Содержание

1. Векторы. Линейная зависимость системы векторов. Базис линейного пространства. Скалярное произведение векторов.

Вектор — это направленный отрезок, который характеризует величину и направление. Геометрически его можно представить как стрелку от начала координат до некоторой точки.

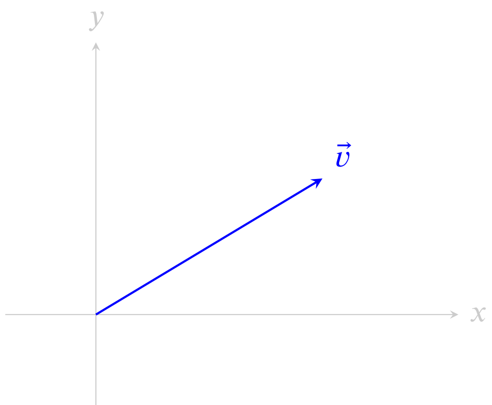
Алгебраически вектор в n-мерном пространстве $\mathbb{R}^n$ — это упорядоченный набор чисел:

$$\vec{v} = (v_1, v_2, \dots, v_n)$$

Примеры:

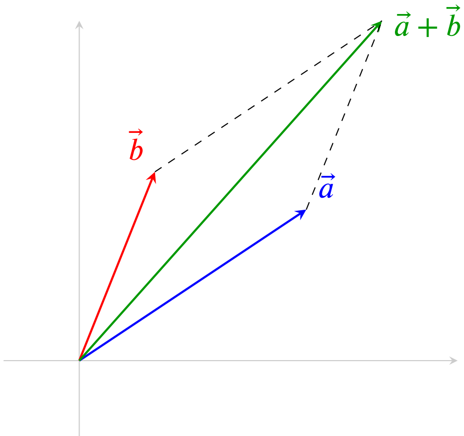
$$\vec{a} = (3, -1), \quad \vec{b} = (1, 2, 4)$$

Геометрическая интерпретация:



Основные операции:

- **Сложение:** $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$
- **Умножение на скаляр:** $\lambda \cdot (x, y) = (\lambda x, \lambda y)$
- **Нулевой вектор:** $\vec{0} = (0, 0, \dots, 0)$



Линейная зависимость и независимость векторов:

Рассмотрим векторы $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$. Если существует набор коэффициентов $\lambda_1, \dots, \lambda_k$ (не все нули), такой что:

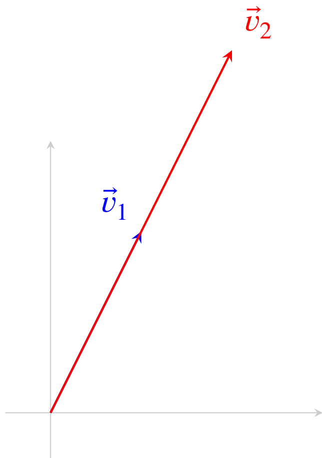
$$\lambda_1 \vec{v}_1 + \dots + \lambda_k \vec{v}_k = \vec{0}$$

то векторы — **линейно зависимы**.

Если единственное решение — тривиальное ($\lambda_1 = \dots = \lambda_k = 0$), то векторы — **линейно независимы**.

Пример:

$$\vec{v}_1 = (1, 2), \quad \vec{v}_2 = (2, 4) \Rightarrow \vec{v}_2 = 2\vec{v}_1 \Rightarrow \text{зависимы}$$

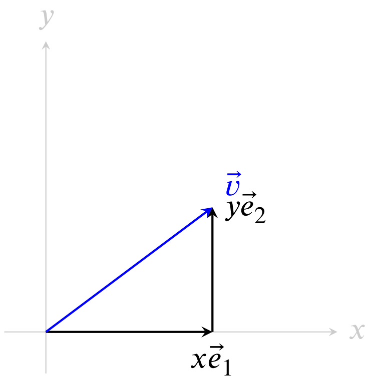


Базис линейного пространства:

Базис — это система линейно независимых векторов, которая порождает всё пространство. Любой вектор пространства выражается через базис как линейная комбинация.

В $\mathbb{R}^2$ стандартный базис:

$$\vec{e}_1 = (1, 0), \quad \vec{e}_2 = (0, 1) \Rightarrow \vec{v} = x\vec{e}_1 + y\vec{e}_2$$



Размерность пространства — это количество векторов в базисе. Например, в $\mathbb{R}^3$ базис содержит 3 вектора.

Скалярное произведение векторов:

Скалярное произведение двух векторов $\vec{a} = (a_1, \dots, a_n)$ и $\vec{b} = (b_1, \dots, b_n)$:

$$\langle \vec{a}, \vec{b} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

Пример:

$$\vec{a} = (1, 2), \quad \vec{b} = (3, 4) \Rightarrow \langle \vec{a}, \vec{b} \rangle = 1 \cdot 3 + 2 \cdot 4 = 11$$

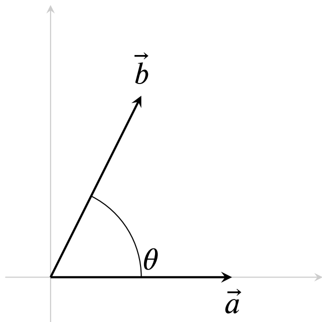
Свойства:

- Симметрия: $\langle \vec{a}, \vec{b} \rangle = \langle \vec{b}, \vec{a} \rangle$
- Линейность: $\langle \alpha\vec{a} + \beta\vec{c}, \vec{b} \rangle = \alpha\langle \vec{a}, \vec{b} \rangle + \beta\langle \vec{c}, \vec{b} \rangle$
- $\langle \vec{a}, \vec{a} \rangle = \|\vec{a}\|^2$

Геометрически:

$$\langle \vec{a}, \vec{b} \rangle = \|\vec{a}\| \|\vec{b}\| \cos\theta$$

где θ — угол между векторами. Если $\langle \vec{a}, \vec{b} \rangle = 0$ — векторы ортогональны (перпендикулярны).



Длина вектора (норма):

$$\|\vec{a}\| = \sqrt{\langle \vec{a}, \vec{a} \rangle}$$

Угол между векторами:

$$\cos\theta = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\| \, \|\vec{b}\|}$$

—

Итоги:

- Векторы — фундаментальные объекты в линейной алгебре.
- Линейная зависимость помогает понимать структуру пространства.
- Базис — минимальный набор независимых векторов, порождающих всё пространство.
- Скалярное произведение связывает векторы с геометрией — углами и длинами.

2. Матрицы. Их свойства. Транспонированная матрица. Ранг матрицы

Матрица — это прямоугольная таблица чисел, организованная в строки и столбцы. Она записывается в виде:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

где a_{ij} — элемент матрицы на i-й строке и j-м столбце.

Обозначения и размерность

Матрицу обозначают заглавной латинской буквой (A, B, C и т.д.). Размерность матрицы — это количество строк и столбцов. Если в матрице m строк и n столбцов, её размер обозначают как $m \times n$.

Примеры:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 3 & 5 \end{pmatrix}$$

Здесь A — квадратная матрица 2×2, B — прямоугольная матрица 2×3.

—

Основные типы матриц

- **Нулевая матрица:** все элементы равны нулю.
- **Единичная матрица I_n:** квадратная матрица с единицами на главной диагонали и нулями вне её.

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- **Диагональная матрица:** все элементы вне главной диагонали равны нулю.
- **Квадратная матрица:** одинаковое число строк и столбцов.
- **Столбец (вектор-столбец):** матрица размером $m \times 1$.
- **Строка (вектор-строка):** матрица размером $1 \times n$.

—

Операции с матрицами

1. **Сложение:** складываются поэлементно. Возможно только для матриц одинакового размера.

$$A + B = \left(a_{ij} + b_{ij}\right)$$

2. **Умножение на скаляр:**

$$\lambda A = \left(\lambda \cdot a_{ij}\right)$$

3. **Умножение матриц:** если A — матрица размера $m \times n$, а B — $n \times k$, то их произведение $C = AB$ будет размером $m \times k$:

$$c_{ij} = \sum_{r=1}^{n} a_{ir} \cdot b_{rj}$$

4. **Транспонирование (см. ниже)** — замена строк и столбцов.

—

Свойства операций

- Коммутативность сложения: $A + B = B + A$
- Ассоциативность: $(A + B) + C = A + (B + C)$
- Дистрибутивность: $\lambda(A + B) = \lambda A + \lambda B$
- $(AB)^T = B^T A^T$ — важное свойство транспонирования

—

Транспонированная матрица

Матрица A^T (читается: «A транспонированная») получается из A заменой строк на столбцы. Формально:

$$(A^T)_{ij} = A_{ji}$$

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \quad \Rightarrow \quad A^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$$

Свойства транспонирования:

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AB)^T = B^T A^T$

—

Ранг матрицы

Ранг матрицы — это максимальное число линейно независимых строк (или столбцов) в матрице.

