Математическая статистика

smvfe

Данный конспект не является официальным учебным материалом и не претендует на полноту и истинность изложенного. Использование данного материала подразумевает понимание контекста и использование критического мышления.

Содержание

1 Предмет математической статистики. Задачи

1.1 Введение и определение

Математическая статистика — раздел математики, посвящённый методам сбора, анализа и интерпретации статистических данных. В отличие от теории вероятностей, где известно распределение и изучаются свойства случайных величин, в математической статистике распределение *неизвестно* (полностью или частично), и задача состоит в его *восстановлении* по наблюдениям.

Ключевое отличие от теории вероятностей:

Теория вероятностей	Математическая статистика
Распределение известно	Распределение неизвестно
$P \to$ свойства X	Выборка $\to$ восстановление P
Дедуктивный подход	Индуктивный подход

1.2 Формальная постановка

Пусть $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ — вероятностное пространство с неизвестным параметром $\theta \in \Theta$, где Θ — *параметрическое пространство*. Наблюдаем реализации случайной величины $X : \Omega \to \mathbb{R}$ с неизвестным распределением $F_\theta(x) = \mathbb{P}_\theta(X \leq x)$.

Цель: по выборке $X_1, \ldots, X_n$ сделать выводы о неизвестном θ или F_θ.

1.3 Основные задачи математической статистики

1. **Точечное оценивание параметров**

   Задача: Найти функцию $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ — **статистику**, которая приближает неизвестный параметр θ.

   Пример: Оценка математического ожидания μ выборочным средним:

$$\hat{\mu}_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

2. **Интервальное оценивание (доверительные интервалы)**

   Задача: Построить интервал $[\hat{\theta}_L, \hat{\theta}_U]$, который содержит истинное значение θ с заданной вероятностью $1 - \alpha$ (уровень доверия):

$$\mathbb{P}_\theta\big(\hat{\theta}_L(X_1, \ldots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, \ldots, X_n)\big) \geq 1 - \alpha$$

3. **Проверка статистических гипотез**

   Задача: На основе выборки принять решение о справедливости гипотезы H_0 (нулевая) против альтернативы H_1.

   Примеры:

   - $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ (двусторонняя)
   - H_0 : выборка из нормального распределения

4. **Непараметрическое оценивание**

   Задача: Оценить функцию распределения $F(x)$ или плотность $f(x)$ без предположений о параметрической форме.

5. **Регрессионный анализ**

   Задача: Установить зависимость $Y = f(X, \theta) + \varepsilon$ и оценить параметры θ.

6. **Классификация и кластеризация**

   Задача: Разбиение объектов на группы по их статистическим характеристикам.

1.4 Требования к оценкам

Хорошая оценка $\hat{\theta}_n$ должна обладать свойствами:

- **Несмещённость:** $\mathbb{E}[\hat{\theta}_n] = \theta$ для всех $\theta \in \Theta$
- **Состоятельность:** $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ при $n \to \infty$
- **Эффективность:** минимальная дисперсия среди несмещённых оценок
- **Асимптотическая нормальность:** $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

2 Модель простейшей выборки

2.1 Определение выборки

Простейшей выборкой объёма n из распределения F называется набор $X_1, X_2, \ldots, X_n$ независимых одинаково распределённых (i.i.d.) случайных величин с функцией распределения F. Формально:

1. $X_1, \ldots, X_n$ — независимы
2. $X_i \sim F$ для всех $i = 1, \ldots, n$

Обозначение: $X_1, \ldots, X_n \stackrel{\text{i.i.d.}}{\sim} F$ или $X_1, \ldots, X_n \sim F$ (n i.i.d. копий).
Интерпретация:

- X_i — случайная величина (до проведения эксперимента)
- x_i — реализация X_i (конкретное наблюдённое значение)
- $(x_1, \ldots, x_n)$ — реализация выборки

2.2 Эмпирическая функция распределения

Пусть $X_1, \dots, X_n$ — выборка. **Эмпирическая функция распределения** (ЭФР):

$$\boxed{F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \le x} = \frac{\#\{i : X_i \le x\}}{n}}$$

где $\mathbf{1}_{X_i \le x}$ — индикатор события $\{X_i \le x\}$.

Геометрически: $F_n(x)$ — доля наблюдений, не превосходящих x.
Альтернативная запись через порядковые статистики:
Пусть $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ — *вариационный ряд* (упорядоченная выборка). Тогда:

$$F_n(x) = \begin{cases} 0, & x < X_{(1)} \\ \dfrac{k}{n}, & X_{(k)} \le x < X_{(k+1)}, \quad k = 1, \dots, n-1 \\ 1, & x \ge X_{(n)} \end{cases}$$

2.3 Свойства эмпирической функции распределения

Пусть $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$. Тогда:

(i) Несмещённость: $\mathbb{E}[F_n(x)] = F(x)$ для любого $x \in \mathbb{R}$

(ii) Дисперсия: $\mathbb{D}[F_n(x)] = \dfrac{F(x)(1 - F(x))}{n}$

(iii) Состоятельность (поточечная): $F_n(x) \xrightarrow{\mathbb{P}} F(x)$ при $n \to \infty$ для любого x

(iv) Асимптотическая нормальность:

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, F(x)(1 - F(x)))$$

(v) Теорема Гливенко—Кантелли (равномерная состоятельность):

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\text{п.н.}} 0 \quad \text{при } n \to \infty$$

(i) Несмещённость:

$$\mathbb{E}[F_n(x)] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \le x}\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\mathbf{1}_{X_i \le x}] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(X_i \le x) = \frac{1}{n} \cdot n \cdot F(x) = F(x)$$

(ii) Дисперсия:
Заметим, что $Y_i = \mathbf{1}_{X_i \le x} \sim \text{Bernoulli}(F(x))$, поэтому $\mathbb{D}[Y_i] = F(x)(1 - F(x))$.

По независимости:

$$\mathbb{D}[F_n(x)] = \mathbb{D}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{D}[Y_i] = \frac{1}{n^2}\cdot n \cdot F(x)(1-F(x)) = \frac{F(x)(1-F(x))}{n}$$

(iii) Состоятельность: Следует из (i), (ii) и неравенства Чебышёва:

$$\mathbb{P}(|F_n(x) - F(x)| > \varepsilon) \le \frac{\mathbb{D}[F_n(x)]}{\varepsilon^2} = \frac{F(x)(1-F(x))}{n\varepsilon^2} \to 0$$

(iv) Асимптотическая нормальность: Применяем ЦПТ к сумме $\sum_{i=1}^{n} \mathbf{1}_{X_i \le x}$.

(v) Теорема Гливенко—Кантелли:
Идея доказательства: Разбиваем $\mathbb{R}$ на конечное число интервалов $(-\infty, t_1], (t_1, t_2], \dots, (t_k, +\infty)$ так, чтобы F менялась на каждом не более чем на ε. На каждом интервале поточечная сходимость даёт оценку $|F_n(t_j) - F(t_j)| < \varepsilon$ п.н. при больших n. Монотонность F_n и F позволяет контролировать отклонение между точками разбиения. ■

Теорема Гливенко—Кантелли — фундаментальный результат, обосновывающий использование ЭФР как оценки истинной функции распределения.

2.4 Способы визуализации выборки

2.4.1 Гистограмма

Пусть $\mathbb{R}$ разбит на интервалы $\Delta_1, \dots, \Delta_k$ длины h (ширина бина). **Гистограмма**:

$$\hat{f}_n(x) = \frac{\nu_j}{n \cdot h}, \quad x \in \Delta_j$$

где $\nu_j = \#\{i : X_i \in \Delta_j\}$ — число наблюдений в j-м интервале.

