

Шамин Роман Вячеславович

# Лекции по теории вероятностей

Москва — 2017

УДК 517.98  
ББК 22.16  
Ш19

**Шамин Р.В.**

Лекции по теории вероятностей. М.: 2017. — —

Книга представляет собой конспект лекций по теории вероятности. Лекции написаны максимально доступно с большими комментариями для неспециалистов.

Книга будет полезна студентам и всем желающим изучить основы теории вероятности.



# Оглавление

<b>Введение</b>	<b>6</b>
<b>Глава I. Вероятностное пространство</b>	<b>7</b>
1. Пространство событий . . . . .	7
2. Вероятность . . . . .	9
3. Условная вероятность и теорема Байеса . . . . .	11
4. Независимость . . . . .	12
<b>Глава II. Случайные величины</b>	<b>14</b>
1. Определение случайной величины . . . . .	14
2. Математическое ожидание и дисперсия . . . . .	16
3. Дискретные случайные величины . . . . .	18
3.1. Распределение Бернулли . . . . .	18
3.2. Биноминальное распределение . . . . .	18
3.3. Распределение Пуассона . . . . .	19
4. Непрерывные распределения . . . . .	19
4.1. Равномерное распределение . . . . .	19
4.2. Нормальное распределение . . . . .	20
4.3. Экспоненциальное распределение . . . . .	20
<b>Глава III. Предельные теоремы</b>	<b>21</b>
1. Неравенство Чебышева . . . . .	21
2. Закон больших чисел . . . . .	22
3. Центральная предельная теорема . . . . .	24
4. Предельная теорема Пуассона . . . . .	25
<b>Глава IV. Случайные процессы</b>	<b>27</b>
1. Определение случайного процесса . . . . .	27

Оглавление	5
2. Сравнение случайных процессов . . . . .	28
3. Моменты случайных процессов . . . . .	29
4. Классы случайных процессов . . . . .	30
4.1. Гауссовские случайные процессы . . . . .	30
4.2. Стационарные случайные процессы . . . . .	31
5. Марковские цепи . . . . .	31
<b>Глава V. Математическая статистика</b>	<b>34</b>
1. Выборочное распределение . . . . .	34
2. Выборочные моменты . . . . .	36
3. Интервальные оценки . . . . .	37
4. Линейная регрессия . . . . .	38
<b>Литература</b>	<b>40</b>

# Введение

Настоящий курс лекций по теории вероятности, читается автором в Российском университете дружбы народов для бакалавров.

Стиль изложения ориентирован на студентов, изучающих высшую математику впервые, поэтому многие понятия и доказательства разъясняются и подробно комментируются.

*Примечание.* Данный текст лекций является предварительным и незаконченным. Лекции пополняются еженедельно. Обратитесь на сайт [calcs.ru](http://calcs.ru) за последней версией лекций.

# Глава I

## Вероятностное пространство

### 1. Пространство событий

Теория вероятности занимается изучением случайных событий. Не очень просто дать определение понятию «случайности». Важнейшим открытием XX века было построение А.Н. Колмогоровым аксиоматических основ теории вероятности. Наша первая цель состоит в изучении этого подхода.

Прежде чем говорить о вероятности, нужно дать определение события, поскольку мы будем говорить лишь о вероятности событий.

Прежде всего нужно понять, что в теории вероятности изучают повторяющиеся события. Каждое такое повторение будем называть опытом. При этом предполагается, что мы можем независимо наблюдать одинаковые опыты любое количество раз.

Каждый опыт будет заканчиваться наблюдением определенного исхода. Множество всех возможных исходов (элементарных исходов) мы будем обозначать через  $\Omega$ , при этом будем считать, что это множество непустое, что мы не будем впредь оговаривать. Если множество элементарных исходов конечно, то можно каждый исход считать событием, но для случая бесконечного множества  $\Omega$  это сделать не всегда возможно. Кроме того, часто нас интересует не элементарные исходы, а события. Событиями мы будем называть элементы множества  $\mathcal{A}$ , которые являются подмножествами множества  $\Omega$  и удовлетворяют следующим условиям:

1.  $\Omega \in \mathcal{A}$ ;

2.  $A \cup B \in \mathcal{A}$ , для всех  $A, B \in \mathcal{A}$ ;
3.  $A \cap B \in \mathcal{A}$ , для всех  $A, B \in \mathcal{A}$ ;
4.  $\bar{A} \in \mathcal{A}$ , для всех  $A \in \mathcal{A}$ .

Из этого определения следует, что  $\emptyset \in \mathcal{A}$ .

Мы будем понимать, что если в опыте реализовался какой-либо элементарный исход  $\tilde{\omega}$ , то возникли все события  $A \in \mathcal{A}$ , для которых

$$\tilde{\omega} \in A.$$

Таким образом, событие  $\Omega$  реализуется всегда, а событие  $\emptyset$  не возникает никогда.

Пусть  $A \in \mathcal{A}$  некоторое событие, тогда событие  $\bar{A}$  состоит в том, что не произошло событие  $A$ . То есть событие  $\bar{A}$  возникает тогда и только тогда, когда не возникает событие  $A$ .

Пусть  $A, B \in \mathcal{A}$  некоторые события, тогда событие  $A \cup B$  состоит в том, что произошло событие  $A$  или произошло событие  $B$ . С другой стороны, событие  $A \cap B$  есть событие, состоящее в том, что произошло событие  $A$  и событие  $B$ .

Рассмотрим примеры.

**Пример 1.1.** Пусть опыт состоит в том, что мы бросаем кубик (шестигранный). Множество элементарных исходов будет следующим

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

В качестве событий мы рассмотрим множество всех подмножеств множества  $\Omega$ .

**Пример 1.2.** Пусть теперь мы бросаем не один кубик а сразу два. При этом будем считать кубики идентичными. Множество элементарных исходов в этом примере:

$$\Omega = \{\{1, 1\}, \{1, 2\}, \dots, \{6, 5\}, \{6, 6\}\}.$$

Это множество состоит из всех пар (неупорядоченных) значений от 1 до 6.