Обозначается: $\mathrm{rank}(A)$.

Интуитивно: ранг показывает, сколько "уникальной" информации содержится в строках или столбцах.

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \Rightarrow \text{строки линейно зависимы} \Rightarrow \mathrm{rank}(A) = 1$$

Другой пример:

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow \mathrm{rank}(B) = 3$$

Как находить ранг?

Обычно с помощью преобразования матрицы к **ступенчатому виду** методом Гаусса. Количество ненулевых строк после преобразования и будет рангом.

Пример пошагово:

Дана матрица:

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 3 & 6 & 3 \end{pmatrix}$$

Видим: вторая и третья строки — кратные первой. После приведения:

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \mathrm{rank}(A) = 1$$

—

Геометрическая интерпретация ранга

Векторы-строки (или столбцы) матрицы можно представить как векторы в пространстве. Ранг говорит о том, какое пространство они натягивают:

- Ранг 1: все лежат на одной прямой

- Ранг 2: в одной плоскости
- Ранг 3: в трёхмерном пространстве и т.д.

—

Важность ранга

Ранг используется в:

- Исследовании решений линейных систем: число решений зависит от ранга матрицы коэффициентов.
- Анализе линейной зависимости строк/столбцов.
- Проверке обратимости матрицы: квадратная матрица обратима $\iff$ её ранг равен размерности.

—

Выводы по теме

- Матрицы — основа линейной алгебры. Они обобщают векторы, храня данные и операции.
- Транспонирование меняет строки и столбцы местами.
- Ранг показывает, сколько независимых строк/столбцов содержит матрица.
- Если ранг меньше полной размерности — значит, матрица ”выражает” только подпространство.

3. Сложение, умножение матрицы на число, умножение матриц, транспонирование матриц. Обратная матрица

Сложение матриц

Две матрицы A и B одинакового размера ($m \times n$) можно сложить, если у них совпадают размеры. Сложение происходит поэлементно:

$$(A + B)_{ij} = A_{ij} + B_{ij}$$

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \Rightarrow A + B = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}$$

Свойства сложения:

- Коммутативность: $A + B = B + A$
- Ассоциативность: $(A + B) + C = A + (B + C)$
- Существование нулевой матрицы O (нулевая поэлементно): $A + O = A$

—

Умножение матрицы на число

Если $\lambda \in \mathbb{R}$ — число (скаляр), то умножение $\lambda \cdot A$ означает умножение каждого элемента матрицы на это число:

$$(\lambda A)_{ij} = \lambda \cdot A_{ij}$$

Пример:

$$A = \begin{pmatrix} 2 & -1 \\ 0 & 3 \end{pmatrix}, \quad \lambda = 4 \Rightarrow 4A = \begin{pmatrix} 8 & -4 \\ 0 & 12 \end{pmatrix}$$

Свойства:

- $\lambda(\mu A) = (\lambda\mu)A$
- $(\lambda + \mu)A = \lambda A + \mu A$
- $\lambda(A + B) = \lambda A + \lambda B$

—

Умножение матриц

Матрицы A и B можно перемножить, если **число столбцов в** A равно **числу строк в** B.

Если A — размера $m \times n$, а B — $n \times k$, то произведение $C = AB$ — это матрица $m \times k$, где:

$$C_{ij} = \sum_{r=1}^{n} A_{ir} \cdot B_{rj}$$

То есть: элемент C_{ij} получается как скалярное произведение i-й строки A и j-го столбца B.

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}$$

$$AB = \begin{pmatrix} 1 \cdot 0 + 2 \cdot 2 & 1 \cdot 1 + 2 \cdot 3 \\ 3 \cdot 0 + 4 \cdot 2 & 3 \cdot 1 + 4 \cdot 3 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ 8 & 15 \end{pmatrix}$$

Важно: $AB \neq BA$ в общем случае! Умножение матриц **не коммутативно**.

Свойства:

- Ассоциативность: $A(BC) = (AB)C$
- Дистрибутивность: $A(B + C) = AB + AC$
- $(AB)^T = B^T A^T$ — транспонирование произведения

—

Транспонирование матрицы

Транспонирование — это операция, при которой строки становятся столбцами, а столбцы — строками.

Для любой матрицы A, её транспонированная матрица A^T определяется как:

$$(A^T)_{ij} = A_{ji}$$

Пример:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Свойства транспонирования:

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AB)^T = B^T A^T$

Эта операция часто используется при симметризации, а также в определениях симметрических и ортогональных матриц.

—

Обратная матрица

Обратная матрица A^{-1} к квадратной матрице A определяется как:

$$A \cdot A^{-1} = A^{-1} \cdot A = I$$

где I — единичная матрица той же размерности.

Условия существования:

- Матрица должна быть **квадратной**.
- Её **определитель не должен равняться нулю** ($\det A \neq 0$).
- Ранг A должен равняться её размерности: $\operatorname{rank}(A) = n$

Пример:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{\det A} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \frac{1}{(1 \cdot 4 - 2 \cdot 3)} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1.5 & -0.5 \end{pmatrix}$$

Способы нахождения обратной матрицы

1. Для 2×2-матриц:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

2. Для больших матриц:
   - Через присоединённую матрицу (алгебраические дополнения + транспонирование + деление на определитель)
   - Метод Гаусса: расширение A до $[A|I]$ и приведение к $[I|A^{-1}]$

—

Свойства обратной матрицы

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^T)^{-1} = (A^{-1})^T$

Важно: не все матрицы имеют обратную. Такие матрицы называются **вырожденными**.

—

Применения обратной матрицы

- Решение систем уравнений: $A\vec{x} = \vec{b} \Rightarrow \vec{x} = A^{-1}\vec{b}$
- Вывод формул в статистике и машинном обучении
- Нормализация линейных преобразований
- Преобразование координат

—

Выводы

- Операции над матрицами (сложение, умножение, транспонирование) формируют алгебраическую структуру.
- Умножение матриц — основа линейных отображений и систем уравнений.
- Транспонирование — полезная симметризующая операция.
- Обратная матрица существует только у невырожденных квадратных матриц и даёт способ обращения линейных операторов.

4. Аппроксимация и интерполяция функций

Аппроксимация и **интерполяция** — это два метода приближённого описания функций, основанные на наборе дискретных точек.

Интерполяция — это построение функции, которая точно проходит через заданные точки. **Аппроксимация** — это построение функции, которая приближённо описывает данные, но может не проходить через все точки.

—

Постановка задачи

Пусть дана таблица значений:

$$(x_0, y_0),\ (x_1, y_1),\ \dots,\ (x_n, y_n)$$

Наша цель — построить функцию $f(x)$ такую, что:

- **Для интерполяции:** $f(x_i) = y_i$ для всех i
- **Для аппроксимации:** $f(x_i) \approx y_i$

—

Интерполяция: идея и цель

Интерполяция позволяет восстанавливать значение функции в промежуточных точках, не выходя за пределы интервала $[x_0, x_n]$.

Пример: если известно, что

$$f(1) = 2, \quad f(2) = 4, \quad f(3) = 6$$

можем интерполировать $f(x)$, скажем, через многочлен второй степени и вычислить $f(1.5)$.

—

Линейная интерполяция

Между двумя точками (x_0, y_0) и (x_1, y_1) интерполяционная функция задаётся по формуле:

$$f(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0)$$

Это уравнение прямой, проходящей через две точки. Очень просто, но недостаточно точно для сложных функций.

—

Полиномиальная интерполяция

Если заданы $n + 1$ точек, можно построить единственный многочлен степени не выше n, который проходит через все точки.

Формула Лагранжа:

$$P_n(x) = \sum_{i=0}^{n} y_i \cdot L_i(x), \quad \text{где } L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j}$$

Каждое $L_i(x)$ — базисный многочлен Лагранжа, равный 1 в точке x_i и 0 в остальных x_j.

Проблема: при увеличении числа узлов интерполяция может сильно колебаться (эффект Рунге), особенно на концах интервала.