Свойства:

- $\int_{-\infty}^{+\infty} \hat{f}_n(x)\, dx = 1$ (нормировка)
- При $n \to \infty$, $h \to 0$, $nh \to \infty$: $\hat{f}_n(x) \to f(x)$ (состоятельность)

Выбор числа интервалов:

- Правило Стёрджеса: $k = 1 + \log_2 n \approx 1 + 3.322 \lg n$
- Правило Фридмана—Диакониса: $h = 2 \cdot \mathrm{IQR}/n^{1/3}$

2.4.2 Ядерная оценка плотности

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

где K — *ядро* (симметричная плотность, например, гауссовское), $h > 0$ — *ширина окна* (bandwidth).

2.4.3 Box-plot (ящик с усами)

Отображает:

- Медиану (центральная линия)
- Первый и третий квартили Q_1, Q_3 (границы ящика)
- Усы: до $Q_1 - 1.5 \cdot \text{IQR}$ и $Q_3 + 1.5 \cdot \text{IQR}$
- Выбросы — точки за пределами усов

где $\text{IQR} = Q_3 - Q_1$ — межквартильный размах.

2.4.4 Q-Q plot (квантиль-квантильный график)

Сравнение эмпирических квантилей $X_{(k)}$ с теоретическими $F^{-1}((k - 0.5)/n)$.

Если точки ложатся на прямую $y = x$, то распределение выборки соответствует F.

3 Начальные выборочные моменты

3.1 Определения

Пусть $X \sim F$ — случайная величина. **Начальный (теоретический) момент порядка** k:

$$\alpha_k = \mathbb{E}[X^k] = \int_{-\infty}^{+\infty} x^k \, dF(x)$$

(если интеграл абсолютно сходится).

Пусть $X_1, \ldots, X_n$ — выборка. **Выборочный начальный момент порядка** k:

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

Частный случай: $A_1 = \overline{X}_n = \dfrac{1}{n} \sum_{i=1}^{n} X_i$ **— выборочное среднее.**

3.2 Свойства выборочных начальных моментов

Пусть $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$, $\mathbb{E}[|X|^k] < \infty$. Тогда:

(i) Несмещённость: $\mathbb{E}[A_k] = \alpha_k$

(ii) Дисперсия: $\mathbb{D}[A_k] = \dfrac{\alpha_{2k} - \alpha_k^2}{n} = \dfrac{\mathbb{D}[X^k]}{n}$

(iii) Состоятельность: $A_k \xrightarrow{\mathbb{P}} \alpha_k$ при $n \to \infty$ (ЗБЧ)

(iv) Асимптотическая нормальность (при $\mathbb{E}[X^{2k}] < \infty$):

$$\sqrt{n}(A_k - \alpha_k) \xrightarrow{d} \mathcal{N}(0, \alpha_{2k} - \alpha_k^2)$$

(i) Несмещённость:

$$\mathbb{E}[A_k] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i^k] = \frac{1}{n}\cdot n \cdot \alpha_k = \alpha_k$$

(ii) Дисперсия: Пусть $Y_i = X_i^k$. Тогда $\mathbb{D}[Y_i] = \mathbb{E}[Y_i^2] - (\mathbb{E}[Y_i])^2 = \alpha_{2k} - \alpha_k^2$.

$$\mathbb{D}[A_k] = \mathbb{D}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n^2}\cdot n \cdot \mathbb{D}[Y_i] = \frac{\alpha_{2k} - \alpha_k^2}{n}$$

(iii) Следует из ЗБЧ (закона больших чисел) для X_i^k.
(iv) Следует из ЦПТ для X_i^k. ■

3.3 Свойства выборочного среднего $\overline{X}_n$

Пусть $\mathbb{E}[X] = \mu$, $\mathbb{D}[X] = \sigma^2 < \infty$. Тогда:

(i) $\mathbb{E}[\overline{X}_n] = \mu$ (несмещённость)

(ii) $\mathbb{D}[\overline{X}_n] = \dfrac{\sigma^2}{n}$

(iii) $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$ (состоятельность)

(iv) $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$ (ЦПТ)

(v) Если $X_i \sim \mathcal{N}(\mu, \sigma^2)$, то $\overline{X}_n \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$ **точно** (не асимптотически)

Выборочное среднее — **эффективная** оценка μ в классе несмещённых оценок для нормального распределения (достигает границы Рао—Крамера).

4 Выборочные центральные моменты. Дельта-метод

4.1 Определения

Теоретический центральный момент порядка k:

$$\mu_k = \mathbb{E}[(X - \mathbb{E}[X])^k] = \mathbb{E}[(X - \mu)^k]$$

В частности: $\mu_1 = 0$, $\mu_2 = \sigma^2$ (дисперсия).

Выборочный центральный момент порядка k:

$$\boxed{M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^k}$$

Есть два варианта определения:
1. Смещённая выборочная дисперсия:

$$S_n^2 = M_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

2. Несмещённая (исправленная) выборочная дисперсия:

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{n}{n-1}S_n^2}$$

4.2 Свойства выборочной дисперсии

Пусть $\mathbb{E}[X] = \mu$, $\mathbb{D}[X] = \sigma^2$, $\mathbb{E}[X^4] < \infty$. Тогда:

(i) Смещение S_n^2:

$$\mathbb{E}[S_n^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

(ii) Несмещённость S^2:

$$\mathbb{E}[S^2] = \sigma^2$$

(iii) Состоятельность: $S_n^2 \xrightarrow{\mathbb{P}} \sigma^2$ и $S^2 \xrightarrow{\mathbb{P}} \sigma^2$

(iv) Асимптотическая нормальность:

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, \mu_4 - \sigma^4)$$

где $\mu_4 = \mathbb{E}[(X-\mu)^4]$ — четвёртый центральный момент.

Разложим сумму квадратов отклонений:

$$\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \sum_{i=1}^{n}[(X_i - \mu) - (\overline{X}_n - \mu)]^2$$

$$= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\overline{X}_n - \mu)\sum_{i=1}^{n}(X_i - \mu) + n(\overline{X}_n - \mu)^2$$

Заметим, что $\sum_{i=1}^{n}(X_i - \mu) = n(\overline{X}_n - \mu)$, поэтому:

$$\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X}_n - \mu)^2$$

Берём математическое ожидание:

$$\mathbb{E}\left[\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right] = \sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)^2] - n\,\mathbb{E}[(\overline{X}_n - \mu)^2]$$

$$= n\sigma^2 - n \cdot \frac{\sigma^2}{n} = n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

Следовательно:

$$\mathbb{E}[S^2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right] = \frac{(n-1)\sigma^2}{n-1} = \sigma^2$$

■

Делим на $(n-1)$, а не на n, потому что при оценке центрального момента теряется одна «степень свободы» — мы используем оценку $\overline{X}_n$ вместо истинного μ.

