ . Следовательно, используя свойства определенного интеграла и свойство плотности  $f(x)$ , будем иметь

$$EX = \int_R xf(x) dx = \int_0^{\infty} xf(x) dx \geq \int_{\varepsilon}^{\infty} xf(x) dx \geq \varepsilon \int_{\varepsilon}^{\infty} f(x) dx = \varepsilon P(X \geq \varepsilon).$$

Отсюда, поделив обе части на  $\varepsilon > 0$  получаем неравенство (7.1). Для дискретной случайной величины доказательство аналогично. В общем случае вместо несобственного риманова интеграла нужно использовать интеграл Лебега-Стилтьеса по мере  $P_X(dx) = dF(x)$ . Все выкладки при этом повторяются.

**Замечание 7.1** Неравенство (7.1) иногда называется неравенством Маркова. Если применить неравенство (7.1) к  $|X|$ , то оно справедливо для любой случайной величины

$$P(|X| \geq \varepsilon) \leq \frac{E|X|}{\varepsilon} \quad \text{и} \quad P(|X| \geq \varepsilon) = P(|X|^s \geq \varepsilon^s) \leq \frac{E(|X|^s)}{\varepsilon^s}, \quad s > 0.$$

**Следствие 7.1** (Неравенство Чебышева.) Пусть  $X$  - случайная величина у которой существует второй начальный момент  $E(X^2) < \infty$ , тогда для любого  $\varepsilon > 0$  справедливы следующие неравенства

$$P(|X| \geq \varepsilon) \leq \frac{E(X^2)}{\varepsilon^2}, \quad P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}.$$

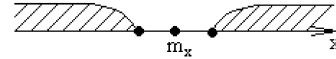
Второе из этих неравенств называется неравенством Чебышева.

*Доказательство.* Во-первых, заметим, что можно доказать, что из существования второго начального момента следует существование математического ожидания и дисперсии.

Первое из приведенных неравенств следует из равенства множеств  $\{|X| \geq \varepsilon\} = \{X^2 \geq \varepsilon^2\}$  и неравенства Маркова. Второе следует из первого, если применить его к случайной величине  $Y = X - E(X)$ .

В приложениях часто используются неравенства Маркова и оба приведенных в следствии неравенства. Поясним геометрический смысл, например, неравенства Чебышева.

Вероятность попадания случайной величины  $X$  в заштрихованную область не превосходит  $D(X)/\varepsilon^2$ .



**Пример 7.1** Для случайной величины  $X$  с заданными математическим ожиданием  $m_x$  и дисперсией  $\sigma_x^2$  оценить сверху вероятность того, что величина  $X$  отклонится от своего математического ожидания  $m_x$  не меньше, чем на  $3\sigma$ . Используя неравенство Чебышева будем иметь

$$P(|X - m_x| \geq 3\sigma_x) \leq \frac{D(X)}{9\sigma_x^2} = \frac{1}{9}.$$

Доказанное неравенство является обобщением правила трех сигм (см. следствие 4.2). Из неравенства Чебышева следует, что  $P(|X - m_x| < \varepsilon) > 1 - D_x/\varepsilon^2$ , откуда  $P(|X - m_x| < 3\sigma_x) > 8/9$ . То есть при наличии математического ожидания и дисперсии с вероятностью большей  $8/9$  (в среднем - 8 значений из 9) все значения случайной величины лежат на промежутке  $(E(X) - 3\sigma(X), E(X) + 3\sigma(X))$ . Более того  $P(|X - m_x| < n\sigma_x) > 1 - 1/n^2$ .

## 7.2 Закон больших чисел. Теорема Чебышева

Рассмотрим различные виды сходимости последовательностей случайных величин.

**Определение 7.1** Говорят, что последовательность случайных величин  $X_n$  сходится при  $n \rightarrow \infty$  к величине  $X$  по вероятности, если для любого  $\varepsilon > 0$

$$P(|X_n - X| \geq \varepsilon) \rightarrow 0.$$

Этот факт обозначают  $X_n \xrightarrow{P} X, n \rightarrow \infty$ .

**Определение 7.2** Говорят, что последовательность случайных величин  $X_n$  сходится при  $n \rightarrow \infty$  к величине  $X$  с вероятностью 1, если  $P\{\omega : X_n(\omega) \rightarrow X(\omega)\} = 1$ , то есть вероятность множества тех исходов, для которых последовательность случайных величин  $X_n$  сходится к случайной величине  $X$  равна 1.

**Теорема 7.2** (Теорема Чебышева, закон больших чисел.) Пусть задана последовательность независимых случайных величин  $X_1, X_2, \dots, X_n, \dots$ , с конечными математическими ожиданиями  $E(X_i)$ ,  $i = 1, 2, \dots, n, \dots$  и ограниченными сверху по совокупности дисперсиями  $D(X_i)$ . То есть существует такое число  $C > 0$ , что  $D(X_i) \leq C$  для любого  $i = 1, 2, \dots$ . Тогда при  $n \rightarrow \infty$  для последовательности средних арифметических  $\bar{X}_n = \sum_{i=1}^n X_i/n$  справедливо соотношение

$$\left| \frac{X_1 + \dots + X_n}{n} - \frac{E(X_1) + \dots + E(X_n)}{n} \right| = |\bar{X}_n - E(\bar{X}_n)| \xrightarrow{P} 0. \quad (7.2)$$

*Доказательство.* Найдем математическое ожидание и дисперсию случайных величин  $\bar{X}_n$ .

$$E(\bar{X}_n) = E\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{E(X_1) + \dots + E(X_n)}{n}.$$

Используя свойство  $D2$  дисперсии получим

$$D(\bar{X}_n) = D\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2} D(X_1 + \dots + X_n).$$

Отсюда, используя независимость случайных величин и ограниченность дисперсий имеем

$$D(\bar{X}_n) = (D(X_1) + D(X_2) + \dots + D(X_n))/n^2 \leq nC/n^2 = C/n.$$

Применяя неравенство Чебышева и переходя к пределу при  $n \rightarrow \infty$  получим

$$P(|\bar{X}_n - E(\bar{X}_n)| \geq \varepsilon) \leq \frac{D(\bar{X}_n)}{\varepsilon^2} \leq \frac{C}{n\varepsilon^2} \rightarrow 0.$$

Это и означает (см. определение (7.1)), что  $|\bar{X}_n - E(\bar{X}_n)|$  сходится по вероятности к нулю.

**Следствие 7.2** Пусть в условиях теоремы Чебышева  $E(X_i) = m$  для любого  $i = 1, 2, \dots$ . Тогда

$$\overline{X}_n \xrightarrow{P} m \quad \text{при } n \rightarrow \infty.$$

Чтобы доказать следствие достаточно заметить, что в этом случае

$$E(\overline{X}_n) = \sum_{i=1}^n E(X_i)/n = m.$$

В частности, условия следствия выполнены для последовательности независимых одинаково распределенных случайных величин:  $X_i \sim X$ ,  $i = 1, 2, \dots$ , если существуют  $E(X)$  и  $D(X)$ . При этом случайную величину  $X$  часто называют генеральной совокупностью, а последовательность  $X_1, X_2, \dots$  - выборкой из генеральной совокупности. Можно интерпретировать  $X_1, X_2, \dots$  как последовательность независимых измерений неизвестной случайной величины  $X$ . Мы получили, что при существовании математического ожидания и дисперсии среднее арифметическое последовательности независимых измерений сходится по вероятности к математическому ожиданию (среднему) измеряемой случайной величины.

**Теорема 7.3 (Бернулли)** Пусть производится  $n$  независимых опытов, в каждом из которых может появиться событие  $A$  с вероятностью  $p = p(A)$ . Тогда при неограниченном увеличении  $n$  - числа опытов, частота (то есть отношение числа успехов к числу опытов) события  $A$  сходится по вероятности к его вероятности  $p = p(A)$ .

**Доказательство.** Рассмотрим независимые случайные величины  $X_i$  - число появлений события  $A$  в  $i$ -ом опыте (бернуллиевские случайные величины). Тогда случайная величина  $X_i$  принимают при  $i = 1, 2, \dots$  ровно два значения - 0 и 1. Ряд распределения случайной величины  $X_i$  при любом  $i$  имеет вид

$$\begin{array}{c|cc} x_i & 0 & 1 \\ \hline p_i & 1-p & p \end{array}.$$

Математическое ожидание -  $E(X_i) = p$ , дисперсия -  $D(X_i) = p(1-p)$ . Пусть  $k = X_1 + X_2 + \dots + X_n$ , тогда  $k$  - число успехов в  $n$  испытаниях является случайной величиной имеющей биномиальное распределение. Последовательность  $X_1, X_2, \dots, X_n, \dots$  удовлетворяет условиям теоремы Чебышева и, следовательно, при  $n \rightarrow \infty$

$$\frac{k}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{P} \frac{m_1 + m_2 + \dots + m_n}{n} = p.$$

Обобщением теоремы Бернулли является теорема Пуассона.

**Теорема 7.4 (Пуассона.)** Если производится  $n$  независимых опытов и вероятность появления события  $A$  в  $i$ -ом опыте равна  $p_i$ , то при увеличении  $n$  модуль разности частоты события  $A$  и среднего арифметического вероятностей  $p_i$  сходится по вероятности к 0.

Аналогично теореме 7.3 теорема 7.4 выводится из теоремы Чебышева 7.2.

Сформулируем усиленный закон больших чисел в форме Колмогорова.

**Теорема 7.5** Пусть  $X_1, X_2, \dots, X_n, \dots$  - последовательность независимых одинаково распределенных случайных величин. Тогда существование математических ожиданий  $E(X_i) = m$  является необходимым и достаточным для того, чтобы с вероятностью 1 среднее арифметическое случайных величин стремилось к их общему математическому ожиданию:

$$\bar{X}_n = \sum_{i=1}^n X_i/n \rightarrow m.$$

### 7.3 Центральная предельная теорема.

Объектом изучения центральных предельных теорем являются предельные законы распределения последовательностей случайных величин. Мы сформулируем одну из них, на которой базируется изучение выборочной модели в математической статистике.

**Теорема 7.6** (Центральная предельная теорема) Пусть  $X_1, X_2, \dots, X_n, \dots$  - независимые одинаково распределенные случайные величины, с конечными ожиданиями  $E(X_i) = m$  и дисперсиями  $D(X_i) = \sigma^2$ . Пусть, как и раньше,  $\bar{X}_n = (\sum_{i=1}^n X_i)/n$ . Рассмотрим последовательность центрированных нормированных величин  $S_n = (\bar{X}_n - E(\bar{X}_n))/\sigma(\bar{X}_n)$ . Тогда при  $n \rightarrow \infty$  последовательность  $F_{S_n}(x)$  функций распределения случайных величин  $S_n$  равномерно по  $x \in R$  сходится к  $\Phi(x)$  - функции распределения стандартного нормального закона.

Если ввести понятие точной верхней грани множества -  $\sup\{D\}$ ,  $D \subset R$  - наименьшего из таких чисел  $C$ , что для всех  $x \in D$ ,  $x \leq C$ , то утверждение теоремы примет вид

$$\sup_{x \in R} |F_{S_n}(x) - \Phi(x)| \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Заметим, что  $E(\bar{X}_n) = m$ ,  $D(\bar{X}_n) = \sigma^2/n$  и, следовательно,  $S_n = \sqrt{n}(\bar{X}_n - m)/\sigma$ .

Рассмотренная в подразделе 3.4 интегральная теорема Муавра-Лапласа является частным случаем центральной предельной теоремы для последовательности бернуллиевских случайных величин.

Этот вариант центральной предельной теоремы демонстрирует значимость нормального распределения – среднее арифметическое случайных величин, распределенных по любым законам в условиях теоремы при больших  $n$  имеет распределение близкое к нормальному. Если изменить условия на  $X_i$ , то предельные распределения могут и не быть нормальными.

## 8 Элементарная теория процессов.

### 8.1 Понятие о случайной функции.

Пусть имеется вещественная переменная  $t$ , которая принимает свои значения на некотором множестве  $T$ . Если каждому значению  $t$  из  $T$  соответствует случайная



величина  $X$ , то будем говорить, что мы имеем случайную функцию вещественной переменной.

Эту функцию будем обозначать  $X(t)$ .

**Пример 8.1** Игральная кость бросается  $n$  раз.  $X(K)$  – число очков, выпадающее на кости при  $K$ -ом бросании. Очевидно, что  $T$  – множество натуральных чисел  $\{1, 2, \dots, n\}$ . При каждом  $K$  случайная величина  $X$  дискретна и принимает целые значения от 1 до 6 с равными вероятностями.

Очевидно также, что случайные величины получающиеся при различных бросаниях  $X(K_1)$  и  $X(K_2)$  ( $K_1 \neq K_2$ ) – независимы. Область определения этой функции дискретна.

**Определение 8.1** Случайные функции с дискретной областью определения называются случайными последовательностями.

**Пример 8.2** Пусть  $X(t)$  – количество сигналов, поступивших на телефонную станцию за время  $t$ . При этом предполагается, что моменты поступления сигналов случайны, независимы и вероятность поступления одного сигнала на малом интервале пропорциональна длине интервала.

Очевидно, что функция  $X(t)$  принимает целые значения от 1 до  $\infty$ , то есть множество её значений дискретно, тогда как область определения  $T = [0, \infty)$  – непрерывна.

**Определение 8.2** Случайные функции, аргументом которых является время, как это было в последнем примере, встречаются довольно часто и обычно называются случайными процессами.

**Пример 8.3** Несколько раз в день измеряется температура воздуха. Пусть  $T = \{t_1, t_2, \dots, t_n\}$  – множество моментов времени, в которые производят измерения и  $X(t_k)$  – непрерывная случайная величина, равная зарегистрированной температуре.

Полученный случайный процесс  $X(t)$  можно назвать процессом с непрерывными состояниями и дискретным временем.

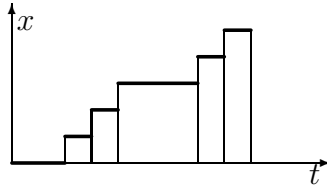
**Пример 8.4** Пусть  $\Theta(t)$  – угол крена корабля. Очевидно, что это случайный процесс с непрерывным временем и непрерывными состояниями.

**Определение 8.3** Сечением случайной функции будем называть случайную величину, которая получается при некотором фиксированном значении  $t$ . Если же, наоборот, изменять  $t$  и для каждого  $t$  брать полученное при опыте значение случайной величины, то получится неслучайная функция  $x(t)$ , которую мы будем называть реализацией случайного процесса.

Например, для примера 8.1 реализацию можно представить в виде таблицы:

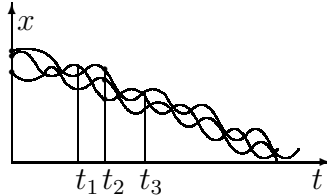
$K$	1	2	3	4	5
$X(K)$	5	1	2	1	6

Для примера 8.2 эту реализацию удобно изобразить на графике:



и так далее.

Иногда случайный процесс изображают в виде набора реализаций.



Здесь каждая кривая является реализацией случайного процесса, а точки пересечения этих кривых с прямыми вида  $t = t_m$  дают значения случайной величины в её сечении.

## 8.2 Характеристики случайных функций.

Очевидно, что одной из наиболее важных характеристик случайной функции является распределение случайной величины, получаемой в сечении функции.

*Вернемся к примерам первого параграфа.*

В примере 8.1 случайная величина  $X(K)$  – дискретна и распределение её вероятностей описывается таблицей:

$X(K)$	1	2	3	4	5	6
$P$	1/6	1/6	1/6	1/6	1/6	1/6

Полезно отметить, что вероятности значений случайной величины  $X(K)$  не зависят от сечения, то есть от значения аргумента  $K$ .

*Второй пример 8.2* более интересен. Зафиксируем некоторый момент  $\tau$  и выведем закон распределения вероятностей случайной величины  $X(\tau)$ . Обозначим через  $P(K, \tau)$  – вероятность того, что за время  $\tau$  поступит  $K$  сигналов. По условию  $P(1, \Delta\tau) = a \cdot \Delta\tau$ , где  $a$  – данный коэффициент пропорциональности.

Будем считать, что  $\Delta\tau$  настолько мал, что

$$P(0, \Delta\tau) + P(1, \Delta\tau) = 1$$

(Это равенство означает, что за время  $\Delta\tau$  поступает не более одного сигнала). Тогда

$$P(0, \Delta\tau) = 1 - a\Delta\tau \quad (8.1)$$

Вычислим  $P(0, \tau)$  – вероятность того, что за время  $\tau$  не поступит ни одного сигнала. Так как событие "не поступило ни одного сигнала за промежуток времени от 0

до  $\tau + \Delta\tau$  является произведением событий "не поступило ни одного сигнала за промежуток от 0 до  $\tau$ " и "не поступило ни одного сигнала за промежуток от  $\tau$  до  $\tau + \Delta\tau$ " и два последние независимы, то

$$P(0, \tau + \Delta\tau) = P(0, \tau)P(0, \Delta\tau).$$

Используя (8.1), получаем

$$P(0, \tau + \Delta\tau) = P(0, \tau)(1 - a\Delta\tau),$$

откуда

$$P(0, \tau + \Delta\tau) - P(0, \tau) = -aP(0, \tau)\Delta\tau.$$

Деля обе части последнего равенства на  $\Delta\tau$  и переходя к пределу при  $\Delta\tau \rightarrow 0$ , получим

$$P'_\tau(0, \tau) = -aP(0, \tau).$$

Интегрируя это уравнение, получим искомую вероятность  $P(0, \tau) = Ce^{-a\tau}$ .

Так как в начальный момент сигналов не было,  $P(0, 0) = 1$  и отсюда  $C = 1$ . Окончательно  $P(0, \tau) = e^{-a\tau}$ . Теперь найдем  $P(1, \tau)$  – вероятность того, что за время  $\tau$  поступит 1 сигнал.

Для этого опять рассмотрим сначала промежуток времени от 0 до  $\tau + \Delta\tau$ . Очевидно, что событие "поступил 1 сигнал за время от 0 до  $\tau + \Delta\tau$ " можно представить в виде суммы произведений  $A_{1\tau+\Delta\tau} = A_{1\tau}A_{0\Delta\tau} + A_{0\tau}A_{1\Delta\tau}$ , где  $A_{it}$  означает событие, состоящее в том, что за время  $t$  поступило  $i$  сигналов. Так как события  $A_{0\tau}$  и  $A_{1\tau}$ , а также события  $A_{0\Delta\tau}$  и  $A_{1\Delta\tau}$  попарно несовместны, а события  $A_{i\tau}$  и  $A_{i\Delta\tau}$  попарно независимы при  $i = 0, 1$  будем иметь

$$\begin{aligned} P(1, \tau + \Delta\tau) &= P(A_{1\tau}A_{0\Delta\tau} + A_{0\tau}A_{1\Delta\tau}) = P(A_{1\tau})P(A_{0\Delta\tau}) + \\ &+ P(A_{0\tau})P(A_{1\Delta\tau}) = P(1, \tau)P(0, \Delta\tau) + P(0, \tau)P(1, \Delta\tau) = \\ &= P(1, \tau)(1 - a\Delta\tau) + e^{-a\tau}a\Delta\tau. \end{aligned}$$

Отсюда

$$P'_\tau(1, \tau) = -aP(1, \tau) + ae^{-a\tau}, \quad \text{где } P(1, 0) = 0.$$

Решая это дифференциальное уравнение, получим  $P(1, \tau) = (a\tau) \cdot e^{-a\tau}$ .

Аналогично можно получить

$$P'(K, \tau) + aP(K, \tau) = aP(K - 1, \tau),$$

откуда по индукции имеем  $P(K, \tau) = \frac{(a\tau)^K}{K!}e^{-a\tau}$ .

Таким образом, мы имеем в каждом сечении случайную величину, распределенную по закону Пуассона. При этом вероятности значений этой величины существенно зависят от значения  $t$ .

*Вернемся к общему случаю.*

Если случайная величина в сечении случайного процесса дискретна, то задается распределение вероятностей её значений, как это было в рассмотренных примерах. Если же эта величина непрерывна, то задается функция её распределения  $F(x, t)$

или плотность распределения  $f(x, t) = F'_x(x, t)$ . Такой закон распределения называется одномерным.

Итак, одной из характеристик случайной функции является ее *одномерный* закон распределения, то есть закон распределения вероятностей значений этой функции в каждый момент времени. Для более полной характеристики вводится *многомерный* закон.

Рассмотрим несколько сечений случайной функции  $X(t_1), \dots, X(t_n)$ . Совокупность этих сечений можно рассматривать, как случайный вектор. Распределение такого вектора называют конечномерным распределением случайной функции.

Например, может быть задана двумерная функция распределения случайной функции  $F(x_1, x_2, t_1, t_2)$ , трехмерная функция и так далее. Но эти характеристики на практике далеко не всегда известны и, даже если известны, то очень трудоемки для обработки. Поэтому чаще используются не законы распределения, а их числовые характеристики.

### 8.3 Числовые характеристики случайных функций.

Пусть  $X(t)$  - случайная функция. Фиксируя  $t$ , мы получим случайную величину. Можно говорить о вычислении моментов этой случайной величины, например, математического ожидания и дисперсии.

**Определение 8.4** Математическим ожиданием случайной функции  $X(t)$  называется неслучайная функция, которая каждому значению аргумента  $t$  сопоставляет математическое ожидание случайной величины, образующейся в сечении случайной функции при данном  $t$ :  $m_x(t) = E(X(t))$ .

**Определение 8.5** Дисперсией случайной функции  $X(t)$  называется неслучайная функция, которая каждому значению аргумента  $t$  сопоставляет дисперсию случайной величины, образующейся в сечении случайной функции при данном  $t$ .

**Пример 8.5** Пусть случайная функция имеет вид  $X(t) = vt + b$ , где  $v$  - случайная величина распределенная по нормальному закону с параметрами  $m_v, \sigma_v$ ,  $b$  - не случайная величина. Найти одномерную плотность распределения  $f(x, t)$  случайной функции и ее характеристики  $m_x(t)$ ,  $D_x(t)$ .

Решение. При фиксированном  $t$  случайная величина  $vt + b$  является линейной функцией случайного аргумента  $v$  распределенного по нормальному закону и, следовательно, она сама будет нормально распределенной с математическим ожиданием

$$m_x(t) = E(vt + b) = tE(v) + b = m_v t + b,$$

и дисперсией

$$D_x(t) = D(vt + b) = t^2 D(v) = \sigma_v^2 t^2.$$

Плотность распределения будет иметь вид

$$f(x, t) = \frac{1}{\sigma_v |t| \sqrt{2\pi}} e^{-(x - (m_v t + b))^2 / (2t^2 \sigma_v^2)}.$$

**Пример 8.6** Рассмотрим пример 8.2 из подразд. 8.1 (см. также подразд. 8.2) и найдем параметры этого случайного процесса.

Решение. Как показано в подразд. 8.2, в этом примере закон распределения сечения  $X(t)$  есть закон Пуассона с параметром  $\lambda t$ . Значит  $m_x(t) = D_x(t) = \lambda t$ .

Очевидно, что, если случайная функция  $X(t)$  в каждом сечении представляет собой случайную величину с одномерной плотностью распределения  $f(x, t)$ , то

$$m_x(t) = \int_{-\infty}^{\infty} x f(x, t) dx \text{ и } D_x(t) = \int_{-\infty}^{\infty} (x - m_x(t))^2 f(x, t) dx.$$

**Определение 8.6** Корреляционной функцией случайной функции  $X(t)$  называется неслучайная функция двух аргументов  $K_x(t_1, t_2)$ , которая при каждой паре значений аргументов  $t_1, t_2$  равна корреляционному моменту соответствующих сечений случайной функции

$$K_x(t_1, t_2) = E[(X(t_1) - m_x(t_1))(X(t_2) - m_x(t_2))].$$

При  $t_1 = t_2 = t$  корреляционная функция превращается в дисперсию случайной функции  $K_x(t, t) = D_x(t) = \sigma_x^2(t)$ .

Наряду с корреляционной функцией рассматривают нормированную корреляционную функцию, то есть коэффициент корреляции сечений  $X(t_1)$  и  $X(t_2)$

$$r_x(t_1, t_2) = \frac{K_x(t_1, t_2)}{\sigma_x(t_1)\sigma_x(t_2)}.$$

Найдем корреляционные функции в примерах, рассмотренных выше.

1. В примере 8.5

$$\begin{aligned} K_x(t_1, t_2) &= E[((vt_1 + b) - (m_v t_1 + b))((vt_2 + b) - (m_v t_2 + b))] = \\ &= E((v - m_v)^2) t_1 t_2 = D(v) t_1 t_2. \end{aligned}$$

2. В примере 8.2 имеем

$$K_x(t_1, t_2) = E[(X(t_1) - m_x(t_1))(X(t_2) - m_x(t_2))].$$

Допустим, что  $t_2 > t_1$ . Тогда  $X(t_2)$  - число сигналов на промежутке от 0 до  $t_2$  будет равно  $X(t_1) + Y(t_2 - t_1)$  - сумма числа событий на промежутке от 0 до  $t_1$  и числа событий на промежутке от  $t_1$  до  $t_2$ . Случайная функция  $Y(t_2 - t_1)$ , равная числу событий на промежутке от  $t_1$  до  $t_2$ , по условию задачи имеет то же распределение, что и  $X(t)$ , случайные функции  $X(t_1)$  и  $Y(t_2 - t_1)$  - некоррелированы. Поэтому

$$\begin{aligned} K_x(t_1, t_2) &= E[(\dot{X}(t_1)(\dot{X}(t_1) + \dot{Y}(t_2 - t_1)))] = \\ &= E[(\dot{X}(t_1))^2] + E[\dot{X}(t_1) \cdot \dot{Y}(t_2 - t_1)] = D_x(t_1) = \lambda t_1, \end{aligned}$$

где  $\dot{X}(t) = X(t) - m_x(t)$  и  $\dot{Y}(t) = Y(t) - m_y(t)$  - центрированные случайные величины, соответствующие данным.