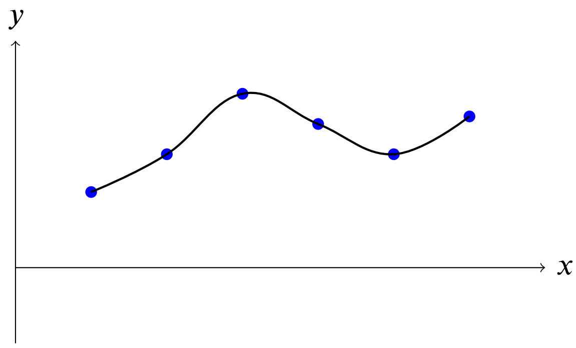
—

Сплайны (кубическая интерполяция)

Сплайн-интерполяция делит интервал на участки и на каждом строит многочлен степени 3 (кубический сплайн), с условием сглаженности в стыках.

- Гладкость первого и второго порядка: C^2-непрерывность
- Сплайны хорошо подходят для графиков, траекторий и данных с шумом

Визуально:



—

Аппроксимация: общая идея

Аппроксимация применяется, когда функция неизвестна, но имеются измеренные значения с шумом. Здесь уже не требуется точное прохождение через точки.

Идея: найти «наилучшую» функцию $f(x)$, которая *приблизительно* соответствует данным.

Часто ищут функцию в виде:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

Аппроксимация методом наименьших квадратов (МНК)

Пусть есть точки (x_i, y_i), и нужно найти параметры $a_0, a_1, \dots, a_n$, минимизирующие отклонение:

$$S(a_0, a_1, \dots, a_n) = \sum_{i=0}^{n} \big(f(x_i) - y_i\big)^2$$

Минимум достигается при решении системы нормальных уравнений, которая получается из частных производных S по параметрам a_k.

Частный случай — линейная аппроксимация:

$$f(x) = a_0 + a_1 x$$

Тогда минимизируется:

$$S(a_0, a_1) = \sum_{i=1}^{n} \left(a_0 + a_1 x_i - y_i\right)^2$$

Решение:

$$\begin{cases} na_0 + a_1 \sum x_i = \sum y_i \\ a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i \end{cases}$$

—

Сравнение: интерполяция vs аппроксимация

Критерий	Интерполяция	Аппроксимация
Проходит через точки	Да	Не обязательно
Чувствительность к шуму	Высокая	Устойчивая
Сложность вычислений	Средняя–высокая	Низкая–средняя
Гладкость	Может не быть	Обычно есть

—

Практические применения

- **Интерполяция**:
  - Таблицы и справочники
  - Заполнение пропущенных значений
  - Графическая визуализация
- **Аппроксимация**:
  - Обработка измерений с шумом
  - Моделирование реальных процессов
  - Предсказания, тренды

—

Выводы по теме

- Интерполяция позволяет точно восстановить функцию внутри диапазона, но может колебаться.
- Аппроксимация — более устойчивая техника, особенно с шумными или неточными данными.

- Полиномы Лагранжа и кубические сплайны — мощные методы интерполяции.
- Метод наименьших квадратов — классический способ аппроксимации, широко используемый в статистике и машинном обучении.

5. Производные. Необходимое и достаточное условия дифференцируемости функции. Частные и полные производные

Понятие производной функции одной переменной

Пусть $f(x)$ определена в окрестности точки x_0. Производная функции f в точке x_0 — это предел отношения приращения функции к приращению аргумента:

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

Если предел существует, то говорят, что функция **дифференцируема** в точке x_0.

Геометрический смысл: производная — это угловой коэффициент касательной к графику функции в данной точке.

Дифференцируемость и непрерывность

- Если функция дифференцируема в точке x_0, то она непрерывна в этой точке.
- Обратное неверно: непрерывность не гарантирует существование производной.

Пример (не дифференцируема):

$$f(x) = |x| \Rightarrow f'(0) \text{ не существует, хотя } f \text{ непрерывна в } 0$$

—

Производная по направлению и частные производные

Пусть $f(x, y)$ — функция двух переменных.

Производную по направлению можно определить через вектор направления $\vec{l} = (l_1, l_2)$:

$$D_{\vec{l}} f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + hl_1, y_0 + hl_2) - f(x_0, y_0)}{h}$$

Частные производные — это производные по отдельным переменным:

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}$$

Обозначения:

$$f_x(x, y), \quad f_y(x, y), \quad \text{или } \partial_x f,\ \partial_y f$$

Пример: пусть $f(x, y) = x^2 y + \sin(xy)$

$$\frac{\partial f}{\partial x} = 2xy + y\cos(xy), \quad \frac{\partial f}{\partial y} = x^2 + x\cos(xy)$$

—

Необходимое и достаточное условия дифференцируемости функции многих переменных

Функция $f(x, y)$ называется **дифференцируемой в точке** (x_0, y_0), если приращение можно представить в виде:

$$\Delta f = f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) = A\Delta x + B\Delta y + o(\sqrt{\Delta x^2 + \Delta y^2})$$

где A и B — постоянные (зависят от точки), а $o(\rho)$ — бесконечно малая по сравнению с ρ.

Формально: f дифференцируема в (x_0, y_0), если существует линейное отображение L, приближающее Δf.

Необходимое условие дифференцируемости

Если f дифференцируема в (x_0, y_0), то:

- Частные производные $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ существуют
- $\Delta f = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + o(\sqrt{\Delta x^2 + \Delta y^2})$

Достаточное условие дифференцируемости

Если:

- Частные производные существуют в окрестности точки
- И они непрерывны в точке (x_0, y_0)

то f дифференцируема в этой точке.

—

Градиент и направление наибольшего роста

Градиент — это вектор, составленный из всех частных производных:

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

Производная функции по направлению вектора $\vec{l}$ выражается как скалярное произведение:

$$D_{\vec{l}} f = \nabla f \cdot \vec{l}$$

Свойства:

- Направление градиента — направление наибольшего возрастания функции.
- Если $\nabla f = \vec{0}$, то это критическая точка (возможно максимум, минимум или седло).

—

Полный дифференциал

Если функция $f(x, y)$ дифференцируема, то её полное приращение можно выразить через полный дифференциал:

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy$$

Пример: $f(x, y) = x^2 y + \sin(xy)$

$$df = (2xy + y\cos(xy))dx + (x^2 + x\cos(xy))dy$$

Полный дифференциал используется:

- В оценке приращений функции
- В дифференциальной геометрии и анализе ошибок
- При переходе к новым координатам

—

Итоги

- Производная — это мера изменения функции.
- Частные производные — изменения по осям координат.
- Дифференцируемость функции двух переменных требует не только существования производных, но и их «согласованного» поведения.
- Градиент показывает направление роста функции.
- Полный дифференциал — обобщение производной на многомерные функции.

6. Частные производные. Градиент функции. Производная по направлению

Частные производные функции двух переменных

Пусть $f(x, y)$ — функция двух переменных, определённая в окрестности точки (x_0, y_0). Тогда:

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$
$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}$$

Частные производные — это производные функции по одной переменной при фиксированной другой.

Геометрический смысл: производная по x — это скорость изменения функции вдоль оси x, при фиксированном y.

Обозначения:

$$f_x = \frac{\partial f}{\partial x}, \quad f_y = \frac{\partial f}{\partial y}$$

Пример: Пусть $f(x, y) = x^2 y + \sin(xy)$. Тогда:

$$\frac{\partial f}{\partial x} = 2xy + y\cos(xy)$$

$$\frac{\partial f}{\partial y} = x^2 + x\cos(xy)$$

—

Частные производные высших порядков

Можно вычислять производные второго порядка и выше:

$$\frac{\partial^2 f}{\partial x^2}, \quad \frac{\partial^2 f}{\partial y^2}, \quad \frac{\partial^2 f}{\partial x \partial y}, \quad \frac{\partial^2 f}{\partial y \partial x}$$

Если f дважды непрерывно дифференцируема, то:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

(теорема Шварца о равенстве смешанных производных)

—

Градиент функции

Пусть $f(x, y)$ — дифференцируемая функция. Тогда **градиент** функции f в точке (x_0, y_0) — это вектор:

$$\nabla f(x_0, y_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right)$$

Если f зависит от n переменных, то градиент — вектор из n компонент:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$

Геометрический смысл:

- Направление градиента — это направление *наибольшего роста функции*.
- Его длина — скорость наибольшего изменения.
- Если $\nabla f = \vec{0}$, то точка является критической (возможный экстремум).

Пример: Пусть $f(x, y) = x^2 + y^2 \Rightarrow \nabla f = (2x, 2y)$ — вектор, указывающий от начала координат.