4.3 Формула для вычисления

Удобная вычислительная формула:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 = A_2 - A_1^2$$

4.4 Дельта-метод

Пусть $\sqrt{n}(T_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ и функция g дифференцируема в точке θ с $g'(\theta) \neq 0$. Тогда:

$$\boxed{\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, [g'(\theta)]^2\sigma^2)}$$

По теореме Тейлора:

$$g(T_n) \approx g(\theta) + g'(\theta)(T_n - \theta)$$

Следовательно:

$$\sqrt{n}(g(T_n) - g(\theta)) \approx g'(\theta) \cdot \sqrt{n}(T_n - \theta) \xrightarrow{d} g'(\theta) \cdot \mathcal{N}(0, \sigma^2) = \mathcal{N}(0, [g'(\theta)]^2\sigma^2)$$

Строгое обоснование: теорема Слуцкого и свойства сходимости по распределению. ■

Найдём асимптотическое распределение выборочного стандартного отклонения $S_n = \sqrt{S_n^2}$.
Имеем: $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, \mu_4 - \sigma^4)$.
Применяем $g(x) = \sqrt{x}$, $g'(x) = \dfrac{1}{2\sqrt{x}}$, $g'(\sigma^2) = \dfrac{1}{2\sigma}$.
По дельта-методу:

$$\sqrt{n}(S_n - \sigma) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mu_4 - \sigma^4}{4\sigma^2}\right)$$

Пусть $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$ в $\mathbb{R}^k$ и $g : \mathbb{R}^k \to \mathbb{R}^m$ дифференцируема в $\boldsymbol{\theta}$. Тогда:

$$\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \nabla g(\boldsymbol{\theta})^\top \Sigma \, \nabla g(\boldsymbol{\theta}))$$

где ∇g — матрица Якоби.

5 Выборочные квантили и порядковые статистики

5.1 Порядковые статистики

Пусть $X_1, \dots, X_n$ — выборка. **Порядковыми статистиками** называются элементы вариационного ряда:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

где $X_{(k)}$ — k-я порядковая статистика.
В частности:

- $X_{(1)} = \min\{X_1, \dots, X_n\}$ — минимум
- $X_{(n)} = \max\{X_1, \dots, X_n\}$ — максимум
- $R_n = X_{(n)} - X_{(1)}$ — размах выборки

5.2 Распределение k-й порядковой статистики

Пусть $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$ с плотностью f. Тогда плотность $X_{(k)}$:

$$\boxed{f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)}$$

или, используя бета-функцию:

$$f_{(k)}(x) = \frac{1}{B(k, n-k+1)}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)$$

Событие $\{x < X_{(k)} \leq x + dx\}$ означает:

- Ровно $(k-1)$ наблюдений $\leq x$
- Ровно одно наблюдение в $(x, x+dx]$
- Ровно $(n-k)$ наблюдений $> x + dx$

Число способов выбрать, какие именно наблюдения попадают в каждую группу:

$$\binom{n}{k-1, 1, n-k} = \frac{n!}{(k-1)! \cdot 1! \cdot (n-k)!}$$

Вероятности:

$$\mathbb{P}(X_i \leq x) = F(x)$$
$$\mathbb{P}(X_i \in (x, x+dx]) \approx f(x)\,dx$$
$$\mathbb{P}(X_i > x + dx) \approx 1 - F(x)$$

Итого:
$$\mathbb{P}(x < X_{(k)} \le x + dx) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1} f(x)\, dx\, [1 - F(x)]^{n-k}$$

■

Частные случаи:
Минимум ($k = 1$):
$$f_{(1)}(x) = n[1 - F(x)]^{n-1} f(x), \quad F_{(1)}(x) = 1 - [1 - F(x)]^n$$

Максимум ($k = n$):
$$f_{(n)}(x) = n[F(x)]^{n-1} f(x), \quad F_{(n)}(x) = [F(x)]^n$$

5.3 Совместное распределение порядковых статистик

Совместная плотность вектора $(X_{(1)}, \ldots, X_{(n)})$:
$$\boxed{f_{(1),\ldots,(n)}(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i), \quad x_1 < x_2 < \ldots < x_n}$$
и 0 иначе.

Рассмотрим преобразование $(X_1, \ldots, X_n) \mapsto (X_{(1)}, \ldots, X_{(n)})$.
Совместная плотность независимых $X_1, \ldots, X_n$: $f(x_1) \cdots f(x_n)$.
Существует ровно $n!$ перестановок $(x_1, \ldots, x_n)$, дающих одну и ту же упорядоченную последовательность. Поэтому:
$$f_{(1),\ldots,(n)}(x_1, \ldots, x_n) = n! \cdot f(x_1) \cdots f(x_n)$$
на области $x_1 < x_2 < \ldots < x_n$. ■

Для $i < j$ совместная плотность $(X_{(i)}, X_{(j)})$:
$$f_{(i),(j)}(x, y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times [F(x)]^{i-1}[F(y) - F(x)]^{j-i-1}[1 - F(y)]^{n-j} f(x) f(y)$$
для $x < y$.

5.4 Выборочные квантили

Квантиль уровня p $(0 < p < 1)$ распределения F:

$$\xi_p = F^{-1}(p) = \inf\{x : F(x) \geq p\}$$

Выборочный квантиль уровня p:

$$\boxed{\hat{\xi}_p = X_{(\lceil np \rceil)}}$$

или с интерполяцией:

$$\hat{\xi}_p = X_{(\lfloor np \rfloor + 1)} \quad \text{или} \quad \hat{\xi}_p = (1-\gamma)X_{(j)} + \gamma X_{(j+1)}$$

где $j = \lfloor np \rfloor$, $\gamma = np - j$.

Важные квантили:

- **Медиана:** $\hat{\xi}_{0.5}$ (или $\text{Med} = \dfrac{X_{(k)} + X_{(k+1)}}{2}$ при чётном $n = 2k$)
- **Первый квартиль:** $Q_1 = \hat{\xi}_{0.25}$
- **Третий квартиль:** $Q_3 = \hat{\xi}_{0.75}$

5.5 Выборочная медиана

$$\text{Med}_n = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & n \text{ — нечётное} \\ \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}, & n \text{ — чётное} \end{cases}$$

5.6 Асимптотическое распределение выборочных квантилей

Пусть F имеет плотность f, непрерывную и положительную в точке $\xi_p = F^{-1}(p)$. Тогда:

$$\boxed{\sqrt{n}(\hat{\xi}_p - \xi_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{[f(\xi_p)]^2}\right)}$$

Используем представление:

$$F_n(\hat{\xi}_p) \approx p, \quad F(\xi_p) = p$$

По теореме об асимптотической нормальности ЭФР:

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, F(x)(1 - F(x)))$$

Применяя дельта-метод к функции F^{-1} и учитывая $(F^{-1})'(p) = 1/f(\xi_p)$:

$$\sqrt{n}(\hat{\xi}_p - \xi_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f(\xi_p)^2}\right)$$

■

Для медианы ($p = 1/2$):

$$\sqrt{n}(\mathrm{Med}_n - \xi_{1/2}) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4f(\xi_{1/2})^2}\right)$$

Для нормального распределения $\mathcal{N}(\mu, \sigma^2)$:

- $\mathbb{D}[\overline{X}_n] = \dfrac{\sigma^2}{n}$
- $\mathbb{D}[\mathrm{Med}_n] \approx \dfrac{\pi\sigma^2}{2n} \approx 1.57 \cdot \dfrac{\sigma^2}{n}$

То есть медиана менее эффективна для нормального распределения.
Однако медиана **робастна** — устойчива к выбросам, тогда как среднее сильно зависит от экстремальных значений.