Так как  $K_x(t_1, t_2) = K_x(t_2, t_1)$ , то при  $t_1 > t_2$  получим  $K_x(t_1, t_2) = \lambda t_2$ . Таким образом, для пуассоновского процесса

$$K_x(t_1, t_2) = \lambda t, \text{ где } t = \min(t_1, t_2).$$

## 8.4 Важнейшие классы случайных процессов

I. Случайная функция  $X(t)$  называется гауссовской, если все ее конечномерные распределения нормальны.

Примером такой функции является винеровский процесс.

II. Случайный процесс  $X(t)$  называется процессом с независимыми приращениями, если для  $t_0 \leq t_1 \leq \dots \leq t_n$ , где  $t_k \in T$  случайные величины  $X(t_1) - X(t_0)$ ,  $X(t_2) - X(t_1)$ , ...,  $X(t_n) - X(t_{n-1})$  независимы.

Например, рассмотренный выше пуассоновский процесс является процессом с независимыми приращениями, так как разность  $X(t_k) - X(t_{k-1})$  равна числу сигналов, полученных за время от  $t_{k-1}$  до  $t_k$ , а это число не зависит от числа сигналов, полученных до момента  $t_{k-1}$ , а зависит только от величины промежутка  $(t_{k-1}, t_k)$ .

III. Случайный процесс  $X(t)$  - называется стационарным, если для любого  $h$  конечномерные распределения случайной функции  $X(t)$ , не меняются при сдвиге на  $h$ , то есть для плотностей распределения

$$f(x, t) = f(x, t + h), \quad f(x_1, x_2, t_1, t_2) = f(x_1, x_2, t_1 + h, t_2 + h) \text{ и т. д.}$$

III-а. Случайный процесс называется стационарным в широком смысле, если у него существуют первые и вторые моменты и они не меняются при сдвиге

$$E(X(t + h)) = E(X(t)) \text{ и } K_x(t_1 + h, t_2 + h) = K_x(t_1, t_2).$$

Это означает, что  $E[X(t)] = E(X(0)) = \text{const}$ , а  $K_x(t_1, t_2) = K_x(0, t_2 - t_1) = K_x(t_2 - t_1)$  - зависит только от разности  $t_2 - t_1$ , то есть является функцией одного аргумента. Если стационарный процесс имеет первые два момента, то он является стационарным в широком смысле.

Мы будем в дальнейшем иметь дело только со стационарными в широком смысле процессами и, поэтому, будем называть их просто стационарными.

Корреляционную функцию такого процесса будем обозначать  $K(\tau)$ , где  $\tau = t_2 - t_1$ . Так как  $K_x(t_1, t_2) = K_x(t_2, t_1)$ , то  $K(\tau) = K(-\tau)$ , а дисперсия  $D_x = K_x(t, t) = K(0)$  - постоянна.

## 8.5 Линейные операторы.

Напомним, что мы говорим, что в некотором классе функций задан оператор, если каждой функции этого класса соответствует (единственным образом) функция того же (или другого) класса. Например, каждой дифференцируемой на некотором промежутке функции можно сопоставить ее производную:

$$x(t) \rightarrow x'(t).$$

Оператор будем обозначать заглавными буквами:  $y(t) = A(x(t))$ . Оператор  $L$  будем называть линейным однородным, если

- 1)  $L(x_1 + x_2) = L(x_1) + L(x_2)$
- 2)  $L(kx) = kL(x)$ .

Примерами таких преобразований являются операторы дифференцирования и интегрирования:

$$y(t) = \left( a_n \frac{d^n}{dx^n} + a_{n-1} \frac{d^{n-1}}{dx^{n-1}} + \dots + a_0 \right) x(t), \quad y(t) = \int_0^t x(\tau) d\tau.$$

На практике часто встречаются с подобными преобразованиями случайных процессов, поэтому определим понятия производной и интеграла от случайной функции.

**Определение 8.7** Будем говорить, что случайная функция  $X(t)$  сходится в среднеквадратичном к случайной величине  $A$  при  $t \rightarrow t_0$ , если  $E(X - A)^2 \rightarrow 0$  при  $t \rightarrow t_0$ . Этот факт будем записывать в виде  $A \stackrel{cp.кв.}{=} \lim_{t \rightarrow t_0} X(t)$ .

**Определение 8.8** Случайную величину  $X'(t)$  будем называть производной от случайной величины  $X(t)$  в точке  $t$ , если

$$X'(t) \stackrel{cp.кв.}{=} \lim_{\Delta t \rightarrow 0} \frac{X(t + \Delta t) - X(t)}{\Delta t}.$$

**Определение 8.9** Интегралом от случайной функции  $X(t)$  будем называть предел интегральных сумм

$$\int_0^t X(t) dt = \lim \sum_i X(t_i) \Delta t_i, \quad \text{при условии } \max_i |\Delta t_i| \rightarrow 0,$$

где интегральная сумма составляется по правилам, принятым в анализе.

Таким образом, можно рассматривать линейные операторы

$$Y_1(t) = \int_0^t X(\tau) d\tau \quad \text{и} \quad Y_2(t) = \frac{dX(t)}{dt}.$$

Без доказательства отметим, что

$$E \left( \int_0^t X(\tau) d\tau \right) = \int_0^t m_x(\tau) d\tau, \quad E \left( \frac{dX(t)}{dt} \right) = \frac{dm_x(t)}{dt};$$

$$K_{Y_1}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} K_x(\tau_1, \tau_2) d\tau_1 d\tau_2, \quad K_{Y_2}(t_1, t_2) = \frac{\partial^2 K_x(t_1, t_2)}{\partial t_1 \partial t_2}.$$

В практических приложениях операторы называемые функциями времени обычно называют динамическими системами. Тогда в записи  $Y = LX$  оператор  $L$  будем называть динамической системой,  $X$  - сигналом, поступающим на ее вход, а  $Y$  - выходом системы.

**Пример 8.7** Пусть  $Y = dX(t)/dt$ . На вход этой системы поступает стационарный случайный процесс  $X(t)$  с математическим ожиданием  $m_x(t) = \sin t$  и корреляционной функцией  $K_x(\tau) = De^{-\alpha\tau^2}$ , где  $\tau = t_2 - t_1$ . Найти математическое ожидание и дисперсию на выходе системы.

Решение.  $m_y(t) = \sin' t = \cos t$ ,

$$K_y(t_1, t_2) = \frac{\partial^2 De^{-\alpha(t_2-t_1)^2}}{\partial t_1 \partial t_2} = 2D\alpha e^{-\alpha(t_2-t_1)^2} \cdot [1 - 2\alpha(t_2 - t_1)^2].$$

При  $t_2 = t_1$  получим  $D_Y(t) = K_Y(t, t) = 2D\alpha$ .

Если случайная функция представима в виде

$$X(t) = m_x(t) + \sum_{i=1}^n V_i \varphi_i(t), \quad (8.2)$$

где  $m_x(t)$  – математическое ожидание  $X(t)$ ,  $\varphi_1(t), \dots, \varphi_n(t)$  – неслучайные функции,  $V_1, \dots, V_n$  – некоррелированные случайные величины с нулевыми математическими ожиданиями, то её корреляционная функция находится совсем просто: так как  $K_{ij} = 0$ , получим

$$\begin{aligned} K_x(t_1, t_2) &= E(\overset{\circ}{X}(t_1) \cdot \overset{\circ}{X}(t_2)) = E\left(\sum_{i,j} V_i V_j \varphi_i(t_1) \varphi_j(t_2)\right) = \\ &= \sum_{i,j} \varphi_i(t_1) \varphi_j(t_2) E(V_i V_j) = \sum_{i=1}^n \varphi_i(t_1) \varphi_i(t_2) D(V_i). \end{aligned}$$

Если функция вида (8.2) поступает на вход линейной динамической системы, то

$$Y(t) = L(X(t)) = L(m_x(t)) + \sum_{i=1}^K V_i L(\varphi_i(t)),$$

то есть на выходе имеем функцию того же вида, что и  $X(t)$  с математическим ожиданием  $m_y(t) = L(m_x(t))$ .

Представление типа (8.2) называется каноническим разложением  $X(t)$ .

## 8.6 Спектральное разложение стационарной случайной функции.

Если стационарная случайная функция  $X(t)$ , заданная на конечном промежутке  $[0, T]$ , имеет каноническое разложение вида

$$X(t) = m_x(t) + \sum_{K=1}^{\infty} (U_K \cos \omega_K t + V_K \sin \omega_K t),$$



где  $U_K, V_K$  – некоррелированные случайные величины с нулевым математическим ожиданием, причем  $D(U_K) = D(V_K) = D_K$ , то её корреляционная функция может быть представлена в виде ряда Фурье

$$K_x(\tau) = \sum_{K=0}^{\infty} D_K \cos \omega_K \tau, \quad \omega_K = K\pi/T.$$

Аналогично, если стационарная функция  $X(t)$  задана на бесконечном промежутке, то рассматривают преобразование Фурье её корреляционной функции, которое называют спектральной плотностью стационарного случайного процесса.

**Определение 8.10** *Спектральной плотностью стационарного случайного процесса называется преобразование Фурье её корреляционной функции.*

Обозначается спектральная плотность через  $S_x(\omega)$ . Тогда

$$S_x(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K_x(\tau) e^{-i\omega\tau} d\tau \text{ и } K_x(\tau) = \int_{-\infty}^{\infty} S_x(\omega) e^{i\omega\tau} d\omega.$$

Используя свойства преобразования Фурье, получим, что спектральная плотность производной  $X'(t)$  связана с  $S_x(\omega)$  формулой

$$S'_x(\omega) = \omega^2 S_x(\omega).$$

Пользуясь этим и другими свойствами этого преобразования, можно находить характеристики на выходе динамической системы, зная характеристики входного процесса.

**Пример 8.8** *Пусть стационарные случайные процессы  $X(t)$  и  $Y(t)$  связаны соотношением*

$$a_0 X' + a_1 X = Y,$$

*причем известна корреляционная функция процесса  $Y(t)$ . Найти корреляционную функцию  $X(t)$ .*

*Можно доказать, что в условиях задачи*

$$a_1^2 K_{x'}(\tau) + a_0^2 K_x(\tau) = K_y(\tau).$$

*Тогда для спектральных плотностей будет выполнено*

$$a_1^2 \omega^2 S_x(\omega) + a_0^2 S_x(\omega) = S_y(\omega), \quad \text{откуда } S_x(\omega) = \frac{S_y(\omega)}{a_1^2 \omega^2 + a_0^2}.$$

*Пользуясь обратным преобразованием Фурье, можно найти корреляционную функцию  $X(t)$*

$$K_x(\tau) = \int_{-\infty}^{\infty} \frac{S_y(\omega)}{a_1^2 \omega^2 + a_0^2} e^{i\omega\tau} d\omega.$$

## 9 Введение

Задачи математической статистики в определенном смысле обратны задачам теории вероятностей. Теория вероятностей на основе вероятностной модели позволяет предсказывать (прогнозировать) различные характеристики случайных явлений (получать их вероятностное описание), в то время как математическая статистика на основе наблюдения случайных явлений строит или уточняет их вероятностную модель. В практических задачах методы теории вероятностей и математической статистики тесно взаимодействуют и дополняют друг друга.

В математической статистике вероятностная модель строится на базе экспериментальных данных, то есть на наборе наблюдений реализаций случайных событий, случайных величин, векторов или процессов, а также на некоторых первоначальных предположениях (обычно достаточно общего характера) о виде этой модели.

Производится сбор экспериментальных данных, которые рассматриваются как реализация случайного вектора  $X^{(n)} = (X_1, \dots, X_n)$ , а величины  $X_i$  обычно рассматриваются как независимые реализации изучаемой случайной величины  $X$ . На основе этих данных с помощью методов математической статистики изучаются распределения случайной величины  $X$ : оценивается функция или плотность распределения, различные числовые характеристики или параметры, проверяются различные предположения (гипотезы) об этих характеристиках.

Принципиальная возможность решения таких задач основана на предельных теоремах теории вероятностей. Например, при росте числа наблюдений в предположении их независимости эмпирическая частота события (то есть доля наблюдений, в которых случайное событие реализуется) стремится к его вероятности, а выборочное среднее (среднее арифметическое значений случайных величин или векторов) стремится к математическому ожиданию. Разработка и исследование математических методов и алгоритмов (статистических процедур) решения подобных задач составляют предмет математической статистики.

## 10 Статистические модели

### 10.1 Вероятностные и статистические модели

Статистическая модель – это описание информации, на основе которой решается та или иная задача математической статистики. Статистическую модель составляют две части:

1) *экспериментальная (опытная) информация* – данные, полученные в результате проведенного эксперимента (опыта). Эти данные представляют собой элемент  $X$  некоторого множества  $\mathcal{X}$  и рассматриваются как реализация случайного эксперимента. Множество  $\mathcal{X}$  может быть произвольной природы;

2) *априорная (имеющаяся до проведения эксперимента) информация* – это информация о характере "случайности" эксперимента.

Приведем общие определения, которые поясним более подробно далее.

В теории вероятностей для описания случайности эксперимента (вероятностной модели) используется понятие вероятностного пространства. Напомним, что *вероятностное пространство* – это тройка объектов  $(\mathcal{X}, \mathcal{A}, P)$ . Здесь  $\mathcal{X} = \{X\}$  – это множество возможных результатов случайного эксперимента (оно обычно называется множеством элементарных исходов). Множество  $\mathcal{A}$  представляет собой  $\sigma$ -алгебру подмножеств множества элементарных исходов. Пару объектов  $(\mathcal{X}, \mathcal{A})$  называют *измеримым пространством*. Наконец,  $P$  есть вероятностная мера (распределение вероятностей, закон распределения вероятностей), то есть числовая функция  $P(A)$ , заданная на множестве  $\mathcal{A}$  со значениями в интервале  $[0, 1]$ , обладающая свойством счетной аддитивности и нормированности:  $P(\mathcal{X}) = 1$ .

В статистической модели информация о случайности эксперимента также характеризуется измеримым пространством  $(\mathcal{X}, \mathcal{A})$ . Однако в отличие от вероятностной модели вероятностная мера (распределение вероятностей) неизвестна. Вместо этого задано некоторое множество  $\mathcal{P} = \{P\}$  вероятностных мер, включающее в себя неизвестную меру  $P$ .

Обычно результаты эксперимента есть набор числовых данных:  $X = X^{(n)} = (X_1, \dots, X_n)^\top$ ,  $X_i \in R^1$ , где  $R^1$  – множество вещественных чисел. При этом  $X^{(n)}$  можно рассматривать как  $n$ -мерный вектор, а множество  $\mathcal{X} = \mathcal{X}_n$  – как некоторое подмножество  $n$ -мерного векторного пространства  $R^n$ .

Часто распределение вероятностей  $P = P_\theta$  задано с точностью до некоторого параметра  $\theta \in \Theta \subset L$ , где  $\Theta = \{\theta\}$  – допустимое множество параметров, причем различным значениям параметра  $\theta$  соответствуют различные распределения вероятностей  $P_\theta$ . Параметр  $\theta$  может быть числом:  $\theta \in R^1 = L$ , вектором:  $\theta = (\theta_1, \dots, \theta_m)^\top \in R^m = L$  или функцией:  $\theta = \theta(t)$ ,  $t \in T$ ;  $L$  – некоторое множество функций на множестве  $T$ . В этом случае статистический эксперимент  $(R^n, \mathcal{B}^n, \mathcal{P})$  полностью характеризуется множеством функций распределения

$$\{F_\theta^{(n)}(t_1, \dots, t_n) = P_\theta(X_1 < t_1, \dots, X_n < t_n), \theta \in \Theta\}.$$

В статистике важно измерять близость (расстояние) между различными значениями параметров, а также между значениями параметров и их различными оценками. Если параметр  $\theta$  – число, близость значений  $\theta$  и  $\theta_0$  измеряют абсолютным отклонением  $|\theta - \theta_0|$ .

В случае векторных и функциональных параметров нужно задать меру близости. Например, если  $\theta = (\theta_1, \dots, \theta_k)$  –  $k$ -мерный вектор, то обычно рассматривается евклидова длина вектора  $\|\theta\|^2 = \sum_{i=1}^k \theta_i^2$ .

Тройка  $(\mathcal{X}, \mathcal{A}, \mathcal{P} = \{P_\theta, \theta \in \Theta\})$  называется *статистической моделью* (часто ее называют также *статистическим экспериментом* или *статистической структурой*).

Другими словами, *статистическая модель* – это описание законов распределения  $P_\theta$  случайных экспериментальных данных, заданное с точностью до неизвестного параметра  $\theta \in \Theta$ , причем информацию о параметре  $\theta$  нужно получить на основе наблюдаемых результатов случайного эксперимента  $x \in \mathcal{X}$ .

Пусть  $\Theta \subset R^k$  есть  $k$ -мерное множество. В этом случае статистическая модель называется *конечномерной* (употребляют также термин *параметрическая статистическая модель*). Если  $(\mathcal{X}, \mathcal{A}) = (R^n, \mathcal{B}^n)$ , то параметрическая статистическая модель полностью характеризуется параметрическим множеством функций распределения  $\{F_\theta^{(n)}(t_1, \dots, t_n), \theta \in \Theta\}$ . Если  $\theta \in \Theta$  есть функциональный параметр и  $\Theta$  есть бесконечномерное функциональное множество, то говорят о *бесконечномерной статистической модели* (употребляют также термин *непараметрическая статистическая модель*).

Различные статистические процедуры должны быть выражены через некоторые функции от наблюдаемых данных. Чтобы можно было изучать вероятностные характеристики статистических процедур, эти функции должны быть измеримыми<sup>1</sup>. Измеримая функция  $(X)$ ,  $X \in \mathcal{X}$ , называется *статистикой*. Аналогично определяется *векторная статистика* (со значениями в  $R^n$ ) и *статистика со значениями в  $\Theta$* . В дальнейшем мы не будем оговаривать требование измеримости, предполагая все рассматриваемые функции измеримыми.

С точки зрения теории вероятностей статистика – это случайная величина; для нее определены вероятности  $P_T(B) = P(T(x) \in B)$ , в частности функция распределения  $F_T(t) = P(T < t)$ , где  $P$  – заданная вероятностная мера, а также различные числовые характеристики, например математическое ожидание  $E(T)$ . С точки зрения математической статистики эти вероятности и числовые характеристики есть функции от неизвестного параметра  $\theta \in \Theta$ :  $F_{T,\theta}(t) = P_\theta(T < t)$ ,  $E_\theta(T)$ , поэтому их значения неизвестны.

Точность статистических результатов возрастает с ростом объема наблюдаемых данных, и достаточно хорошие результаты можно получить при достаточно большом объеме экспериментальных данных. Поэтому в математической статистике широко распространен *асимптотический подход*. В рамках этого подхода конкретная статистическая модель рассматривается как элемент *последовательности статистических моделей*  $(\mathcal{X}_n, \mathcal{A}_n, P_n = \{P_\theta, \theta \in \Theta\})$ ,  $n \rightarrow \infty$ . Увеличение  $n$  соответствует росту объема информации о наблюдаемых данных. Множество  $\Theta$  неизвестных значений параметра обычно не зависит от  $n$ .

При асимптотическом подходе изучаются предельные свойства различных статистических процедур при  $n \rightarrow \infty$ . Математическая статистика имеет дело с различными статистическими моделями. Мы остановимся на одной из самых простых и стандартных моделей.

## 10.2 Модель независимой однородной выборки с неизвестной функцией распределения

Важнейшим случаем статистической модели является модель *независимой однородной выборки*. В этом случае результат эксперимента рассматривается как реализация  $n$  независимых значений  $X^{(n)} = (X_1, \dots, X_n)$ , случайной величины  $X$  с

<sup>1</sup>Напомним, что числовая функция  $(X)$  от наблюдаемых данных  $X \in \mathcal{X}$  называется *измеримой*, если прообразы всех подмножеств  $B(t) = (-\infty, t)$ ;  $t \in R^1$  измеримы:  $A(t) = \{X : (X) \in B(t)\} \in \mathcal{A}$ ; в этом случае и прообразы всех борелевских множеств измеримы.

общей функцией распределения

$$F_{X_i}(t) = F_X(t) = F(t), \quad F(t) = F \in \mathcal{F}, \quad t \in R^1,$$

где  $\mathcal{F}$  есть некоторое множество функций распределения.

Случайную величину  $X$  иногда называют *генеральной совокупностью*, а набор измерений  $X^{(n)} = (X_1, \dots, X_n)$  – *выборкой из генеральной совокупности  $X$  с неизвестной функцией распределения  $F \in \mathcal{F}$* . Набор данных можно рассматривать как реализацию случайного вектора  $X^{(n)}$  с функцией распределения

$$F^{(n)}(t_1, \dots, t_n) = F(t_1) \times \dots \times F(t_n), \quad F \in \mathcal{F}, \quad t \in R^k.$$

Множество  $\mathcal{F}$  может быть, например, множеством всех функций распределения, множеством функций распределения, имеющих моменты заданного порядка и т.д.

Если все функции распределения  $F \in \mathcal{F}$  непрерывны и имеют плотности распределения  $f = f(t) \in \mathfrak{F}$ , то в качестве функционального параметра может рассматриваться плотность распределения, а  $\mathfrak{F}$  может быть множеством непрерывных плотностей распределения, множеством плотностей распределения, имеющих ограниченные производные заданного порядка и т.д.

При  $n \rightarrow \infty$  мы имеем *последовательность статистических моделей независимой однородной выборки*.

В рамках модели независимой однородной выборки рассматриваются различные задачи, связанные с получением по выборке той или иной информации о неизвестной функции распределения, а также о различных вероятностных характеристиках  $g_X = g(F_X)$  генеральной совокупности  $X$ , то есть о характеристиках  $g(F)$  неизвестной функции распределения  $F_X = F \in \mathcal{F}$ .

## 10.3 Важнейшие параметрические семейства

Часто закон распределения  $P_\theta$  генеральной совокупности (то есть функция распределения  $F(t)$  или плотность распределения  $f(t)$ ) лежат в заданном конечнопараметрическом семействе:  $F(t) = F(t, \theta)$  или  $f(t) = f(t, \theta)$ ,  $\theta \in \Theta$ . В этом случае мы часто будем использовать обозначения:  $X \sim P_\theta$ ,  $X \sim F(t, \theta)$  или  $X \sim f(t, \theta)$ . Тогда последовательность статистических моделей является конечнопараметрической:

$$\begin{aligned} F^{(n)}(t_1, \dots, t_n; \theta) &= F(t_1, \theta) \times \dots \times F(t_n, \theta); \\ f^{(n)}(t_1, \dots, t_n; \theta) &= f(t_1, \theta) \times \dots \times f(t_n, \theta), \quad \theta \in \Theta. \end{aligned}$$

Приведем наиболее важные примеры.

### 10.3.1 Семейство бернуллиевских распределений

Это семейство  $B_p$ ,  $p \in (0, 1)$  соответствует дискретной случайной величине  $X$  с двумя возможными значениями 0 и 1, причем  $P(X = 1) = p$ ,  $P(X = 0) = 1 - p$ . Если  $A \in \mathcal{A}$  – случайное событие и  $p = P(A)$  – его вероятность, то индикаторная функция  $X = \Pi_A$  события  $A$ , принимающая значение  $X = 1$  в случае успеха – появления события  $A$  или значение  $X = 0$  при неудаче испытания, имеет распределение  $B_p$ :  $X \sim B_p$ .

Для бернуллиевской случайной величины  $E_p(X) = p$ ,  $D_p(X) = p(1 - p)$ .

### 10.3.2 Семейство биномиальных распределений

Это семейство  $B_p^n$ , где  $p \in (0, 1)$ ,  $n \geq 1$  – натуральное число, соответствует дискретной случайной величине  $X$  с возможными значениями  $0, 1, \dots, n$ , причем

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n, \quad p \in (0, 1).$$

Если  $A \in \mathcal{A}$  – случайное событие и  $p = P(A)$  – его вероятность, а  $X$  – число успехов в серии из  $n$  независимых испытаний, в котором регистрируется событие  $A$ , то  $X$  имеет биномиальное распределение  $B_p^n$ . Таким образом, случайная величина  $X \sim B_p^n$  может быть представлена в виде  $X = \sum_{i=1}^n X_i$ , где  $X_i \sim B_p$  и независимы (случайная величина  $X_i$  соответствует успеху или неудаче  $i$ -го испытания).

Для биномиальной случайной величины

$$E_{n,p}(X) = np, \quad D_{n,p}(X) = np(1 - p).$$

### 10.3.3 Семейство показательных распределений

Семейство  $\text{Exp}_\lambda$ ,  $\lambda > 0$ , соответствует непрерывной случайной величине  $X \geq 0$ , функция распределения  $F(t, \lambda)$  и плотность распределения  $f(t, \lambda)$  которой имеют вид:

$$F(t, \lambda) = \begin{cases} 0 & \text{при } t < 0; \\ 1 - e^{-\lambda t} & \text{при } t \geq 0; \end{cases} \quad f(t, \lambda) = \begin{cases} 0 & \text{при } t < 0; \\ \lambda e^{-\lambda t} & \text{при } t \geq 0. \end{cases}$$

Часто показательное распределение удобно характеризовать параметром  $u = 1/\lambda$ ; в этом случае будем использовать обозначение  $\text{Exp}^u$ . Для показательного распределения

$$E_\lambda(X) = 1/\lambda = u, \quad D_\lambda(X) = 1/\lambda^2 = u^2.$$

### 10.3.4 Семейство нормальных (гауссовских) распределений

Нормальные распределения  $N(a, d)$ ,  $a \in \mathbb{R}^1$ ,  $d = \sigma^2 > 0$ , характеризуются плотностью распределения  $f(t; a, d)$  и функцией распределения  $F(t; a, d)$ :

$$f(t; a, d) = \frac{1}{\sqrt{2\pi d}} \exp\left(-\frac{(t - a)^2}{2d}\right), \quad F(t; a, d) = \Phi((t - a)/\sigma),$$

где

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

Функция  $\Phi(t)$  есть *функция распределения стандартного нормального закона*  $N(0, 1)$ . Стандартная нормальная случайная величина  $X \sim N(0, 1)$  симметрична, то есть  $\Phi(-t) = 1 - \Phi(t)$ . Для нормальной случайной величины  $X \sim N(a, d)$  среднее и дисперсия есть  $E_{a,d}(X) = a$ ,  $D_{a,d}(X) = d$ ; отметим также, что

$$E_{a,d}((X - a)^3) = 0; \quad E_{a,d}((X - a)^4) = 3d^2.$$

Если  $X \sim N(a, d)$ ,  $Y = v + bX$ , то  $Y \sim N(v + ba, b^2d)$ ; в частности, при  $X \sim N(0, 1)$  имеем:  $Y = v + bX \sim N(v, b^2)$  и, если  $X \sim N(a, d)$ ,  $d > 0$ , то  $Y = (X - a)/\sqrt{d} \sim N(0, 1)$ .