—

Производная функции по направлению

Пусть функция $f(x, y)$ определена в точке (x_0, y_0), и задан единичный вектор направления:

$$\vec{l} = (\cos\alpha, \cos\beta), \quad \|\vec{l}\| = 1$$

Производная функции f по направлению вектора $\vec{l}$ в точке (x_0, y_0):

$$D_{\vec{l}} f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h\cos\alpha, y_0 + h\cos\beta) - f(x_0, y_0)}{h}$$

Свойство: Если функция f дифференцируема, то производная по направлению вычисляется через градиент:

$$D_{\vec{l}} f = \nabla f \cdot \vec{l} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \cdot (\cos\alpha, \cos\beta)$$

Это **скалярное произведение** векторов: градиента и направления.
Следствия:

- Наибольшая производная по направлению достигается в направлении градиента.
- Если $\vec{l} \perp \nabla f$, то производная по направлению равна нулю (функция не меняется вдоль этого направления).

Пример:

Пусть $f(x, y) = x^2 y + y$, точка $M = (1, 2)$, направление $\vec{l} = \frac{1}{\sqrt{2}}(1, 1)$.

$$\nabla f = \left(2xy, x^2 + 1\right) \Rightarrow \nabla f(1, 2) = (4, 2)$$

$$D_{\vec{l}} f = \nabla f \cdot \vec{l} = \frac{1}{\sqrt{2}}(4 + 2) = \frac{6}{\sqrt{2}} = 3\sqrt{2}$$

—

Итоги

- Частные производные показывают, как функция меняется по каждой координате.
- Градиент — главный вектор изменения, указывает направление наибольшего роста.

- Производная по направлению обобщает понятие производной на произвольное направление.
- Всё вместе используется в оптимизации, градиентном спуске, анализе поверхности.

7. Численные методы решения задач оптимизации. Метод Ньютона и секущей. Методы покоординатного и градиентного спуска

Оптимизация — это процесс нахождения минимума или максимума функции при заданных ограничениях (или без них). В задачах прикладной математики часто встречаются функции, которые слишком сложны для аналитического нахождения экстремума, поэтому используют численные методы.

В данной секции мы рассмотрим:

- Метод Ньютона;
- Метод секущей;
- Метод покоординатного спуска;
- Метод градиентного спуска.

7.1. Постановка задачи оптимизации

Пусть дана функция:

$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$$

Необходимо найти точку $\mathbf{x}^* \in \mathbb{R}^n$, такую что:

$$f(\mathbf{x}^*) \le f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$$

(для задачи минимизации).

Для поиска минимума часто используют производные:

- **Необходимое условие экстремума:** $\nabla f(\mathbf{x}^*) = 0$
- **Достаточное условие минимума:** матрица Гессе $H_f(\mathbf{x}^*)$ положительно определена.

7.2. Метод Ньютона

Метод Ньютона — это численный метод, который использует разложение функции в ряд Тейлора второго порядка для приближения к экстремуму.

7.2.1. Алгоритм в одномерном случае

Для уравнения $f'(x) = 0$:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Здесь:

- $f'(x_k)$ — первая производная функции в точке x_k;
- $f''(x_k)$ — вторая производная.

7.2.2. Многомерный случай

Для векторной функции:

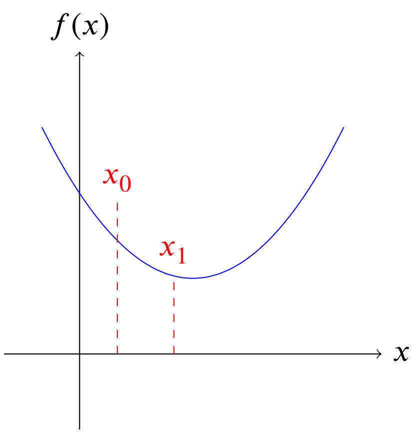
$$\mathbf{x}_{k+1} = \mathbf{x}_k - H_f^{-1}(\mathbf{x}_k)\nabla f(\mathbf{x}_k)$$

где $H_f(\mathbf{x}_k)$ — матрица Гессе.

7.2.3. Плюсы и минусы

- **Плюсы:** Быстрая сходимость (обычно квадратичная).
- **Минусы:** Нужно вычислять и инвертировать матрицу Гессе, что дорого для больших n.

7.2.4. Графическая иллюстрация



7.3. Метод секущей

Метод секущей — это упрощённая версия метода Ньютона, в которой вторую производную заменяют приближением по разностям.

$$x_{k+1} = x_k - f'(x_k) \cdot \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})}$$

- **Плюс:** Не нужно вычислять вторую производную.
- **Минус:** Скорость сходимости ниже, чем у метода Ньютона.

7.4. Метод покоординатного спуска

Метод покоординатного спуска оптимизирует функцию, изменяя одну координату за раз, оставляя остальные фиксированными.

7.4.1. Алгоритм

1. Выбираем начальную точку $\mathbf{x}_0$.
2. Для каждой координаты i ищем минимум функции по x_i при фиксированных остальных координатах.
3. Повторяем процесс до сходимости.

7.4.2. Особенности

- Подходит для задач, где минимизация по одной переменной проста.
- Может сходиться медленно, если переменные сильно связаны.

7.5. Метод градиентного спуска

Метод градиентного спуска — один из самых популярных методов оптимизации.

7.5.1. Основная идея

Движемся из текущей точки в направлении, противоположном градиенту функции (т.к. градиент указывает направление наибольшего роста).

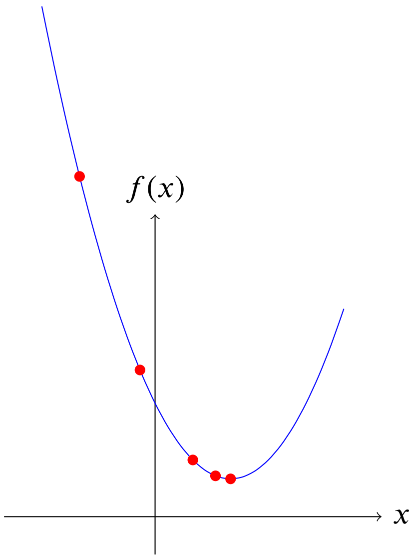
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha \nabla f(\mathbf{x}_k)$$

где $\alpha > 0$ — шаг обучения.

7.5.2. Выбор шага

- Слишком большой α — можем «перепрыгнуть» минимум.
- Слишком маленький α — сходимость медленная.

7.5.3. Графическая иллюстрация



7.5.4. Варианты

- **С постоянным шагом**
- **С адаптивным шагом** (например, Adam, RMSprop)

7.6. Заключение

Выбор метода оптимизации зависит от свойств задачи:

- Метод Ньютона — быстрый, но требует вычислений второй производной.
- Метод секущей — компромисс, подходит для случаев, когда вторая производная недоступна.
- Покоординатный спуск — полезен при раздельной оптимизации переменных.
- Градиентный спуск — универсальный, особенно в задачах машинного обучения.

8. Основные понятия теории вероятностей. Определение вероятности. Вероятность случайных событий. Формула полной вероятности

Введение и мотивация

Теория вероятностей изучает случайные явления и формализует интуицию о «шансах» наступления событий. Классические примеры: подбрасывание монеты, бросок игральной кости, выбор случайного человека в опросе. Цель — построить строгую математику для рассуждений о вероятности событий, их сочетаниях и последствиях.

В этой секции мы подробно разберём:

- базовые понятия: пространство исходов, события;
- определения вероятности (классическое, частотное, аксиоматическое);
- свойства вероятности (аддитивность, монотонность и т.д.);
- условную вероятность и независимость;
- формулу полной вероятности и практические примеры;
- теорему Байеса и применение формулы полной вероятности при вычислении апостериорных вероятностей;
- включение иллюстраций и задач на проверку.

8.1. Пространство элементарных исходов и события

Определение. *Пространство элементарных исходов* (универсум, sample space) обозначается Ω и содержит все возможные исходы случайного эксперимента. Каждый элемент $\omega \in \Omega$ называется *элементарным исходом*.

События. Событие A — любое подмножество Ω (в базовом подходе). Если при проведении эксперимента полученный исход ω лежит в A, то говорят, что событие A произошло.

Примеры.

- Подбрасывание монеты: $\Omega = \{\text{Орел},\ \text{Решка}\}$.
- Бросок правильного шестигранного кубика: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Случайный выбор человека из группы: Ω — множество людей группы.

8.2. Классическое определение вероятности

Если Ω состоит из конечного числа равновозможных элементарных исходов (классическая ситуация), то вероятность события $A \subseteq \Omega$ определяется как

$$P(A) = \frac{|\{\omega \in \Omega : \ \omega \in A\}|}{|\Omega|}.$$

То есть — отношение числа благоприятных исходов к общему числу исходов.