5.7 Совместная асимптотика порядковых статистик

Пусть $0 < p_1 < p_2 < 1$, $\xi_{p_i} = F^{-1}(p_i)$, $f(\xi_{p_i}) > 0$. Тогда:

$$\sqrt{n}\begin{pmatrix} \hat{\xi}_{p_1} - \xi_{p_1} \\ \hat{\xi}_{p_2} - \xi_{p_2} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Sigma\right)$$

где

$$\Sigma = \begin{pmatrix} \dfrac{p_1(1-p_1)}{f(\xi_{p_1})^2} & \dfrac{p_1(1-p_2)}{f(\xi_{p_1})f(\xi_{p_2})} \\ \dfrac{p_1(1-p_2)}{f(\xi_{p_1})f(\xi_{p_2})} & \dfrac{p_2(1-p_2)}{f(\xi_{p_2})^2} \end{pmatrix}$$

Сводная таблица основных результатов

Статистика	Формула	$\mathbb{E}[\cdot]$	Асимптотика
$F_n(x)$	$\frac{1}{n}\sum \mathbf{1}_{X_i \le x}$	$F(x)$	$\mathcal{N}\left(F(x), \frac{F(x)(1-F(x))}{n}\right)$
$\overline{X}_n$	$\frac{1}{n}\sum X_i$	μ	$\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
S_n^2	$\frac{1}{n}\sum (X_i - \overline{X})^2$	$\frac{n-1}{n}\sigma^2$	$\mathcal{N}\left(\sigma^2, \frac{\mu_4 - \sigma^4}{n}\right)$
S^2	$\frac{1}{n-1}\sum (X_i - \overline{X})^2$	σ^2	—
$\hat{\xi}_p$	$X_{(\lceil np \rceil)}$	$\approx \xi_p$	$\mathcal{N}\left(\xi_p, \frac{p(1-p)}{nf(\xi_p)^2}\right)$

6 Асимптотики среднего и крайних членов вариационного ряда

6.1 Асимптотика среднего члена вариационного ряда

Средний член вариационного ряда связан с выборочной медианой. Рассмотрим более общий результат для центральных порядковых статистик.

Пусть $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} F$ с плотностью f, непрерывной и положительной в точке $\xi_p = F^{-1}(p)$. Пусть $k = k(n)$ такое, что $k/n \to p$ при $n \to \infty$. Тогда:

$$\sqrt{n}(X_{(k)} - \xi_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f(\xi_p)^2}\right)$$

Шаг 1. Рассмотрим величину $U_i = F(X_i)$. Тогда $U_i \sim \text{Uniform}(0,1)$ (интегральное преобразование).

Шаг 2. Порядковые статистики $U_{(k)} = F(X_{(k)})$. Известно, что:

$$U_{(k)} \sim \text{Beta}(k, n-k+1)$$

с $\mathbb{E}[U_{(k)}] = \frac{k}{n+1} \approx p$ и $\mathbb{D}[U_{(k)}] = \frac{k(n-k+1)}{(n+1)^2(n+2)} \approx \frac{p(1-p)}{n}$.

Шаг 3. По ЦПТ для бета-распределения:

$$\sqrt{n}(U_{(k)} - p) \xrightarrow{d} \mathcal{N}(0, p(1-p))$$

Шаг 4. Применяем дельта-метод к $X_{(k)} = F^{-1}(U_{(k)})$:

$$\sqrt{n}(X_{(k)} - \xi_p) = \sqrt{n}(F^{-1}(U_{(k)}) - F^{-1}(p)) \approx (F^{-1})'(p) \cdot \sqrt{n}(U_{(k)} - p)$$

Так как $(F^{-1})'(p) = 1/f(\xi_p)$, получаем:

$$\sqrt{n}(X_{(k)} - \xi_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f(\xi_p)^2}\right)$$

■

6.2 Асимптотика крайних членов вариационного ряда

Для экстремальных порядковых статистик $X_{(1)}$ и $X_{(n)}$ асимптотика существенно отличается и зависит от поведения хвостов распределения.

Пусть $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} F$. Если существуют нормирующие последовательности $a_n > 0$, $b_n \in \mathbb{R}$ такие, что:

$$\frac{X_{(n)} - b_n}{a_n} \xrightarrow{d} G$$

для некоторой невырожденной функции распределения G, то G принадлежит одному из трёх типов:

Тип I (Гумбель):

$$G(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}$$

Тип II (Фреше):

$$G(x) = \begin{cases} 0, & x \le 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases}, \quad \alpha > 0$$

Тип III (Вейбулл):

$$G(x) = \begin{cases} \exp(-(-x)^{\alpha}), & x \le 0 \\ 1, & x > 0 \end{cases}, \quad \alpha > 0$$

Пусть $X_i \sim \text{Exp}(\lambda)$, $F(x) = 1 - e^{-\lambda x}$ для $x \ge 0$.
Тогда:

$$F_{(n)}(x) = [F(x)]^n = (1 - e^{-\lambda x})^n$$

Выберем $b_n = \dfrac{\ln n}{\lambda}$, $a_n = \dfrac{1}{\lambda}$. Тогда:

$$\mathbb{P}\left(\frac{X_{(n)} - b_n}{a_n} \le x\right) = \left(1 - \frac{e^{-x}}{n}\right)^n \to e^{-e^{-x}}$$

Это распределение Гумбеля (тип I).

Пусть $X_i \sim \text{Uniform}(0, 1)$. Тогда $F(x) = x$ для $x \in [0, 1]$.

$$\mathbb{P}(n(1 - X_{(n)}) \le x) = \mathbb{P}\left(X_{(n)} \ge 1 - \frac{x}{n}\right) = 1 - \left(1 - \frac{x}{n}\right)^n \to 1 - e^{-x}$$

Следовательно, $n(1 - X_{(n)}) \xrightarrow{d} \text{Exp}(1)$, что соответствует типу III (Вейбулл).

Для минимума $X_{(1)}$ аналогичные результаты получаются заменой $X \to -X$. Если:

$$\frac{X_{(1)} - b'_n}{a'_n} \xrightarrow{d} G'$$

то G' также принадлежит одному из трёх типов.

Ключевая идея: Предельное распределение G должно быть *max-устойчивым*, т.е. удовлетворять функциональному уравнению:

$$G^n(a_n x + b_n) = G(x)$$

для некоторых $a_n > 0$, b_n.
Это означает, что максимум n независимых копий из G после линейной нормировки снова имеет распределение G.
Решение этого функционального уравнения приводит ровно к трём типам распределений, перечисленным в теореме. ■

6.3 Связь типов с хвостами распределения

- **Тип I (Гумбель):** Экспоненциально убывающие хвосты (нормальное, экспоненциальное, логнормальное)
- **Тип II (Фреше):** Тяжёлые хвосты, степенное убывание (Парето, Коши, t-распределение)
- **Тип III (Вейбулл):** Ограниченный носитель (равномерное, бета)

7 Точечное оценивание параметров

7.1 Постановка задачи

Параметрическая статистическая модель — семейство распределений:

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

где $\Theta \subseteq \mathbb{R}^d$ — *параметрическое пространство*.
Наблюдаем выборку $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} P_\theta$ с неизвестным θ.

Точечная оценка (статистика) параметра θ — любая измеримая функция выборки:

$$\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n) : \mathbb{R}^n \to \Theta$$

не зависящая от θ.

Задача: Построить «хорошую» оценку $\hat{\theta}_n$ по выборке.

7.2 Свойства оценок

7.2.1 Несмещённость

[redacted]

Оценка $\hat{\theta}_n$ называется **несмещённой**, если:

$$\mathbb{E}_\theta[\hat{\theta}_n] = \theta \quad \text{для всех } \theta \in \Theta$$

Смещение оценки:

$$\text{bias}(\hat{\theta}_n) = \mathbb{E}_\theta[\hat{\theta}_n] - \theta$$

[redacted]

Оценка $\hat{\theta}_n$ **асимптотически несмещена**, если:

$$\lim_{n\to\infty} \mathbb{E}_\theta[\hat{\theta}_n] = \theta$$

7.2.2 Состоятельность

[redacted]

Оценка $\hat{\theta}_n$ **состоятельна**, если:

$$\hat{\theta}_n \xrightarrow{\mathbb{P}_\theta} \theta \quad \text{при } n \to \infty$$

для всех $\theta \in \Theta$.
Эквивалентно: $\forall \varepsilon > 0$: $\mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \varepsilon) \to 0$.