# 11 Задачи математической статистики

Задачи, которые изучает математическая статистика, состоят в том, чтобы по данным наблюдений  $X \in \mathcal{X}$  на основе статистической модели получить ту или иную информацию о законе распределения  $P = P_\theta$  наблюдаемых данных. Основные статистические задачи – это задачи оценивания и доверительного оценивания параметров и различных характеристик неизвестного распределения, а также задачи проверки различных статистических гипотез о неизвестном распределении.

## 11.1 Задачи оценивания

### 11.1.1 Задача оценивания параметра

Естественно попытаться восстановить (оценить) распределение  $P$ , порождающее экспериментальные данные, по результатам наблюдений. Поскольку это распределение  $P = P_\theta$  полностью характеризуется значением параметра  $\theta \in \Theta$ , эта задача сводится к задаче построения оценок неизвестного параметра распределения.

*Оценкой параметра  $\theta \in \Theta$  называется статистика  $\hat{\theta}$  со значениями в множестве  $\Theta$ .* Таким образом, оценка есть способ (алгоритм) построения некоторых значений  $\hat{\theta}(X)$  для любых конкретных реализаций данных наблюдений  $X \in \mathcal{X}$ .

Задача оценивания параметра состоит в построении оценок  $\hat{\theta}$ , достаточно близких к неизвестному значению параметра  $\theta \in \Theta$ .

Желательно построить наилучшие в том или ином смысле оценки и изучить различные характеристики их качества. При асимптотическом подходе изучаются предельные свойства последовательностей оценок  $\hat{\theta}_n = \hat{\theta}_n(X_n)$ ,  $X_n \in \mathcal{X}_n$ , при  $n \rightarrow \infty$ .

### 11.1.2 Задача оценивания числовых характеристик

В случае многомерного или бесконечномерного параметра  $\theta \in \Theta$  вида  $\theta = (\theta^\Pi, \theta^M)$ ,  $\theta^\Pi \in \Theta^\Pi$ ,  $\theta^M \in \Theta^M$ , часто требуется оценить лишь часть его компонент  $g(\theta) = \theta^\Pi \in \Theta^\Pi$ . Параметр  $\theta^\Pi \in \Theta^\Pi$  обычно называют *полезным*, а параметр  $\theta^M \in \Theta^M$  – *мешающим*. Например, при измерении неизвестного параметра  $a$  со случайными ошибками  $\xi \sim N(0, \sigma)$  при неизвестном  $\sigma > 0$  модель характеризуется параметром  $\theta = (a, \sigma)$ ,  $a \in R^1, \sigma \in R_+^1$ ; здесь  $a = \theta^\Pi$  – полезный, а  $\sigma = \theta^M$  – мешающий параметр.

Во многих задачах требуется оценить лишь некоторый набор числовых характеристик  $g(P)$  распределения данных  $P = P_\theta$  (например среднее значение, дисперсию и т.д.). Такие характеристики можно рассматривать как функции  $g(\theta)$ ,  $\theta \in \Theta$ , неизвестного параметра, принимающие значения в некоторой области  $G$ .

Обе эти задачи связаны с частичным восстановлением вероятностной модели и называются *задачами оценки характеристик распределения*.

*Оценкой характеристики  $g(\theta)$ ,  $\theta \in \Theta$ , называется статистика  $\hat{g} = \hat{g}(X)$  со значениями в множестве  $G$ .*

Задача оценивания характеристики состоит в построении оценок  $\hat{g}$ , достаточно близких к неизвестному значению  $g(\theta)$ . При асимптотическом подходе изучаются

предельные свойства последовательностей оценок  $\hat{g}_n = \hat{g}_n(X_n)$ ,  $X_n \in \mathcal{X}_n$ , при  $n \rightarrow \infty$ .

### 11.1.3 Функция потерь и функция риска

Формально задача оценивания характеристик содержит в себе задачу оценивания параметра при  $g(\theta) = \theta$ . Поэтому основные определения приведем для задачи оценивания характеристик распределения.

Близость оценки  $\hat{g}$  к значению  $g(\theta)$  измеряется случайной величиной  $\|\hat{g} - g(\theta)\|$ , где  $\|\cdot\|$  – выбранная норма (длина вектора) в пространстве параметров. Чтобы количественно измерять качество оценок, вводятся понятия *функции потерь* и *функции риска оценки*. Функция потерь  $l(\hat{g}, g(\theta))$  – это неотрицательная функция, характеризующая потери при использовании оценки  $\hat{g}$  для случая, когда истинное значение параметра есть  $\theta$ . Функция риска оценки есть математическое ожидание потерь:  $R(\hat{g}, \theta) = E_\theta(l(\hat{g}, g(\theta)))$ .

Функция потерь обычно имеет вид  $l(\hat{g}, g(\theta)) = w(\|\hat{g} - g(\theta)\|)$ , где  $w(t)$  – неубывающая числовая функция неотрицательного аргумента  $t$ ,  $w(0) = 0$ . Функция потерь  $l(\hat{g}, g(\theta))$  оценки  $\hat{g}$  при заданном  $\theta$  есть случайная величина, а риск  $R(\hat{g}, \theta)$  оценки есть функция от параметра  $\theta \in \Theta$  и оценки (способа оценивания)  $\hat{\theta}$ . Наибольший интерес обычно представляют два типа функций риска, соответствующие двум видам функций потерь.

**1. Вероятности отклонений.** Пусть функция потерь  $w(t) = \mathbb{I}_\delta(t)$  есть функция, равная 0 при  $|t| < \delta$  и равная 1 при  $|t| \geq \delta$ . Тогда функция риска есть вероятность получить отклонение оценки  $\hat{g}$  от оцениваемого параметра  $g(\theta)$  не меньше, чем  $\delta$ :

$$R^\delta(\hat{g}, \theta) = E_\theta w(\|\hat{g} - g(\theta)\|) = P_\theta(\|\hat{g} - g(\theta)\| \geq \delta).$$

При асимптотическом подходе последовательность оценок  $\hat{g}_n = \hat{g}_n(X_n)$ ,  $X_n \in \mathcal{X}_n$  называется *состоятельной*, если для любого  $\delta > 0$

$$R^\delta(\hat{g}_n, \theta) = P_{n,\theta}(\|\hat{g}_n - g(\theta)\| \geq \delta) \rightarrow 0 \text{ при } n \rightarrow \infty.$$

**2. Квадратичный риск** соответствует квадратичной функции потерь  $w(t) = t^2$ ,  $l(\hat{g}, g(\theta)) = \|\hat{g} - g(\theta)\|^2$  и имеет вид

$$R_2(\hat{g}, \theta) = E_\theta(\|\hat{g} - g(\theta)\|^2).$$

**Теорема 11.1** *Справедливо неравенство, связывающее квадратичный риск и вероятности отклонений:*

$$R^\delta(\hat{g}, \theta) \leq R_2(\hat{g}, \theta)/\delta^2.$$

Доказательство.

Используя неравенство Маркова  $P(|X| \geq t) \leq E(|X|)/t$ ,  $t > 0$ , получим:

$$P_\theta(\|\hat{g} - g(\theta)\| \geq \delta) = P_\theta(\|\hat{g} - g(\theta)\|^2 \geq \delta^2) \leq E_\theta(\|\hat{g} - g(\theta)\|^2)/\delta^2.$$

При асимптотическом подходе отсюда следует, что для состоятельности последовательности оценок  $\hat{g}_n$  достаточно выполнения соотношения  $R_2(\hat{g}_n, \theta) \rightarrow 0$  при  $n \rightarrow \infty$ .



Для широкого класса задач статистического оценивания типичное поведение квадратичного риска для достаточно хороших оценок – убывание со скоростью  $1/n$ , то есть

$$R_2(\hat{g}_n, \theta) \sim u(\theta)/n,$$

где  $u(\theta)$  – положительная функция. Отсюда следует, что  $R^{\delta_n}(\hat{g}, \theta) \rightarrow 0$  для любой последовательности  $\delta_n \rightarrow 0$  такой, что  $\sqrt{n}\delta_n \rightarrow \infty$ . Это свойство называют  $\sqrt{n}$ -состоятельностью оценок. Оно означает, что типичный порядок отклонения оценки  $\hat{g}_n$  от оцениваемой характеристики  $g(\theta)$  есть  $1/\sqrt{n}$ .

**Замечание 11.1** В конечнопараметрических задачах оценивания обычно хорошие оценки обладают свойствами состоятельности и  $\sqrt{n}$ -состоятельности. Это верно и для ряда бесконечномерных задач оценивания (например при оценивании неизвестной функции распределения  $F = F_X$  независимой однородной выборки). Вместе с тем для широкого класса бесконечномерных задач оценивания (например при оценивании неизвестной плотности распределения  $f = f_X$ ) обычно не существует  $\sqrt{n}$ -состоятельных оценок: порядок точности оценивания зависит от заданного класса  $\mathfrak{F} = \{f\}$  неизвестных плотностей и рассматриваемой меры близости  $\|\hat{f}_n - f\|$ .

#### 11.1.4 Задача доверительного оценивания

При решении задачи оценивания параметров или числовых характеристик распределения мы получаем некоторое случайное значение  $\hat{\theta} \in \Theta$  или  $\hat{g} \in G$  (то есть точку в множестве  $\Theta$  или  $G$ ; такие оценки называют также *точечными*), причем заранее не ясно, сколь сильно отличается оценка  $\hat{\theta}$  или  $\hat{g}$  от оцениваемой величины  $\theta$  или  $g(\theta)$ . Во многих задачах желательно указать не точку, а область  $\tilde{\Theta}$  в множестве  $\Theta$  или область  $\tilde{G}$  в множестве  $G$ , в которой лежит неизвестное значение  $\theta$  или  $g(\theta)$ . Достоверно такую область нельзя указать (кроме, разумеется, всего множества  $\tilde{\Theta} = \Theta$  или  $\tilde{G} = G$ ). Поэтому задача доверительного оценивания ставится следующим образом.

Задается величина  $\gamma \in (0, 1)$ , которую называют *уровнем надежности*. Требуется на основе данных наблюдений  $X \in \mathcal{X}$  построить такую область  $\tilde{\Theta}_\gamma = \tilde{\Theta}_\gamma(X) \subset \Theta$  или  $\tilde{G}_\gamma = \tilde{G}_\gamma(X) \subset G$ , чтобы выполнялось соотношение

$$P_\theta(\theta \in \tilde{\Theta}_\gamma) \geq \gamma \quad \text{или} \quad P_\theta(g(\theta) \in \tilde{G}_\gamma) \geq \gamma \quad \text{для любого } \theta \in \Theta. \quad (11.1)$$

Такая область  $\tilde{\Theta}_\gamma$  (или  $\tilde{G}_\gamma$ ) называется *доверительной областью надежности  $\gamma$* .

Обычно величину  $\gamma$  выбирают близкой к 1:  $\gamma = 0,9; 0,95; 0,99$ . Соотношения (11.1) означают, что неизвестное значение  $\theta$  или  $g(\theta)$  лежит в области  $\tilde{\Theta}_\gamma$  или  $\tilde{G}_\gamma$  с близкой к 1 вероятностью.

В случае одномерного параметра  $\theta \in R^1$  или одной числовой характеристики  $g(\theta)$  в качестве доверительной области обычно выбирают *доверительный интервал*  $\tilde{\Theta}_\gamma = [\theta_\gamma^-, \theta_\gamma^+]$  или  $\tilde{G}_\gamma = [g_\gamma^-, g_\gamma^+]$ , где  $\theta_\gamma^\pm = \theta_\gamma^\pm(X)$  или  $g_\gamma^\pm = g_\gamma^\pm(X)$  – статистики, представляющие из себя концы доверительного интервала.

При асимптотическом подходе задача заключается в том, чтобы на основе данных наблюдений  $X_n \in \mathcal{X}_n$  построить такую последовательность областей  $\tilde{\Theta}_{n,\gamma} =$

$\tilde{\Theta}_{n,\gamma}(X_n) \subset \Theta$  или  $\tilde{G}_{n,\gamma} = \tilde{G}_{n,\gamma}(X_n) \subset G$ , чтобы при  $n \rightarrow \infty$  выполнялись предельные соотношения

$$\lim P_{n,\theta}(\theta \in \tilde{\Theta}_\gamma) \geq \gamma \quad \text{или} \quad \lim P_{n,\theta}(g(\theta) \in \tilde{G}_\gamma) \geq \gamma \quad \text{для любого } \theta \in \Theta.$$

Такие области называются *асимптотическими доверительными областями уровня надежности  $\gamma$* ; аналогично определяется и понятие *асимптотического доверительного интервала уровня надежности  $\gamma$* .

## 11.2 Задача проверки статистических гипотез

### 11.2.1 Гипотеза и альтернатива

Часто на основе данных наблюдения нужно проверить те или иные предположения о распределении вероятностей экспериментальных данных  $P = P_\theta$ , например о том, что это распределение совпадает с заданным заранее:  $\theta = \theta_0$ , что оно имеет заданные характеристики (среднее, дисперсию и т.д.):  $g(\theta) = g_0$ , или принадлежит заданному классу распределений (является равномерным, нормальным и т.д.). Любое такое предположение называется *статистической гипотезой  $H$*  и выражается соотношением  $H : \theta \in \Theta_H$ . Здесь  $\Theta_H$  – некоторое подмножество  $\Theta$ .

Если множество  $\Theta_H$  состоит из одного элемента  $\Theta_H = \{\theta_H\}$ , то гипотеза  $H$  называется *простой*:  $H : \theta = \theta_H$ . Если множество  $\Theta_H$  состоит более чем из одного элемента, то гипотеза  $H$  называется *сложной*.

Обычно имеется несколько возможных гипотез  $H_1, \dots, H_m$ , которым соответствуют непересекающиеся подмножества  $\Theta_1 = \Theta_{H_1}, \dots, \Theta_m = \Theta_{H_m}$ . Требуется на основе данных наблюдений  $X \in \mathcal{X}$  принять решение о справедливости одной из гипотез.

Мы будем рассматривать двухальтернативные задачи проверки гипотез, то есть  $m = 2$ . В этом случае одна из гипотез  $H_0 : \theta \in \Theta_0$  называется *основной (нулевой) гипотезой*, а другая  $H_1 : \theta \in \Theta_1$  называется *альтернативой*;  $\Theta_0, \Theta_1 \subset \Theta$ ,  $\Theta_0 \cap \Theta_1 = \emptyset$ .

Часто в задачах проверки гипотез задается лишь основная гипотеза  $H_0 : \theta \in \Theta_0$ . При этом обычно имеется в виду, что альтернатива  $H_1$  соответствует случаю, когда основная гипотеза не выполнена:  $H_1 : \theta \notin \Theta_0$ , то есть  $\Theta_1 = \Theta \setminus \Theta_0$  – дополнение множества  $\Theta_0$  в  $\Theta$ . Такие задачи называются *задачами проверки согласия* (с гипотезой  $H_0$ ).

### 11.2.2 Тесты проверки гипотез

Правило принятия или отклонения основной гипотезы называется *тестом проверки гипотез* (его называют также *критерием проверки гипотез*). Тест можно рассматривать как функцию наблюдений  $\psi(X)$ ,  $X \in \mathcal{X}$ , принимающую значения  $\psi(X) = 0$  (это соответствует *принятию основной гипотезы*) или  $\psi(X) = 1$  (это соответствует *отклонению основной гипотезы, то есть принятию альтернативы*). Тест однозначно определяется одним из двух непересекающихся и дополняющих друг друга подмножеств

$$\mathcal{X}_0 = \{X \in \mathcal{X} : \psi(X) = 0\}; \quad \mathcal{X}_1 = \{X \in \mathcal{X} : \psi(X) = 1\}; \quad \mathcal{X}_0 = \mathcal{X} \setminus \mathcal{X}_1.$$

Множество  $\mathcal{X}_0$  называется *допустимым*, а множество  $\mathcal{X}_1$  называется *критическим*. Если наблюдаемые данные  $x$  попадают в  $\mathcal{X}_1$ , то основная гипотеза  $H_0$  отвергается (принимается альтернатива  $H_1$ ); в противном случае ( $x \in \mathcal{X}_0$ ) основная гипотеза  $H_0$  принимается. Обычно критерий задается с помощью *статистики критерия*  $L = L(X)$ ,  $X \in \mathcal{X}$ , и числового *порога критерия*  $T$ :

$$\psi(X) = \begin{cases} 0 & \text{при } L(X) < T, \\ 1 & \text{при } L(X) \geq T. \end{cases}$$

Алгоритм принятия решения составляют два этапа:

- 1) вычисление по наблюдениям  $X$  значения статистики  $L = L(X)$ ;
  - 2) сравнение  $L = L(X)$  с порогом  $T$ ;
- основная гипотеза принимается при  $L < T$  и отвергается при  $L \geq T$ .

### 11.2.3 Ошибки I и II рода и их вероятности

Решения, принимаемые на основе случайных данных с помощью того или иного теста  $\psi$ , могут быть как правильными, так и ошибочными. В задачах проверки гипотез различают ошибки двух типов.

1. *Ошибки I рода: отклонение основной гипотезы  $H_0$ , в то время как она справедлива*, то есть  $\psi(X) = 1$  при  $\theta \in \Theta_0$ .

2. *Ошибки II рода: принятие основной гипотезы  $H_0$ , в то время как имеет место альтернатива*, то есть  $\psi(X) = 0$  при  $\theta \in \Theta_1$ .

Количественной характеристикой ошибок I и II рода являются их вероятности. Они характеризуют достоверность решений, принимаемых с помощью того или иного теста проверки гипотез.

Вероятность ошибок I рода теста  $\psi$  обозначается  $\alpha(\psi, \theta)$  и зависит от значения параметра  $\theta \in \Theta_0$ :

$$\alpha(\psi, \theta) = P_\theta(\mathcal{X}_1), \quad \theta \in \Theta_0.$$

*Уровнем значимости  $\alpha(\psi)$  теста  $\psi$*  называется величина максимальной вероятности ошибок I рода:

$$\alpha(\psi) = \sup_{\theta \in \Theta_0} \alpha(\psi, \theta).$$

Вероятность ошибок II рода теста  $\psi$  обозначается  $\beta(\psi, \theta)$  и зависит от значения параметра  $\theta \in \Theta_1$ :

$$\beta(\psi, \theta) = P_\theta(\mathcal{X}_0), \quad \theta \in \Theta_1.$$

Часто рассматривают также *функцию мощности*  $\gamma(\psi, \theta)$  теста  $\psi$ :

$$\gamma(\psi, \theta) = P_\theta(\mathcal{X}_1) = 1 - \beta(\psi, \theta), \quad \theta \in \Theta_1.$$

Функция мощности  $\gamma(\psi, \theta)$  отличается от вероятности ошибок I рода  $\alpha(\psi, \theta)$  лишь областью изменения аргумента  $\theta$ . Тест называется *несмещенным*, если  $\gamma(\psi, \theta_1) \geq \alpha(\psi, \theta_0)$  для всех  $\theta_0 \in \Theta_0, \theta_1 \in \Theta_1$ .

Для теста  $\psi$ , заданного с помощью статистики  $L$  и порога  $T$ ,

$$\begin{aligned} \alpha(\psi, \theta_0) &= 1 - G_{\theta_0}(T), \quad \theta_0 \in \Theta_0; \\ \beta(\psi, \theta_1) &= G_{\theta_1}(T), \quad \gamma(\psi) = 1 - G_{\theta_1}(T), \quad \theta_1 \in \Theta_1, \end{aligned}$$

где  $G_\theta(T) = P_\theta(L < T)$  есть значение функции распределения статистики  $L$  для порога  $T$ , вычисленное при значении параметра  $\theta \in \Theta$ .

#### 11.2.4 Подход Неймана – Пирсона

В задачах проверки гипотез желательно построить такие тесты, у которых вероятности ошибок как I рода, так и II рода были бы минимальны. Это требование, однако, противоречиво: обычно уменьшение вероятностей ошибок I рода влечет увеличение вероятностей ошибок II рода (уменьшение мощности) и наоборот. Например, для теста, основанного на некоторой статистике  $L$ , уменьшение вероятностей ошибок I рода требует увеличения порога  $T$ , а уменьшение вероятностей ошибок II рода требует уменьшения порога  $T$ .

В этой связи обычно используют *подход Неймана – Пирсона*, состоящий в следующем. Задается малая величина  $\alpha \in (0, 1)$  – максимально допустимая величина вероятностей ошибок I рода (ее называют также *допустимым уровнем значимости*) и рассматриваются тесты  $\psi = \psi_\alpha$ , для которых  $\alpha(\psi) \leq \alpha$ . Выбор величины  $\alpha$  зависит от конкретной задачи. Часто выбирают  $\alpha = 0, 1, 0, 05$  или  $0, 01$ ; если ошибочное отклонение основной гипотезы сопряжено с большими потерями, выбирают  $\alpha = 0, 001$  и менее.

В теории проверки гипотез разрабатываются методы построения тестов проверки различных гипотез при различных альтернативах и изучаются различные характеристики достоверности принимаемых решений, то есть вероятности ошибок и мощность тестов. Наиболее общие результаты удастся получить в рамках асимптотического подхода.

#### 11.2.5 Асимптотические задачи проверки гипотез

При асимптотическом подходе последовательность тестов  $\psi = \psi_n$  часто называется просто тестом и проводится исследование асимптотических (предельных) свойств тестов  $\psi = \{\psi_n\}$  при  $n \rightarrow \infty$ .

Для того чтобы характеризовать вероятности ошибок I рода, вводятся следующие определения. Говорят, что *тест  $\psi = \{\psi_n\}$  имеет асимптотический уровень значимости  $\alpha(\psi)$ ,  $\alpha(\psi) \in [0, 1]$ , если*

$$\alpha_n(\psi_n) = \sup_{\theta \in \Theta_0} \alpha(\psi_n, \theta) \rightarrow \alpha(\psi), \quad n \rightarrow \infty.$$

При использовании подхода Неймана – Пирсона в асимптотическом варианте ограничение накладывается на асимптотический уровень значимости:  $\alpha(\psi) \leq \alpha$ .

Важнейшие асимптотические свойства, характеризующие вероятности ошибок II рода, – это свойства состоятельности тестов. При асимптотическом подходе *тест  $\psi = \{\psi_n\}$  называется состоятельным, если  $\beta(\psi_n, \theta) \rightarrow 0$  при  $n \rightarrow \infty$  для любого  $\theta \in \Theta_1$* . Тест  $\psi = \psi_n$  называется *состоятельным для последовательности  $\theta_n \in \Theta_1$ , если  $\beta(\psi_n, \theta_n) \rightarrow 0$  при  $n \rightarrow \infty$* .

Состоятельность теста  $\psi = \{\psi_n\}$  означает, что для любой простой альтернативы, соответствующей фиксированному значению параметра  $\theta \in \Theta_1$ , вероятность ошибок II рода при достаточно больших  $n$  может быть сделана произвольно малой.

Однако это не значит, что тест  $\psi = \{\psi_n\}$  будет состоятельным для любой последовательности  $\theta_n \in \Theta_1$ , то есть вероятность ошибок II рода будет мала *при всех*  $\theta \in \Theta_1$  для достаточно больших  $n$ . Напротив, в задачах проверки согласия, как правило, часто оказывается, что для любого  $n$  можно найти такие значения  $\theta_n \in \Theta_1$ , при которых вероятность ошибок II рода будет достаточно большой:  $\beta(\psi_n, \theta_n) \geq \beta_0$ . Это имеет место для последовательностей  $\theta_n \in \Theta$ , слишком близких к множеству  $\Theta_0$ , соответствующему нулевой гипотезе.

Для заданной нормы  $\|\cdot\|$  в  $L$ ,  $\Theta \subset L$ , введем меру близости альтернативы  $\theta \in \Theta_1$  и гипотезы  $H_0$ :

$$\rho(\theta, \Theta_0) = \inf_{\theta_0 \in \Theta_0} \|\theta - \theta_0\|.$$

Тест  $\psi = \psi_n$  называется  $\sqrt{n}$ -состоятельным, если  $\beta(\psi_n, \theta_n) \rightarrow 0$  при  $n \rightarrow \infty$  для такой последовательности  $\theta_n \in \Theta_1$ , что  $\sqrt{n}\rho(\theta_n, \Theta_0) \rightarrow \infty$ .

Для задач проверки гипотез справедливо замечание, аналогичное замечанию 11.1, с заменой слова "оценка" словом "тест".

Наряду с состоятельностью в асимптотической теории проверки гипотез изучаются свойства тестов, связанные со скоростью убывания вероятностей ошибок I и II рода, а также методы построения наилучших в том или ином смысле тестов.

Отметим, что различные тесты состоятельны для различных последовательностей альтернатив  $\theta_n$  и имеют разную скорость убывания вероятностей ошибок II рода. Этим объясняется многообразие различных критериев, используемых в статистике.