Пример. При броске справедливой кости вероятность того, что выпадет четное число:

$$P(\text{четное}) = \frac{3}{6} = \frac{1}{2}.$$

8.3. Частотная (эмпирическая) интерпретация

В частотном подходе вероятность события A понимается как предел относительной частоты при многократном повторении эксперимента:

$$P(A) = \lim_{n \to \infty} \frac{N_n(A)}{n},$$

где $N_n(A)$ — количество опытов из n, в которых событие A произошло (при предположении существования предела).

8.4. Аксиоматическое определение (Колмогоров)

Для общей теории наиболее строгой и удобной является аксиоматическая постановка.

Определение. Пусть Ω — множество исходов, $\mathscr{F}$ — множество событий (обычно σ-алгебра над Ω). Функция $P : \mathscr{F} \to [0, 1]$ называется вероятностной мерой, если выполняются аксиомы Колмогорова:

1. (Неотрицательность) $\forall A \in \mathscr{F}$: $P(A) \geq 0$;
2. (Нормировка) $P(\Omega) = 1$;
3. (Счётная аддитивность) Для любых попарно несовместных событий $A_1, A_2, \dots \in \mathscr{F}$ выполняется

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

Примечание. Для дискретных задач достаточно конечной аддитивности; для непрерывных процессов нужна счётная аддитивность.

8.5. Следствия из аксиом (свойства вероятности)

Из аксиом Колмогорова легко выводятся важные свойства:

1. $P(\varnothing) = 0$.
2. Для любого $A \in \mathscr{F}$: $0 \leq P(A) \leq 1$.
3. Моночленость: если $A \subseteq B$, то $P(A) \leq P(B)$.
4. Формула суммы для двух событий:

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

   (Доказательство: разложить объединение на попарно несовместные части.)
5. Формула включения—исключения для трёх событий:

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

8.6. Условная вероятность

Определение. Пусть $P(B) > 0$. *Условная вероятность* события A при условии B определяется как

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

Интуиция: мы рассматриваем пространство исходов, ограниченное тем, что произошло событие B, и оцениваем долю тех исходов в B, при которых произошло также A.

Свойства.

- Для фиксированного B с $P(B) > 0$, $P(\cdot \mid B)$ является вероятностной мерой на $\mathscr{F}$.
- Из определения следует формула для совместной вероятности:

  $$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A).$$

Простейший пример условной вероятности

Бросаем две монеты. Событие A — «вторая монета — орёл», событие B — «хотя бы одна монета — орёл». Найдём $P(A \mid B)$.

Исходы: $\Omega = \{HH, HT, TH, TT\}$. $A = \{HH, TH\}$, $B = \{HH, HT, TH\}$. Тогда

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{HH, TH\})}{P(\{HH, HT, TH\})} = \frac{2/4}{3/4} = \frac{2}{3}.$$

8.7. Независимость событий

Два события. События A и B называются (статистически) *независимыми*, если

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

Эквивалентно: $P(A \mid B) = P(A)$ (при $P(B) > 0$).

Несколько событий. Система событий $\{A_i\}_{i \in I}$ называется взаимно независимой (или попарно и всесторонне независимой) если для любой конечной подсемьи $A_{i_1}, \dots, A_{i_k}$ выполняется:

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_k}).$$

Важно — попарная независимость не влечёт взаимной независимости для трёх и более событий (пример с братьями и сестрами и т.п.).

8.8. Формула полной вероятности (закон полной вероятности)

Формулировка. Пусть $\{H_1, H_2, \dots, H_n\}$ — разбиение пространства Ω (то есть попарно несовместные события с $\bigcup_{i=1}^n H_i = \Omega$) и $P(H_i) > 0$ для всех i. Тогда для любого события A выполнена формула полной вероятности:

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

Доказательство (простой): Так как $\{H_i\}$ — разбиение, имеем представление множества A как объединение попарно несовместных множеств:

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

поэтому по аддитивности:

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

Интуиция. Формула полной вероятности разлагает вероятность события A по сценариям H_i — каждому сценарию придаётся вероятность $P(H_i)$ и свой вклад $P(A \mid H_i)$.

Пример: заводы и брак (развернуто)

Пусть изделие может быть произведено на одном из трёх заводов H_1, H_2, H_3 с вероятностями выпуска $P(H_1) = 0.6$, $P(H_2) = 0.3$, $P(H_3) = 0.1$. Пусть вероятность брака на каждом заводе: $P(A \mid H_1) = 0.01$, $P(A \mid H_2) = 0.02$, $P(A \mid H_3) = 0.05$. Найдём общую вероятность брака:

По формуле полной вероятности:

$$P(A) = 0.6 \cdot 0.01 + 0.3 \cdot 0.02 + 0.1 \cdot 0.05 = 0.006 + 0.006 + 0.005 = 0.017.$$

8.9. Теорема Байеса (формула для апостериорных вероятностей)

Формулировка. При тех же условиях, что и для формулы полной вероятности, для любого j имеем:

$$P(H_j \mid A) = \frac{P(H_j) P(A \mid H_j)}{\sum_{i=1}^{n} P(H_i) P(A \mid H_i)}.$$

Это следует непосредственно из определения условной вероятности и формулы полной вероятности:

$$P(H_j \mid A) = \frac{P(H_j \cap A)}{P(A)} = \frac{P(H_j) P(A \mid H_j)}{P(A)}.$$

Интуиция. Теорема Байеса позволяет обновить априорные вероятности $P(H_j)$ на основе наблюдения события A и получить апостериорную вероятность того, что гипотеза H_j истинна.

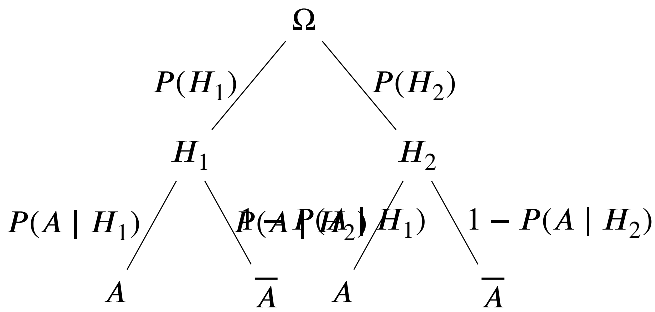
Пример (обратный пример завода)

Используем данные предыдущего примера и предположим, что обнаружен брак (событие A). Найдём вероятность того, что изделие произведено на заводе H_3:

$$P(H_3 \mid A) = \frac{0.1 \cdot 0.05}{0.017} = \frac{0.005}{0.017} \approx 0.294.$$

То есть, при обнаруженном браке вероятность, что изделие с завода №3, существенно увеличилась с 0.1 до ≈0.294.

8.10. Дерево вероятностей (иллюстрация)



Дерево помогает визуально аккумулировать произведения вероятностей вдоль ветвей (например, $P(H_1 \cap A) = P(H_1)P(A \mid H_1)$).

8.11. Формула полной вероятности для непрерывного случая

Если пространство условно разбивается по непрерывному параметру (или H — событие с непрерывным индексом), то формула принимает интегральную форму. Например, если случайная величина Θ имеет плотность $p_\Theta(\theta)$ и при фиксированном θ наблюдается событие A с условной вероятностью $P(A \mid \Theta = \theta)$, то

$$P(A) = \int_{-\infty}^{\infty} P(A \mid \Theta = \theta)\, p_\Theta(\theta)\, d\theta.$$

8.12. Некоторые важные следствия и полезные формулы

- **Комментирование условной вероятности:** $P(A \mid B) + P(\overline{A} \mid B) = 1$ при $P(B) > 0$.

- **Закон умножения для нескольких событий:**

$$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B).$$

- **Полная формула включения—исключения** (общая) позволяет вычислить вероятность объединения любого конечного числа событий.

8.13. Типичные ошибки и ловушки

- Ошибочно приравнивать несвязанные понятия: равновероятность исходов — это сильное условие, не выполняющееся в реальных задачах без проверки.

- Путаница между условной вероятностью $P(A \mid B)$ и $P(B \mid A)$ — они, как правило, не равны.

- Пренебрежение условием $P(B) > 0$ при использовании условных вероятностей.

- Не различать попарную независимость и взаимную независимость.

8.14. Упражнения для самопроверки (с ответами)

1. (**Простой**) Подбрасывают две честные монеты. Какова вероятность того, что выпадет ровно один орёл?
   Решение: исходы $\{HH, HT, TH, TT\}$, благоприятные: $\{HT, TH\}$, $P = 2/4 = 1/2$.

2. (**Формула полной вероятности**) Два источника генерируют сообщения: источник 1 — с вероятностью 0.7, источник 2 — с вероятностью 0.3. Вероятность ошибки в сообщении для 1-го — 0.01, для 2-го — 0.05. Какова общая вероятность ошибки?
   Решение: $P(\text{ошиб.}) = 0.7 \cdot 0.01 + 0.3 \cdot 0.05 = 0.007 + 0.015 = 0.022$.