[redacted]

Оценка $\hat{\theta}_n$ **сильно состоятельна**, если:

$$\hat{\theta}_n \xrightarrow{\text{п.н.}} \theta$$

[redacted]

Если $\hat{\theta}_n$ несмещена и $\mathbb{D}_\theta[\hat{\theta}_n] \to 0$ при $n \to \infty$, то $\hat{\theta}_n$ состоятельна.

[redacted]

По неравенству Чебышёва:

$$\mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \varepsilon) = \mathbb{P}_\theta(|\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]| > \varepsilon) \leq \frac{\mathbb{D}[\hat{\theta}_n]}{\varepsilon^2} \to 0$$

■

7.2.3 Эффективность

Несмещённая оценка $\hat{\theta}_n^*$ называется **эффективной** в классе несмещённых оценок, если для любой другой несмещённой оценки $\hat{\theta}_n$:

$$\mathbb{D}_\theta[\hat{\theta}_n^*] \leq \mathbb{D}_\theta[\hat{\theta}_n] \quad \text{для всех } \theta \in \Theta$$

Состоятельная оценка $\hat{\theta}_n$ **асимптотически эффективна**, если:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, I(\theta)^{-1})$$

где $I(\theta)$ — информация Фишера (см. далее).

7.2.4 Среднеквадратичная ошибка

Среднеквадратичная ошибка:

$$\mathrm{MSE}_\theta(\hat{\theta}_n) = \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2]$$

Разложение:

$$\boxed{\mathrm{MSE} = \mathbb{D}[\hat{\theta}_n] + (\mathrm{bias})^2}$$

$$\begin{aligned}
\mathrm{MSE} &= \mathbb{E}[(\hat{\theta}_n - \theta)^2] = \mathbb{E}[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n] + \mathbb{E}[\hat{\theta}_n] - \theta)^2] \\
&= \mathbb{E}[(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])^2] + 2(\mathbb{E}[\hat{\theta}_n] - \theta)\,\mathbb{E}[\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n]] + (\mathbb{E}[\hat{\theta}_n] - \theta)^2 \\
&= \mathbb{D}[\hat{\theta}_n] + 0 + (\mathrm{bias})^2
\end{aligned}$$

■

Иногда выгоднее использовать смещённую оценку с меньшей дисперсией, если общий MSE меньше. Пример: S_n^2 vs S^2.

7.2.5 Асимптотическая нормальность

Оценка $\hat{\theta}_n$ **асимптотически нормальна**, если существует $\sigma^2(\theta) > 0$:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$$

$\sigma^2(\theta)$ называется **асимптотической дисперсией**.

8 Метод моментов

8.1 Идея метода

Метод моментов — один из старейших и простейших способов построения оценок.

Идея: Приравнять теоретические моменты к выборочным и решить получившуюся систему уравнений относительно параметров.

8.2 Формальное описание

Пусть $\theta = (\theta_1, \ldots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$ — вектор параметров.

Шаг 1. Выразить первые d моментов через параметры:

$$\alpha_k(\theta) = \mathbb{E}_\theta[X^k], \quad k = 1, \ldots, d$$

Шаг 2. Вычислить выборочные моменты:

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

Шаг 3. Решить систему уравнений:

$$\begin{cases} \alpha_1(\theta) = A_1 \\ \alpha_2(\theta) = A_2 \\ \vdots \\ \alpha_d(\theta) = A_d \end{cases}$$

Решение $\hat{\theta}_n^{MM}$ — **оценка метода моментов**.

Можно использовать центральные моменты вместо начальных, или их комбинацию.

8.3 Свойства оценок метода моментов

При регулярных условиях оценки метода моментов:

(i) Состоятельны: $\hat{\theta}_n^{MM} \xrightarrow{\mathbb{P}} \theta$

(ii) Асимптотически нормальны:

$$\sqrt{n}(\hat{\theta}_n^{MM} - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

где Σ определяется дельта-методом.

Состоятельность следует из ЗБЧ для $A_k \xrightarrow{\mathbb{P}} \alpha_k$ и непрерывности обратного отображения $(\alpha_1, \ldots, \alpha_d) \mapsto \theta$.
Асимптотическая нормальность — из ЦПТ для $(A_1, \ldots, A_d)$ и многомерного дельта-метода. ■

8.4 Примеры

Параметры: $\theta = (\mu, \sigma^2)$.
Моменты:

$$\alpha_1 = \mu, \quad \alpha_2 = \mu^2 + \sigma^2$$

Система уравнений:

$$\begin{cases} \mu = A_1 = \overline{X}_n \\ \mu^2 + \sigma^2 = A_2 \end{cases}$$

Решение:

$$\boxed{\hat{\mu}^{MM} = \overline{X}_n, \quad \hat{\sigma}^{2,MM} = A_2 - \overline{X}_n^2 = S_n^2}$$

Плотность: $f(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, $x > 0$.
Моменты:

$$\mathbb{E}[X] = \frac{\alpha}{\beta}, \quad \mathbb{E}[X^2] = \frac{\alpha(\alpha+1)}{\beta^2}$$

Дисперсия: $\mathbb{D}[X] = \dfrac{\alpha}{\beta^2}$.
Из системы $\mathbb{E}[X] = A_1$, $\mathbb{D}[X] = S_n^2$:

$$\hat{\alpha}^{MM} = \frac{\overline{X}_n^2}{S_n^2}, \quad \hat{\beta}^{MM} = \frac{\overline{X}_n}{S_n^2}$$

$\mathbb{E}[X] = \theta/2$, поэтому:

$$\hat{\theta}^{MM} = 2\overline{X}_n$$

Заметим: это *не* оптимальная оценка (ср. с ОМП $\hat{\theta}^{MLE} = X_{(n)}$).

8.5 Достоинства и недостатки

Достоинства:

- Простота вычисления
- Не требует знания точной формы распределения
- Всегда состоятельны

Недостатки:

- Могут быть неэффективны (большая дисперсия)
- Оценки могут выходить за пределы Θ
- Не единственны (зависят от выбора моментов)

9 Метод максимального правдоподобия

9.1 Функция правдоподобия

Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью (или функцией вероятности) $f(x;\theta)$.

Функция правдоподобия:

$$\boxed{L(\theta) = L(\theta; X_1, \ldots, X_n) = \prod_{i=1}^{n} f(X_i;\theta)}$$

Логарифмическая функция правдоподобия:

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(X_i;\theta)$$

Интерпретация: $L(\theta)$ — «правдоподобие» параметра θ при данных наблюдениях. Чем больше $L(\theta)$, тем лучше θ объясняет данные.

$L(\theta)$ — функция от θ при фиксированных $X_1, \ldots, X_n$. Это **не** плотность по θ!

9.2 Оценка максимального правдоподобия

Оценка максимального правдоподобия (ОМП, MLE):

$$\boxed{\hat{\theta}_n^{MLE} = \arg\max_{\theta\in\Theta} L(\theta) = \arg\max_{\theta\in\Theta} \ell(\theta)}$$

Алгоритм нахождения ОМП:

Шаг 1. Записать $\ell(\theta) = \sum_{i=1}^{n} \ln f(X_i;\theta)$.

Шаг 2. Найти критические точки из **уравнения правдоподобия**:

$$\frac{\partial \ell}{\partial \theta} = 0 \quad (\text{или } \nabla_\theta \ell = 0 \text{ для вектора } \theta)$$

Шаг 3. Проверить, что найденная точка — максимум (вторая производная < 0).
Шаг 4. Проверить граничные точки Θ, если множество ограничено.