**Пример 1.3.** Теперь мы будем бросать кубик два раза последовательно, отмечая какой номер выпал в первый раз, а какой во второй. В этом случае множество исходов есть множество всех упорядоченных пар

$$\Omega = \{(1, 1), (1, 2), (2, 1), \dots, (5, 6), (6, 5), (6, 6)\}.$$

Количество событий в последнем примере на много больше количества событий в предыдущем, поскольку мы получаем дополнительную информацию о порядке номеров.

## 2. Вероятность

До сих пор мы рассматривали события только с точки зрения их возможности наступления, но ничего не говорили о вероятности.

Что значит событие  $A$  имеет вероятность  $p$ ? С бытовой точки зрения мы говорим, например, «вероятность купить неспелый арбуз равна 0.25<sup>1</sup>» Эта фраза означает, что из ста купленных арбузов примерно 25 будут неспелыми. То есть мы понимаем вероятность события, как предельную частоту возникновения этого события при многократном повторении.

В некоторых опытах такое определение оправдано, более того, на этом основании можно вычислять вероятность. Если сделать некоторые предположения относительно проводимых опытов.

**Пример 2.1.** Пусть мы бросаем последовательно два кубика. Если мы будем верить, что эти кубики честные, то можно считать, что любой исход равновероятен. Какова тогда вероятность, что у нас выпадет дубль?

Всего возможных элементарных исходов равно  $n = 6 \cdot 6 = 36$ , а количество дублей  $m = 6$ , следовательно, вероятность выпадания дублей

$$p = \frac{m}{n} = \frac{6}{36} = \frac{1}{6}.$$

На этом принципе построена комбинаторная теория вероятности. Однако для полноценной теории вероятности такой подход недостаточен. Согласно подходу Колмогорова дадим следующее определение вероятности.

---

<sup>1</sup>На самом деле в быту говорят о 25% вероятности.

Пусть мы имеем множество элементарных исходов  $\Omega$  и множество событий  $\mathcal{A}$ . Пусть теперь каждому событию, т.е. каждому элементу множества  $\mathcal{A}$ , поставлено в соответствие число  $P(A)$  из отрезка  $[0, 1]$ . Эта функция

$$P : \mathcal{A} \rightarrow [0, 1]$$

удовлетворяет следующим аксиомам

1.  $P(\Omega) = 1$ ;
2. для любых  $A, B \in \mathcal{A}$  таких, что  $A \cap B = \emptyset$  имеет место

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

Число  $P(A)$  называется вероятностью события  $A$ .

Дадим несколько определений. Будем считать, что  $A, B, A_1, A_2, \dots$  — события.

Если  $P(A) = 1$ , то событие  $A$  называется достоверным, если  $P(A) = 0$ , то событие называется невозможным. Видим, что событие  $\emptyset$  является невозможным событием, поскольку

$$P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset).$$

Если

$$\bigcap_{k=1}^N A_k = \emptyset,$$

то говорят, что события  $\{A_k\}$  несовместны, что означает, вероятность одновременного наступления этих событий равна нулю.

Событие  $\bar{A}$  называется противоположным событием. Поскольку

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

то

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

Тройку  $\langle \Omega, \mathcal{A}, P \rangle$  будем называть вероятностным пространством. Отныне всегда, когда речь будет заходить о вероятности или случайных событиях, мы будем предполагать, что задано вероятностное пространство.

### 3. Условная вероятность и теорема Байеса

Для двух событий  $A$  и  $B$  определим произведение этих событий следующим образом

$$AB = A \cap B.$$

Пусть событие  $A$  такое, что  $P(A) > 0$ . Тогда условную вероятность события  $B$  при условии  $A$  будем обозначать  $P(B|A)$ , которая определяется по формуле

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

По определению видно, что  $P(B|A) \geq 0$ . Кроме того,

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

Пусть теперь события  $B$  и  $C$  таковы, что  $B \cap C = \emptyset$  тогда

$$\begin{aligned} P(B \cup C|A) &= \frac{P(A(B \cup C))}{P(A)} = \frac{P(AB \cup AC)}{P(A)} = \\ &= \frac{P(AB) + P(AC)}{P(A)} = P(B|A) + P(C|A). \end{aligned}$$

Таким образом, для вероятностного пространства  $\langle \Omega, \mathcal{A}, P \rangle$  и фиксированного события  $A \in \mathcal{A}$  мы имеем новое вероятностное пространство  $\langle \Omega, \mathcal{A}, P(\cdot|A) \rangle$ .

Рассмотрим два события  $A, B$  вероятности которых отличны от нуля. Имеем

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

или

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

Аналогично мы имеем

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

Из этих равенств получаем, что

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

Следовательно, верна следующая формула

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

Эта формула называется формулой Байеса. Формула Байеса имеет огромное прикладное значение, в частности, в задачах машинного обучения.

Пусть у нас есть несколько гипотез, которые состоят в реализации событий  $A_1, \dots, A_N$ . Предположим, что мы знаем их априорные вероятности  $P(A_k)$ . В случае, когда мы узнали, что произошло событие  $B$ . Какая гипотеза становится более вероятной? Разумеется, та гипотеза у которой условное вероятность  $P(A_k|B)$  больше. С помощью формулы Байеса это можно вычислить, если мы знаем условные вероятности  $P(B|A_k)$ . Пусть  $k_*$  такое, что

$$\max_k P(A_k)P(B|A_k) = P(A_{k_*})P(B|A_{k_*}),$$

тогда гипотезу (гипотезы, если этот максимум реализуется при нескольких  $k$ )  $A_{k_*}$  следует предпочесть.

## 4. Независимость

Принципиальным моментом в теории вероятности является понятие независимости двух событий. Мы будем говорить, что два события  $A$  и  $B$  независимы относительно вероятности  $P$ , если

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

Если два события  $A$  и  $B$  независимы, то для их условных вероятностей (предполагаем, что вероятности этих событий положительны) имеет место формулы

$$P(A|B) = P(A),$$

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

Это говорит о том, что для независимых событий факт возникновения одно события никак не влияет на вероятность другого.