3. (**Байес**) С учётом предыдущей задачи: найдите вероятность того, что сообщение пришло от источника 2, если оно оказалось ошибочным.
   Решение: $P(H_2 \mid \text{ошиб.}) = \dfrac{0.3 \cdot 0.05}{0.022} = \dfrac{0.015}{0.022} \approx 0.6818$.

4. (**Независимость**) Два броска правильной кости: событие A — «в первом броске выпало 6», событие B — «во втором выпало 6». Независимы ли A и B?
   Решение: Да, $P(A) = 1/6$, $P(B) = 1/6$, $P(A \cap B) = 1/36 = P(A)P(B)$.

5. (**Усложнённая**) В урне 10 шаров: 4 белых, 6 чёрных. Два шара извлекают без возвращения. Найдите вероятность того, что оба белые.
   Решение: $P = \dfrac{4}{10} \cdot \dfrac{3}{9} = \dfrac{12}{90} = \dfrac{2}{15}$.

8.15. Короткие доказательства (наиболее нужные для экзамена)

Доказательство: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Разложим объединение на непересекающиеся части:

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

где A и $B \setminus A$ попарно несовместны. Следовательно,

$$P(A \cup B) = P(A) + P(B \setminus A) = P(A) + \big(P(B) - P(A \cap B)\big).$$

Откуда требуется равенство. □

Доказательство формулы полной вероятности. Дано разбиение $\{H_i\}$. Так как $A = \bigcup_i (A \cap H_i)$ — объединение попарно несовместных множеств, применяем аддитивность и получаем

$$P(A) = \sum_i P(A \cap H_i) = \sum_i P(H_i) P(A \mid H_i).$$

□

8.16. Резюме — что запомнить точно

- Аксиомы Колмогорова — основа теории.
- Условная вероятность: $P(A \mid B) = P(A \cap B)/P(B)$.
- Формула полной вероятности для разбиения пространства.
- Теорема Байеса для обратных вероятностей.
- Различие между несовместностью и независимостью.

8.17. Если хочешь — углубимся дальше

Могу дополнить секцию подробными темами:

- теория случайных величин (дискретные и непрерывные), плотности и функции распределения;
- математическое ожидание, дисперсия, ковариация;
- законы больших чисел и центральная предельная теорема;

- байесовский вывод и примеры с непрерывными априорными распределениями;
- более сложные практические задачи (моделирование, имитация Монте-Карло).

Если хочешь — сейчас разверну каждую подпункту ещё глубже (больше доказательств, задач, графиков и примеров) и подготовлю версию формата «тезисы + примеры + тесты» — говори, в каком виде удобнее: учебный конспект, задания с решениями или шпаргалка.

9. Понятие случайной величины. Функция плотности распределения и её свойства. Дискретные и непрерывные законы распределения, их свойства

1. Интуиция и формальное определение случайной величины

Интуитивно **случайная величина** — это правило, которое каждому исходу случайного эксперимента сопоставляет некоторое числовое значение. Например, при броске монеты можно сопоставить "орёл" → 1, "решка" → 0; при броске кости — само выпавшее число.

Формально: пусть $(\Omega, \mathscr{F}, P)$ — вероятностное пространство (пространство элементарных исходов Ω, σ-алгебра событий $\mathscr{F}$ и вероятность P). Тогда случайная величина — это измеримая функция

$$X : \Omega \to \mathbb{R},$$

то есть для любого промежутка (или борелевского множества) $B \subseteq \mathbb{R}$ множество $\{\omega \in \Omega : \ X(\omega) \in B\}$ должно быть событием (лежать в $\mathscr{F}$).

Классификация по типу значений:

- **Дискретная** случайная величина принимает не более счётного множества значений.
- **Непрерывная** (абсолютно непрерывная) имеет плотность распределения относительно меры Лебега (нет атомов точной массы).
- **Смешанная** — содержит как дискретную, так и непрерывную составляющие.

2. Функция распределения (CDF) — основа описания закона

Для произвольной случайной величины X её **функция распределения** (Cumulative Distribution Function, CDF) определяется как

$$F_X(x) := P(X \le x), \quad x \in \mathbb{R}.$$

Функция распределения полностью задаёт закон случайной величины (включая дискретные и непрерывные части).

Ключевые свойства $F_X(x)$:

1. F_X монотонно неубывает: если $x_1 \le x_2$ то $F_X(x_1) \le F_X(x_2)$.
2. Правосторонняя непрерывность: $\lim\limits_{t \downarrow x} F_X(t) = F_X(x)$.
3. Пределы на бесконечностях: $\lim\limits_{x \to -\infty} F_X(x) = 0, \quad \lim\limits_{x \to +\infty} F_X(x) = 1.$
4. Для любых $a < b$ выполнено $P(a < X \le b) = F_X(b) - F_X(a)$.

Атомы (точечные массы). Если в точке x_0 сразу возникает положительный скачок $p_0 = F_X(x_0) - \lim_{x \to x_0^-} F_X(x)$, то $P(X = x_0) = p_0 > 0$ — это дискретная (атомная) часть распределения.

3. Дискретные законы распределения

Определение. Случайная величина X называется дискретной, если существует (счётное) набор значений $\{x_k\}$ такой, что $P(X \in \{x_k\}) = 1$. Тогда её закон задаётся **функцией вероятности** (PMF):

$$p_X(x_k) := P(X = x_k), \qquad \sum_k p_X(x_k) = 1, \quad p_X(x_k) \ge 0.$$

Ожидание и дисперсия. Если суммы сходятся абсолютно, определяются математическое ожидание и дисперсия:

$$\mathbb{E}[X] = \sum_k x_k p_X(x_k), \qquad \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \sum_k (x_k - \mathbb{E}X)^2 p_X(x_k).$$

Типичные примеры (с формулами и краткой интерпретацией):

- **Бернуллиевская (Bernoulli)**: $X \in \{0, 1\}$, $P(X = 1) = p$, $P(X = 0) = 1 - p$.
  $\mathbb{E}X = p$, $\mathrm{Var}(X) = p(1 - p)$.

- **Биномиальная (Binomial)** $X \sim Bin(n, p)$: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ для $k = 0, \dots, n$.
  $\mathbb{E}X = np$, $\mathrm{Var}(X) = np(1-p)$. Модель: n независимых испытаний Бернулли.

- **Геометрическая (Geometric)**: $P(X = k) = (1-p)^{k-1} p$ (номер первого успеха). $\mathbb{E}X = \frac{1}{p}$.

- **Пуассоновская (Poisson)** $X \sim Pois(\lambda)$: $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \geq 0$.
  Это предел биномиального при $n \to \infty$, $p \to 0$, $np \to \lambda$. $\mathbb{E}X = \lambda$, $\mathrm{Var}(X) = \lambda$.

Генерирующие функции (полезный инструмент).

- **Моментная функция (MGF):** $M_X(t) = \mathbb{E}[e^{tX}] = \sum_k e^{tx_k} p_X(x_k)$.

- **Функция порождающая (PGF) для неотрицательных целых:** $G_X(s) = \mathbb{E}[s^X] = \sum_{k \geq 0} s^k p_X(k)$.

MGF и PGF удобны для вычисления моментов и сумм независимых случайных величин.

4. Непрерывные законы распределения и функция плотности (PDF)

Определение. В случае, когда X непрерывна, её закон, как правило, задаётся **функцией плотности** $f_X(x)$ (PDF), такой что для любых множеств

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx.$$

При этом $f_X(x) \geq 0$ почти везде и

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

Связь CDF и PDF. Если F_X дифференцируема в точке x, то $f_X(x) = F'_X(x)$. В общем случае F_X может содержать дискретные скачки и непрерывную часть; тогда F_X раскладывается в сумму атомов и интегральной части.

Ожидание и дисперсия (непрерывный случай):

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx, \qquad \mathrm{Var}(X) = \int (x - \mathbb{E}X)^2 f_X(x)\,dx.$$

Типичные непрерывные распределения:

- **Равномерное** $U(a,b)$: $f(x) = \dfrac{1}{b-a}$ при $a \le x \le b$, иначе 0. $\mathbb{E}X = \dfrac{a+b}{2}$, $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$.
- **Экспоненциальное** $Exp(\lambda)$: $f(x) = \lambda e^{-\lambda x}$ при $x \ge 0$. Памятьlessness: $P(X > t+s \mid X > t) = P(X > s)$. $\mathbb{E}X = 1/\lambda$, $\mathrm{Var}(X) = 1/\lambda^2$.
- **Нормальное** $N(\mu, \sigma^2)$: $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\bigl(-\dfrac{(x-\mu)^2}{2\sigma^2}\bigr)$ для $x \in \mathbb{R}$. $\mathbb{E}X = \mu$, $\mathrm{Var}(X) = \sigma^2$. Центральная предельная теорема делает нормальное распределение фундаментальным.
- **Гамма, Бета, Коши** и др. — семейства с разными формами плотностей, полезные в разных задачах.