Score-функция (функция счёта):

$$S(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln f(X_i;\theta)$$

9.3 Инвариантность ОМП

Если $\hat{\theta}^{MLE}$ — ОМП для θ, то для любой функции g:

$$\widehat{g(\theta)}^{MLE} = g(\hat{\theta}^{MLE})$$

Если $\eta = g(\theta)$ и g — взаимно однозначная, то:

$$L(\eta) = L(g^{-1}(\eta))$$

максимизируется при $\eta = g(\hat{\theta}^{MLE})$. ■

9.4 Примеры

$f(x;p) = p^x(1-p)^{1-x}$, $x \in \{0,1\}$.

$$\ell(p) = \sum_{i=1}^{n}[X_i \ln p + (1 - X_i)\ln(1-p)] = n\overline{X}_n \ln p + n(1-\overline{X}_n)\ln(1-p)$$

$$\frac{d\ell}{dp} = \frac{n\overline{X}_n}{p} - \frac{n(1-\overline{X}_n)}{1-p} = 0 \implies \hat{p}^{MLE} = \overline{X}_n$$

$f(k;\lambda) = \dfrac{\lambda^k e^{-\lambda}}{k!}$.

$$\ell(\lambda) = \sum_{i=1}^{n}[X_i \ln\lambda - \lambda - \ln(X_i!)] = n\overline{X}_n \ln\lambda - n\lambda - \sum \ln(X_i!)$$

$$\frac{d\ell}{d\lambda} = \frac{n\overline{X}_n}{\lambda} - n = 0 \implies \hat{\lambda}^{MLE} = \overline{X}_n$$

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

$$\ell(\lambda) = n\ln\lambda - \lambda\sum_{i=1}^{n} X_i = n\ln\lambda - n\lambda\overline{X}_n$$

$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - n\overline{X}_n = 0 \implies \hat{\lambda}^{MLE} = \frac{1}{\overline{X}_n}$$

$f(x;\theta) = \dfrac{1}{\theta}\cdot \mathbf{1}_{0\le x\le\theta}$.

$$L(\theta) = \frac{1}{\theta^n}\cdot \mathbf{1}_{\theta \ge X_{(n)}}$$

$L(\theta)$ убывает по θ при $\theta \ge X_{(n)}$, поэтому:

$$\hat{\theta}^{MLE} = X_{(n)} = \max\{X_1,\dots,X_n\}$$

Внимание: ОМП смещена вниз! $\mathbb{E}[X_{(n)}] = \dfrac{n}{n+1}\theta < \theta$.

10 ОМП для нормальной и полиномиальной моделей

10.1 Нормальная модель

Пусть $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Тогда:

$$\boxed{\hat{\mu}^{MLE} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

$$\boxed{\hat{\sigma}^{2,MLE} = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}$$

Плотность:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Логарифм правдоподобия:

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

По μ:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu) = 0 \implies \hat{\mu} = \overline{X}_n$$

По σ^2:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(X_i - \mu)^2 = 0$$

Подставляя $\hat{\mu} = \overline{X}_n$:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = S_n^2$$

■

$\hat{\sigma}^{2,MLE} = S_n^2$ — смещённая оценка. Несмещённая оценка: $S^2 = \dfrac{n}{n-1} S_n^2$.

10.1.1 Распределения оценок в нормальной модели

Для $X_i \sim \mathcal{N}(\mu, \sigma^2)$:

(i) $\overline{X}_n \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

(ii) $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{nS_n^2}{\sigma^2} \sim \chi^2_{n-1}$

(iii) $\overline{X}_n$ и S^2 независимы (теорема Фишера)

(iv) $\dfrac{\overline{X}_n - \mu}{S/\sqrt{n}} \sim t_{n-1}$ (распределение Стьюдента)

10.2 Полиномиальная (мультиномиальная) модель

Пусть имеется k категорий с вероятностями $p_1, \ldots, p_k$, $\sum_{j=1}^{k} p_j = 1$.

При n независимых испытаниях вектор частот $(N_1, \ldots, N_k)$ имеет **полиномиальное распределение**:

$$\mathbb{P}(N_1 = n_1, \ldots, N_k = n_k) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}$$

где $\sum_{j=1}^{k} n_j = n$.

$$\boxed{\hat{p}_j^{MLE} = \frac{N_j}{n} = \frac{\text{число наблюдений категории } j}{n}}$$

Логарифм правдоподобия (без константы):

$$\ell(p_1, \ldots, p_k) = \sum_{j=1}^{k} n_j \ln p_j$$

при ограничении $\sum_{j=1}^{k} p_j = 1$.

Используем метод множителей Лагранжа:

$$\mathcal{L} = \sum_{j=1}^{k} n_j \ln p_j + \lambda \left(1 - \sum_{j=1}^{k} p_j\right)$$

$$\frac{\partial \mathcal{L}}{\partial p_j} = \frac{n_j}{p_j} - \lambda = 0 \implies p_j = \frac{n_j}{\lambda}$$

Из ограничения: $\sum p_j = 1 \implies \lambda = n$.

Следовательно: $\hat{p}_j = \dfrac{n_j}{n}$. ■

10.3 Свойства ОМП в полиномиальной модели

$$\sqrt{n}(\hat{p}_j - p_j) \xrightarrow{d} \mathcal{N}(0, p_j(1-p_j))$$

Совместно:

$$\sqrt{n}(\hat{\mathbf{p}} - \mathbf{p}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$$

где $\Sigma_{jj} = p_j(1-p_j)$, $\Sigma_{jl} = -p_j p_l$ при $j \neq l$.

$$\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j} = n \sum_{j=1}^{k} \frac{(\hat{p}_j - p_j)^2}{p_j} \xrightarrow{d} \chi^2_{k-1}$$

при $n \to \infty$ (если все $p_j > 0$).

11 Информация Фишера

11.1 Определение и мотивация

Информация Фишера — фундаментальное понятие, измеряющее количество информации о параметре θ, содержащееся в наблюдениях.

Статистическая модель $\{f(x;\theta) : \theta \in \Theta\}$ называется **регулярной**, если:

(R1) $\Theta \subseteq \mathbb{R}$ — открытое множество

(R2) Носитель $\{x : f(x;\theta) > 0\}$ не зависит от θ

(R3) $f(x;\theta)$ дважды дифференцируема по θ

(R4) Можно менять порядок дифференцирования и интегрирования:

$$\frac{\partial}{\partial\theta} \int f(x;\theta)\,dx = \int \frac{\partial f(x;\theta)}{\partial\theta}\,dx$$

(R5) $\mathbb{E}_\theta\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right] < \infty$

Score-функция (функция счёта) для одного наблюдения:

$$s(X;\theta) = \frac{\partial}{\partial\theta} \ln f(X;\theta) = \frac{\partial_\theta f(X;\theta)}{f(X;\theta)}$$

При условиях регулярности:

$$\mathbb{E}_\theta[s(X;\theta)] = 0$$

$$\mathbb{E}_\theta[s(X;\theta)] = \int \frac{\partial_\theta f(x;\theta)}{f(x;\theta)} f(x;\theta)\,dx = \int \partial_\theta f(x;\theta)\,dx = \frac{\partial}{\partial\theta}\int f(x;\theta)\,dx = \frac{\partial}{\partial\theta}1 = 0$$

■

Информация Фишера для одного наблюдения:

$$\boxed{I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right] = \mathbb{E}_\theta[s(X;\theta)^2] = \mathbb{D}_\theta[s(X;\theta)]}$$

При условиях регулярности:

$$\boxed{I(\theta) = -\,\mathbb{E}_\theta\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right]}$$