### 11.3 Оценивание и проверка гипотез

Задачи оценивания тесно связаны с задачами проверки гипотез. Пусть, например, гипотезе и альтернативе соответствуют множества  $\Theta_0$  и  $\Theta_1$  в  $\Theta$ , причем

$$d(\Theta_0, \Theta_1) = \inf_{\theta_1 \in \Theta_1, \theta_0 \in \Theta_0} \|\theta_1 - \theta_0\| > 0$$

(это условие означает, что множества  $\Theta_0$  и  $\Theta_1$  отделены друг от друга). Пусть  $\hat{\theta}_n$  – состоятельная оценка параметра  $\theta \in \Theta$ . Тогда при гипотезе  $H_0$  значение оценки  $\hat{\theta}_n$  с большой вероятностью будет близко к  $\Theta_0$ , а при альтернативе  $H_1$  – к  $\Theta_1$ . Поэтому, сравнивая между собой расстояния  $d(\hat{\theta}_n, \Theta_1)$  и  $d(\hat{\theta}_n, \Theta_0)$  оценки  $\hat{\theta}_n$  от множеств  $\Theta_0$  и  $\Theta_1$ , где

$$d(\hat{\theta}_n, \Theta_j) = \inf_{\theta \in \Theta_j} \|\hat{\theta}_n - \theta\| \geq 0, \quad j = 0, 1$$

(то есть принимая гипотезу  $H_0$  при  $d(\hat{\theta}_n, \Theta_1) > d(\hat{\theta}_n, \Theta_0)$  и альтернативу  $H_1$  в противном случае), мы получим состоятельный тест проверки гипотез.

С другой стороны, пусть  $\tilde{G}_\gamma$  – доверительная область для характеристики  $g(\theta)$  с уровнем доверия  $\gamma \in (0, 1)$ . Рассмотрим задачу проверки гипотезы  $H_0 : g(\theta) \in G_0$ , где  $G_0$  – заданное множество значений характеристики  $g(\theta)$ . Эту гипотезу можно представить в виде  $H_0 : \theta \in \Theta_0$ , где  $\Theta_0$  – множество таких значений параметра, при которых  $g(\theta) \in G_0$ .

Рассмотрим тест проверки согласия с гипотезой  $H_0$  следующего вида: гипотеза  $H_0$  принимается, если доверительная область  $\tilde{G}_\gamma$  пересекается с множеством  $G_0$ , то

есть при  $\tilde{G}_\gamma \cap G_0 \neq \emptyset$ , и отвергается в противном случае. Поскольку при  $\theta \in \Theta_0$  событие  $g(\theta) \in \tilde{G}_\gamma$  влечет событие  $\tilde{G}_\gamma \cap G_0 \neq \emptyset$ , для вероятности ошибок первого рода для этого теста имеем:

$$\alpha(\psi, \theta) = 1 - P_\theta(\tilde{G}_\gamma \cap G_0 \neq \emptyset) \leq 1 - P_\theta(\theta \in \tilde{G}_\gamma) \leq 1 - \gamma, \quad \theta \in \Theta_0,$$

так что мы получаем тест уровня значимости не больше  $\alpha = 1 - \gamma$ .

Аналогично, имея последовательность доверительных областей  $\tilde{G}_{n,\gamma}$  с асимптотическим уровнем доверия  $\gamma \in (0, 1)$ , мы получаем последовательность тестов асимптотического уровня значимости не больше  $\alpha = 1 - \gamma$ .

Во многих задачах описанные конструкции приводят к оптимальным или асимптотически оптимальным тестам.

## 12 Статистические задачи, связанные с неизвестной функцией распределения

В этом разделе мы рассмотрим модель независимой однородной выборки с неизвестной функцией распределения  $F = F_X$ , где  $X$  – случайная величина.

Неизвестная функция распределения  $F_X = F$  может рассматриваться как функциональный параметр  $F \in \mathcal{F}$ , где  $\mathcal{F}$  – некоторое множество функций распределения.

Мы рассмотрим методы решения основных статистических задач: оценивания, доверительного оценивания неизвестной функции распределения, а также проверки некоторых гипотез о функции распределения  $F = F_X$ .

### 12.1 Эмпирическая функция распределения

Пусть  $X$  – случайная величина. С выборкой  $X^{(n)} = (X_1, \dots, X_n)$  можно связать эмпирическую функцию распределения (ЭФР)  $F_n(t) = F_n(t, X^{(n)})$ :

$$F_n(t) = \frac{1}{n} \left( \text{число элементов } X_i \text{ выборки } X^{(n)}, \text{ меньших } t \right).$$

Эмпирическую функцию распределения можно описать иначе.

1. Обозначим через  $\mathbb{I}_A(X)$ ,  $X \in \mathcal{X}$ , индикаторную функцию (индикатор) множества  $A \subset \mathcal{X}$ :

$$\mathbb{I}_A(X) = \begin{cases} 1, & \text{если } X \in A, \\ 0, & \text{если } X \notin A. \end{cases}$$

Пусть  $A(t) = \{X < t\}$  – событие, состоящее в том, что случайная величина примет значение, меньшее  $t$ , где  $t \in R^1$  – вещественное число. Тогда

$$F_n(t) = F_n(t, X^{(n)}) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{A(t)}(X_i). \quad (12.1)$$

2. Пусть среди элементов выборки нет равных. Обозначим через  $X_n^*$  дискретную случайную величину, принимающую значения  $X_1, \dots, X_n$  с вероятностями  $1/n$ :

$$P(X_n^* = X_i) = \frac{1}{n}; \quad i = 1, \dots, n.$$

Если среди элементов выборки есть равные (значение  $X_i$  встречается  $n_i$  раз), то  $P(X_n^* = X_i) = n_i/n$ . Тогда  $F_n(t) = F_{X_n^*}(t)$  – функция распределения случайной величины  $X_n^*$ .

3. Вариационным рядом выборки  $X^{(n)} = (X_1, \dots, X_n)$  называется упорядоченный набор  $X_{(1)} \leq \dots \leq X_{(n)}$ , составленный из элементов выборки  $X^{(n)}$ . Величины  $X_{(k)}$  называются порядковыми статистиками порядка  $k$ ,  $1 \leq k \leq n$ . Величины  $X_{(1)} = \min X_i$ ,  $X_{(n)} = \max X_i$  называются крайними членами вариационного ряда. Формально положим  $X_{(0)} = -\infty$ ,  $X_{(n+1)} = \infty$ . Тогда

$$F_n(t) = \frac{k}{n} \text{ при } X_{(k)} < t < X_{(k+1)}, \quad k = 0, \dots, n. \quad (12.2)$$

При заданной выборке  $X^{(n)} = (X_1, \dots, X_n)$  эмпирическая функция распределения  $F_n(t, X^{(n)})$  есть кусочно-постоянная функция, равная 0 при  $t < X_{(1)}$ , равная 1 при  $t \geq X_{(n)}$  и со скачками  $n_k/n$  в точках  $X_{(k)}$  вариационного ряда выборки.

Соотношение (12.2) определяет функцию  $F_n(t)$  везде, кроме точек вариационного ряда. В точках вариационного ряда значения  $F_n(X_{(k)})$  определяются требованием непрерывности слева. Если все члены вариационного ряда различны (среди элементов выборки нет равных), то  $F_n(X_{(k+1)}) = k/n$ ,  $k = 0, \dots, n-1$ . Если среди элементов выборки есть равные (значение  $X_{(k)}$  встречается  $n_k$  раз), то  $F_n(X_{(k+1)}) = F_n(X_{(k)}) + n_k/n$  при  $X_{(k+1)} > X_{(k)}$  и  $F_n(X_{(k+1)}) = F_n(X_{(k)})$  при  $X_{(k+1)} = X_{(k)}$ ,  $k = 0, \dots, n-1$ ;  $F_n(X_{(0)}) = 0$ .

Можно показать, что эти три определения эквивалентны.

## 12.2 Статистические свойства ЭФР

Статистические свойства эмпирической функции распределения связаны с трактовкой наблюдаемых данных как реализации набора  $X^{(n)} = (X_1, \dots, X_n)$  независимых случайных величин с функцией распределения  $F(t) = F_X(t)$ . В этом случае  $F_n(t) = F_n(t, X^{(n)})$  при фиксированном  $t$  есть случайная величина, а при переменном  $t$  – случайная функция от  $t$ .

**Замечание.** Для непрерывной функции распределения  $F = F_X$  случайная величина  $\chi = F(X)$  имеет равномерное распределение на интервале  $[0, 1]$ . Поэтому совместное  $P_{n,F}$ -распределение значений  $F(X_{(1)}), \dots, F(X_{(n)})$  совпадает с совместным распределением вариационного ряда  $\chi_{(1)}, \dots, \chi_{(n)}$  выборки из равномерного распределения на интервале  $[0, 1]$  и не зависит от функции распределения  $F$ .

### 12.2.1 Свойства ЭФР при фиксированном значении аргумента

Зафиксируем  $t$  и рассмотрим свойства случайной величины  $\hat{g}_n = F_n(t) = F_n(t, X^{(n)})$  как оценки значения  $g(F) = F(t)$ . Заметим, что входящие в (12.1) случайные ве-

личины  $Y_i(t) = \mathbb{I}_{A(t)}(X_i)$  независимы, одинаково распределены и имеют бернуллиевское распределение  $B_p$  с параметром  $p = F(t)$ , то есть  $Y(t) = \mathbb{I}_{A(t)}(X) \sim B_p$ . Поэтому

$$E_F Y(t) = F(t); \quad D_F Y(t) = F(t)(1 - F(t)).$$

В силу (12.1) отсюда имеем:

$$E_{n,F}(F_n(t)) = \frac{1}{n} E_{n,F} \left( \sum_{i=1}^n Y_i(t) \right) = E_F Y(t) = F(t); \quad (12.3)$$

$$\begin{aligned} D_{n,F}(F_n(t)) &= \frac{1}{n^2} D_{n,F} \left( \sum_{i=1}^n Y_i(t) \right) = \frac{1}{n} D_F Y(t) = \\ &= \frac{1}{n} F(t)(1 - F(t)). \end{aligned} \quad (12.4)$$

Соотношение (12.3) показывает, что значение эмпирической функции распределения  $F_n(t) = F_n(t, X^{(n)})$  является *несмещенной оценкой значения неизвестной функции распределения*  $g(F) = F(t)$  (см. подразд. 13.1).

Из несмещенности оценки следует, как отмечено в подразд. 13.1, что квадратичный риск совпадает с дисперсией и с учетом (12.4) есть

$$R_2(F_n(t), F) = D_{n,F}(F_n(t)) = F(t)(1 - F(t))/n \rightarrow 0, \quad n \rightarrow \infty.$$

Отсюда следует (см. п. 11.1.3), что эта оценка является *состоятельной и  $\sqrt{n}$ -состоятельной*.

## 12.2.2 Свойства ЭФР "в целом"

Для того чтобы рассматривать свойства эмпирической функции распределения как оценки функции распределения "в целом", то есть одновременно при всевозможных  $t$ , нужно ввести меру близости  $F_n$  и  $F$ . Можно считать, что  $\mathcal{F} = \Theta \subset L$ , где в качестве линейного пространства  $L$  рассматривается множество ограниченных измеримых функций  $u = u(t)$  на  $R^1$ .

Есть разные способы измерения расстояния между функциями распределения, основанные на различных нормах в линейном пространстве  $L$ . Одним из наиболее важных в статистике является *расстояние Колмогорова*

$$\rho_\infty(F, F_0) = \sup_t |F(t) - F_0(t)|, \quad (12.5)$$

основанное на норме максимального отклонения

$$\|u\|_\infty = \sup_t |u(t)|, \quad u \in L.$$

Сформулируем некоторые теоремы о расстояниях Колмогорова.

**Теорема 12.1 (Гливенко – Кантелли).** Пусть множество  $\mathcal{F}$  состоит из всех функций распределения. Тогда для любой  $F \in \mathcal{F}$  с вероятностью 1 справедливо предельное равенство  $\rho_\infty(F_n, F) \rightarrow 0$  при  $n \rightarrow \infty$ .



Теорема Гливенко – Кантелли утверждает, что с вероятностью 1 числовая последовательность  $\rho_\infty(F_n, F)$  стремится к нулю, то есть с вероятностью 1 эмпирическая функция распределения  $F_n(t)$  сходится к теоретической функции распределения  $F(t) = F_X(t)$  случайной величины  $X$ .

Из теории вероятностей известно, что сходимость с вероятностью 1 влечет за собой сходимость по вероятности: для любого  $\varepsilon > 0$

$$P_{n,F}(\rho_\infty(F_n, F) > \varepsilon) \rightarrow 0, \text{ если } n \rightarrow \infty.$$

Это соотношение означает состоятельность эмпирической функции распределения как оценки функции распределения при использовании расстояния Колмогорова для измерения близости. Более точные оценки близости основаны на следующей теореме.

**Теорема 12.2 (Колмогорова).** Пусть множество  $\mathcal{F}_H$  состоит из непрерывных функций распределения. Тогда для любой  $F \in \mathcal{F}_H$  справедливо предельное равенство

$$P_{n,F}(\sqrt{n}\rho_\infty(F_n, F) < u) \rightarrow K(u), \text{ если } n \rightarrow \infty,$$

где функция  $K(u)$  определяется равенством

$$K(u) = \begin{cases} 0, & \text{если } u \leq 0, \\ \sum_{j=-\infty}^{\infty} (-1)^j e^{-2(ju)^2}, & \text{если } u > 0. \end{cases}$$

Функция  $K(u)$  называется *функцией распределения Колмогорова*. Она непрерывна, строго возрастает при  $u > 0$ , и  $K(u) \rightarrow 1$  при  $u \rightarrow \infty$ . Таблицы функции распределения Колмогорова содержатся, например, в [2].

Случайные величины  $d_n = \sqrt{n}\rho_\infty(F_n, F)$  могут быть выражены в явном виде через значения  $F(X_{(i)})$  неизвестной функции распределения в точках  $X_{(i)}$  вариационного ряда выборки:

$$d_n = \sqrt{n} \max_{1 \leq i \leq n} \max\{|F(X_{(i)}) - i/n|; |F(X_{(i)}) - (i-1)/n|\}. \quad (12.6)$$

Из замечания п. 12.2 следует, что не только предельные, но и допредельные  $P_{n,F}$ -распределения статистик  $d_n$  не зависят от неизвестной непрерывной функции распределения  $F \in \mathcal{F}_H$  при всех  $n$ . Статистики, обладающие таким свойством, называются *свободными от распределения*.

Из теоремы Колмогорова и из свойств функции  $K(u)$  вытекает следующее соотношение: для любой числовой последовательности  $u_n$  и любой последовательности непрерывных функций распределения  $F^{(n)} \in \mathcal{F}_H$

$$P_{n,F^{(n)}}(\sqrt{n}\rho_\infty(F_n, F^{(n)}) < u_n) = K(u_n) + o(1). \quad (12.7)$$

## 12.3 Доверительные области для функции распределения

Используя теорему Колмогорова 12.2, можно построить *доверительную полосу для неизвестной функции распределения*. Положим

$$F_{n,\gamma}^-(t) = \max(0, F_n(t) - u_\gamma/\sqrt{n}), \quad F_{n,\gamma}^+(t) = \min(1, F_n(t) + u_\gamma/\sqrt{n}),$$

где величина  $u_\gamma$  определяется из условия  $K(u_\gamma) = \gamma$ .

Рассмотрим множество  $\mathcal{F}_{n,\gamma,\infty}$ , состоящее из непрерывных функций распределения  $F \in \mathcal{F}_H$ , график которых лежит внутри полосы, ограниченной графиками ступенчатых функций  $F_{n,\gamma}^\pm(t)$ ,  $-\infty < t < \infty$ . Тогда теорема Колмогорова означает, что для любой непрерывной функции распределения  $F \in \mathcal{F}_H$

$$P_{n,F}(F \in \mathcal{F}_{n,\gamma,\infty}) = P_{n,F}(d_n < u_\gamma) \rightarrow K(u_\gamma) = \gamma, \quad n \rightarrow \infty.$$

Другими словами, множества  $\mathcal{F}_{n,\gamma,\infty}$  представляют собой асимптотическую доверительную полосу для неизвестной функции распределения с уровнем доверия  $\gamma$ : при больших  $n$  с вероятностью, близкой к  $\gamma$ , неизвестная функция распределения будет лежать внутри полосы  $\mathcal{F}_{n,\gamma,\infty}$ .

## 12.4 Критерии согласия Колмогорова

Пусть  $F_0 \in \mathcal{F}_H$  – заданная непрерывная функция распределения. Рассмотрим задачу проверки простой гипотезы  $H_0 : F_X = F_0$  о функции распределения генеральной совокупности  $X$ . Эту задачу можно трактовать как задачу проверки согласия и в качестве множества  $\mathcal{F}$  рассматривать множество  $\mathcal{F}_H$  всех непрерывных функций распределения, то есть  $H_1 : F_X \in \mathcal{F}_H, F_X \neq F_0$ .

Назовем критерием Колмогорова асимптотического уровня значимости  $\alpha$  последовательность тестов

$$\psi_n(X_1, \dots, X_n) = \begin{cases} 1, & \text{если } d_n = \sqrt{n}\rho_\infty(F_n, F_0) \geq u^\alpha, \\ 0, & \text{если } d_n = \sqrt{n}\rho_\infty(F_n, F_0) < u^\alpha, \end{cases}$$

где величина  $u^\alpha = u_{1-\alpha}$  определяется из условия  $K(u^\alpha) = 1 - \alpha$ .

Для критерия Колмогорова вероятность ошибок I рода есть

$$\alpha(\psi_n) = P_{n,F_0}(d_n \geq u^\alpha) = 1 - P_{n,F_0}(d_n < u^\alpha) \rightarrow 1 - K(u^\alpha) = \alpha$$

при  $n \rightarrow \infty$  в силу теоремы Колмогорова.

Для оценки вероятностей ошибок II рода критерия Колмогорова зафиксируем альтернативу  $F \neq F_0$ . В силу теоремы Гливленко – Кантелли  $F_n(t) \rightarrow F(t)$  при  $n \rightarrow \infty$  равномерно по  $t \in R^1$ , так что  $\rho_\infty(F_n, F_0) \rightarrow \rho_\infty(F, F_0) > 0$ . Следовательно,  $\sqrt{n}\rho_\infty(F_n, F_0) \rightarrow \infty$  при  $n \rightarrow \infty$  по  $P_{n,F}$ -вероятности. Поэтому

$$\beta(\psi_n, F) = P_{n,F}(\sqrt{n}\rho_\infty(F_n, F_0) < u^\alpha) \rightarrow 0, \quad n \rightarrow \infty.$$

Отметим также неравенство

$$\rho_\infty(F_n, F_0) \geq \rho_\infty(F, F_0) - \rho_\infty(F_n, F),$$

используя которое, а также (12.7), для любой такой последовательности альтернатив  $F^{(n)} \in \mathcal{F}_H$ , что  $\sqrt{n}\rho_\infty(F^{(n)}, F_0) \rightarrow \infty$ , имеем:

$$\begin{aligned} \beta(\psi_n, F^{(n)}) &= P_{n, F^{(n)}}(\sqrt{n}\rho_\infty(F_n, F_0) < u^\alpha) \leq \\ &\leq P_{n, F^{(n)}}(\sqrt{n}(\rho_\infty(F^{(n)}, F_0) - \rho_\infty(F_n, F^{(n)})) < u^\alpha) = \\ &= P_{n, F^{(n)}}(\sqrt{n}\rho_\infty(F_n, F^{(n)}) > \sqrt{n}\rho_\infty(F^{(n)}, F_0) - u^\alpha) = \\ &= 1 - K(\sqrt{n}\rho_\infty(F^{(n)}, F_0) - u^\alpha) + o(1) \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Таким образом, критерий Колмогорова обеспечивает вероятность ошибок I рода, стремящуюся к  $\alpha$  при увеличении объема выборки  $n$ , а также является состоятельным в классе всех функций распределения и  $\sqrt{n}$ -состоятельным в классе всех непрерывных функций распределения с мерой близости  $\rho_\infty(F, F_0)$ .

## 13 Задачи статистического оценивания

Как отмечалось выше, задача оценивания характеристик  $g(\theta)$ ,  $\theta \in \Theta$ , содержит в себе задачу оценивания параметра при  $g(\theta) = \theta$ . Поэтому основные понятия и определения мы приведем для задачи оценивания характеристик распределения.

### 13.1 Одномерная характеристика

В случае одномерной характеристики  $g(\theta)$  квадратичный риск можно представить в виде

$$R_2(\hat{g}, \theta) = E_\theta((\hat{g} - g(\theta))^2) = b_\theta^2(\hat{g}) + D_\theta(\hat{g}), \quad (13.1)$$

где величину  $b_\theta(\hat{g}) = E_\theta(\hat{g}) - g(\theta)$  называют *смещением оценки*  $\hat{g}$ , а  $D_\theta(\hat{g})$  есть дисперсия оценки  $\hat{g}$ . Смещение характеризует систематическое отклонение оценки  $\hat{g}$  от значения характеристики  $g(\theta)$ , а дисперсия оценки  $\hat{g}$  – случайные колебания оценки относительно ее среднего значения  $E_\theta(\hat{g})$ . Оценку  $\hat{g}$  называют *несмещенной*, если  $b_\theta(\hat{g}) = 0$ . Таким образом, для несмещенной оценки

$$R_2(\hat{g}, \theta) = D_\theta(\hat{g}). \quad (13.2)$$

Условие несмещенности означает отсутствие систематических отклонений оценки  $\hat{g}$  от значения  $g(\theta)$  и представляется достаточно естественным. В конечнопараметрических задачах хорошие оценки обычно оказываются несмещенными или асимптотическое смещение мало по сравнению со среднеквадратическим разбросом оценки:

$$b_{n,\theta}(\hat{g}_n)/\sigma_{n,\theta}(\hat{g}_n) \rightarrow 0, \quad n \rightarrow \infty; \quad \sigma_{n,\theta}(\hat{g}_n) = \sqrt{D_{n,\theta}(\hat{g}_n)}.$$

Однако во многих сложных непараметрических задачах "хорошие" оценки часто оказываются смещенными, так что отказ от требования несмещенности позволяет получать лучшие оценки в этих задачах.

**Пример 13.1 Оценка вероятности по частоте.** Пусть проводится серия  $n$  независимых экспериментов, в каждом из которых может произойти или не произойти случайное событие  $A$ , вероятность которого  $P(A) = p$  неизвестна. Пусть  $k$  – число экспериментов, в которых событие  $A$  произошло. В качестве оценки неизвестной вероятности естественно выбрать наблюдаемую частоту  $\hat{p}_n = k/n$  события  $A$ . Случайная величина  $k$  имеет биномиальное распределение  $B_{n,p}$ ,  $p \in (0, 1)$ , и в силу свойств биномиального распределения

$$E_{n,p}\hat{p}_n = p, \quad D_{n,p}\hat{p}_n = p(1-p)/n. \quad (13.3)$$

Первое равенство (13.3) означает несмещенность оценки  $\hat{p}_n$ , и в силу второго равенства

$$R_2(\hat{p}_n, p) = p(1-p)/n.$$

**Пример 13.2 Оценка среднего нормального распределения.** Задачу оценки неизвестной величины  $a$  по измерениям  $X_i = a + \xi_i$ ,  $\xi_i \sim N(0, \sigma^2)$ , упомянутую во введении, можно трактовать как задачу оценки неизвестного среднего  $a$  по независимой выборке  $X_1, \dots, X_n$  из нормального распределения  $X \sim N(a, \sigma^2)$ ,  $\theta = (a, \sigma^2) \in R^1 \times R_+^1$ . Рассмотрим в качестве оценки  $\hat{a}_n$  выборочное среднее

$$\hat{a}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (13.4)$$

Эта оценка имеет нормальное распределение  $\hat{a}_n \sim N(a, \sigma^2/n)$ . Следовательно,  $\bar{X}_n$  есть несмещенная оценка и

$$R_2(\bar{X}_n, \theta) = D_{n,\theta}\bar{X}_n = \sigma^2/n. \quad (13.5)$$

**Пример 13.3 Оценка дисперсии нормального распределения.** Пусть в примере 13.2 при  $n > 1$  требуется оценить параметр  $d = \sigma^2$ . В качестве оценки  $d_n$  обычно используется

$$\sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (13.6)$$

Известно [3], [2], [7], что случайная величина  $\tau_n = (n-1)\sigma_n^2/\sigma^2$  имеет распределение хи-квадрат с  $n-1$  степенями свободы, которое не зависит от  $\theta$ :

$$P_{n,\theta}(\tau_n < t) = \chi_{n-1}^2(t). \quad (13.7)$$

Здесь и далее  $\chi_n^2(t) = P(z_n < t)$  есть функция распределения хи-квадрат с  $n$  степенями свободы: это функция распределения случайной величины  $z_n = \sum_{i=1}^n \xi_i^2$ , где  $\xi_1, \dots, \xi_n$  – независимые стандартные нормальные случайные величины. Таблицы этой функции имеются во всех учебниках и задачниках по математической статистике. Среднее и дисперсия случайной величины  $z_n$  есть  $Ez_n = n$ ,  $Dz_n = 2n$ , так что оценка  $\sigma_n^2$  является несмещенной и

$$R_2(\sigma_n^2, \theta) = D_{n,\theta}\sigma_n^2 = \frac{2\sigma^4}{n-1}. \quad (13.8)$$

В силу центральной предельной теоремы распределение хи-квадрат является асимптотически нормальным:

$$\sup_t |\chi_n^2(t) - \Phi((t - n)/\sqrt{2n})| \rightarrow 0 \text{ при } n \rightarrow \infty. \quad (13.9)$$

## 13.2 Асимптотически нормальные оценки

В примерах 13.1 – 13.3 квадратичный риск убывает со скоростью  $1/n$ :

$$R_2(\hat{g}_n, \theta) \sim u(\theta)/n, \quad u(\theta) > 0 \quad (13.10)$$

при, соответственно,  $u(\theta) = \theta(1 - \theta)$ ;  $u(\theta) = \sigma^2$ ,  $u(\theta) = 2\sigma^4$ . Отсюда следует (см. раздел 11.1.3)  $\sqrt{n}$ -состоятельность этих оценок, то есть типичный порядок отклонения оценки  $\hat{g}_n$  от оцениваемой характеристики  $g(\theta)$  есть  $1/\sqrt{n}$ .