Пример вычислений (равномерное и экспоненциальное).

- $X \sim U(0,1)$: $\mathbb{E}[X] = \int_0^1 x\,dx = \frac{1}{2}$, $\mathrm{Var}(X) = \int_0^1 (x - \frac{1}{2})^2 dx = \frac{1}{12}$.
- $X \sim Exp(\lambda)$: $\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \lambda \cdot \frac{1}{\lambda^2} = \frac{1}{\lambda}$ (полезно интегрировать по частям).

5. Смешанные распределения

Реально встречаются законы, у которых есть и дискретная часть (атомы), и плотность. Тогда CDF распадается:

$$F_X(x) = \sum_{x_k \le x} p_k + \int_{-\infty}^{x} f_{\mathrm{cont}}(t)\,dt,$$

где $p_k = P(X = x_k)$ — массы в точках, а f_{cont} — плотность непрерывной части. Часто такие случаи возникают, например, при моделировании с выпадением особого события (атом) плюс «обычный» непрерывный шум.

6. Характеристики распределения: мода, медиана, квантили, моментные характеристики

- **Мода** — значение x (необязательно единственное), в котором плотность (или PMF) достигает максимума.

- **Медиана** m — решение $F_X(m) \geq 1/2$ и $F_X(m-) \leq 1/2$; для непрерывных распределений часто единственна.
- **Квантили:** $q_\alpha = \inf\{x : F_X(x) \geq \alpha\}$.
- **Моменты:** $\mathbb{E}[X^k]$ при существовании; центральные моменты $\mathbb{E}[(X - \mathbb{E}X)^k]$; моментная функция $M_X(t) = \mathbb{E}[e^{tX}]$ (если существует в окрестности 0).

Эти характеристики используются для описания асимметрии (скос), крутизны (эксцесс) и т.д.

7. Преобразования случайных величин

Дискретный случай. Если X дискретна с $P(X = x_k) = p_k$, а $Y = g(X)$, то

$$P(Y = y) = \sum_{k:\ g(x_k)=y} p_k.$$

Непрерывный случай (монотонная функция). Пусть X имеет плотность f_X и $Y = g(X)$, где g монотонна и дифференцируема. Тогда плотность f_Y для $y = g(x)$ даётся формулой

$$f_Y(y) = f_X\big(g^{-1}(y)\big) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

Для многомерного случая используется якобиан преобразования.

Пример (известный): Если $X \sim U(0,1)$ и $Y = -\ln X$, то $Y \sim Exp(1)$. Действительно, $P(Y \leq y) = P(-\ln X \leq y) = P(X \geq e^{-y}) = 1 - e^{-y}$ для $y \geq 0$, дифференцируя получаем плотность $f_Y(y) = e^{-y}$.

8. Совместные распределения, маргинальные и условные законы (кратко)

Хотя основной вопрос — одномерные законы, важно упомянуть:

- Для вектора случайных величин (X, Y) задаётся **совместная** PMF или PDF $p_{X,Y}(x, y)$ или $f_{X,Y}(x, y)$.
- **Маргинальная** плотность: $f_X(x) = \int f_{X,Y}(x, y)\, dy$ (аналогично для дискретного: суммирование).

- **Условная** плотность: $f_{Y|X}(y \mid x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$ при $f_X(x) > 0$.
- Независимость: X и Y независимы □ $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ (или для дискретного — $p_{X,Y}(x,y) = p_X(x) p_Y(y)$).

9. Заключение — практические советы и «чек-лист» для экзамена

- **Запомнить определения:** CDF F_X, PMF p_X для дискретных, PDF f_X для непрерывных.
- **Свойства:** F монотонен, правосторонне непрерывен, пределы 0 и 1; $p \geq 0$, $\sum p = 1$; $f \geq 0$, $\int f = 1$.
- **Переходы:** $F'(x) = f(x)$ (когда существует), $P(a < X \leq b) = F(b) - F(a)$.
- **Частые формулы:** $\mathbb{E}[X] = \sum x_k p_k$ (дискретно) или $\int x f(x)\,dx$ (непрерывно).
- **Преобразования:** дискретный — суммирование по прообразам; непрерывный — замена переменной с модулем якобиана.
- **Типовые распределения:** знать формулы PMF/PDF, ожидание и дисперсию для Bernoulli, Binomial, Poisson, Geometric, Uniform, Exponential, Normal.

10. Задачи для закрепления (с краткими подсказками)

1. Докажите, что функция распределения F_X права-непрерывна и монотонна. (*Подсказка: используйте свойства вероятности и представление* $F(x) = P(X \leq x)$.)
2. Пусть $X \sim Bin(n,p)$. Найдите MGF $M_X(t)$ и используйте её, чтобы получить $\mathbb{E}X$ и $\mathrm{Var}(X)$. (*Подсказка:* $M_X(t) = (1 - p + pe^t)^n$.)
3. Для $X \sim N(0,1)$ доказать, что плотность интегрируется в 1 (можно сослаться на табличный интеграл или заменить в полярных координатах). (*Подсказка: вычислите* $I^2 = \left(\int e^{-x^2/2} dx\right)^2$ *через двойной интеграл.*)
4. Пусть X имеет CDF F. Покажите, что $P(X = a) = F(a) - \lim_{x \to a^-} F(x)$. (*Подсказка: используйте определение вероятности точки как разности* $P(X \leq a) - P(X < a)$.)

Если хочешь, я могу:

- развернуть этот материал ещё глубже (формулы для моментов более высоких порядков, характеристические функции, теорема Лебега о разложении мер в дискретную и абсолютно непрерывную части);
- подготовить набор типовых экзаменационных задач с полными решениями;
- сгенерировать иллюстрации в TikZ для PMF (столбиковые диаграммы) и PDF (кривые) для каждого приведённого распределения.

Скажи, что делаем дальше — рисуем графики или сразу задачи с подробными решениями?

10. Наиболее употребимые теоретические законы распределения вероятностей. Примеры и свойства распределений для дискретных и непрерывных величин

В теории вероятностей **закон распределения** случайной величины описывает, какие значения она может принимать и с какой вероятностью. Все законы делятся на два больших класса: **дискретные** и **непрерывные**.

10.1. Дискретные законы распределения

Дискретная случайная величина может принимать конечное или счётное число значений. Закон распределения задаётся таблицей или функцией вероятности:

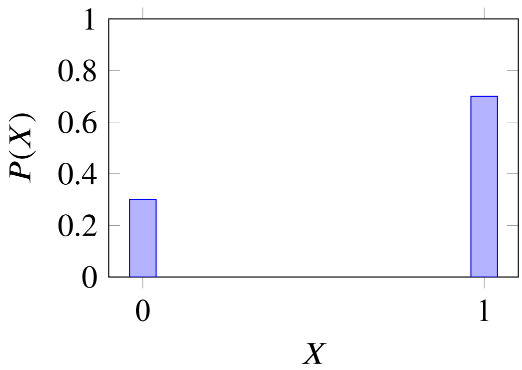
$$P(X = x_i) = p_i, \quad p_i \geq 0, \quad \sum_i p_i = 1.$$

10.1.1. Распределение Бернулли

Описывает исход одного эксперимента с двумя результатами: “успех” (1) с вероятностью p и “неудача” (0) с вероятностью $q = 1 - p$.

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

Математическое ожидание: $E[X] = p$. **Дисперсия:** $D[X] = p(1 - p)$.

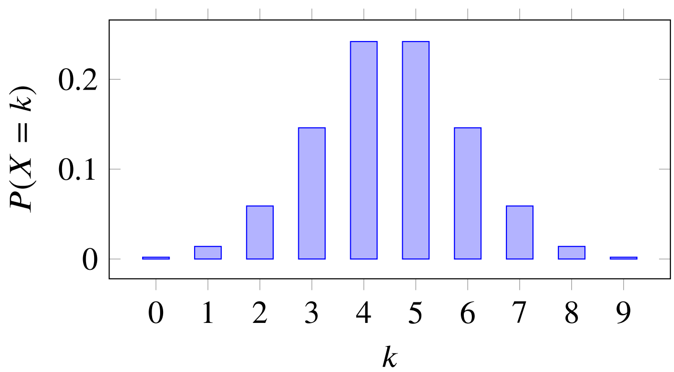


10.1.2. Биномиальное распределение

Описывает количество успехов в n независимых испытаниях Бернулли с вероятностью успеха p.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Математическое ожидание: $E[X] = np$. **Дисперсия:** $D[X] = np(1-p)$.