Набор событий  $A_k$ ,  $k = 1, 2, \dots, N$  будем называть независимым, если

$$P(A_1 A_2 \dots A_N) = P(A_1)P(A_2) \dots P(A_N).$$

Важно понимать, что из попарной независимости событий  $A_k$  не следует независимость событий. Действительно, пусть  $\Omega = \{a_1, a_2, a_3, a_4\}$  и все исходы равновероятны. Рассмотрим три события

$$A = \{a_1, a_2\}, \quad B = \{a_1, a_3\}, \quad C = \{a_1, a_4\}.$$

Эти события попарно независимы, например,

$$P(AB) = P(\{a_1\}) = \frac{1}{4} = P(A)P(B),$$

поскольку  $P(A) = P(B) = \frac{1}{2}$ , аналогично для  $AC$  и  $CB$ . Но

$$P(ABC) = P(\{a_1\}) = \frac{1}{4}$$

тогда, когда

$$P(A)P(B)P(C) = \frac{1}{8}.$$

# Глава II

## Случайные величины

### 1. Определение случайной величины

Пусть мы имеем вероятностное пространство  $\langle \Omega, \mathcal{A}, P \rangle$ . Работать с элементарными исходами — элементами множества  $\Omega$  не очень удобно поскольку в этом множестве нет операций порядка (сравнения) или арифметических операций. Поэтому каждому элементу  $\omega \in \Omega$  сопоставляют некоторое действительное число. В результате мы получаем функцию

$$\xi : \Omega \rightarrow \mathbb{R}.$$

Для того, чтобы описывать вероятности событий (элементов множества  $\mathcal{A}$ ) необходимо, чтобы функция  $\xi$  удовлетворяла дополнительному требованию

$$\{\omega \in \Omega : \xi(\omega) < x\} \in \mathcal{A}$$

для любого  $x \in \mathbb{R}$ . Такая функция называется случайной величиной. Заметим, что сама функциональная зависимость является детерминированной, а случайным является элементарный исход  $\omega$ .

Задать случайную величину — это означает задать функцию распределения вероятности, которая определяется по следующей формуле

$$F_\xi(x) = P(\{\omega \in \Omega : \xi(\omega) < x\}), \quad x \in \mathbb{R}.$$

Непосредственно из определения функции распределения вероятности следует, что

$$\lim_{x \rightarrow -\infty} F_\xi(x) = 0,$$

$$\lim_{x \rightarrow \infty} F_{\xi}(x) = 1.$$

Кроме того, функция распределения вероятности всегда неубывающая функция.

Из свойств  $\mathcal{A}$  следует, что не только множества  $\{\omega : \xi(\omega) < c\}$  принадлежат  $\mathcal{A}$ , но и множества  $\{\omega : a < \xi(\omega) < b\}$  принадлежат  $\mathcal{A}$ .

По определению функции распределения

$$P(a \leq \xi(\omega) < b) = F_{\xi}(b) - F_{\xi}(a).$$

Если функция распределения вероятности дифференцируема по  $x$

$$f_{\xi}(x) = \frac{dF_{\xi}(x)}{dx}.$$

Тогда функция  $f_{\xi}(x)$  называется плотностью распределения вероятности случайной величины.

При этом имеем представление

$$F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(s) ds.$$

Плотность распределения вероятности  $f(x)$  удовлетворяет следующим свойствам

1.  $f(x) \geq 0$ .
2. Для любых  $x' < x < x''$  имеет место

$$P(a \leq x < b) = \int_a^b f(x) dx.$$

- 3.

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Две случайные величины  $X$  и  $Y$  называются независимыми случайными величинами, если

$$F_{XY}(x) = F_X(x)F_Y(x),$$

где  $F_X(x)$  и  $F_Y(x)$  суть функции распределения случайных величин  $X$  и  $Y$  соответственно, а  $F_{XY}(x)$  — функция распределения вероятности случайной величины  $XY$ .

## 2. Математическое ожидание и дисперсия

Пусть дискретная случайная величина  $\xi$  принимает значения  $x_k$  с вероятностью  $p_k$ , тогда если абсолютно сходится ряд (или сумма)

$$E[\xi] = \sum_k x_k p_k,$$

то число  $E[\xi]$  называется математическим ожиданием.

Для случайной величины  $\xi$ , имеющей плотность распределения вероятности, математическое ожидание  $E[\xi]$  существует, если абсолютно сходится ряд

$$E[\xi] = \int_{-\infty}^{\infty} x f(x) dx.$$

Математическое ожидание имеет смысл среднего значения случайной величины.

В дальнейшем мы будем рассматривать математическое ожидание только дискретных или непрерывных случайных величин.

Рассмотрим свойства математического ожидания. Если случайная величина не является случайной, то есть  $\xi = a$ , то математическое ожидание

$$E[a] = a.$$

Далее, математическое ожидание линейно

$$E[a\xi + b\eta] = aE[\xi] + bE[\eta].$$

Если случайные величины  $X, Y$  независимые, то

$$E[XY] = E[X]E[Y],$$

предполагается, что все математические ожидания существуют.

Если с вероятностью 1 имеет место  $X \leq Y$ , то

$$E[X] \leq E[Y].$$

Если математическое ожидание характеризует среднее значение случайной величины, то дисперсия случайной величины является мерой разброса значений случайной величины относительно среднего



значения. Пусть  $X$  некоторая случайная величина, которая имеет математическое ожидание. Тогда дисперсия этой случайной величины определяется по формуле

$$D[X] = E[(X - E[X])^2],$$

если соответствующее математическое ожидание существует.

Можно получить другую формулу для вычисления дисперсии

$$\begin{aligned} D[X] &= E[(X - E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2] = \\ &= E[X^2] - 2E[X]E[X] + (E[X])^2 = E[X^2] - (E[X])^2. \end{aligned}$$

Из определения дисперсии и определения случайной величины мы получаем следующие свойства дисперсии.