Дифференцируем тождество $\mathbb{E}_\theta[s(X;\theta)] = 0$ по θ:

$$0 = \frac{\partial}{\partial\theta}\int s(x;\theta) f(x;\theta)\,dx = \int \left[\frac{\partial s}{\partial\theta} f + s\cdot\frac{\partial f}{\partial\theta}\right] dx$$

Заметим, что $\dfrac{\partial f}{\partial\theta} = s\cdot f$, поэтому:

$$0 = \int \frac{\partial s}{\partial\theta} f\,dx + \int s^2 f\,dx = \mathbb{E}\left[\frac{\partial s}{\partial\theta}\right] + \mathbb{E}[s^2]$$

Следовательно:

$$I(\theta) = \mathbb{E}[s^2] = -\,\mathbb{E}\left[\frac{\partial s}{\partial\theta}\right] = -\,\mathbb{E}\left[\frac{\partial^2 \ln f}{\partial\theta^2}\right]$$

■

11.2 Информация Фишера для выборки

Для i.i.d. выборки $X_1, \ldots, X_n$:

$$I_n(\theta) = n \cdot I(\theta)$$

$$I_n(\theta) = \mathbb{E}\left[\left(\sum_{i=1}^{n} s(X_i;\theta)\right)^2\right] = \sum_{i=1}^{n} \mathbb{E}[s(X_i;\theta)^2] + \sum_{i\neq j} \mathbb{E}[s(X_i;\theta)]\,\mathbb{E}[s(X_j;\theta)]$$

Так как $\mathbb{E}[s(X_i;\theta)] = 0$, получаем $I_n(\theta) = n \cdot I(\theta)$. ∎

11.3 Примеры вычисления информации Фишера

$$\ln f(x;\mu) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$s(x;\mu) = \frac{\partial}{\partial\mu}\ln f = \frac{x-\mu}{\sigma^2}$$

$$I(\mu) = \mathbb{E}\left[\frac{(X-\mu)^2}{\sigma^4}\right] = \frac{\sigma^2}{\sigma^4} = \boxed{\frac{1}{\sigma^2}}$$

$$\ln f(x;\sigma^2) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ln f}{\partial\sigma^2} = -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4}$$

$$\frac{\partial^2 \ln f}{\partial(\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

$$I(\sigma^2) = -\,\mathbb{E}\left[\frac{1}{2\sigma^4} - \frac{(X-\mu)^2}{\sigma^6}\right] = -\frac{1}{2\sigma^4} + \frac{\sigma^2}{\sigma^6} = \boxed{\frac{1}{2\sigma^4}}$$

$$\ln f(x;p) = x\ln p + (1-x)\ln(1-p)$$

$$s(x;p) = \frac{x}{p} - \frac{1-x}{1-p} = \frac{x-p}{p(1-p)}$$

$$I(p) = \mathbb{E}\left[\frac{(X-p)^2}{p^2(1-p)^2}\right] = \frac{p(1-p)}{p^2(1-p)^2} = \boxed{\frac{1}{p(1-p)}}$$

$$\ln f(k;\lambda) = k\ln\lambda - \lambda - \ln(k!)$$
$$\frac{\partial^2 \ln f}{\partial\lambda^2} = -\frac{k}{\lambda^2}$$
$$I(\lambda) = -\,\mathbb{E}\left[-\frac{X}{\lambda^2}\right] = \frac{\mathbb{E}[X]}{\lambda^2} = \boxed{\frac{1}{\lambda}}$$

11.4 Многомерный случай

Для $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$ **матрица информации Фишера**:

$$I(\theta)_{jk} = \mathbb{E}_\theta\left[\frac{\partial \ln f}{\partial\theta_j}\cdot\frac{\partial \ln f}{\partial\theta_k}\right] = -\,\mathbb{E}_\theta\left[\frac{\partial^2 \ln f}{\partial\theta_j\partial\theta_k}\right]$$

В матричной форме:
$$I(\theta) = \mathbb{E}[\nabla_\theta \ln f\cdot(\nabla_\theta \ln f)^\top] = -\,\mathbb{E}[\nabla^2_\theta \ln f]$$

12 Неравенство Рао—Крамера

12.1 Формулировка

Неравенство Рао—Крамера (также: неравенство Крамера—Рао, информационное неравенство) устанавливает нижнюю границу для дисперсии несмещённых оценок.

Пусть выполнены условия регулярности, $\hat\theta_n$ — несмещённая оценка θ. Тогда:

$$\boxed{\mathbb{D}_\theta[\hat\theta_n] \ge \frac{1}{I_n(\theta)} = \frac{1}{n\cdot I(\theta)}}$$

Величина $\dfrac{1}{nI(\theta)}$ называется **границей Рао—Крамера**.

Рассмотрим выборку $X_1, \dots, X_n$ с совместной плотностью:

$$f_n(\mathbf{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Обозначим:
$$S_n(\theta) = \frac{\partial}{\partial\theta}\ln f_n(\mathbf{X};\theta) = \sum_{i=1}^{n} s(X_i;\theta)$$

Шаг 1. Заметим, что $\mathbb{E}[S_n(\theta)] = 0$ и $\mathbb{D}[S_n(\theta)] = I_n(\theta) = nI(\theta)$.
Шаг 2. Из несмещённости $\hat{\theta}_n$:

$$\mathbb{E}_\theta[\hat{\theta}_n] = \int \hat{\theta}_n(\mathbf{x}) f_n(\mathbf{x};\theta)\, d\mathbf{x} = \theta$$

Дифференцируем по θ (используя условия регулярности):

$$1 = \int \hat{\theta}_n(\mathbf{x}) \frac{\partial f_n}{\partial \theta}\, d\mathbf{x} = \int \hat{\theta}_n(\mathbf{x}) \cdot S_n(\theta) \cdot f_n(\mathbf{x};\theta)\, d\mathbf{x} = \mathbb{E}[\hat{\theta}_n \cdot S_n(\theta)]$$

Шаг 3. Так как $\mathbb{E}[S_n] = 0$:

$$\mathrm{Cov}(\hat{\theta}_n, S_n) = \mathbb{E}[\hat{\theta}_n \cdot S_n] - \mathbb{E}[\hat{\theta}_n] \cdot \mathbb{E}[S_n] = 1 - \theta \cdot 0 = 1$$

Шаг 4. По неравенству Коши—Буняковского:

$$|\mathrm{Cov}(\hat{\theta}_n, S_n)|^2 \le \mathbb{D}[\hat{\theta}_n] \cdot \mathbb{D}[S_n]$$

$$1 \le \mathbb{D}[\hat{\theta}_n] \cdot I_n(\theta)$$

Следовательно:

$$\mathbb{D}[\hat{\theta}_n] \ge \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}$$

■

12.2 Эффективные оценки

Несмещённая оценка $\hat{\theta}_n^*$ называется **эффективной**, если она достигает границы Рао—Крамера:

$$\mathbb{D}_\theta[\hat{\theta}_n^*] = \frac{1}{nI(\theta)}$$

Несмещённая оценка $\hat{\theta}_n$ эффективна тогда и только тогда, когда:

$$\hat{\theta}_n = \theta + \frac{S_n(\theta)}{I_n(\theta)}$$

или эквивалентно: $\hat{\theta}_n$ линейно зависит от score-функции $S_n(\theta)$.