10.1.3. Геометрическое распределение

Вероятность того, что первый успех произойдёт на k-м испытании:

$$P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, \dots$$

Математическое ожидание: $E[X] = \frac{1}{p}$. **Дисперсия:** $D[X] = \frac{1-p}{p^2}$.

10.2. Непрерывные законы распределения

Непрерывная случайная величина может принимать любое значение на отрезке или на всей числовой прямой. Её распределение задаётся функцией плотности вероятности $f(x)$, удовлетворяющей условиям:

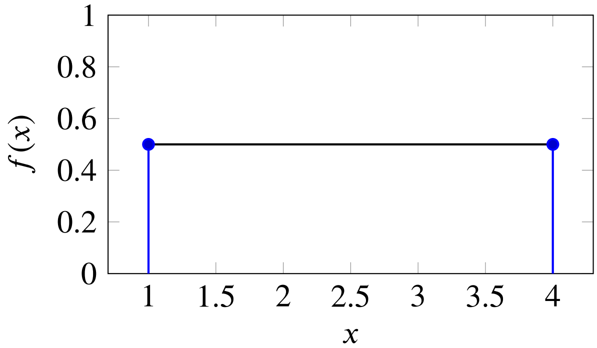
$$f(x) \geq 0, \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

10.2.1. Равномерное распределение

Если случайная величина X равновероятно принимает значения на отрезке $[a, b]$, то

$$f(x) = \frac{1}{b-a}, \qquad a \leq x \leq b.$$

Математическое ожидание: $E[X] = \frac{a+b}{2}$. **Дисперсия:** $D[X] = \frac{(b-a)^2}{12}$.

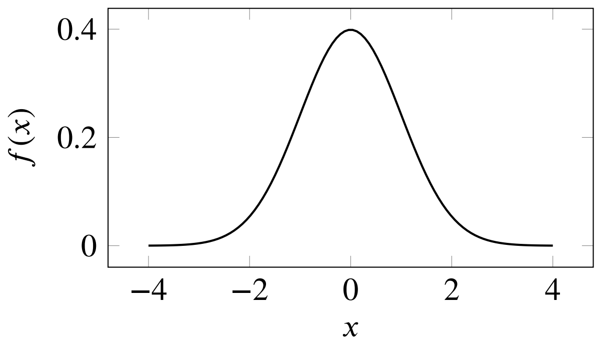


10.2.2. Нормальное распределение

Наиболее распространённое в природе и статистике. Функция плотности:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty.$$

Математическое ожидание: $E[X] = \mu$. **Дисперсия:** $D[X] = \sigma^2$.

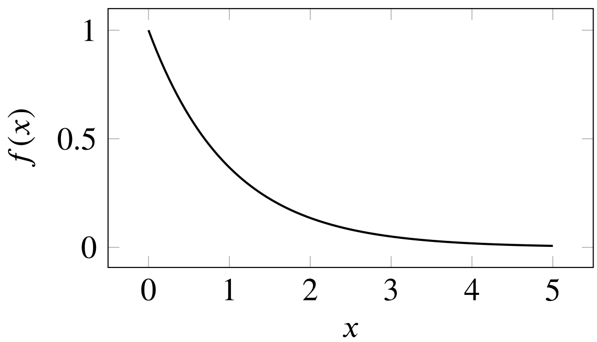


10.2.3. Экспоненциальное распределение

Часто описывает время ожидания между событиями.

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

Математическое ожидание: $E[X] = \frac{1}{\lambda}$. **Дисперсия:** $D[X] = \frac{1}{\lambda^2}$.



10.3. Выводы и сравнение

Каждое распределение имеет свои особенности и применяется в определённых задачах:

- Бернулли и биномиальное — для дискретных экспериментов с успехами и неудачами.
- Геометрическое — для моделирования числа попыток до первого успеха.
- Равномерное — когда все значения равновероятны.
- Нормальное — в большинстве природных и социальных явлений.
- Экспоненциальное — для моделирования времени ожидания.

11. Выборочные характеристики разброса и центральной тенденции дискретных и непрерывных случайных величин

11.1. Введение в выборочные характеристики

В статистике важную роль играет описание данных с помощью характеристик, которые позволяют сделать выводы о распределении случайной величины на основе выборки. Под **выборочными характеристиками** понимают числовые показатели, вычисленные по данным выборки, которые используются для оценки свойств генеральной совокупности.

Все характеристики можно условно разделить на две большие группы:

- **Характеристики центральной тенденции** — показывают, вокруг каких значений сосредоточены наблюдения.
- **Характеристики разброса** — показывают, насколько сильно наблюдения отклоняются от центра.

11.2. Центральная тенденция

Центральная тенденция описывает ”середину” данных, то есть значение, около которого сконцентрированы результаты.

11.2.1. Выборочное среднее

Выборочное среднее $\overline{x}$ — это сумма всех элементов выборки, делённая на их количество:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Где:

- n — объём выборки
- x_i — i-е наблюдение

Смысл: выборочное среднее — оценка математического ожидания генеральной совокупности.

11.2.2. Медиана

Медиана — значение, которое делит упорядоченные данные на две равные части:

- Половина значений меньше или равна медиане
- Половина — больше или равна медиане

Для нечётного n: медиана — это значение с индексом $\frac{n+1}{2}$. Для чётного n: медиана — среднее арифметическое двух средних элементов.

11.2.3. Мода

Мода — значение, которое встречается чаще всего. Если все значения встречаются одинаково часто, то мода может отсутствовать или быть не единственной.

11.3. Характеристики разброса

Разброс характеризует, насколько сильно значения данных отклоняются от среднего.

11.3.1. Размах

Размах R — разница между максимальным и минимальным значениями:

$$R = x_{\max} - x_{\min}$$

Показывает диапазон значений, но не учитывает их распределение.

11.3.2. Выборочная дисперсия

Выборочная дисперсия S^2 — среднее квадратов отклонений значений от выборочного среднего:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Используем $n-1$ в знаменателе для получения несмещённой оценки дисперсии.

11.3.3. Выборочное стандартное отклонение

Стандартное отклонение S — квадратный корень из дисперсии:

$$S = \sqrt{S^2}$$

Показывает среднее отклонение данных от среднего значения в тех же единицах, что и сами данные.

11.3.4. Коэффициент вариации

Коэффициент вариации V — относительная мера разброса:

$$V = \frac{S}{\overline{x}} \cdot 100\%$$

Позволяет сравнивать вариацию в разных выборках, даже если их средние сильно различаются.

11.4. Дискретные и непрерывные случайные величины

11.4.1. Дискретная случайная величина

Если случайная величина X принимает конечное или счётное множество значений $x_1, x_2, \dots, x_k$, то выборочные характеристики вычисляются по прямым формулам, приведённым выше.

11.4.2. Непрерывная случайная величина

Для непрерывных величин выборочные характеристики вычисляются так же, но на практике используют дискретное приближение (интервалы значений). При больших объёмах данных гистограммы и интегральные графики помогают визуализировать тенденции.

11.5. Пример вычислений

Пусть имеется выборка: $3, 5, 7, 5, 9$.

- Выборочное среднее:

$$\overline{x} = \frac{3+5+7+5+9}{5} = \frac{29}{5} = 5.8$$

- Медиана: 5 (середина упорядоченного ряда $3, 5, 5, 7, 9$)

- Мода: 5 (встречается дважды)
- Размах: $R = 9 - 3 = 6$
- Дисперсия:

$$S^2 = \frac{(3-5.8)^2 + (5-5.8)^2 + (7-5.8)^2 + (5-5.8)^2 + (9-5.8)^2}{5-1} = \frac{7.84 + 0.64 + 1.44 + 0.64}{4}$$

- Стандартное отклонение:

$$S = \sqrt{5.2} \approx 2.28$$

- Коэффициент вариации:

$$V \approx \frac{2.28}{5.8} \cdot 100\% \approx 39.3\%$$

11.6. Заключение

Выборочные характеристики центральной тенденции и разброса — это основа описательной статистики. Они позволяют понять структуру данных, выявить закономерности и сравнить разные выборки. В дискретных и непрерывных случаях подход к вычислению одинаков, но в непрерывных задачах часто используют аппроксимации и графические методы.