Дисперсия всегда неотрицательна

$$D[X] \geq 0.$$

Кроме дисперсии рассматривают среднеквадратичное отклонение случайной величины

$$\sigma_X = \sqrt{D[X]}.$$

Для неслучайной величины дисперсия равна нулю

$$D[a] = 0.$$

Если  $X$  некоторая случайная величина, а  $a$  константа, то

$$D[aX] = a^2 D[X],$$

$$D[X + a] = D[X].$$

Пусть  $X$  и  $Y$  некоторые случайные величины, каждая из которых имеет конечную дисперсию. Тогда ковариацией этих случайных определяется следующим образом

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

Для независимых случайных величин  $X$  и  $Y$  их ковариация равна нулю

$$\text{cov}(X, Y) = 0.$$

Ковариация представляет собой симметричную функцию

$$\text{cov}(X, Y) = \text{cov}(Y, X).$$

Ковариация случайной величины с собой равна дисперсии

$$\text{cov}(X, X) = D[X].$$

Через ковариацию выражается дисперсия суммы случайных величин

$$D[X + Y] = D[X] + D[Y] + \text{cov}(X, Y).$$

Для двух случайных величин  $X$  и  $Y$ , имеющих не нулевые дисперсии, введем коэффициент корреляции

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

### 3. Дискретные случайные величины

Рассмотрим примеры дискретных случайных величин

#### 3.1. Распределение Бернулли

Случайная величина  $X$  имеет распределение Бернулли, если она принимает только два значения 1 с вероятностью  $0 < p < 1$  и значение 0 с вероятностью  $q = 1 - p$ .

Математическое ожидание и распределения Бернулли

$$E[X] = p,$$

$$D[X] = pq.$$

#### 3.2. Биномиальное распределение

Пусть  $X_1, X_2, \dots, X_n$  — независимые одинаково распределенные случайные величины Бернулли. Тогда случайная величина

$$Y = X_1 + X_2 + \dots + X_n$$

имеет биномиальное распределение. Это распределение имеет следующие характеристики

$$P_Y(k) = P(Y \leq k) = C_k^n p^k q^{n-k},$$

где  $C_k^n = \frac{n!}{k!(n-k)!}$  — биномиальный коэффициент.

Математическое ожидание и дисперсия

$$E[Y] = np,$$

$$D[Y] = npq.$$

### 3.3. Распределение Пуассона

Пусть  $\lambda > 0$  фиксированное число. Определим случайную величину  $X$  следующим образом

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Такая случайная величина имеет распределение Пуассона. Распределение Пуассона моделирует ситуацию с подсчетом числа случайных событий, происходящих в фиксированное время.

Математическое ожидание и дисперсия распределения Пуассона совпадают

$$E[X] = E[Y] = \lambda.$$

## 4. Непрерывные распределения

### 4.1. Равномерное распределение

Пусть  $X$  — случайная величина, принимающая значения из отрезка  $[a, b]$ . Эта величина имеет равномерное распределение, если ее плотность распределения задана формулой

$$f(x) = \frac{1}{b-a}, \quad x \in [a, b]$$

и

$$f(x) = 0, \quad x \notin [a, b].$$

Математическое ожидание и дисперсия равны

$$E[X] = \frac{a + b}{2},$$

$$D[X] = \frac{(b - a)^2}{12}.$$

Будем использовать следующее обозначение  $X \sim R[a, b]$ .

## 4.2. Нормальное распределение

Принципиальное значение имеет нормальное или гаусово распределение случайной величины  $X$ . Это распределение имеет два параметра  $a \in \mathbb{R}$  и  $\sigma > 0$ . Плотность распределения вероятности имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Математическое ожидание и дисперсия равны

$$E[X] = m,$$

$$D[X] = \sigma^2.$$

Нормальное распределение возникает как сумма независимых случайных величин в предельных законах.

## 4.3. Экспоненциальное распределение

Случайная величина  $X$  имеет экспоненциальное распределение с параметром  $\lambda > 0$ , если ее плотность вероятности имеет вид

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$

$$f(x) = 0, \quad x < 0.$$

Математическое ожидание и дисперсия равны

$$E[X] = \frac{1}{\lambda},$$

$$D[X] = \frac{1}{\lambda^2}.$$

# Глава III

## Предельные теоремы

### 1. Неравенство Чебышева

Пусть  $X$  — некоторая случайная величина, которая имеет математическое ожидание  $E[X]$ . Для любого  $\varepsilon > 0$  имеет место неравенство

$$P\{|X| \geq \varepsilon\} \leq \frac{1}{\varepsilon} E|X|. \quad (\text{III.1})$$

Докажем это неравенство для случая непрерывной случайной величины. Пусть  $f(x)$  — плотность распределения вероятности случайной величины  $X$ . Выразим вероятность  $P\{|X| \geq \varepsilon\}$  через плотность распределения вероятности

$$\begin{aligned} P\{|X| \geq \varepsilon\} &= \int_{-\infty}^{-\varepsilon} f(x) dx + \int_{\varepsilon}^{\infty} f(x) dx \leq \\ &\leq \int_{-\infty}^{-\varepsilon} \frac{|x|}{\varepsilon} f(x) dx + \int_{\varepsilon}^{\infty} \frac{|x|}{\varepsilon} f(x) dx \leq \\ &\leq \int_{-\infty}^{-\varepsilon} \frac{|x|}{\varepsilon} f(x) dx + \int_{\varepsilon}^{\infty} \frac{|x|}{\varepsilon} f(x) dx + \int_{-\varepsilon}^{\varepsilon} \frac{|x|}{\varepsilon} f(x) dx = \\ &= \frac{1}{\varepsilon} \int_{-\infty}^{\infty} |x| f(x) dx = \frac{1}{\varepsilon} E|X|. \end{aligned}$$

Аналогично, заменяя интегралы рядами, доказывается неравенство (III.1) для случая дискретной случайной величины.

Событие  $\{|X| \geq \varepsilon\}$ , разумеется, равносильно событию  $\{X^2 \geq \varepsilon^2\}$ . Поэтому применяя неравенство (III.1) к случайной величине  $|X - E[X]|$ , получаем

$$P\{|X - E[X]| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} E[(X - E[X])^2] = \frac{1}{\varepsilon^2} D[X].$$

Таким образом, для случайной величины  $X$ , которая имеет дисперсию, имеет место неравенство Чебышева

$$P\{|X - E[X]| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} D[X]. \quad (\text{III.2})$$

## 2. Закон больших чисел

С помощью неравенства Чебышева оценим вероятность события  $\{|X - E[X]| < \varepsilon\}$ , которое является противоположным событию  $\{|X - E[X]| \geq \varepsilon\}$ , поэтому имеем

$$P\{|X - E[X]| < \varepsilon\} = 1 - P\{|X - E[X]| \geq \varepsilon\} \geq 1 - \frac{1}{\varepsilon} D[X]. \quad (\text{III.3})$$