Для более детального анализа отклонений порядка  $1/\sqrt{n}$  важную роль играет свойство *асимптотической нормальности оценок*.

Последовательность оценок  $\hat{g}_n$  характеристики  $g(\theta)$  называется *асимптотически нормальной с асимптотической дисперсией*  $\Delta^2(\theta) > 0$ , если случайная величина  $Y_n = \sqrt{n}(\hat{g}_n - g(\theta))$  сходится по  $P_{n,\theta}$ -распределению к нормальной случайной величине с нулевым средним и дисперсией  $\Delta^2(\theta)$ :

$$Y_n \xrightarrow{P_{n,\theta}} Y \sim N(0, \Delta^2(\theta)). \quad (13.11)$$

Соотношение (13.11) можно переписать в виде

$$\hat{g}_n = g(\theta) + n^{-1/2}Y_n, \quad Y_n \xrightarrow{P_{n,\theta}} Y \sim N(0, \Delta^2(\theta)), \quad (13.12)$$

то есть отклонение оценки от неизвестного значения оцениваемой характеристики имеет приближенно нормальное распределение с нулевым средним и дисперсией  $\Delta^2(\theta)/n$ .

Величину  $\Delta(\theta) > 0$  называют *нормирующим множителем*. Напомним, что (13.11) равносильно выполнению соотношения

$$P_{n,\theta}(Y_n/\Delta(\theta) < t) \rightarrow P(\xi < t) = \Phi(t), \quad n \rightarrow \infty,$$

где  $\Phi(t)$  – функция распределения стандартного нормального закона. Отсюда следует, что для любых чисел  $a < b$  выполнено соотношение

$$\begin{aligned} P_{n,\theta}(a\Delta(\theta)/\sqrt{n} < \hat{g}_n - g(\theta) < b\Delta(\theta)/\sqrt{n}) = \\ = P_{n,\theta}(a < Y_n/\Delta(\theta) < b) \rightarrow \Phi(b) - \Phi(a), \quad n \rightarrow \infty. \end{aligned} \quad (13.13)$$

Очевидно, что в примере 13.2 выполнено условие нормальности  $Y_n \sim N(0, \sigma^2)$  (даже не асимптотической) и  $\Delta(\theta) = \sigma$ . В примерах 13.1 и 13.3 условие асимптотической нормальности выполнено при  $\Delta^2(p) = p(1 - p)$  и  $\Delta^2(\theta) = 2\sigma^4$  в силу центральной предельной теоремы.

Выбирая некоторое  $t > 0$ , полагая  $\delta_n = t\Delta(\theta)/\sqrt{n}$  и используя (13.13), получим для асимптотически нормальных оценок асимптотику вероятностей отклонения:

$$\begin{aligned} R_n^\delta(\hat{g}_n, \theta) &= P_{n,\theta}(|\hat{g}_n - g(\theta)| \geq \delta_n) = P_{n,\theta}(|Y_n/\Delta(\theta)| \geq t) = \\ &= 1 - P_{n,\theta}(-t < Y_n/\Delta(\theta) < t) \rightarrow 1 - \Phi(t) + \Phi(-t) = 2(1 - \Phi(t)). \end{aligned}$$

Обычно, помимо условия асимптотической нормальности (13.11), выполнено дополнительное условие сходимости моментов 2-го порядка

$$E_{n,\theta}(Y_n^2) \rightarrow E(Y^2) = \Delta^2(\theta), \quad (13.14)$$

из которого следует сходимость моментов 1-го порядка:  $E_{n,\theta}(Y_n) \rightarrow E(Y) = 0$ . Эти соотношения позволяют описать асимптотическое поведение смещения  $b_{n,\theta}(\hat{g}_n)$ , дисперсии  $D_{n,\theta}(\hat{g}_n)$  и квадратичного риска  $R_2(\hat{g}_n, \theta)$  оценки  $\hat{g}_n$ , удовлетворяющей (13.11), (13.14):

$$\sqrt{n}b_{n,\theta}(\hat{g}_n) = E_{n,\theta}(Y_n) \rightarrow 0; \quad (13.15)$$

$$nD_{n,\theta}(\hat{g}_n) = D_{n,\theta}(Y_n) \rightarrow \Delta^2(\theta); \quad (13.16)$$

$$nR_2(\hat{g}_n, \theta) = E_{n,\theta}(Y_n^2) \rightarrow \Delta^2(\theta). \quad (13.17)$$

Соотношение (13.15) называют  $\sqrt{n}$ -несмещенностью оценки  $\hat{g}_n$ . Соотношения (13.16), (13.17) означают, что дисперсия и квадратичный риск оценки  $\hat{g}_n$  асимптотически совпадают и имеют порядок  $\Delta^2(\theta)/n$ . Сопоставление этих соотношений с (13.10) показывает, что в последнем соотношении  $u(\theta) = \Delta^2(\theta)$  (сравни с примерами 13.1 – 13.3).

### 13.3 Выборочные характеристики и их свойства

С генеральной совокупностью  $X$  связаны различные вероятностные характеристики  $g_X = g(F_X)$ , то есть характеристики  $g(F)$  неизвестной функции распределения  $F_X = F \in \mathcal{F}$ , где  $\mathcal{F}$  – некоторое множество функций распределения. Многие вероятностные характеристики  $g_X$  можно определить и для эмпирической функции распределения  $g_{X_n^*} = g(F_n)$ . *Выборочной характеристикой называется статистика  $g_n = g_n(X_1, \dots, X_n) = g(F_n)$ , то есть значение оцениваемой характеристики для эмпирической функции распределения* (при условии, что это значение определено).

Выборочные характеристики  $g(X_n^*) = g(F_n)$  естественно рассматривать как оценки характеристик  $g(X) = g(F)$  в модели независимой однородной выборки (или выборки из генеральной совокупности  $X$  с неизвестной функцией распределения  $F_X = F \in \mathcal{F}$ ).

#### 13.3.1 Выборочное математическое ожидание

Напомним, что выборочным математическим ожиданием (или выборочным средним) называется статистика

$$\bar{X}_n = m(X_n^*) = m(F_n) = \frac{1}{n} \sum_{i=1}^n X_i. \quad (13.18)$$

Математическое ожидание – важная числовая характеристика случайной величины или функции распределения.

Обозначим через  $\mathcal{F}_k$  множество таких функций распределения, для которых  $E_F(|X|^k) = \int_{-\infty}^{\infty} |x|^k dF(x) < \infty$ . Математическое ожидание определено для функции распределения  $F \in \mathcal{F}_1$ . В силу свойств математического ожидания при  $F \in \mathcal{F}_1$

$$E(aX) = aE(X); E\left(\sum_i X_i\right) = \sum_i E(X_i) \quad (13.19)$$

и, поскольку случайные величины  $X_1, \dots, X_n$  имеют общее распределение,

$$E_F(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E_F(X_i) = E_F(X) = m(F). \quad (13.20)$$

Это свойство означает *несмещенность оценки*  $\bar{X}_n$ .

Из закона больших чисел следует, что при  $n \rightarrow \infty$

$$\bar{X}_n \rightarrow m(F) \text{ по } P_F\text{-вероятности.} \quad (13.21)$$

То есть для любого  $\delta > 0$

$$P_F(|\bar{X}_n - m(F)| > \delta) \rightarrow 0, \quad n \rightarrow \infty. \quad (13.22)$$

Это соотношение означает *состоятельность оценки*  $\bar{X}_n$  (см. раздел 11.1.3).

Пусть существует дисперсия  $D(F)$ , то есть  $F \in \mathcal{F}_2$ . В силу свойств дисперсии

$$D(aX + b) = a^2 D(X); D\left(\sum_i X_i\right) = \sum_i D(X_i), \quad (13.23)$$

если  $X_i$  независимы, и, поскольку случайные величины  $X_1, \dots, X_n$  независимы и имеют общее распределение,

$$D_F(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n D_F(X_i) = \frac{1}{n} D_F(X) = \frac{1}{n} D(F). \quad (13.24)$$

В силу неравенства Чебышева отсюда вытекает уточнение соотношений (13.22): для любого  $\delta > 0$

$$P_F(|\bar{X}_n - m(F)| > \delta) \leq D(F)/n\delta^2. \quad (13.25)$$

В частности, если  $\delta = \delta_n \rightarrow 0$ ,  $\sqrt{n}\delta_n \rightarrow \infty$ , то

$$P_F(|\bar{X}_n - m(F)| > \delta_n) \leq \frac{D(F)}{n\delta_n^2} \rightarrow 0. \quad (13.26)$$

Соотношение (13.26) показывает, что, если отклонение статистики  $\bar{X}_n$  от оцениваемой величины  $m(F)$  по порядку больше чем  $1/\sqrt{n}$ , то его вероятность стремится к нулю при  $n \rightarrow \infty$ . Это свойство означает  *$\sqrt{n}$ -состоятельность оценки*

$\bar{X}_n$ . Для оценки вероятностей отклонений порядка  $1/\sqrt{n}$  воспользуемся центральной предельной теоремой (ЦПТ) теории вероятностей. Рассмотрим последовательность случайных величин  $Z_n = \sqrt{n}(\bar{X}_n - m(F))/\sigma(F)$ . Согласно ЦПТ, последовательность  $Z_n$  сходится по  $P_F$ -распределению к стандартной нормальной случайной величине  $\xi \sim N(0, 1)$ :

$$Z_n \xrightarrow{P_F} \xi, \quad n \rightarrow \infty. \quad (13.27)$$

Это означает, что для любого  $t \in R$  при  $n \rightarrow \infty$

$$P_F(Z_n < t) = P_F\left(\bar{X}_n < m(F) + t \frac{\sigma(F)}{\sqrt{n}}\right) \rightarrow P(\xi < t) = \Phi(t); \quad (13.28)$$

более того,

$$\begin{aligned} \sup_t (P_F(Z_n < t) - \Phi(t)) = \\ = \sup_u \left( P_F(\bar{X}_n < u) - \Phi\left((u - m(F)) \frac{\sqrt{n}}{\sigma(F)}\right) \right) \rightarrow 0. \end{aligned} \quad (13.29)$$

Свойства (13.27)–(13.29) показывают *асимптотическую нормальность оценки  $\bar{X}_n$  с асимптотической дисперсией  $\Delta^2(F) = \sigma^2(F)$* .

Из соотношений (13.27)–(13.29) следует, что

$$\begin{aligned} P_F(|\bar{X}_n - m(F)| \geq \delta_n) &\rightarrow P(|\xi| \geq t) = \\ &= \Phi(-t) + 1 - \Phi(t) = 2(1 - \Phi(t)) \end{aligned} \quad (13.30)$$

при  $\delta_n = t\sigma(F)/\sqrt{n}$ ,  $t > 0$ . В частности, для любого  $\alpha \in (0, 1)$  выбирая величину  $t = t^\alpha$  из условия  $\Phi(t^\alpha) = 1 - (\alpha/2)$ , получим:

$$P_F\left(|\bar{X}_n - m(F)| \geq t^\alpha \frac{\sigma(F)}{\sqrt{n}}\right) \rightarrow \alpha, \quad n \rightarrow \infty. \quad (13.31)$$

Соотношение (13.31) описывает вероятности отклонений порядка  $n^{-1/2}$ .

### 13.3.2 Выборочные моменты

*Начальным моментом порядка  $k = 1, 2, \dots$  называется величина*

$$m^k(X) = m^k(F) = E_F(X^k) = \int_{-\infty}^{\infty} x^k dF(x).$$

*Центральным моментом порядка  $k = 1, 2, \dots$  называется величина*

$$\mu^k(X) = \mu^k(F) = E_F((X - m(X))^k) = \int_{-\infty}^{\infty} (x - m(X))^k dF(x).$$

*Абсолютным начальным моментом порядка  $k > 0$  называется величина*

$$m_a^k(X) = m_a^k(F) = E_F(|X|^k) = \int_{-\infty}^{\infty} |x|^k dF(x).$$



Абсолютным центральным моментом порядка  $k > 0$  называется величина

$$\mu_a^k(X) = \mu_a^k(F) = E_F(|X - m(X)|^k) = \int_{-\infty}^{\infty} |x - m(X)|^k dF(x).$$

Математическое ожидание есть начальный момент порядка 1, дисперсия есть центральный момент порядка 2. Моменты порядка  $k > 0$  определены для случайных величин  $X$  с функцией распределения  $F_X = F \in \mathcal{F}_k$ .

Центральные моменты нечетного порядка  $k \geq 3$  характеризуют степень несимметрии отклонений случайной величины от ее среднего значения. Абсолютные центральные моменты больших порядков  $k > 2$  характеризуют поведение "хвостов" распределения случайной величины (то есть больших отклонений от среднего значения).

Рассмотрим соответствующие выборочные характеристики.

*Выборочным (начальным) моментом порядка  $k$  называется статистика*

$$m_n^k = m^k(X_n^*) = m^k(F_n) = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots$$

*Абсолютным выборочным моментом порядка  $k$  называется статистика*

$$m_{n,a}^k = m_a^k(X_n^*) = m_a^k(F_n) = \frac{1}{n} \sum_{i=1}^n |X_i|^k, \quad k > 0.$$

*Выборочным центральным моментом порядка  $k$  называется статистика*

$$\mu_n^k = \mu^k(X_n^*) = \mu^k(F_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^k, \quad k = 1, 2, \dots$$

Начальные моменты случайной величины  $X$  можно рассматривать как среднее значение другой случайной величины  $Z = X^k$  или  $Z = |X|^k$ , то же относится к выборочным моментам. В частности, при  $F \in \mathcal{F}_k$  выборочные моменты есть несмещенные состоятельные оценки моментов  $E_F(Z)$  случайной величины  $X$ , а при  $F \in \mathcal{F}_{2k}$  они также  $\sqrt{n}$ -состоятельны и асимптотически нормальны с асимптотической дисперсией  $\Delta^2(F) = D_F(Z) = E_F(Z^2) - (E_F(Z))^2$ .

Итак, выборочные начальные моменты порядка  $k$  при больших  $n$  приближают неизвестные начальные моменты с точностью порядка  $1/\sqrt{n}$ , однако для более детального анализа точности требуется оценивать моменты порядка  $2k$ .

### 13.3.3 Выборочная дисперсия

Несколько более сложная ситуация с выборочными центральными моментами. Рассмотрим выборочный центральный момент 2-го порядка, то есть выборочную дисперсию:

$$S_n^2 = D(X_n^*) = D(F_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \left( \frac{1}{n} \sum_{i=1}^n X_i^2 \right) - (\bar{X}_n)^2. \quad (13.32)$$

Пусть  $F \in \mathcal{F}_2$ . Вычислим математическое ожидание статистики  $S_n^2$ . Переходя к центрированным случайным величинам  $\tilde{X}_i = X_i - m(F)$  (в силу свойств дисперсии это не влияет ни на дисперсию, ни на выборочную дисперсию), можно считать, что

$$m(F) = 0, \quad D(F) = E_F(X^2)$$

и с учетом (13.24) имеем:

$$E_F(\bar{X}_n) = 0; \quad E_F((\bar{X}_n)^2) = D_F(\bar{X}_n) + (E_F(\bar{X}_n))^2 = D(F)/n.$$

Отсюда следует, что

$$E_F S_n^2 = E_F \left( \frac{1}{n} \sum_{i=1}^n X_i^2 \right) - E_F((\bar{X}_n)^2) = D(F) - \frac{1}{n} D(F) = \frac{n-1}{n} D(F). \quad (13.33)$$

Соотношение (13.33) показывает, что выборочная дисперсия – смещенная оценка. Ее, однако, легко исправить, перейдя к *несмещенной оценке*:

$$\sigma_n^2 = \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (13.34)$$

которую обычно и используют (см. пример 13.3). Поскольку выборочная дисперсия  $S_n^2$  и "исправленная" выборочная дисперсия  $\sigma_n^2$  выражаются аналогично дисперсии через выборочные моменты 1-го и 2-го порядков, которые сходятся с вероятностью 1 к соответствующим моментам, то и статистики  $S_n^2$  и  $\sigma_n^2$  также сходятся с вероятностью 1 к дисперсии  $D(F)$ . Следовательно, *выборочная дисперсия  $S_n^2$  и "исправленная" выборочная дисперсия  $\sigma_n^2$  есть состоятельные оценки дисперсии  $D(F)$ .*

Аналогично можно показать, что *при  $F \in \mathcal{F}_k$  выборочные центральные моменты порядка  $k$*

$$\mu_n^k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^k; \quad \mu_{n,a}^k = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}_n|^k$$

*есть состоятельные оценки центральных моментов  $\mu^k(F)$  и  $\mu_a^k(F)$ .*

Отметим без доказательства, что *при  $F \in \mathcal{F}_4$  оценки  $S_n^2$  и  $\sigma_n^2$  асимптотически нормальны с дисперсией  $\Delta^2(F) = \mu^4(F) - D^2(F)$ .* Отсюда следует, что *для анализа точности выборочной дисперсии как оценки дисперсии  $D(F)$  требуется оценивать моменты четвертого порядка.*

### 13.3.4 Выборочные медиана и квантили

Пусть функция  $F(t) = F_X(t)$  непрерывна и строго возрастает по  $t$  для всех  $t$ , для которых  $0 < F(t) < 1$ . Тогда для любого  $\alpha \in (0, 1)$  однозначно определена величина  $t_\alpha = t_\alpha(F)$ , для которой

$$F(t_\alpha) = \alpha. \quad (13.35)$$

Величина  $t_\alpha(F) = t_\alpha(X)$  называется  $\alpha$ -*квантилью* функции распределения  $F$  или случайной величины  $X$ .

Если функция распределения имеет участки постоянства, то решения уравнения (13.35) могут образовывать интервал  $(t_\alpha^-, t_\alpha^+]$ . В этом случае полагают  $t_\alpha = (t_\alpha^+ + t_\alpha^-)/2$ . Если функция распределения имеет разрывы, то уравнение (13.35) может не иметь решений. Тогда полагают  $t_\alpha = \max\{t : F(t) \leq \alpha\}$ .

Величина  $\text{med}(F) = \text{med}(X) = t_{1/2}(F)$  называется *медианой* функции распределения  $F$  или случайной величины  $X$ .

Таким образом, для непрерывной  $F(t)$  медиана  $\text{med}(F) = \text{med}(X)$  – это такое значение, что вероятности получить значение случайной величины  $X$  как большее, так и меньшее  $\text{med}(F)$  равны  $1/2$ .

Медиана также является своеобразной характеристикой среднего значения случайной величины  $X$ . Если функция распределения  $F$  симметрична относительно некоторой точки  $t^*$ , то  $t^*$  является медианой и совпадает с математическим ожиданием (если оно существует).

*Выборочной квантилью*  $t_{n,\alpha} = t_\alpha(F_n)$  называют квантиль порядка  $\alpha$  эмпирической функции распределения. *Выборочной медианой*  $\text{med}_n = \text{med}(F_n)$  называют медиану эмпирической функции распределения.

При  $\alpha \in ((k-1)/n, k/n)$  выборочная квантиль  $t_{n,\alpha}$  порядка  $\alpha$  есть член вариационного ряда  $X_{(k)}$  с номером  $k$ , а при  $\alpha = k/n$  выборочная квантиль есть  $t_{n,\alpha} = (X_{(k)} + X_{(k+1)})/2$ . Выборочная медиана имеет вид:

$$\text{med}_n = \begin{cases} X_{(k)} & \text{при } n = 2k - 1 - \text{четном;} \\ (X_{(k)} + X_{(k+1)})/2 & \text{при } n = 2k - \text{нечетном.} \end{cases}$$

Отметим асимптотические свойства выборочных квантилей и выборочной медианы (см. [2]). Пусть плотность распределения  $f(t) = F'(t) > 0$  непрерывна и положительна для таких  $t$ , что  $0 < F(t) < 1$ . Тогда для любого  $\alpha \in (0, 1)$  *выборочная квантиль*  $t_{n,\alpha}$  *есть*  $\sqrt{n}$ -*несмещенная и асимптотически нормальная оценка квантили*  $t_\alpha$  *с дисперсией*  $\Delta^2(F) = \alpha(1-\alpha)/f^2(t_\alpha)$ .

В частности, *выборочная медиана*  $\text{med}_n$  *асимптотически нормальна с дисперсией*  $\Delta^2(F) = 1/(4f^2(t_{1/2}))$ .

### 13.3.5 Выборочная ковариация и корреляция

Выборочная ковариация и корреляция имеют важное значение при анализе зависимостей между компонентами случайного вектора  $X = (X^1, \dots, X^k)^\top$ .

Для простоты обозначений рассмотрим двумерный случайный вектор, то есть систему из двух случайных величин  $(X, Y)$ . Тогда выборочная ковариация  $X$  и  $Y$  имеет вид:

$$k_n = \frac{1}{n} \sum_{i=1}^n ((X_i - \bar{X}_n)(Y_i - \bar{Y}_n)),$$

где  $\bar{X}_n$  и  $\bar{Y}_n$  выборочные средние величин  $X$  и  $Y$ . Аналогично (13.33) можно показать, что

$$E_F(k_n) = \frac{n-1}{n} k_F(X, Y),$$

то есть  $k_n$  – смещенная оценка. Можно перейти к несмещенной оценке  $\tilde{k}_n = nk_n/(n-1)$ . Выборочная корреляция  $X$  и  $Y$  имеет вид:

$$r_n = \frac{k_n}{\sqrt{S_{n,X}^2 S_{n,Y}^2}} = \frac{\tilde{k}_n}{\sigma_{n,X} \sigma_{n,Y}},$$

где  $S_{n,X}^2, S_{n,Y}^2$  – выборочные дисперсии, а  $\sigma_{n,X}^2, \sigma_{n,Y}^2$  – "исправленные" выборочные дисперсии компонент  $X, Y$ .

Можно показать, что *выборочные ковариация и корреляция есть состоятельные,  $\sqrt{n}$ -состоятельные и асимптотически нормальные оценки ковариации  $k_F$  и корреляции  $r_F$  с дисперсиями, соответственно:*

$$\Delta^2(F, k_n) = \mu_F^{22} - k_F^2, \quad \Delta^2(F, r_n) = (\mu_F^{22} - k_F^2)/(\sigma_F^2(X)\sigma_F^2(Y)),$$

где

$$\mu_F^{22} = E_F((X - m(X))^2(Y - m(Y))^2)$$

– смешанный центральный момент 4-го порядка. В частности, если случайные величины  $X$  и  $Y$  независимы, то

$$k_F = r_F = 0, \quad \mu_F^{22} = \sigma_F^2(X)\sigma_F^2(Y); \quad \Delta(F, r) = 1.$$

## 13.4 Метод подстановки в задаче построения оценок

### 13.4.1 Идея метода подстановки

Наиболее распространенные методы построения оценок связаны с различными вариантами *метода подстановки*. Простейший вариант этого метода уже встречался в задаче оценки характеристик распределения генеральной совокупности: он состоит в использовании выборочных характеристик  $g(F_n)$  в качестве оценок характеристики  $g(F)$ , поскольку эмпирическая функция распределения  $F_n$  в соответствии с результатами раздела 12 есть хорошая оценка неизвестной функции распределения  $F$  генеральной совокупности. Аналогично, если  $\hat{\theta}_n$  – достаточно хорошая оценка параметра  $\theta \in \Theta$ , метод подстановки рекомендует использовать  $\hat{g}_n = g(\hat{\theta}_n)$  в качестве оценки характеристики  $g(\theta)$ . Ниже мы увидим, что в случае конечномерного параметра и построения оценок  $\hat{\theta}_n = \theta_n^*$  на основе метода максимального правдоподобия этот прием приводит к асимптотически оптимальным оценкам для широкого класса функций  $g(\theta)$ .

Другой вариант метода подстановки для оценки параметров распределения генеральной совокупности состоит в следующем. Пусть неизвестный параметр распределения можно представить как решение уравнения  $E_\theta H(X, \theta) = 0$ , где  $H(x, \theta)$  – заданная функция. Тогда метод подстановки состоит в замене математического ожидания выборочным средним и в использовании в качестве оценки  $\hat{\theta}_n(X^{(n)})$  решения уравнения

$$\sum_{i=1}^n H(X_i, \theta) = 0.$$

На этом варианте метода подстановки основан *метод моментов*.

Еще один вариант метода подстановки связан с рассмотрением экстремальных задач вида

$$\min_{\tilde{\theta} \in \Theta} E_{\theta} H(X, \tilde{\theta}), \quad (13.36)$$

где  $H(x, \theta)$  – такая функция, что минимум достигается при  $\tilde{\theta} = \theta$ . Тогда эта задача заменяется задачей минимизации

$$\min_{\tilde{\theta} \in \Theta} \sum_{i=1}^n H(X_i, \tilde{\theta}) \quad (13.37)$$

и значение  $\hat{\theta}_n(X^{(n)})$ , на котором достигается минимум, принимается в качестве оценки. По существу на этом варианте метода подстановки основан рассматриваемый далее *метод максимального правдоподобия*.

Отметим, что использование методов подстановки само по себе не гарантирует получения хороших оценок: каждый вариант метода требует отдельного исследования.