Пусть теперь  $X_1, X_2, \dots, X_n$  — независимые случайные величины, которые имеют дисперсии. Образует новую случайную величину

$$Y = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Математическое ожидание и дисперсия этой случайной величины вычисляются по формулам

$$E[Y] = \frac{1}{n} \sum_{i=1}^n E[X_i],$$

$$D[Y] = \frac{1}{n^2} \sum_{i=1}^n D[X_i].$$

Применим к  $Y$  неравенство (III.3) для любого  $\varepsilon > 0$

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^k X_i - \frac{1}{n} \sum_{i=1}^k E[X_i] \right| < \varepsilon \right\} = P\{|Y - E[Y]| < \varepsilon\} \geq$$

$$\geq 1 - \frac{1}{\varepsilon^2} D[Y] = 1 - \frac{1}{n\varepsilon^2} \frac{D[X_1] + \dots + D[X_n]}{n}.$$

Таким образом, мы имеем неравенство

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^k X_i - \frac{1}{n} \sum_{i=1}^k E[X_i] \right| < \varepsilon \right\} \geq 1 - \frac{1}{n\varepsilon^2} \frac{D[X_1] + \dots + D[X_n]}{n}. \quad (\text{III.4})$$

Закон больших чисел устанавливает условия сходимости среднего арифметического  $n$  случайных величин к среднему арифметическому их математических ожиданий. В частности, для одинаково распределенных случайных величин их среднее арифметическое сходится к математическому ожиданию.

В вероятностном мире сходимость случайных величин определяется следующим образом. Последовательность случайных величин  $\{X_k\}$  сходится по вероятности к числу  $b$ , если для любого  $\varepsilon > 0$

$$P\{|X_n - b| < \varepsilon\} \rightarrow 1, \quad n \rightarrow \infty.$$

**Теорема 2.1** (Закон больших чисел). Пусть  $X_1, \dots, X_n$  — последовательность независимых случайных величин, дисперсии которых ограничены числом  $M$

$$D[X_n] \leq M.$$

Тогда для любого  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E[X_i] \right| < \varepsilon \right\} = 1.$$

*Доказательство.* Воспользуемся неравенством (III.4). Имеем

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E[X_i] \right| < \varepsilon \right\} \geq 1 - \frac{1}{n\varepsilon^2} \left( \frac{1}{n} \sum_{i=1}^n D[X_i] \right) \geq 1 - \frac{M}{n\varepsilon^2}.$$

Поскольку вероятность не превышает 1, то имеем

$$1 - \frac{M}{n\varepsilon^2} \leq P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E[X_i] \right| < \varepsilon \right\} \leq 1.$$

Переходя в этом неравенстве к пределу при  $n \rightarrow \infty$  получаем

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E[X_i] \right| < \varepsilon \right\} = 1.$$

□

В частности, для независимых случайных величин одинаково распределенных с математическим ожиданием  $a$  и дисперсией  $\sigma^2$  имеем по закону больших чисел, что арифметическое среднее этих случайных величин

$$\frac{1}{n} \sum_{i=1}^n X_i$$

сходится по вероятности к числу  $a$ .

### 3. Центральная предельная теорема

Центральная предельная теорема устанавливает, что при весьма широких условиях сумма большого числа независимых случайных величин асимптотически имеют нормальное распределение. Эта теорема имеет принципиально важное значение для приложений теории вероятности.

Напомним, что нормально распределенная случайная величина имеет следующую плотность распределения вероятности

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Соответственно, функция нормального распределения обозначается следующим образом

$$\Phi(x) = \int_{-\infty}^x f(s) ds.$$

Рассмотрим сумму  $S_n$  независимых случайных величин  $\xi_{kn}$

$$S_n = \sum_{k=1}^n \xi_{kn},$$

где  $\xi_{kn}$  таковы, что

$$E[\xi_{kn}] = 0,$$

$$D[\xi_{kn}] \rightarrow 0, \quad n \rightarrow \infty$$

равномерно по  $k$  и

$$D[S_n] = \sum_{k=1}^n D[\xi_{kn}] = 1.$$



**Теорема 3.1** (Центральная предельная теорема). Пусть выполнено условие Ляпунова

$$\sum_{k=1}^n E|\xi_{kn}|^3 \rightarrow 0.$$

Тогда

$$\lim_{n \rightarrow \infty} P\{a \leq S_n \leq b\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

#### 4. Предельная теорема Пуассона

Кроме нормального распределения, которое является предельным для суммы независимых случайных величин, есть другой вариант предельных законов, называемых предельными законами Пуассона.

Рассмотрим последовательность  $n$  случайных величин Бернулли, принимающих значения 1 с вероятностью  $p_n \in (0, 1)$  и значение 0 с вероятностью  $q_n = 1 - p_n$ .

**Теорема 4.1** (Предельная теорема Пуассона). Если  $np_n \rightarrow \lambda > 0$  при  $n \rightarrow \infty$ , то вероятность того, что число наступлений исхода 1 равно в точности  $k$ , имеет следующий предел при  $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} P\{\nu^n = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

*Доказательство.* Согласно биномиальному распределению мы имеем

$$\begin{aligned} P\{\nu^n = k\} &= \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} = \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k e^{(n-k)\ln(1-p_n)}. \end{aligned}$$

При фиксированном  $k$  рассмотрим предел

$$\lim_{n \rightarrow \infty} (n-k)\ln(1-p_n) = -\lim_{n \rightarrow \infty} (n-k)p_n = \lim_{n \rightarrow \infty} p_n n \left(1 - \frac{k}{n}\right) = -\lambda.$$

Далее,

$$\lim_{n \rightarrow \infty} n(n-1)\dots(n-k+1)p_n^k = \lim_{n \rightarrow \infty} (np_n)^k = \lambda^k.$$

Таким образом,

$$\lim_{n \rightarrow \infty} P\{\nu^n = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

□

# Глава IV

## Случайные процессы

### 1. Определение случайного процесса

Многие случайные явления развиваются во времени, поэтому часто возникает необходимость рассматривать не только случайные величины, но и случайные процессы.

Случайным процессом называется семейство случайных величин  $\xi(t, \omega)$ ,  $t \in T$ , где  $T \subset \mathbb{R}$ .

В качестве шкалы времени обычно используют либо дискретную шкалу

$$T = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

или

$$T = \mathbb{R}.$$

В первом случае случайный процесс называется процессом с дискретным временем, а во втором — процессом с непрерывным временем.

Пусть  $\xi(t, \omega)$  — случайный процесс. Для фиксированного  $t_0 \in T$  случайная величина  $\xi(t_0, \omega)$  называется сечением случайного процесса в точке  $t_0 \in T$ .

При фиксированном элементе  $\omega_0 \in \Omega$  неслучайная функция  $\xi(t, \omega_0)$  называется траекторией или реализацией случайного процесса.

Рассмотрим пример простейшего случайного процесса. Пусть  $X$  — случайная величина с равномерным распределением  $R[0, 1]$ . Случайный процесс  $\xi(t, \omega)$  определим следующим образом

$$\xi(t, \omega) = t^2 X(\omega).$$

Сечениями этого случайного процесса при  $t_0$  представляют собой случайную величину с равномерным распределением  $R[0, t_0^2]$ . А траектории этого процесса представляют собой параболы

$$\xi(t, \omega_0) = \omega_0 t^2.$$

Мы видели, что случайные величины характеризуются функциями распределения вероятности. Для случайных процессов также можно рассмотреть конечные семейства функций распределения вероятности.

Пусть  $\xi(t)$ ,  $t \in T$  некоторый случайный процесс. Для заданного конечного множества

$$\{t_1, t_2, \dots, t_n\} \subset T$$

рассмотрим множество случайных величин

$$\xi(t_1), \xi(t_2), \dots, \xi(t_n).$$

Эти случайные величины имеют совместную функцию распределения

$$F(x_1, \dots, x_n) = P(\xi(t_1) \leq x_1, \dots, \xi(t_n) \leq x_n). \quad (\text{IV.1})$$

Совокупность всех конечных семейств функций распределений вида (IV.1) называется семейством конечномерных распределений случайного процесса  $\xi(t)$ .

Можно показать (теорема Колмогорова), что семейства конечномерных распределений полностью задают случайный процесс.

## 2. Сравнение случайных процессов

Два случайных процесса  $X$  и  $Y$  называются стохастически эквивалентными в широком смысле, если их семейства конечномерных распределений совпадают.

Случайные процессы называются просто эквивалентными, если

$$P(X(t) = Y(t)) = 1, \quad t \in T.$$

Можно показать, что эквивалентные случайные процессы являются эквивалентными случайными процессами в широком смысле.

Далее два процесса  $X(t)$  и  $Y(t)$  называются неотличимыми если

$$P(X(t) = Y(t), \forall t \in T) = 1.$$

Разница между эквивалентными и неотличимыми случайными процессами состоит в том, что для неотличимого случайного процесса вероятность совпадения  $X(t)$  и  $Y(t)$  равна единицы сразу для всех  $t \in T$ .

### 3. Моменты случайных процессов

Пусть  $\xi(t)$  — случайный процесс. Математическим ожиданием случайного процесса  $\xi(t)$  называется неслучайная функция, определяемая по формуле

$$m_\xi(t) = E[\xi(t)],$$

в случае, когда для всех  $t$  существует математическое ожидание.

Неслучайная функция

$$D_\xi(t) = D[\xi(t)] = E[(\xi(t) - E[\xi])^2]$$

называется дисперсией случайного процесса.

Важнейшей характеристикой случайных процессов является ковариационная функция случайного процесса, которая определяется по формуле

$$R_\xi(t, s) = \text{cov}(\xi(t), \xi(s)).$$

Смысл ковариационной функции состоит в том, что она характеризует степень линейно зависимости между сечениями случайного процесса. Из определения следует, что

$$R_\xi(t, t) = D_\xi(t).$$

Для двух, вообще говоря, различных случайных процессов  $\xi(t)$  и  $\eta(t)$  взаимная ковариационная функция определяется по формуле

$$R_{\xi\eta}(t, s) = \text{cov}(\xi(t), \eta(s)).$$

## 4. Классы случайных процессов

### 4.1. Гауссовские случайные процессы

Случайный процесс  $\xi(t)$ ,  $t \in T$  называется гауссовским случайным процессом, если для любого множества  $\{t_1, \dots, t_n\}$  случайный вектор

$$(\xi(t_1), \xi(t_2), \dots, \xi(t_n))$$

имеет многомерное нормальное распределение. Из определения гауссовского случайного процесса следует, что такие процессы полностью определяются неслучайными функциями: математическим ожиданием  $m(t)$  и ковариационной функцией  $C(t, s)$ .

В силу центральной предельной теоремы гауссовский случайный процесс имеет важнейшее значение.

Заметим, что гауссовские случайные процессы могут иметь совершенно разные формы. Приведем несколько примеров гауссовских случайных процессов.

Пусть  $X_1, X_2, \dots, X_n$  суть гауссовские случайные величины, тогда случайный процесс

$$X(t) = \sum_{k=1}^n X_k t^k$$

является гауссовским случайным процессом.

Наиболее известными гауссовскими процессами являются винеровский процесс и процесс белого шума. Стандартный винеровский процесс  $w(t)$  определяется, как гауссовский случайный процесс, выходящий из нуля, т.е.  $w(0) = 0$ , имеющий следующие моментные характеристики

$$E_w(t) = 0,$$

$$R_w(t, s) = \min(t, s).$$

Стандартный винеровский процесс называется также процессом броуновского движения.

Другим примером гауссовского случайного процесса является процесс белого шума, который имеет следующие моментные функции

$$E_w(t) = 0,$$

$$R_w(t, s) = \begin{cases} 1, & t = s, \\ 0, & t \neq s. \end{cases}$$

Процесс белого шума состоит из независимых случайных величин, имеющих нормальное распределение.

## 4.2. Стационарные случайные процессы

Случайный процесс  $\xi(t)$  называется стационарным случайным процессом в узком смысле, если для любого набора  $t_1, \dots, t_n \in T$  совместное распределение случайных величин

$$\xi(t_1), \xi(t_2), \dots, \xi(t_n)$$

такое же, как у случайных величин

$$\xi(t_1 + \tau), \xi(t_2 + \tau), \dots, \xi(t_n + \tau)$$

для всех таких  $\tau$ , что  $t_k + \tau \in T$ .