### 13.4.2 Метод моментов

Пусть распределение  $F_X(t) = F(t, \theta)$  генеральной совокупности  $X$  известно с точностью до  $k$ -мерного вектора параметров  $\theta \in \Theta \subset R^k$ . Тогда любые числовые характеристики (функции)  $g(F_X) = g(\theta)$  генеральной совокупности можно выразить через параметры распределения  $\theta = (\theta_1, \dots, \theta_k)^\top$ . Выберем  $k$ -мерную векторную характеристику  $g(F_X) = (g_1(F_X), \dots, g_k(F_X))^\top$ ,  $g_j(F_X) = g_j(\theta)$ ,  $j = 1, \dots, k$ , такую, что система уравнений относительно векторного параметра  $\theta$

$$g_j(\theta) = g_j, \quad j = 1, \dots, k, \quad (13.38)$$

где  $g = (g_1, \dots, g_k)^\top$  – теоретическое значение характеристики, имеет единственное решение и обладает свойством устойчивости: при малом изменении вектора  $g$  вектор решений  $\theta$  также мало меняется (для этого достаточно непрерывности и невырожденности матрицы частных производных  $g'(\theta) = \|\partial g_j(\theta)/\partial \theta_l\|$ ,  $j, l = 1, \dots, k$ ).

Заменим в системе уравнений (13.38) теоретическое значение характеристики оценкой  $\hat{g}_n = \hat{g}_n(X^{(n)})$ :

$$g_j(\theta) = \hat{g}_{n,j}, \quad j = 1, \dots, k. \quad (13.39)$$

В качестве оценок  $\hat{\theta}_n$  возьмем решения системы уравнений (13.39).

Как правило, в качестве характеристик  $g_j(\theta)$  выбирают моменты генеральной совокупности (начальные или центральные), а в качестве оценок  $\hat{g}_{n,j}$  – соответствующие выборочные моменты. Поэтому описанный метод построения оценок неизвестных параметров обычно называют *методом моментов*, а уравнения (13.39) – *уравнениями метода моментов*.

#### Свойства метода моментов

1. Если  $\hat{g}_n$  – состоятельные оценки характеристики  $g(\theta)$ , то решения  $\hat{\theta}_n$  системы уравнений (13.39) дают состоятельные оценки векторного параметра  $\theta$ . Это свойство вытекает из устойчивости решений системы уравнений (13.38).

2. Если  $\hat{g}_n$  – асимптотически нормальные оценки характеристики  $g(\theta)$ , то оценки  $\hat{\theta}_n$  также асимптотически нормальны.

Простейший вариант метода моментов при одномерном параметре – выражение  $\theta \in R^1$  через математическое ожидание, а оценки  $\hat{\theta}$  – через выборочное среднее. По существу этот метод и был использован в примерах 13.1 и 13.2. Рассмотрим другие примеры.

**Пример 13.4 Оценка параметра показательного распределения.** Для показательного распределения  $\text{Exp}_\lambda = \text{Exp}^u$ ,  $u = 1/\lambda$ , математическое ожидание есть  $E_u(X) = u$ , так что получаем оценку  $\hat{u}_n = \bar{X}_n$ . Эта оценка не смещена, поэтому, так как  $D_u(X) = u^2$ , имеем:

$$D_{n,u}(\hat{u}_n) = \frac{1}{n} D_u(X) = \frac{u^2}{n}.$$

**Пример 13.5 Оценка параметров гамма-распределения.** Пусть генеральная совокупность имеет гамма-распределение, плотность которого

$$f(x, a, \lambda) = \begin{cases} \frac{a^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-ax} & \text{при } x \geq 0; \\ 0 & \text{при } x < 0; \end{cases} \quad \theta = (a, \lambda)^\top, \quad (13.40)$$

где  $\lambda > 0$ ,  $a > 0$ ,  $\Gamma(\lambda)$  – гамма-функция

$$\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx; \quad \Gamma(n) = (n-1)!$$

Параметр  $a$  называют параметром масштаба, параметр  $\lambda$  – параметром формы. Для гамма-распределения имеют место равенства

$$E_\theta(X) = \lambda/a, \quad D_\theta(X) = \lambda/a^2,$$

заменяя в которых теоретические моменты эмпирическими, получим оценки метода моментов:

$$a_n = \bar{X}_n / \sigma_n^2, \quad \lambda_n = \bar{X}_n^2 / \sigma_n^2. \quad (13.41)$$

### 13.4.3 Метод максимального правдоподобия

Этот метод основан на принципе максимального правдоподобия: в качестве оценки неизвестного параметра  $\theta \in \Theta$  по наблюдениям  $x \in \mathcal{X}$  выбирается такое значение параметра, при котором полученные результаты наблюдений наиболее вероятны (правдоподобны).

Будем считать, что  $\Theta$  – область в  $R^k$  и что выполнено одно из двух предположений:

**А.**  $\mathcal{X} \subset R^N$ , и все распределения  $P_\theta$ ,  $\theta \in \Theta$ , имеют плотности распределения  $f(x, \theta)$ ;  $x \in R^N$ ,  $\theta \in \Theta$ ;  $f(x, \theta) = 0$  при  $x \notin \mathcal{X}$  (то есть  $X$  – случайный вектор с непрерывным распределением для всех  $\theta \in \Theta$ );

**Б.** Множество  $\mathcal{X} = \{x^{(1)}, \dots, x^{(j)}, \dots\}$  дискретно. В этом случае обозначим  $f(x, \theta)$ ,  $x \in \mathcal{X}$ , значение вероятности  $f(x^{(j)}, \theta) = P_\theta(X = x^{(j)})$ .

Функцией правдоподобия называется функция

$$L(\theta, X) = f(X, \theta), \quad \theta \in \Theta, \quad X \in \mathcal{X},$$

рассматриваемая как функция от параметра  $\theta \in \Theta$ .

Логарифмической функцией правдоподобия называется функция

$$l(\theta, X) = \ln L(\theta, X) = \ln f(X, \theta), \quad \theta \in \Theta, \quad X \in \mathcal{X}.$$

Функция правдоподобия и логарифмическая функция правдоподобия при заданном значении  $x \in \mathcal{X}$  есть обычные функции параметра  $\theta \in \Theta$ , а при случайном  $X \in \mathcal{X}$  – это случайные функции  $\theta \in \Theta$ .

Для модели независимой однородной выборки функция правдоподобия и логарифмическая функция правдоподобия имеют вид:

$$L(\theta, X^{(n)}) = \prod_{i=1}^n f(X_i, \theta); \quad l(\theta, X^{(n)}) = \sum_{i=1}^n \ln f(X_i, \theta); \quad X^{(n)} \in \mathcal{X}^n, \quad (13.42)$$

где  $f(x, \theta)$  – плотность распределения или вероятности значений генеральной совокупности  $X$ .

Отметим, что если генеральная совокупность  $X$  имеет дискретное распределение с возможными значениями  $\mathcal{X} = \{x^{(1)}, \dots, x^{(j)}, \dots\}$ , то

$$L(\theta, X^{(n)}) = \prod_j f(x^{(j)}, \theta)^{n_j}, \quad l(\theta, X^{(n)}) = \sum_j n_j \ln f(x^{(j)}, \theta), \quad (13.43)$$

где  $n_j$  – число элементов в выборке  $X^{(n)} = (X_1, \dots, X_n)$ , принимающих значение  $x^{(j)}$ .

Оценкой максимального правдоподобия называется значение параметра  $\theta^* = \theta^*(X) \in \Theta$ , доставляющее максимум функции правдоподобия, то есть такое значение  $\theta^*(X) \in \Theta$ , что

$$L(\theta^*, X) = \max_{\theta \in \Theta} L(\theta, X).$$

Очевидно, что при  $L(\theta, X) > 0$  функцию правдоподобия можно заменить любой возрастающей функцией от нее, например логарифмической функцией правдоподобия, то есть значение  $\theta^*(X)$  можно определить из условия

$$l(\theta^*, X) = \max_{\theta \in \Theta} l(\theta, X); \quad l(\theta, X) = \ln L(\theta, X).$$

Метод максимального правдоподобия можно рассматривать как вариант метода подстановки (13.36), (13.37) при выборе

$$H(X, \tilde{\theta}) = -\ln f(X, \tilde{\theta}).$$

Пусть  $\Theta$  – область в  $R^k$  и логарифмическая функция правдоподобия непрерывно дифференцируема. Тогда оценка максимального правдоподобия (если она существует) есть одно из решений векторного уравнения  $l'_\theta(\theta, X) = 0$ , то есть системы уравнений

$$\frac{\partial}{\partial \theta_j} l(\theta, X) = 0, \quad j = 1, \dots, k,$$

которые называются *уравнениями максимального правдоподобия*.

При оценке параметров распределения по выборке  $X^{(n)} = (X_1, \dots, X_n)$  уравнения максимального правдоподобия имеют вид:

$$l'_\theta(\theta, X^{(n)}) = \sum_{i=1}^n l'_\theta(\theta, X_i) = 0, \quad (13.44)$$

где  $l'(\theta, X)$  – логарифмическая функция правдоподобия генеральной совокупности  $X$ , соответствующая одному наблюдению.

Вообще говоря, решения уравнений максимального правдоподобия могут давать не только максимумы (локальные) функции правдоподобия, но также минимумы и седловые точки. Поэтому среди решений нужно выбрать такие, для которых выполнено условие отрицательной определенности матрицы вторых производных:  $-l''(\theta, X) > 0$ , что дает локальные максимумы и среди них – значение, обеспечивающее глобальный максимум. Часто уравнения максимального правдоподобия имеют единственное решение, что существенно упрощает анализ.

Можно показать, что при достаточно общих предположениях регулярности оценки максимального правдоподобия (ОМП) являются состоятельными, асимптотически  $\sqrt{n}$ -несмещенными и обладают различными свойствами асимптотической оптимальности (подробнее эти вопросы будут рассмотрены ниже). В частности, при оценке параметров распределения генеральной совокупности  $X$  в случае одномерного параметра *ОМП асимптотически нормальны с дисперсией  $\Delta^2(\theta, \theta_n^*) = 1/I(\theta)$* , где

$$I(\theta) = E_\theta(l'(\theta, X))^2.$$

Функция  $I(\theta)$  называется *информацией Фишера*.

**Пример 13.6 Оценка неизвестной вероятности события.** (Ср. с примером 13.1.). Пусть проводится  $n$  независимых экспериментов, в каждом из которых может произойти или не произойти случайное событие  $A$ , вероятность которого  $P(A) = p \in (0, 1)$  неизвестна. Пусть  $X$  – индикаторная функция события  $A$ , то есть  $X = 1$ , если событие  $A$  произошло, и  $X = 0$ , если событие  $A$  не произошло;  $X_i$  – индикаторная функция события  $A$  в  $i$ -м эксперименте,  $i = 1, \dots, n$ . Таким образом, результаты экспериментов мы можем рассматривать как независимую выборку из генеральной совокупности  $X$  с дискретным бернуллиевским распределением:  $P(X = 1) = p$ ,  $P(X = 0) = q = 1 - p$ .

Вычислим информацию Фишера. Для одного наблюдения  $\mathcal{X} = \{0, 1\}$ , и выражение для  $f(x, p)$  можно представить в виде:

$$f(x, p) = p^x(1 - p)^{1-x},$$

где  $x$  принимает значения 0 или 1; откуда

$$\begin{aligned} l(x, p) &= x \ln p + (1 - x) \ln(1 - p); \quad l'_p(x, p) = \frac{x - p}{p(1 - p)}; \\ l''(0, p) &= -(1 - p)^{-2}, \quad l''(1, p) = -p^{-2}; \\ I(p) &= E_p(l'(X, p))^2 = \frac{1}{(p(1 - p))^2} E_p(X - p)^2 = \frac{1}{p(1 - p)}, \end{aligned}$$



поскольку  $E_p(X - p)^2 = D_p(X) = p(1 - p)$ . Таким образом,

$$I(p) = \frac{1}{p(1 - p)}.$$

Уравнение максимального правдоподобия и ОМП  $p_n^*$  имеют вид:

$$\sum_{i=1}^n \frac{X_i - p}{p(1 - p)} = \frac{k - np}{p(1 - p)} = 0, \quad p_n^* = \frac{k}{n},$$

где  $k$  – число успехов в серии из  $n$  испытаний, то есть  $p_n^*$  есть наблюдаемая частота  $k/n$ . Как было показано в подразд. 13.1 (см. формулу (13.3)),  $p_n^*$  есть несмещенная оценка  $p$  и

$$D_{n,p}(p_n^*) = p(1 - p)/n = 1/(nI(p)).$$

**Пример 13.7** Оценка неизвестного среднего нормального распределения с известной дисперсией. (Ср. с примером 13.2). Пусть  $X \sim N(a, \sigma^2)$ , дисперсия  $\sigma^2 > 0$  известна, среднее  $a$  есть неизвестный параметр. Вычислим вначале информацию Фишера. Для одного наблюдения

$$\begin{aligned} f(x, a) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - a)^2}{2\sigma^2}\right), \quad l(a, x) = \text{const} - \frac{(x - a)^2}{2\sigma^2}; \\ l'(a, x) &= \frac{x - a}{\sigma^2}; \quad I(a) = E_a(l'(a, X))^2 = \frac{1}{\sigma^4} E_a(X - a)^2 = \frac{1}{\sigma^2}, \end{aligned}$$

так что  $I(a) = \sigma^{-2}$ . Уравнение максимального правдоподобия и его решение  $a_n^*$  имеют вид:

$$\sum_{i=1}^n \frac{X_i - a}{\sigma^2} = 0, \quad a_n^* = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i;$$

то есть  $a_n^* = \bar{X}_n$  – выборочное среднее. Как было показано выше,  $\bar{X}_n$  – несмещенная оценка, следовательно (см. (13.5)),

$$D_{n,a}(\bar{X}_n) = \sigma^2/n = 1/(nI(a)).$$

## 13.5 Задача доверительного оценивания

### 13.5.1 Принцип построения доверительных областей

Методы построения доверительных областей основаны на следующем принципе. Пусть нам удалось найти такую случайную величину  $Y = h(X, g(\theta))$ , где  $X$  – случайные данные, что  $P_\theta$ -функция распределения  $F_Y(t) = P_\theta(Y < t)$  непрерывна и не зависит от  $\theta \in \Theta$ . Тогда доверительная область  $\tilde{G}_\gamma = \tilde{G}_\gamma(X)$  определяется наблюдаемыми данными  $X$  и состоит из таких значений  $g \in G$ , которые удовлетворяют неравенству:  $h(X, g) \leq H_\gamma$ ; пороговое значение  $H_\gamma$  определяется из условия  $F_Y(H_\gamma) = \gamma$ .

Вернемся к примерам 13.2 и 13.7. Пусть величина  $d = \sigma^2$  известна. Тогда величина  $h(X^{(n)}, a) = \sqrt{n}|\bar{X}_n - a|/\sigma$  распределена как  $|\xi|$ , где  $\xi \sim N(0, 1)$ , так что доверительный интервал  $\tilde{A}_\gamma$  для неизвестного среднего  $a$  нормального распределения при известном  $\sigma$  имеет вид:

$$\tilde{A}_\gamma = [a_{n,\gamma}^-, a_{n,\gamma}^+], \quad a_{n,\gamma}^\pm = \bar{X}_n \pm \delta_{n,\gamma}, \quad \delta_{n,\gamma} = \frac{t_\gamma \sigma}{\sqrt{n}}, \quad (13.45)$$

где величина  $t_\gamma$  выбирается по уровню надежности  $\gamma$  из соотношения  $P(|\xi| < t_\gamma) = \Phi(t_\gamma) - \Phi(-t_\gamma) = 2\Phi(t_\gamma) - 1 = \gamma$ , то есть

$$\Phi(t_\gamma) = (1 + \gamma)/2. \quad (13.46)$$

Если величина  $d$  неизвестна, можно использовать вместо нее оценку  $\sigma_n^2$  вида (13.6). Известно (см. [2], [7]), что при этом распределение величины  $\eta_n = \sqrt{n}(\bar{X}_n - a)/\sigma_n$  не зависит от неизвестных параметров  $\theta = (a, \sigma^2)$ :

$$P_{n,\theta}(\eta_n < t) = S_{n-1}(t),$$

где  $S_{n-1}(t)$  есть функция распределения Стьюдента с  $n - 1$  степенью свободы. Отметим, что распределение Стьюдента симметрично:  $S_{n-1}(t) = 1 - S_{n-1}(-t)$ .

Кроме того, распределение Стьюдента асимптотически нормально при  $n \rightarrow \infty$ . Действительно, в силу закона больших чисел  $\sigma_n \rightarrow \sigma$  по вероятности, поэтому при  $n \rightarrow \infty$

$$P_{n,\theta}(\eta_n < t) = P_{n,\theta}(\sqrt{n}(\bar{X}_n - a)/\sigma < t) + o(1) = P(\xi < t) + o(1) \rightarrow \Phi(t), \quad \xi \sim N(0, 1),$$

так как случайная величина  $\sqrt{n}(\bar{X}_n - a)/\sigma$  имеет стандартное нормальное распределение. Напомним, что через  $o(1)$  обозначают величину, стремящуюся в рассматриваемом предельном переходе к нулю.

Положим  $h(X^{(n)}, a) = |\eta_n| = \sqrt{n}|\bar{X}_n - a|/\sigma_n$ . Тогда доверительный интервал для неизвестного среднего  $a$  нормального распределения при неизвестном  $\sigma$  также имеет вид, аналогичный (13.45):

$$\tilde{A}_\gamma = [a_{n,\gamma}^-, a_{n,\gamma}^+], \quad a_{n,\gamma}^{pm} = \bar{X}_n \pm \delta_{n,\gamma}, \quad \delta_{n,\gamma} = \frac{t_{n,\gamma} \sigma_n}{\sqrt{n}}, \quad (13.47)$$

с тем отличием, что величина  $t_{n,\gamma}$  выбирается из соотношения

$$P(|\eta_n| < t_{n,\gamma}) = S_{n-1}(t_{n,\gamma}) - S_{n-1}(-t_{n,\gamma}) = 2S_{n-1}(t_{n,\gamma}) - 1 = \gamma,$$

то есть

$$S_{n-1}(t_{n,\gamma}) = (1 + \gamma)/2. \quad (13.48)$$

Таблицы величин  $t_{n,\gamma}$ , определяемых (13.48), имеются, например, в [2], [7]. При  $n > 30$  величины  $t_{n,\gamma}$  практически не зависят от  $n$  и совпадают с величиной  $t_\gamma$ , определяемой соотношением (13.46), что и соответствует асимптотической нормальности распределения Стьюдента.

В задаче оценки дисперсии нормального распределения при неизвестном среднем можно положить:

$$h(X^{(n)}, d) = (n-1)|(\sigma_n^2/\sigma^2) - 1| = |\tau_n - n + 1|.$$

Случайная величина  $\tau_n = (n-1)\sigma_n^2/\sigma^2$  имеет распределение хи-квадрат с  $n-1$  степенью свободы. Поэтому пороговое значение  $H_{n,\gamma}$  можно определить из условия:

$$\begin{aligned} P(|\tau_n - n + 1| \leq H_{n,\gamma}) &= P(n-1-H_{n,\gamma} \leq \tau_n \leq n-1+H_{n,\gamma}) = \\ &= \chi_{n-1}^2(n-1+H_{n,\gamma}) - \chi_{n-1}^2(n-1-H_{n,\gamma}) = \gamma. \end{aligned}$$

Это приводит к доверительному интервалу вида

$$\tilde{D}_\gamma = [d_{n,\gamma}^-, d_{n,\gamma}^+], \quad d_{n,\gamma}^\pm = \frac{\sigma_n^2}{1 \mp H_{n,\gamma}/(n-1)}. \quad (13.49)$$

В силу асимптотической нормальности распределения хи-квадрат при  $n \rightarrow \infty$  пороговое значение  $H_{n,\gamma}$  можно представить в виде  $H_{n,\gamma} \sim t_\gamma \sqrt{2(n-1)}$ , где величина  $t_\gamma$  определяется соотношением (13.46). Поэтому границы доверительного интервала при больших  $n$  асимптотически имеют вид:

$$d_{n,\gamma}^\pm = \frac{\sigma_n^2}{1 \mp t_\gamma \sqrt{2/(n-1)}}. \quad (13.50)$$

В общем случае методы выбора функции  $h(X, g(\theta))$  зависят от конкретной модели. Достаточно общие методы можно предложить в рамках асимптотического подхода.

### 13.5.2 Асимптотические доверительные интервалы

Рассмотрим задачу доверительного оценивания одномерного параметра  $\theta \in \Theta \subset R^1$ . Пусть  $\hat{\theta}_n$  – асимптотически нормальная последовательность точечных оценок с дисперсией  $\Delta^2(\theta)$ , то есть последовательность случайных величин  $Y_n = \sqrt{n}(\hat{\theta}_n - \theta)$  асимптотически нормальна по  $P_{n,\theta}$ -распределению:

$$Y_n \xrightarrow{P_{n,\theta}} Y \sim N(0, \Delta^2(\theta)), \quad n \rightarrow \infty.$$

Отсюда следует, что  $Z_n = Y_n/\Delta(\theta) \xrightarrow{P_{n,\theta}} \xi \sim N(0, 1)$ ,  $n \rightarrow \infty$ , так что

$$P_{n,\theta}(|Z_n| \leq t) = P_{n,\theta}(-t \leq Z_n \leq t) \rightarrow P(-t \leq \xi \leq t) = 2\Phi(t) - 1. \quad (13.51)$$

Пусть множество  $\tilde{\Theta}_{n,\gamma}$  состоит из таких  $\theta \in \Theta$ , что выполнено неравенство

$$|Z_n| = \sqrt{n}|\hat{\theta}_n - \theta|/\Delta(\theta) \leq t_\gamma, \quad (13.52)$$

где величина  $t_\gamma$  определяется из условия (13.46). Тогда из соотношения (13.51) следует, что для любого  $\theta \in \Theta$  при  $n \rightarrow \infty$

$$P_{n,\theta}(\theta \in \tilde{\Theta}_{n,\gamma}) = P_{n,\theta}(|Z_n| \leq t_\gamma) \rightarrow \gamma.$$

Это соотношение означает, что  $\tilde{\Theta}_{n,\gamma}$  есть асимптотическое доверительное множество уровня надежности  $\gamma$ .

Эти рассуждения без каких-либо изменений переносятся на задачу доверительного оценивания одной характеристики  $g(\theta)$ , если  $\hat{g}_n$  – асимптотически нормальная последовательность точечных оценок с дисперсией вида  $\Delta^2(\theta) = \Delta^2(g(\theta))$ , которая зависит лишь от оцениваемой характеристики  $g(\theta)$ . При этом асимптотическое доверительное множество  $\tilde{G}_{n,\gamma}$  уровня надежности  $\gamma$  определяется неравенством, аналогичным (13.52):

$$|Z_n| = \sqrt{n}|\hat{g}_n - g(\theta)|/\Delta(g(\theta)) \leq t_\gamma. \quad (13.53)$$

Обычно множества  $\tilde{\Theta}_{n,\gamma}$  и  $\tilde{G}_{n,\gamma}$  представляют собой интервалы:

$$\tilde{\Theta}_{n,\gamma} = [\theta_{n,\gamma}^-, \theta_{n,\gamma}^+]; \quad \tilde{G}_{n,\gamma} = [g_{n,\gamma}^-, g_{n,\gamma}^+],$$

граничные точки которых определяются как решения уравнений:

$$\sqrt{n}(\theta_{n,\gamma}^\pm - \hat{\theta}_n)/\Delta(\theta_{n,\gamma}^\pm) = \pm t_\gamma; \quad \sqrt{n}(g_{n,\gamma}^\pm - \hat{g}_n)/\Delta(g_{n,\gamma}^\pm) = \pm t_\gamma.$$

В примере 13.1 (оценка неизвестной вероятности события) при использовании оценок  $\hat{\theta}_n = k/n$  этот подход приводит к квадратным уравнениям для концов асимптотических доверительных интервалов  $\tilde{\Theta}_{n,\gamma} = [\theta_{n,\gamma}^-, \theta_{n,\gamma}^+]$ :

$$(\hat{\theta}_n - \theta_{n,\gamma}^\pm)^2 = \frac{t_\gamma^2 \theta_{n,\gamma}^\pm (1 - \theta_{n,\gamma}^\pm)}{n},$$

решения которого имеют вид:

$$\theta_{n,\gamma}^\pm = \hat{\theta}_n \frac{n}{n + t_\gamma^2} + t_\gamma \frac{t_\gamma \pm \sqrt{t_\gamma^2 + 4n\hat{\theta}_n(1 - \hat{\theta}_n)}}{2(n + t_\gamma^2)}, \quad \hat{\theta}_n = \frac{k}{n}. \quad (13.54)$$

В примере 13.3 (оценка неизвестной дисперсии нормального распределения при неизвестном среднем) при использовании оценок  $\hat{d}_n = \sigma_n^2$  вида (13.6) мы получаем асимптотические доверительные интервалы с концами

$$d_{n,\gamma}^\pm = \frac{\sigma_n^2}{1 \mp t_\gamma \sqrt{2/n}}. \quad (13.55)$$

Нетрудно видеть, что величины, определяемые (13.55) и (13.50), асимптотически эквивалентны (их отношение и разность есть величины порядка  $n^{-1}$ ).

Рассмотрим задачу доверительного оценивания одной характеристики  $g(\theta)$  в общем случае. Пусть  $g_n$  – асимптотически нормальная последовательность точечных оценок с дисперсией  $\Delta^2(\theta)$ . Пусть также имеется последовательность  $\Delta_n$  состоятельных оценок характеристики  $\Delta(\theta)$ . Тогда последовательность случайных величин  $Z_n = \sqrt{n}(g_n - g(\theta))/\Delta_n$  асимптотически нормальна по  $P_{n,\theta}$ -распределению, и, следовательно,

$$|Z_n| \xrightarrow{P_{n,\theta}} |\xi|, \quad \xi \sim N(0, 1).$$

Поэтому, полагая

$$g_n^\pm = g_n \pm \delta(n, \gamma); \quad \delta(n, \gamma) = t_\gamma \Delta_n / \sqrt{n},$$

где величина  $t_\gamma$  определяется из условия (13.46), имеем:

$$\begin{aligned} P_{n,\theta}(g_n^- < g(\theta) < g_n^+) &= P_{n,\theta}(|Z_n| < t_\gamma) \rightarrow P(|\xi| < t_\gamma) = \\ &= \Phi(t_\gamma) - \Phi(-t_\gamma) = 2\Phi(t_\gamma) - 1 = \gamma. \end{aligned}$$

Это означает, что  $\tilde{G}_\gamma = [g_\gamma^-, g_\gamma^+]$  есть асимптотический доверительный интервал надежности  $\gamma$ .