Случайный процесс  $\xi(t)$  называется стационарным в широком смысле, если

$$E_\xi(t) = \text{const},$$

$$R_\xi(t, s) = C(t - s).$$

Ясно, что стационарный процесс в узком смысле всегда является стационарным и в широком смысле. Обратное не всегда верно. Для гауссовских случайных процессов стационарность в широком смысле влечет за собой стационарность в узком смысле.

## 5. Марковские цепи

Важнейшим классом случайных процессов являются процессы, удовлетворяющие марковскому свойству или марковские процессы. Это такие процессы, для которых будущее зависит только от настоящего, но не зависит от прошлого. Марковские процессы могут быть с непрерывным или дискретным временем. Мы рассмотрим марковские процессы с дискретным временем, которые могут находиться в каждый момент времени только в одном состоянии. Такие случайные процессы называются марковскими цепями.

Рассмотрим систему, которая в каждый дискретный момент времени

$$T = \{0, 1, 2, \dots\}$$

может находиться только в одном из состояний из конечного множества состояний. Эти состояния мы занумеруем и будем рассматривать множество состояний

$$S = \{1, 2, \dots, N\}.$$

В каждый момент времени система переходит из текущего состояния  $S(k)$  в состояние в следующий момент  $S(k+1)$  случайным образом.

Вероятность перехода системы из состояния  $i$  в состояние  $j$  обозначим через  $p_{ij}$ . Для простоты будем рассматривать стационарный случай, когда  $p_{ij}$  не зависят от времени.

Таким образом, мы имеем матрицу переходных вероятностей

$$P = \{p_{ij}\}.$$

Необходимо требовать выполнения условия

$$\sum_{j=1}^N p_{ij} = 1, \quad i = 1, 2, \dots, N.$$

Соответственно матрица переходов за  $k$  шагов представляет собой  $k$ -ую степень матрицы  $P$

$$P(k) = P^k.$$

Чтобы задать эволюцию цепи Маркова необходимо задать начальное распределение состояний. Пусть в какой-то момент система находится в состоянии  $i$  с вероятностью  $q_i$ . При этом

$$\sum_{n=1}^N q_n = 1.$$

Тогда вектор-строка

$$q = (q_1, q_2, \dots, q_N)$$

описывает вероятности нахождения системы в одном из состояний в фиксированный момент времени. Тогда в следующий момент времени вероятность находиться в состоянии  $i$  можно вычислить по следующей формуле

$$P\{S(0) = i, S(1) = j\} = P\{S(1) = j | S(0) = i\}P\{S(0) = i\} = q_i p_{ij}.$$



Распределение вероятностей состояний  $q^*$  называется эргодическим или стационарным, если распределение вероятностей состояний в любой другой момент, остается  $q^*$

$$q_j^* = \sum_i q_i^* p_{ij}.$$

# Глава V

## Математическая статистика

### 1. Выборочное распределение

Если основная задача теории вероятностей состоит в нахождении вероятности событий по известному распределению вероятностей, то математическая статистика решает обратную задачу — нахождение распределений вероятностей по наблюдаемым реализациям случайных величин.

Пусть у нас есть вероятностное пространство  $\langle \Omega, \mathcal{A}, P \rangle$ , в котором функция  $P$  нам не известна. Пусть также мы имеем набор реализаций случайной величины  $\xi$ :

$$(x_1, x_2, \dots, x_n),$$

который мы будем называть случайной выборкой объема  $n$ . Основная задача оценить функцию распределения вероятности случайной величины  $\xi$ .

Имея случайную выборку достаточно большого объема можно построить эмпирическую функцию распределения по следующей формуле

$$F_n^*(x) = \frac{\nu_n(x)}{n},$$

где  $\nu_n(x)$  — число значений в выборке меньших  $x$ . По построению очевидно, что эмпирическая функция распределения удовлетворяет всем условиям на функцию распределения.

Имеет место фундаментальная теорема Гливенко-Кантелли, которая показывает, что эмпирическая функция распределения сходит-

ся по вероятности к настоящей функции распределения вероятности случайной величины  $\xi$ .

**Теорема 1.1** (Гливенко-Кантелли). Пусть  $F(x)$  — функция распределения вероятности случайной величины  $\xi$ , а  $F_n^*(x)$  — эмпирическая функция распределения вероятности, построенная по случайной выборке объема  $n$ . Тогда для любого  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\{|F_n^*(x) - F(x)| < \varepsilon\} = 1.$$

*Доказательство.* Для каждого  $x_k$  можно рассмотреть два противоположны события:  $\{x_k < x\}$  и  $\{x_k \geq x\}$ . Вероятности этих событий

$$p = P\{x_k < x\} = F(x), \quad q = P\{x_k \geq x\} = 1 - F(x).$$

Первое событие назовем успехом, тогда  $\nu_n(x)$  равно числу успехов в серии  $n$  независимых испытаний случайной величины, имеющей распределение Бернулли. При этом величина

$$F_n^*(x) = \frac{\nu_n(x)}{n}$$

есть частота успеха. Согласно закону больших чисел эта частота сходится по вероятности к математическому ожиданию, равному  $p$ .  $\square$

Существует и оценка скорости сходимости эмпирической функции распределения к настоящей функции распределения.

**Теорема 1.2** (Колмогоров). Если  $F(x)$  непрерывна, то при  $n \rightarrow \infty$

$$P\left\{\sqrt{n} \left( \sup_{-\infty < x < \infty} |F_n^*(x) - F(x)| \right) < z\right\} \rightarrow K(z),$$

где  $K(z)$  есть функция распределения Колмогорова

$$K(z) = \begin{cases} 0, & z \leq 0 \\ \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 z^2}, & z > 0 \end{cases}$$

## 2. Выборочные моменты

Хотя теорема Гливенко-Кантелли принципиально решают вопрос о восстановлении функции распределения вероятностей случайной величины по ее случайной выборке, часто возникает задача оценки не всей функции распределения вероятности, а отдельных числовых характеристик. В частности, нас будут интересовать математическое ожидание и дисперсия, построенные по случайной выборке. Такие оценки мы будем называть выборочными.

Начнем с определения выборочного математического ожидания случайной величины, построенной по случайной выборке  $x_k$  объема  $n$

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

Поскольку выборочное математическое ожидание построено по случайной выборке, то величина  $\bar{x}$  сама является случайной величиной. Рассмотрим ее математическое ожидание, при условии, что случайная величина  $\xi$  имеет математическое ожидание

$$E[\bar{x}] = \frac{1}{n} \sum_{k=1}^n E[x_k] = E[\xi].$$

Теперь рассмотрим дисперсию этой оценки, при условии существования дисперсии у случайной величины  $\xi$

$$D[\bar{x}] = \frac{1}{n^2} \sum_{k=1}^n D[x_k] = \frac{1}{n} D[\xi].$$

Выборочная дисперсия определяется по формуле

$$S_n^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2.$$

Оценим математическое ожидание выборочной дисперсии

$$E[S_n] = \frac{1}{n} \sum_{k=1}^n (y_k - \bar{y})^2 = \frac{1}{n} \sum_{k=1}^n y_k^2 - \bar{y}^2,$$

где  $y_k = x_k - E[x_k]$ .

Поскольку  $E[y_k] = 0$ ,  $E[y_k^2] = D[\xi]$ ,  $E[y_k y_l] = 0$  ( $k \neq l$ ), то

$$E[\bar{y}^2] = \frac{1}{n} \sum_{k,l=1}^n E[y_k y_l] = \frac{1}{n} D[\xi].$$

Таким образом,

$$S_n^2 = \frac{n-1}{n} D[\xi].$$

Математическое ожидание и дисперсия являются параметрами функции распределения вероятностей случайной величины, а их выборочные моменты суть оценки этих параметров. Через  $\theta$  обозначим параметр, тогда через  $\hat{\theta}_n$  мы обозначаем выборочную оценку параметра  $\theta$  по случайной выборке объема  $n$ .

Оценка  $\hat{\theta}_n$  называется несмещенной, если при любом  $n$

$$E[\hat{\theta}_n] = \theta.$$

Мы видим, что выборочное математическое ожидание является несмещенной оценкой. С другой стороны, выборочная дисперсия есть смещенная оценка. Поэтому часто вместо  $S_n^2$  используют оценку

$$s^2 = \frac{n}{n-1} S_n^2,$$

которая уже будет несмещенной.

Оценка  $\hat{\theta}_n$  называется состоятельной оценкой, если для любого  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\{|\hat{\theta}_n - \theta| < \varepsilon\} = 1.$$

Можно показать, что выборочное математическое ожидание и выборочная дисперсия представляют собой состоятельные оценки.

### 3. Интервальные оценки

Выборочные оценки параметров функций распределения представляют собой пример точечных оценок. Однако в математической статистике большую роль играют интервальные оценки. Пусть мы по случайной выборке оцениваем параметр  $\theta$ . Интервальной оценкой называется пара точечных оценок  $\underline{\theta}$  и  $\bar{\theta}$  таких, что

$$P\{\underline{\theta} < \theta < \bar{\theta}\} = 1 - 2\alpha.$$

При этом интервал  $(\underline{\theta}, \bar{\theta})$  называется доверительным интервалом неизвестного параметра  $\theta$ , соответствующим доверительной вероятности  $1 - 2\alpha$ .

## 4. Линейная регрессия

Рассмотрим важное приложение математической статистики при построении линейной регрессии. Понятие регрессии означает восстановление функциональной зависимости между множеством пар переменных. В случае линейной регрессии речь идет о восстановлении линейной зависимости.

Пусть мы имеем набор пар двух числовых переменных

$$(x_t, y_t), \quad t = 1, 2, \dots, n.$$

Задача регрессии состоит в том, чтобы найти функциональную зависимость

$$y = f(x),$$

которая будет в каком-либо смысле наилучшим образом аппроксимировать наши исходные данные. Задача построения такой регрессии относится к проблемам машинного обучения.

Мы будем рассматривать линейную регрессию, т.е. будем искать функцию  $f$  в следующем виде

$$y = ax + b.$$

В качестве меры близости к исходным значениям мы будем рассматривать метод наименьших квадратов. Согласно этому методу мы должны найти такие параметры  $a$  и  $b$ , чтобы они минимизировали следующую функцию

$$F(a, b) = \sum_{t=1}^n (y_t - (ax_t + b))^2.$$

Экстремум будем искать в точке, в которой градиент этой функции равен нулю. Дифференцируем по  $a$  и по  $b$  и приравнявая к нулю

$$\frac{\partial F}{\partial a} = -2 \sum_{t=1}^n x_t (y_t - ax_t - b) = 0.$$

$$\frac{\partial F}{\partial b} = -2 \sum_{t=1}^n (y_t - ax_t - b) = 0.$$

Решая эти уравнения относительно  $a$  и  $b$ , получаем

$$a = \frac{n \sum_{t=1}^n x_t y_t - \left( \sum_{t=1}^n x_t \right) \left( \sum_{t=1}^n y_t \right)}{n \sum_{t=1}^n x_t^2 - \left( \sum_{t=1}^n x_t \right)^2},$$

$$b = \frac{1}{n} \sum_{t=1}^n y_t - \frac{1}{n} \sum_{t=1}^n x_t a.$$

Введем обозначения

$$\hat{y}_t = ax_t + b,$$

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t,$$

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t.$$

По построению нашей регрессии мы имеем соотношение

$$\sum_{t=1}^n (y_t - \bar{y})^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 + \sum_{t=1}^n (\hat{y}_t - \bar{y})^2.$$

Эти слагаемые обозначаются

$$TSS = ESS + RSS,$$

TSS (total sum of squares), ESS (error sum of squares), RSS (regression sum of squares).

Показателем качества регрессии является коэффициент детерминации

$$R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}.$$

Этот коэффициент принимает значения от 0 до 1. Нулевое значение означает, что регрессии не дает никакой информации, а значение 1 говорит, что имеет место линейная зависимость между переменными.

**Литература**

- [1] *Шамин Р. В.* Функциональный анализ от нуля до единицы. — М.: ЛЕНАНД/URSS, 2016.
- [2] *Шамин Р. В.* Математические вопросы волн-убийц. — М.: ЛЕНАНД/URSS, 2016.
- [3] *Шамин Р. В.* Полугруппы операторов. — М.: РУДН, 2008.