В примере 13.1 можно взять  $\Delta_n = (\hat{\theta}_n(1 - \hat{\theta}_n))^{1/2}$ ,  $\hat{\theta}_n = k/n$ , откуда получим асимптотический доверительный интервал надежности  $\gamma$  для неизвестной вероятности  $\theta = P(A)$  события  $A$ :

$$\tilde{\Theta}_{n,\gamma} = [\theta_n^- \leq \theta \leq \theta_n^+], \quad \theta_n^\pm = \frac{k}{n} \pm \frac{t_\gamma}{n} \sqrt{\frac{k(n-k)}{n}}. \quad (13.56)$$

Величины, определяемые (13.56) и (13.54), асимптотически эквивалентны (их отношение и разность есть величины порядка  $n^{-1}$ ).

В примере 13.2 при неизвестной величине  $\sigma$ , полагая  $\Delta_n = \sigma_n$ , придем к асимптотическому доверительному интервалу надежности  $\gamma$  для неизвестного среднего при неизвестной дисперсии:

$$\tilde{A}_{n,\gamma} = [a_{n,\gamma}^-, a_{n,\gamma}^+]; \quad a_{n,\gamma}^\pm = \bar{X}_n \pm \frac{t_\gamma \sigma_n}{\sqrt{n}}. \quad (13.57)$$

Этот доверительный интервал отличается от точного доверительного интервала (13.47) лишь заменой величины  $t_{n,\gamma}$  на  $t_\gamma$ , что допустимо, как отмечалось выше, при  $n > 30$ .

Аналогично в примере 13.3 при оценке неизвестной дисперсии  $d = \sigma^2$  нормального распределения при неизвестном среднем  $a$  мы получим доверительный интервал вида

$$\tilde{D}_{n,\gamma} = [d_{n,\gamma}^-, d_{n,\gamma}^+]; \quad d_{n,\gamma}^\pm = d_n(1 \pm t_\gamma \sqrt{2/n}). \quad (13.58)$$

Сопоставляя этот доверительный интервал с (13.55), видим, что они асимптотически эквивалентны, поскольку

$$(1 \mp t_\gamma \sqrt{2/n})^{-1} = 1 \pm t_\gamma \sqrt{2/n} + O(n^{-1}).$$

Рассмотрим задачу доверительного оценивания дисперсии при неизвестной функции распределения  $F = F_X$  генеральной совокупности, предполагая лишь существование моментов 4-го порядка:  $F \in \mathcal{F}_4$ . Используя результаты раздела 13.3.3, в качестве асимптотического доверительного интервала надежности  $\gamma$  можно выбрать интервал

$$\tilde{D}_{n,\gamma} = [d_{n,\gamma}^-, d_{n,\gamma}^+]; \quad d_{n,\gamma}^\pm = d_n \pm t_\gamma \sqrt{(\mu_n^4 - d_n^2)/n},$$

где  $\mu_n^4$  — выборочный центральный момент 4-го порядка. Если генеральная совокупность имеет нормальное распределение  $N(a, d)$ , то при  $n \rightarrow \infty$  с  $P_F$ -вероятностью 1

$$\sqrt{\mu_n^4 - d_n^2} \rightarrow \sqrt{\mu_4(F) - D^2(F)} = \sqrt{3d^2 - d^2} = \sqrt{2}d,$$

так что эти доверительные интервалы асимптотически эквивалентны.

### 13.5.3 Доверительное оценивание на основе ОМП

Результаты п. 13.5.2 позволяют строить асимптотические доверительные интервалы для неизвестных параметров  $\theta$  и характеристик  $g(\theta)$  на основе оценок максимального правдоподобия.

Пусть параметр  $\theta$  одномерный. В качестве асимптотического доверительного множества  $\tilde{\Theta}_{n,\gamma}$  надежности  $\gamma$  можно выбрать множество, состоящее из таких  $\theta \in \Theta$ , что выполнено неравенство

$$\sqrt{nI(\theta)}|\theta_n^* - \theta| \leq t_\gamma, \quad (13.59)$$

где величина  $t_\gamma$  определяется из условия  $\Phi(t_\gamma) = (1 + \gamma)/2$ . Заменяя в (13.59) неизвестное значение  $I(\theta)$  величиной  $I(\theta_n^*) \rightarrow I(\theta)$ , получим доверительный интервал  $\tilde{\Theta}_{n,\gamma} = (\theta_{n,-}^*, \theta_{n,+}^*)$  с центром в точке  $\theta_n^*$  и длиной  $2\delta_n = 2t_\gamma/\sqrt{nI(\theta_n^*)}$ :

$$\theta_{n,\pm}^* = \theta_n^* \pm \delta_n. \quad (13.60)$$

Если  $g(\theta)$  – непрерывно дифференцируемая функция  $\theta$ , то аналогично (13.60) получим асимптотический доверительный интервал для характеристики  $g(\theta)$  надежности  $\gamma$ :

$$\tilde{G}_{n,\gamma} = [g_{n,-}^*, g_{n,+}^*], \quad g_{n,\pm}^* = g(\theta_n^*) \pm \delta_n, \quad \delta_n = \frac{t_\gamma}{g'(\theta_n^*)\sqrt{nI(\theta_n^*)}}. \quad (13.61)$$

Применяя этот метод в задачах доверительного оценивания неизвестного среднего, дисперсии нормального распределения или неизвестной вероятности, мы получим те же доверительные интервалы, что и в п. 13.5.2.

## 14 Оптимальное и асимптотически оптимальное оценивание

В этом разделе мы рассмотрим задачи оценивания конечномерного параметра распределения  $P_\theta$ ,  $\theta \in \Theta \subset R^k$ , а также характеристик (функций)  $g(\theta) \in G \subset R^m$ ,  $\theta \in \Theta$  по наблюдениям  $X \in \mathcal{X}$ .

### 14.1 О сравнении качества оценок

Сравнивать различные оценки можно пытаться с помощью функции риска. *Говорят, что оценка  $\hat{g}^1$  не хуже оценки  $\hat{g}^2$ , если  $R(\hat{g}^1, \theta) \leq R(\hat{g}^2, \theta)$  для всех  $\theta \in \Theta$  (это обозначают  $\hat{g}^1 \succeq \hat{g}^2$ ).*

К сожалению, такое сравнение возможно не для всех оценок: может быть  $R(\hat{g}^1, \theta_1) < R(\hat{g}^2, \theta_1)$  для одних значений  $\theta_1 \in \Theta$ , в то время как  $R(\hat{g}^1, \theta_2) > R(\hat{g}^2, \theta_2)$  для других значений  $\theta_2 \in \Theta$ .

Пусть  $\mathcal{G} = \{\hat{g}\}$  – некоторый класс оценок. *Говорят, что оценка  $\hat{g}^*$  эффективна в классе  $\mathcal{G}$ , если  $\hat{g}^* \succeq \hat{g}$  для всех  $\hat{g} \in \mathcal{G}$ .*

В классе *всех* оценок не существует эффективной оценки. Действительно, для такой оценки должно выполняться неравенство:

$$R(\hat{g}^*, \theta) \leq \min_{\hat{g}} R(\hat{g}, \theta), \quad \forall \theta \in \Theta. \quad (14.1)$$

Однако для любого заданного  $\theta_0 \in \Theta$  можно рассмотреть вырожденную оценку  $\hat{g}_0(x) \equiv g(\theta_0)$ , для которой  $R(\hat{g}_0, \theta_0) = 0$ . Поэтому минимум в правой части (14.1) есть 0, так что должно выполняться равенство:  $R(\hat{g}^*, \theta) \equiv 0$  для всех  $\theta \in \Theta$ , а это, как правило, невозможно.

Для поиска эффективных оценок нужны ограничения на класс рассматриваемых оценок. Так, условие несмещенности заведомо исключает "вырожденные" оценки. В классе несмещенных оценок иногда удается найти эффективные оценки.

При асимптотическом подходе естественно рассматривать асимптотически  $\sqrt{n}$ -несмещенные, в частности асимптотически нормальные, оценки. Их можно сравнивать по величине дисперсии или нормирующего множителя  $\Delta(\theta)$ , а именно:  $\hat{g}_n^1 \succeq \hat{g}_n^2$ , если  $\Delta^{(1)}(\theta) \leq \Delta^{(2)}(\theta)$  для всех  $\theta \in \Theta$ . Можно было бы назвать оценку  $\hat{g}^*$  асимптотически эффективной в классе асимптотически нормальных оценок, если  $\hat{g}_n^* \succeq \hat{g}_n$  для всех асимптотически нормальных оценок  $\hat{g}_n$ .

Такой подход к сравнению оценок был предложен одним из основоположников современной статистики Р. Фишером. К сожалению, непосредственно его реализовать не удастся: оказывается, многие естественные оценки можно улучшать в отдельных точках, уменьшая асимптотическую дисперсию (такие оценки называют *суперэффективными*).

Отмеченные трудности сравнения и выбора наилучших оценок на основе функции риска оценки заставляют несколько иначе подойти к проблеме сравнения: нужно сравнивать не функции риска разных оценок, а какие-нибудь числовые величины (функционалы) от функции риска, которые "в целом" характеризуют функцию риска оценки. Опишем один из вариантов такого подхода.

#### 14.1.1 Минимаксный подход

При минимаксном подходе качество оценки характеризуется максимальным значением риска

$$R_{\max}(\hat{g}) = \sup_{\theta \in \Theta} R(\hat{g}, \theta).$$

Оценка  $\hat{g}$  называется *минимаксной*, если  $R_{\max}(\hat{g}) \leq R_{\max}(\tilde{g})$  для любой оценки  $\tilde{g}$ .

Таким образом, минимаксный подход ориентирован на построение оценок с минимальным значением максимального риска. Иногда такой подход считают слишком пессимистичным. Однако в большинстве случаев, особенно при асимптотическом подходе, минимаксные оценки являются вполне разумными.

Минимаксные оценки существуют при достаточно общих предположениях. В общем случае строить минимаксные оценки достаточно сложно.

### 14.1.2 Асимптотически минимаксные оценки

Оценка  $\hat{g}_n$  называется *асимптотически минимаксной*, если

$$\lim_{n \rightarrow \infty} R_{\max}(\hat{g}_n)/R_{\max}(\tilde{g}_n) \leq 1$$

для любой оценки  $\tilde{g}_n$ .

Оценка  $\hat{g}_n$  называется *локально асимптотически минимаксной* в точке  $\theta_0 \in \Theta$ , если для достаточно малых  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \sup_{\|\theta - \theta_0\| \leq \varepsilon} R(\hat{g}_n, \theta) / \sup_{\|\theta - \theta_0\| \leq \varepsilon} R(\tilde{g}_n, \theta) \leq 1$$

для любой оценки  $\tilde{g}_n$  (то есть при локальном подходе сравниваются максимумы риска в малой окрестности точки  $\theta_0$ ).

## 14.2 Свойства функции правдоподобия (одномерный параметр)

Пусть  $\Theta$  – конечный или бесконечный интервал на прямой  $R^1$ . Будем считать также, что выполнено одно из двух предположений **A** или **B** п. 13.4.3. При этом в силу свойств плотности или вероятности

$$f(x, \theta) \geq 0, \int_{\mathcal{X}} f(x, \theta) dx = 1 \text{ или } f(x, \theta) \geq 0, \sum_{x \in \mathcal{X}} f(x, \theta) = 1. \quad (14.2)$$

Также будем предполагать выполненными *условия регулярности*: логарифмическая функция правдоподобия – дважды непрерывно дифференцируемая по  $\theta$  функция для всех  $x \in \mathcal{X}$ , и при дифференцировании интегралов или сумм (14.2) допустимо изменение порядка интегрирования и дифференцирования.

Рассмотрим производную по  $\theta$  логарифмической функции правдоподобия

$$l'_\theta(\theta, x) = \frac{\partial}{\partial \theta} l(\theta, x) = f'_\theta(x, \theta) / f(x, \theta)$$

как случайную функцию от  $\theta \in \Theta$  и вычислим ее математическое ожидание  $E_\theta l'(\theta, X)$ . Для этого при непрерывном распределении  $X$  продифференцируем первое равенство (14.2) и воспользуемся условиями регулярности:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \int_{\mathcal{X}} f(x, \theta) dx = \int_{\mathcal{X}} \frac{\partial}{\partial \theta} f(x, \theta) dx = \\ &= \int_{\mathcal{X}} l'(\theta, x) f(x, \theta) dx = E_\theta l'(\theta, X). \end{aligned} \quad (14.3)$$

Второй раз дифференцируя то же равенство, получим:

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \int_{\mathcal{X}} l'(\theta, x) f(x, \theta) dx = \int_{\mathcal{X}} \frac{\partial}{\partial \theta} (l'(\theta, x) f(x, \theta)) dx = \\ &= \int_{\mathcal{X}} ((l'(\theta, x))^2 + l''(\theta, x)) f(x, \theta) dx, \end{aligned}$$



где  $l''(\theta, x) = \frac{\partial^2}{\partial \theta^2} l(\theta, x)$ . Отсюда получаем равенство:

$$E_{\theta}(l'(\theta, X))^2 = -E_{\theta}l''(\theta, X). \quad (14.4)$$

При дискретном распределении  $X$  мы дифференцируем второе равенство (14.2) и получаем те же результаты.

Напомним, что *информацией Фишера* называется функция  $I(\theta) = E_{\theta}(l'(\theta, X))^2$ . В силу (14.3) информацию Фишера также можно представить в виде дисперсии логарифмической производной

$$I(\theta) = D_{\theta}(l'(\theta, X)). \quad (14.5)$$

Для модели независимой однородной выборки, дифференцируя представление (13.43), получим представление логарифмической производной как суммы независимых случайных величин. Используя свойства дисперсии и (14.5), получим  $I_n(\theta) = nI_1(\theta)$ , где  $I_1(\theta)$  и  $I_n(\theta)$  – информация Фишера для одного и для  $n$  наблюдений. Таким образом, *информация Фишера пропорциональна длине выборки  $n$ .*

### 14.3 Неравенство Рао–Крамера и эффективные оценки

Пусть  $\hat{\theta} = \hat{\theta}(X)$  – несмещенная оценка одномерного параметра, выполнены условия регулярности и  $I(\theta) > 0$  для всех  $\theta \in \Theta$ .

**Неравенство Рао–Крамера.** Для любого  $\theta \in \Theta \subset R^1$

$$D_{\theta}(\hat{\theta}) \geq 1/I(\theta).$$

В частности, для несмещенной оценки  $\hat{\theta}_n$  параметра распределения генеральной совокупности, которая строится по независимой однородной выборке длины  $n$ ,

$$D_{n,\theta}(\hat{\theta}_n) \geq 1/(nI_1(\theta)).$$

*Доказательство.* Будем считать, что наблюдения  $X$  имеют непрерывное распределение (случай дискретного распределения рассматривается аналогично). Запишем условие несмещенности

$$E_{\theta}\hat{\theta} = \int_{\mathcal{X}} \hat{\theta}(x)f(x, \theta)dx = \theta$$

и продифференцируем это равенство:

$$\int_{\mathcal{X}} \hat{\theta}(x)l'(\theta, x)f(x, \theta)dx = 1. \quad (14.6)$$

Умножая равенство (14.3) на  $\theta$  и вычитая из (14.6), получим:

$$\int_{\mathcal{X}} (\hat{\theta}(x) - \theta)l'(\theta, x)f(x, \theta)dx = 1. \quad (14.7)$$

Воспользуемся интегральным неравенством Коши–Буняковского:

$$\left( \int_{\mathcal{X}} g_1(x)g_2(x)dx \right)^2 \leq \int_{\mathcal{X}} g_1^2(x)dx \int_{\mathcal{X}} g_2^2(x)dx, \quad (14.8)$$

справедливым для любых функций  $g_1(x)$ ,  $g_2(x)$ , квадрат которых интегрируем. Возводя в квадрат равенство (14.7) и применяя это неравенство к функциям

$$g_1(x) = (\hat{\theta}(x) - \theta)(f(x, \theta))^{1/2}, \quad g_2(x) = l'(x, \theta)(f(x, \theta))^{1/2},$$

получим требуемое неравенство:

$$1 \leq \int_{\mathcal{X}} (\hat{\theta}(x) - \theta)^2 f(x, \theta) dx \int_{\mathcal{X}} (l'(\theta, x))^2 f(x, \theta) dx = D_{\theta}(\hat{\theta})I(\theta).$$

Неравенство Рао–Крамера дает нижнюю границу для дисперсии и квадратичного риска несмещенных оценок. Оценка, на которой достигается нижняя граница Рао–Крамера, называется *эффективной*. Таким образом, для эффективной несмещенной оценки

$$D_{\theta}(\hat{\theta}) = 1/I(\theta). \quad (14.9)$$

В частности, для несмещенной эффективной оценки  $\hat{\theta}_n$  параметра распределения генеральной совокупности, которая строится по независимой однородной выборке длины  $n$ ,

$$D_{n,\theta}(\hat{\theta}_n) = 1/(nI_1(\theta)). \quad (14.10)$$

**Замечание 14.1** *Неравенство Рао–Крамера носит локальный характер, то есть при нахождении нижней границы для дисперсии оценок используется информация о параметрическом семействе  $f(x, \theta)$ ,  $\theta \in \Theta$ , лишь в любой малой окрестности  $\Theta_0$  значения  $\theta$ .*

С учетом этого замечания можно получить нижние границы для дисперсии несмещенных оценок числовой характеристики  $g(\theta)$ . Пусть функция  $g(\theta)$  непрерывно дифференцируема и  $g'(\theta) \neq 0$ . Тогда для любой несмещенной оценки  $\hat{g} : E_{\theta}(\hat{g}) = g(\theta)$  справедливо неравенство:

$$D_{\theta}(\hat{g}) \geq g'(\theta)^2/I(\theta). \quad (14.11)$$

**Замечание 14.2** *В качестве примера параметрического семейства, не удовлетворяющего условиям регулярности, для которого оценки параметров обладают существенно другими свойствами, можно рассмотреть семейство равномерных распределений на интервале  $[0, \theta]$ ,  $\theta > 0$ . Отметим, что для этого семейства формально вычисленная информация Фишера равна бесконечности.*

**Замечание 14.3** *Неравенство Рао–Крамера справедливо и для смещенных оценок (со смещением  $b'_{\theta}(\hat{\theta})$ ) в следующей форме:*

$$D_{\theta}(\hat{\theta}) \geq (1 + b'_{\theta}(\hat{\theta}))^2/I(\theta).$$

Однако равенство может достигаться только в случае, когда смещение  $b'_\theta(\hat{\theta})$  равно 0, и функция  $f(x, \theta)$  имеет вид (см. [5]):

$$f(x, \theta) = h(x) \exp(C_0(\theta) + C_1(\theta)\hat{\theta}(x)),$$

здесь функция наблюдений  $\hat{\theta}(x)$  называется достаточной статистикой.

Возвращаясь к примерам 13.6 и 13.7 п. 13.4.3, видим, что построенные в них оценки являются эффективными.

## 14.4 Асимптотические свойства ОМП

В этом разделе мы установим различные свойства асимптотической оптимальности оценок максимального правдоподобия. Мы рассмотрим модель независимой однородной выборки из генеральной совокупности  $X$  и будем считать выполненными условия регулярности. Будем обозначать  $I(\theta) = I_1(\theta)$  информацию Фишера, соответствующую одному наблюдению. Будем считать, что  $I(\theta) > 0$  и непрерывна по  $\theta \in \Theta$ .

### 14.4.1 Асимптотическая нормальность ОМП

Пусть  $\theta$  – одномерный параметр. Зафиксируем значение  $\theta_0 \in \Theta$  и пусть  $\theta = \theta_0 + n^{-1/2}u$ . Представим логарифмическую функцию правдоподобия (13.42) с помощью формулы Тейлора:

$$l(\theta, X^{(n)}) = l(\theta_0, X^{(n)}) + n^{-1/2}ul'(\theta_0, X^{(n)}) + \frac{u^2}{2n}l''(\tilde{\theta}, X^{(n)}), \quad (14.12)$$

где  $\tilde{\theta} = \theta_0 + n^{-1/2}u\xi$ ,  $\xi \in [0, 1]$ ,  $\tilde{\theta} \rightarrow \theta_0$  для любого  $u$  при  $n \rightarrow \infty$ . В силу закона больших чисел и с учетом (14.4) имеем:

$$\begin{aligned} n^{-1}l''(\tilde{\theta}, X^{(n)}) &= \frac{1}{n} \sum_{i=1}^n l''(\tilde{\theta}, X_i) \rightarrow E_\theta(l''(\theta_0, X)) = \\ &= -E_\theta(l'(\theta_0, X))^2 = -I(\theta_0). \end{aligned} \quad (14.13)$$

Используя (14.13), выражение (14.12) для логарифмической функции правдоподобия можно представить в виде:

$$l(\theta, X^{(n)}) = l(\theta_0, X^{(n)}) + uY_n - \frac{I(\theta_0)u^2}{2} + \delta_n, \quad Y_n = n^{-1/2}l'(\theta_0, X^{(n)}), \quad (14.14)$$

где остаточный член  $\delta_n \rightarrow 0$  по  $P_{\theta_0}$ -вероятности при  $n \rightarrow \infty$ . Кроме того, в силу центральной предельной теоремы и с учетом (14.3)

$$Y_n = n^{-1/2} \sum_{i=1}^n l'(\theta_0, X_i) \xrightarrow{P_{\theta_0}} Y \sim N(0, I(\theta_0)), \quad n \rightarrow \infty. \quad (14.15)$$

Соотношения (14.14), (14.15) называются условиями *локальной асимптотической нормальности*. Они лежат в основе асимптотического исследования широкого класса статистических задач и моделей (см. [4], [5]).

Оценка максимального правдоподобия соответствует максимизации по  $\theta \in \Theta$  левой части выражения (14.14). Отбрасывая остаточный член в (14.14) и максимизируя оставшуюся квадратичную функцию по  $u$ , получим приближенное выражение для уравнения правдоподобия и оценки максимального правдоподобия:

$$l'_u(\theta, X^{(n)}) \approx Y_n - uI(\theta_0) = 0; \quad \theta_n^* \approx \theta_0 + n^{-1/2}Z_n, \quad (14.16)$$

где

$$Z_n = Y_n/I(\theta_0) \xrightarrow{P_{\theta_0}} Z \sim N(0, 1/I(\theta_0)), \quad n \rightarrow \infty. \quad (14.17)$$

Это соотношение показывает, что *оценки максимального правдоподобия асимптотически нормальны с дисперсией  $I^{-1}(\theta_0)$ , где  $I(\theta)$  – информация Фишера.*

Заметим теперь, что если оценка  $\hat{\theta}_n$  асимптотически нормальна с дисперсией  $\Delta^2(\theta)$ , и характеристика  $g(\theta)$  есть непрерывно дифференцируемая функция  $\theta$ , то оценка характеристики  $\hat{g}_n = g(\hat{\theta}_n)$  также асимптотически нормальна с дисперсией  $\Delta_g^2(\theta) = (\Delta(\theta) g'(\theta))^2$ .

Пусть  $\theta_n^*$  – оценки максимального правдоподобия. Тогда мы получаем, что *оценки  $g_n^* = g(\theta_n^*)$  асимптотически нормальны с дисперсией  $g'(\theta)^2/I(\theta)$ , где  $I(\theta)$  – информация Фишера.*

#### 14.4.2 Асимптотическая минимаксность ОМП

Рассмотрим случай одномерного параметра  $\theta \in \Theta \subset R^1$ , где  $\Theta$  – конечный или бесконечный интервал. Предположим, что априорная плотность  $p(\theta)$  непрерывна и строго положительна, функция  $g(\theta)$  непрерывно дифференцируема,  $g'(\theta) \neq 0$ . В этих предположениях справедливо утверждение

**Теорема 14.1** *Пусть  $\hat{g}_n$  – ОМП характеристики  $g(\theta)$ , тогда она является локально асимптотически минимаксностной: при любых  $\theta_0 \in \Theta$ , любых оценках  $\hat{\theta}_n$  и достаточно малых  $\varepsilon > 0$ , таких что интервал  $\Theta_\varepsilon = \{|\theta - \theta_0| < \varepsilon\}$  содержится в  $\Theta$ , справедливо неравенство Гаека:*

$$\lim_{n \rightarrow \infty} \sup_{|\theta - \theta_0| < \varepsilon} nR_2(\hat{g}_n, \theta) \frac{I(\theta)}{(g'(\theta))^2} \geq 1. \quad (14.18)$$

В частности, для оценок параметра  $\theta \in \Theta$

$$\lim_{n \rightarrow \infty} \sup_{|\theta - \theta_0| < \varepsilon} nR_2(\hat{\theta}_n, \theta) I(\theta) \geq 1. \quad (14.19)$$

Сопоставим неравенства Рао–Крамера и Гаека. Неравенство Рао–Крамера для оценок одномерной характеристики и одномерного параметров можно переписать в виде

$$nR(\hat{\theta}_n, \theta) I(\theta) \geq 1, \quad nR(\hat{g}_n, \theta) \frac{I(\theta)}{(g'(\theta))^2} \geq 1. \quad (14.20)$$

Неравенства (14.20) являются точными и дают нижние границы квадратичного риска для всех значений параметров. Однако они справедливы лишь для несмещенных оценок и равенство в них достигается лишь в специальных случаях. Неравенства (14.18), (14.19) являются, с одной стороны, асимптотическими, с другой стороны, характеризуют максимум квадратичного риска в окрестности любого значения параметра. Однако они справедливы для любых оценок, и соответствующие предельные равенства выполнены для оценок максимального правдоподобия.

## 15 Оценка плотности распределения

К сожалению, метод подстановки нельзя непосредственно применить к задаче оценивания плотности распределения: даже если  $X$  – непрерывная случайная величина, то есть существует плотность распределения

$$f(x) = f_X(t) = F'_X(t) : F_X(t) = \int_{-\infty}^t f(x)dx,$$

нельзя определить выборочную плотность  $f_n(t) = F'_n(t)$  как обычную функцию, так как эмпирическая функция распределения разрывна в точках  $X^{(i)}$  вариационного ряда; формально производная эмпирической функции распределения имеет вид  $F'_n(t) = n^{-1} \sum_{i=1}^n \delta(t - X_i)$ , где  $\delta(t)$  – так называемая обобщенная функция Дирака, равная 0 при  $t \neq 0$ , равная  $\infty$  при  $t = 0$  и интеграл от которой равен 1. Такой функции в обычном смысле не существует. Ее можно рассматривать как формальный предел "пикообразной" последовательности функций вида  $v_h(t) = h^{-1}v(t/h)$ ,  $h > 0$ ,  $h \rightarrow 0$ , где функция  $v(t)$  интегрируема и  $\int_{-\infty}^{\infty} v(t)dt = 1$ .

В этой связи для оценки плотности распределения используются методы *группировки и сглаживания данных*.

### 15.1 Гистограмма как оценка плотности распределения

#### 15.1.1 Построение гистограммы

Метод группировки данных основан на следующей идее. Значение непрерывной плотности распределения  $f(t)$  в некоторой точке  $t \in R^1$  приближенно пропорционально отношению вероятности  $P(\Delta)$  попадания значения случайной величины  $X$  в малый интервал  $\Delta = [z_0, z_1)$ , содержащий точку  $t$ , к длине  $|\Delta| = h = z_1 - z_0$  этого интервала:

$$f(t) \approx \frac{P(\Delta)}{|\Delta|}, \quad P(\Delta) = F(z_1) - F(z_0) = \int_{z_0}^{z_1} f(x)dx. \quad (15.1)$$

Вероятность  $P(\Delta) = F(z_1) - F(z_0)$  можно оценить через эмпирическую функцию распределения:  $p_n(\Delta) = F_n(z_1) - F_n(z_0) = k(\Delta)/n$ , где  $k(\Delta)$  – число элементов выборки, лежащих в  $\Delta$ . Таким образом, в качестве оценки значения плотности распределения  $f(t)$  в точке  $t \in \Delta$  можно взять величину  $f_n = k(\Delta)/n|\Delta|$ .

Эту процедуру можно провести одновременно для всех  $t \in R^1$  или  $t \in [a, b]$ , где  $[a, b]$  – заданный (конечный или бесконечный) интервал, содержащий возможные

значения случайной величины  $X$ . Пусть  $\{\Delta_0, \Delta_{\pm 1}, \dots, \Delta_{\pm m}, \dots\}$  – разбиение интервала  $[a, b]$  на интервалы длины  $h = h_n > 0$  (их называют *интервалами группировки*). Тогда *гистограммой называется функция  $f_n(t)$ , принимающая на интервале  $\Delta_m$  значение  $f_{n,m} = k(\Delta_m)/nh$* .

Таким образом, гистограмма есть кусочно-постоянная функция, интервалы постоянства которой есть интервалы группировки  $\Delta_m$ . Гистограмма  $f_n(t)$  обладает всеми свойствами плотности распределения:  $f_n(t) \geq 0$ ,

$$\int_{-\infty}^{\infty} f_n(t) dt = \sum_m \int_{\Delta_m} f_n(t) dt = \sum_m h f_{n,m} = n^{-1} \sum_m k(\Delta_m) = 1,$$

так как  $\sum_m k(\Delta_m) = n$ .

На практике часто в качестве границ интервала  $[a, b]$  выбирают крайние члены вариационного ряда  $a = \min_i \{X_i\} = X_{(1)}$ ,  $b = \max_i \{X_i\} = X_{(n)}$ , а длину интервалов группировки выбирают так, чтобы в каждый интервал попадало не менее 5–10 элементов выборки.

### 15.1.2 Статистические свойства гистограммы

Изучим статистические свойства гистограммы как оценки плотности распределения. Для этого рассмотрим квадратичное отклонение  $R_{n,2}(t) = E_F(f_n(t) - f(t))^2$ , которое при  $t \in \Delta_m$  можно представить в виде  $R_{n,2}(t) = b_n^2(t) + \sigma_n^2(t)$ . Здесь  $b_n(t) = E_F(f_n(t) - f(t)) = E_F(f_n(t)) - f(t)$  – смещение оценки (систематическая ошибка), а  $\sigma_n^2(t) = D_F(f_n(t))$  – дисперсия оценки, характеризующая случайную ошибку оценивания.

Поскольку  $k(\Delta_m)$  – случайная величина, имеющая биномиальное распределение  $B(p, n)$ ,  $p = p_{n,m} = P_F(\Delta_m)$ , то

$$E_F(k(\Delta_m)) = np, \quad D_F(k(\Delta_m)) = np(1 - p),$$

откуда имеем:

$$b_n(t) = \left(\frac{p}{h} - f(t)\right), \quad \sigma_n^2(t) = \frac{p(1-p)}{nh^2} \leq \frac{p}{h} \frac{1}{nh}. \quad (15.2)$$

Заметим теперь, что для плотности  $f(x)$ , непрерывной на интервале  $\Delta_m$ , по теореме о среднем

$$\frac{p}{h} = \frac{1}{h} \int_{\Delta_m} f(x) dx = f(\tilde{t}), \quad \tilde{t} \in \Delta_m,$$

откуда следует, что

$$b_n(t) = f(\tilde{t}) - f(t) \rightarrow 0; \quad \sigma_n^2(t) = f(\tilde{t})/nh \rightarrow 0; \quad R_{n,2}(t) \rightarrow 0, \quad (15.3)$$

если  $h = h_n \rightarrow 0$ ,  $nh_n \rightarrow \infty$  при  $n \rightarrow \infty$ . Наконец, из неравенства Маркова следует, что при этом для любого  $\varepsilon > 0$

$$P_F(|f_n(t) - f(t)| > \varepsilon) \leq R_{n,2}(t)/\varepsilon^2 \rightarrow 0, \quad n \rightarrow \infty.$$

Таким образом, мы получили следующий результат.

Пусть  $h = h_n \rightarrow 0$ ,  $nh_n \rightarrow \infty$  при  $n \rightarrow \infty$ . Тогда для любой плотности распределения  $f$  и любой точки  $t \in R^1$ , в окрестности которой плотность  $f$  непрерывна, значение гистограммы  $f_n(t)$  является состоятельной оценкой значения плотности  $f(t)$ .

Этот результат показывает, что при построении гистограммы длины  $h = h_n$  интервалов группировки должны убывать с ростом объема выборки  $n$ , но не слишком быстро: должно выполняться соотношение  $nh_n \rightarrow \infty$ .

### 15.1.3 Интегральный квадратичный риск и наилучший выбор длин интервалов группировки

Для того чтобы оценить порядок точности оценивания "в целом", рассмотрим интегральное среднеквадратичное расстояние

$$\rho^2(f, f_n) = \int_{-\infty}^{\infty} (f(t) - f_n(t))^2 dt$$

и интегральный квадратичный риск

$$R_{n,2} = E_F \rho^2(f, f_n) = \int_{-\infty}^{\infty} E_F (f(t) - f_n(t))^2 dt,$$

который можно представить как сумму двух слагаемых:

$$R_{n,2} = \int_{-\infty}^{\infty} R_{n,2}(t) dt = \int_{-\infty}^{\infty} (b_n^2(t) + \sigma_n^2(t)) dt = b_n^2 + \sigma_n^2, \quad (15.4)$$

где первое слагаемое соответствует квадрату интегрального смещения, второе – интегральной дисперсии случайной ошибки. При этом

$$b_n^2 = \sum_m b_{n,m}^2; \quad \sigma_n^2 = \sum_m \sigma_{n,m}^2; \quad b_{n,m}^2 = \int_{\Delta_m} b_n^2(t) dt; \quad \sigma_{n,m}^2 = \int_{\Delta_m} \sigma_n^2(t) dt; \quad (15.5)$$

каждое  $m$ -е слагаемое соответствует интервалу группировки  $\Delta_m$ .

Для того чтобы изучить свойства гистограммы как оценки неизвестной плотности распределения  $f$ , нужно ввести предположения о классе  $\mathfrak{F} = \{f\}$  оцениваемых плотностей.

Будем считать, что неизвестные плотности распределения лежат в множестве  $\mathfrak{F}_C^1$ , которое состоит из плотностей распределения  $f(t)$  с ограниченной в среднеквадратичном производной  $f'(t)$ :

$$\int_{-\infty}^{\infty} (f'(t))^2 dt \leq C^2.$$

В этом предположении для любого интервала  $\Delta_m$  и для любых  $t, \tilde{t} \in \Delta_m$ , используя представление 15.3 и интегральное неравенство Коши–Буняковского (14.8), имеем неравенство

$$b_n^2(t) = (f(\tilde{t}) - f(t))^2 = \left( \int_{[t, \tilde{t}]} f'(u) du \right)^2 \leq |t - \tilde{t}| \int_{[t, \tilde{t}]} (f'(u))^2 du \leq h C_m^2,$$

где

$$C_m^2 = \int_{\Delta_m} (f'(u))^2 du; \quad \sum_m C_m^2 = \int_{-\infty}^{\infty} (f'(u))^2 du \leq C^2.$$

Отсюда с учетом (15.5) получаем оценки для квадрата интегрального смещения:

$$b_{n,m}^2 = \int_{\Delta_m} b_n^2(t) dt \leq h^2 C_m^2; \quad b_n^2 = \sum_m b_{n,m}^2 \leq h^2 C^2. \quad (15.6)$$

Для оценки интегральной дисперсии случайной ошибки в силу (15.2) имеем неравенство  $\sigma_n^2(t) \leq p_{n,m}/nh^2$  при  $t \in \Delta_m$ ,  $p_{n,m} = P_F(\Delta_m)$ . Поскольку

$$\sum_m p_{n,m} = \sum_m P_F(\Delta_m) = P_F(R^1) = 1,$$

с учетом (15.5) получаем:

$$\sigma_{n,m}^2 = \int_{\Delta_m} \sigma_n^2(t) dt \leq \frac{p_{n,m}}{nh}; \quad \sigma_n^2 = \sum_m \sigma_{n,m}^2 \leq \frac{1}{nh}. \quad (15.7)$$

Таким образом, из (15.6), (15.7) в силу (15.4) мы получаем следующую оценку интегрального квадратичного риска неизвестной плотности распределения  $f$  с помощью гистограммы  $f_n$ , равномерную по классу плотностей распределения  $\mathfrak{F}_C^1$ : для любой плотности распределения  $f \in \mathfrak{F}_C^1$  при длине интервалов группировки  $h$

$$R_{n,2} \leq h^2 C^2 + \frac{1}{nh}. \quad (15.8)$$

Первое слагаемое в правой части (15.8) соответствует квадрату интегрального смещения и убывает при уменьшении  $h$ , второе – интегральной дисперсии случайной ошибки и возрастает при уменьшении  $h$ . Можно определить наилучшую скорость убывания величин  $h = h_n$  с ростом  $n$  исходя из "баланса" между квадратом интегрального смещения и интегральной дисперсией, минимизируя правую часть (15.8) по  $h$ . Приравнявая к 0 производную по  $h$ , получаем выражение для  $h_n$  и соответствующую верхнюю границу для интегрального квадратичного риска:

$$2C^2 h = 1/nh^2; \quad h_n = (2C^2 n)^{-1/3}; \quad R_{n,2} \leq B(C/n)^{2/3}, \quad (15.9)$$

где  $B = 2^{-2/3} + 2^{1/3} \approx 0,63 + 1,26 = 1,89$ .

Из неравенства Маркова следует оценка вероятностей отклонений гистограммы от плотности с точки зрения интегрального среднеквадратичного расстояния  $\rho(f, f_n)$ : для любой последовательности  $T_n \rightarrow \infty$

$$P_F(\rho(f, f_n) > T_n/n^{1/3}) \leq R_{n,2} n^{2/3}/T_n^2 \rightarrow 0, \quad n \rightarrow \infty. \quad (15.10)$$

Это соотношение означает  $n^{1/3}$ -состоятельность оценки: типичный порядок интегрального среднеквадратичного расстояния гистограммы от неизвестной плотности распределения не превосходит  $n^{-1/3}$ .



На основании соотношений (15.9), (15.10) получаем следующее утверждение. *Наилучшая скорость убывания длины интервалов группировки  $h = h_n$  с ростом  $n$  в классе плотностей распределения  $f \in \mathfrak{Z}_C^1$  имеет порядок  $n^{-1/3}$ . При этом интегральный квадратичный риск убывает со скоростью  $n^{-2/3}$ . Такой же порядок имеют скорости убывания как квадрата интегрального смещения, так и интегральной дисперсии ошибки. Гистограмма является  $n^{1/3}$ -состоятельной оценкой плотности распределения  $f \in \mathfrak{Z}_C^1$  с точки зрения интегрального среднеквадратичного расстояния.*

Можно также показать (см. [5]), что не существует оценок неизвестной плотности распределения, которые обеспечивали бы более быстрый порядок убывания интегрального среднеквадратичного риска или расстояния для всех плотностей  $f \in \mathfrak{Z}_C^1$ . При указанном выборе длин интервалов группировки гистограмма есть наилучшая по порядку оценка неизвестной плотности распределения из класса  $\mathfrak{Z}_C^1$ .

Отметим два важных отличия этих результатов от аналогичных результатов в задачах оценки функции распределения или таких числовых характеристик, как среднее или дисперсия.

1. Порядок точности наилучших оценок есть не  $n^{1/2}$ , а ниже  $n^{1/3}$ , то есть оценить плотность распределения сложнее, чем функцию распределения или ее числовые характеристики. Этот порядок точности зависит от класса  $\mathfrak{Z}$  рассматриваемых плотностей.

2. Наилучшие оценки являются смещенными, причем имеет место баланс между порядками убывания смещения и случайной ошибки.

## 16 Критерий хи-квадрат.

### 16.1 Дискретная случайная величина.

Пусть генеральная совокупность  $X$  является дискретной с конечным множеством значений  $\bar{t} = \{t_1, \dots, t_k\}$ , и  $\bar{p} = \{p_1, \dots, p_k\}$  - набор неизвестных вероятностей значений. Пусть основная гипотеза является простой и характеризуется набором  $\bar{p}_0 = \{p_{1,0}, \dots, p_{k,0}\}$ ;  $p_{j,0} > 0$ . Обозначим через  $\mathbf{P}$  множество векторов

$$\mathbf{P} = \{\bar{p} = \{p_1, \dots, p_k\} : p_j \geq 0; \sum_j p_j = 1\}.$$

Тогда  $\bar{p}_0$  можно рассматривать как параметр  $\theta$  из  $(k-1)$ -мерного множества  $\Omega = \mathbf{P}$  и рассматривать параметрическую задачу проверки согласия с  $H_0 : \bar{p} = \bar{p}_0$ . Пусть  $X^{(n)} = (X_1, \dots, X_n)$  - выборка из генеральной совокупности  $X$ , обозначим через  $n_j$  количество элементов этой выборки равных  $t_j$ , а через  $F_0(t)$  - функцию распределения генеральной совокупности  $X$  при условии  $H_0$ , то есть при  $\bar{p} = \bar{p}_0$ .

*Критерий хи-квадрат* для проверки этой гипотезы основан на статистике

$$\chi_{n,k-1}^2(x_1, \dots, x_n) = \sum_{j=1}^k \frac{(n_j - np_{j,0})^2}{np_{j,0}}, \quad (16.1)$$

которая называется *статистикой хи-квадрат* с  $k - 1$  степенью свободы. Для этой статистики справедлив следующий результат, который мы приведем без доказательства

**Теорема 16.1 (Пирсона)** Пусть справедливо равенство  $\bar{p} = \bar{p}_0$ , тогда для функции распределения статистики  $\chi^2_{n,k-1}$  справедливо при  $n \rightarrow \infty$  соотношение

$$\sup_{u \in (0, \infty)} |P_{F_0}(\chi^2_{n,k-1} < u) - \chi^2_{k-1}(u)| \rightarrow 0,$$

где  $\chi^2_{k-1}(u)$  – функция распределения хи-квадрат распределения с  $k - 1$  степенью свободы (см. пример 13.3, формула (13.7)).

Назовем критерием хи-квадрат асимптотического уровня значимости  $\alpha$  для проверки согласия с гипотезой  $H_0 : \bar{p} = \bar{p}_0$  последовательность тестов

$$\psi_n(X_1, \dots, X_n) = \begin{cases} 1 & \text{при } \chi^2_{n,k-1} \geq t_{k-1,\alpha}, \\ 0 & \text{при } \chi^2_{n,k-1} < t_{k-1,\alpha}, \end{cases}$$

где величина  $t_{k-1,\alpha}$  определяется из условия  $\chi^2_{k-1}(t_{k-1,\alpha}) = 1 - \alpha$ .

Для критерия хи-квадрат найдем вероятность ошибок I рода

$$\begin{aligned} \alpha(\psi_n) &= P_{n,F_0}(\chi^2_{n,k-1} \geq t_{k-1,\alpha}) = \\ &= 1 - P_{n,F_0}(\chi^2_{n,k-1} < t_{k-1,\alpha}) \rightarrow 1 - \chi^2_{k-1}(t_{k-1,\alpha}) = \alpha \end{aligned}$$

при  $n \rightarrow \infty$  в силу теоремы Пирсона.

Для оценки вероятностей ошибок II рода критерия хи-квадрат зафиксируем альтернативу  $\bar{p}_1 \neq \bar{p}_0$ , пусть  $p_{j_0,1} \neq p_{j_0,0}$ ,  $|p_{j_0,1} - p_{j_0,0}| = a > 0$ . Тогда в силу закона больших чисел частота  $n_{j_0}/n \rightarrow p_{j_0,1}$  при  $n \rightarrow \infty$  с  $P_F$ -вероятностью 1. Следовательно,  $(n_{j_0} - np_{j_0,0})^2 \sim n^2 a^2$  при  $n \rightarrow \infty$ , а определенная (16.1) статистика  $\chi^2_{n,k-1} \rightarrow \infty$ . Отсюда

$$\beta(\psi_n, F) = P_{n,F}(\chi^2_{n,k-1} < t_{k-1,\alpha}) \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Таким образом, критерий хи-квадрат является состоятельным критерием асимптотического уровня значимости  $\alpha$ .

## Критерий хи-квадрат. Дискретизация.

Рассмотрим теперь случайную величину (генеральную совокупность)  $X$  общего вида. Пусть основная гипотеза является простой, то есть  $H_0 : F_X(x) = F_0(x)$ . Чтобы применить критерий хи-квадрат в задаче проверки гипотезы  $H_0$ , используют *дискретизацию* данных: множество возможных значений случайной величины  $X$  разбивается на  $k$  непересекающихся подмножеств  $I_1, \dots, I_k$ , обычно это интервалы  $I_j = [a_{j-1}, a_j)$ ,  $j = 1, \dots, k$ , где  $a_1 < \dots < a_{k-1}$  – точки деления;  $a_0 = -\infty$ ,  $a_k = +\infty$ .

Для каждого элемента выборки  $X_i$  регистрируется номер  $j = j(X_i)$  интервала  $I_j$ , в который попадает  $X_i$ , то есть исходная случайная величина  $X$  заменяется

новой дискретной случайной величиной  $\xi_k$ , принимающей значения  $1, \dots, k$ , и при основной гипотезе

$$p_{j,0} = P(\xi_k = j) = P(X \in I_j) = F_0(a_j) - F_0(a_{j-1}), \quad j = 1, \dots, k. \quad (16.2)$$

Для случайной величины  $\xi_k$  основная гипотеза имеет вид  $H_0 : \bar{p} = \bar{p}_0 = (p_{1,0}, \dots, p_{k,0})$  и для ее проверки можно использовать критерий хи-квадрат, описанный в предыдущем пункте и основанный на статистике  $\chi_{n,k-1}^2$  вида (16.1).

## 16.2 Проверка параметрической гипотезы.

Пусть основная гипотеза является сложной параметрической, то есть  $H_0 : F_X(t) \in \{F(x, \theta), \theta \in \Omega\}$ , где  $\Omega$  – область в  $R^r$ ,  $r \geq 0$ . Тогда при выполнении достаточно общих предположений "регулярности" в статистике хи-квадрат неизвестные в этом случае вероятности  $p_{j,0}$  можно заменить их оценками, вычисленными по формуле

$$\hat{p}_{j,0} = F(a_j, \hat{\theta}_n) - F(a_{j-1}, \hat{\theta}_n), \quad j = 1, \dots, k, \quad (16.3)$$

где  $\hat{\theta}_n$  – оценки максимального правдоподобия параметра  $\theta \in \Omega$ . При этом пороговое значение статистики хи-квадрат заменяется на  $t_{k-1-r, \alpha}$ .

**Замечание 16.1** Если распределение генеральной совокупности имеет плотность –  $f_X(x)$ , то при помощи критерия хи-квадрат можно проверять гипотезы о плотности как простую так и сложную –  $H_0 : f_X(x) = f_0(x)$  или  $H_0 : f_X(x) \in \{f(x, \theta), \theta \in \Omega\}$  при этом

$$p_{j,0} = \int_{I_j} f_X(x) dx \quad \text{или} \quad \hat{p}_{j,0} = \int_{I_j} f_X(x, \hat{\theta}_n) dx.$$

Таким образом, процедура проверки простой или параметрической гипотезы с помощью критерия хи-квадрат состоит из следующих этапов.

1) При сложной основной гипотезе по выборке  $X_1, \dots, X_n$  строятся оценки максимального правдоподобия параметра  $\theta \in \Omega \subseteq R^r$  для гипотезы.

2) Выбирается допустимый уровень значимости  $\alpha$  и определяется порог  $t_{k-1-r, \alpha}$  из таблиц распределения хи-квадрат с  $k-1-r$  степенями свободы.

3) Выбираются точки деления  $a_1 < \dots < a_{k-1}$  для разбиения  $I_1, \dots, I_k$ . Подсчитываются количества  $n_j$  элементов выборки, попадающих в подмножества  $I_j$ ,  $j = 1, \dots, k$ .

4) Рассчитываются вероятности  $p_{j,0}$  или  $\hat{p}_{j,0}$  по формулам (16.2) или (16.3); разбиение выбирается так, чтобы эти вероятности были положительны (желательно, близкими к  $1/k$ ) для всех  $j = 1, \dots, k$ .

5) Рассчитывается значения статистики  $\chi_{n,k-1}^2$  по формуле (16.1) (с заменой  $p_{j,0}$  на  $\hat{p}_{j,0}$  для сложной гипотезы) и сравниваются с порогом  $t_{k-1-r, \alpha}$ , определяемым из таблиц для заданного  $\alpha$ . Основная гипотеза принимается при  $\chi_{n,k-1}^2 < t_{k-1-r, \alpha}$  и отвергается в противном случае.

Иногда рассчитывают "достигаемый уровень значимости  $\alpha_n$ " из соотношения:  $P_{k-1-r}(\chi_{n,k-1}^2) = 1 - \alpha_n$ , характеризующий "степень надежности" решения: если  $\alpha_n$

мало (существенно меньше  $\alpha$ ), то решение отвергнуть гипотезу считается надежным. Отметим, что значения  $\alpha_n$ , близкие к 1, *не говорят в пользу справедливости гипотезы*: для достаточно больших  $k-1-r$  при справедливости гипотезы наиболее вероятны значения  $\alpha_n$ , близкие к  $1/2$ .

## Список литературы

- [1] **Боровков А. А.** Теория вероятностей. – М.: Наука, 1986.
- [2] **Боровков А. А.** Математическая статистика. – М.: Наука, 1984.
- [3] **Бородин А. Н.** Элементарный курс теории вероятностей и математической статистики. – СПб.: Лань, 1998.
- [4] **Гнеденко Б. В.** Курс теории вероятностей. – М.: Наука, 1988.
- [5] **Ивченко Г. И., Медведев Ю. И.** Математическая статистика. – М.: Высшая школа, 1984.
- [6] **Ингстер Ю. И.** Асимптотические методы в статистике. – С.П.: ПГУПС, 2000.
- [7] **Кокс Д., Хинкли Д.** Теоретическая статистика. – М.: Мир, 1978.
- [8] **Колемаев В. А. и др.** Теория вероятностей и математическая статистика / В. А. Колемаев, О. В. Староверов, В. Б. Турундаевский – М.: Высшая школа, 1991.
- [9] **Леман Э.** Проверка статистических гипотез. – М.: Наука, 1979.
- [10] **Пугачев В. С.** Теория вероятностей и математическая статистика. – М.: Наука, 1979.
- [11] **Ротарь В. И.** Теория вероятностей. – М.: Высшая школа, 1992.
- [12] **Севастьянов Б. А.** Вероятностные модели. – М.: Наука, 1992.
- [13] **Севастьянов Б. А.** Курс теории вероятностей и математической статистики. – М.: Наука, 1982.
- [14] **Тюрин Ю. Н., Макаров А. А.** Статистический анализ данных на компьютере. – М.: Имфра-М, 1998.
- [15] **Феллер В.** Введение в теорию вероятностей и ее приложения. Т. 1, 2. – М.: Мир, 1984.
- [16] **Ширяев А. Н.** Вероятность. – М.: Наука, 1989.