Равенство в неравенстве Коши—Буняковского достигается тогда и только тогда, когда $\hat{\theta}_n$ и S_n линейно зависимы:

$$\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n] = c \cdot (S_n - \mathbb{E}[S_n]) = c \cdot S_n$$

для некоторой константы c.
Из условия $\text{Cov}(\hat{\theta}_n, S_n) = 1$:

$$c \cdot \mathbb{D}[S_n] = c \cdot I_n(\theta) = 1 \implies c = \frac{1}{I_n(\theta)}$$

Следовательно: $\hat{\theta}_n = \theta + \dfrac{S_n(\theta)}{I_n(\theta)}$. ■

Эффективная оценка существует тогда и только тогда, когда score-функция имеет вид:

$$S_n(\theta) = I_n(\theta) \cdot (T(\mathbf{X}) - \theta)$$

для некоторой статистики $T(\mathbf{X})$.
В этом случае $\hat{\theta}_n^* = T(\mathbf{X})$ — эффективная оценка.

Для $X_i \sim \mathcal{N}(\mu, \sigma^2)$:

$$S_n(\mu) = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma^2} = \frac{n(\overline{X}_n - \mu)}{\sigma^2} = I_n(\mu) \cdot (\overline{X}_n - \mu)$$

Следовательно, $\overline{X}_n$ — эффективная оценка μ.
Проверка: $\mathbb{D}[\overline{X}_n] = \dfrac{\sigma^2}{n} = \dfrac{1}{n \cdot \frac{1}{\sigma^2}} = \dfrac{1}{nI(\mu)}$ ✓

12.3 Обобщения неравенства Рао—Крамера

Если $\mathbb{E}_\theta[\hat{\theta}_n] = g(\theta)$ (оценка со смещением), то:

$$\mathbb{D}_\theta[\hat{\theta}_n] \geq \frac{[g'(\theta)]^2}{I_n(\theta)}$$

Пусть $\tau = \tau(\theta)$ — функция параметра, $\hat{\tau}_n$ — несмещённая оценка τ. Тогда:

$$\mathbb{D}_\theta[\hat{\tau}_n] \geq \frac{[\tau'(\theta)]^2}{I_n(\theta)}$$

Для $\theta \in \mathbb{R}^d$ и несмещённой оценки $\hat{\theta}_n$:

$$\text{Cov}(\hat{\theta}_n) \geq I_n(\theta)^{-1}$$

в смысле неотрицательной определённости разности матриц.

13 Свойства оценок максимального правдоподобия

13.1 Основные результаты

ОМП обладают рядом замечательных асимптотических свойств, делающих их универсальным инструментом статистического оценивания.

При условиях регулярности ОМП состоятельна:

$$\hat{\theta}_n^{MLE} \xrightarrow{\mathbb{P}} \theta_0$$

где θ_0 — истинное значение параметра.

Рассмотрим нормированный логарифм правдоподобия:

$$\frac{1}{n}\ell(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ln f(X_i;\theta)$$

По ЗБЧ:

$$\frac{1}{n}\ell(\theta) \xrightarrow{\mathbb{P}} \mathbb{E}_{\theta_0}[\ln f(X;\theta)] =: M(\theta)$$

Ключевой факт: $M(\theta)$ достигает максимума в точке $\theta = \theta_0$ (это следует из неравенства Йенсена и свойств KL-дивергенции).
По теореме о сходимости точек максимума:

$$\hat{\theta}_n^{MLE} = \arg\max \frac{1}{n}\ell(\theta) \xrightarrow{\mathbb{P}} \arg\max M(\theta) = \theta_0$$

■

При условиях регулярности:

$$\boxed{\sqrt{n}(\hat{\theta}_n^{MLE} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0)^{-1})}$$

[redacted]

Разложим score-функцию в ряд Тейлора около θ_0:

$$S_n(\hat{\theta}_n) = S_n(\theta_0) + S_n'(\theta^*)(\hat{\theta}_n - \theta_0)$$

для некоторого θ^* между θ_0 и $\hat{\theta}_n$.
Так как $\hat{\theta}_n$ — точка максимума, $S_n(\hat{\theta}_n) = 0$:

$$0 = S_n(\theta_0) + S_n'(\theta^*)(\hat{\theta}_n - \theta_0)$$

Следовательно:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\frac{\frac{1}{\sqrt{n}} S_n(\theta_0)}{\frac{1}{n} S_n'(\theta^*)}$$

Числитель: По ЦПТ:

$$\frac{1}{\sqrt{n}} S_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(X_i; \theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0))$$

Знаменатель: По ЗБЧ:

$$\frac{1}{n} S_n'(\theta^*) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \ln f(X_i; \theta^*)}{\partial \theta^2} \xrightarrow{\mathbb{P}} -I(\theta_0)$$

По теореме Слуцкого:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \frac{\mathcal{N}(0, I(\theta_0))}{I(\theta_0)} = \mathcal{N}(0, I(\theta_0)^{-1})$$

■

[redacted]

ОМП асимптотически эффективна: её асимптотическая дисперсия $I(\theta)^{-1}$ совпадает с границей Рао—Крамера.

13.2 Дополнительные свойства ОМП

[redacted]

Если $\hat{\theta}^{MLE}$ — ОМП для θ, то для любой функции g:

$$\widehat{g(\theta)}^{MLE} = g(\hat{\theta}^{MLE})$$

[redacted]

Пусть $\tau = g(\theta)$, g дифференцируема, $g'(\theta) \neq 0$. Тогда:

$$\sqrt{n}(g(\hat{\theta}_n^{MLE}) - g(\theta_0)) \xrightarrow{d} \mathcal{N}\left(0, \frac{[g'(\theta_0)]^2}{I(\theta_0)}\right)$$

Для $\theta \in \mathbb{R}^d$:

$$\sqrt{n}(\hat{\theta}_n^{MLE} - \theta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I(\theta_0)^{-1})$$

где $I(\theta)$ — матрица информации Фишера.

13.3 Асимптотическое смещение ОМП

ОМП в общем случае **смещена** в конечных выборках. Однако смещение имеет порядок $O(1/n)$:

$$\mathbb{E}[\hat{\theta}_n^{MLE}] = \theta + \frac{b(\theta)}{n} + O(n^{-2})$$

для некоторой функции $b(\theta)$.
Можно построить **исправленную ОМП**:

$$\tilde{\theta}_n = \hat{\theta}_n^{MLE} - \frac{b(\hat{\theta}_n^{MLE})}{n}$$

которая будет асимптотически несмещённой с точностью $O(n^{-2})$.

13.4 Сравнение ОМП и метода моментов

Свойство	ОМП	Метод моментов
Состоятельность	Да	Да
Асимптотическая нормальность	Да	Да
Асимптотическая эффективность	Да	Не всегда
Инвариантность	Да	Нет
Простота вычисления	Может быть сложно	Обычно просто
Существование в замкнутой форме	Не всегда	Обычно да

14 Байесовские оценки

14.1 Байесовский подход к статистике

В байесовском подходе параметр θ рассматривается как **случайная величина** с некоторым **априорным распределением**.

Байесовская статистическая модель задаётся:

1. **Априорное распределение** $\pi(\theta)$ — плотность распределения θ до наблюдения данных
2. **Условное распределение данных** $f(\mathbf{x}|\theta)$ — функция правдоподобия

Апостериорное распределение θ при данных $\mathbf{X} = \mathbf{x}$:

$$\boxed{\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta) \cdot \pi(\theta)}{\int f(\mathbf{x}|\theta')\pi(\theta')\, d\theta'} \propto f(\mathbf{x}|\theta) \cdot \pi(\theta)}$$

Словами: Апостериорное $\propto$ Правдоподобие $\times$ Априорное.

14.2 Функция потерь и байесовский риск

Функция потерь $L(\theta, a)$ измеряет «штраф» за принятие решения a, когда истинный параметр равен θ.
Популярные функции потерь:

- **Квадратичная:** $L(\theta, a) = (\theta - a)^2$
- **Абсолютная:** $L(\theta, a) = |\theta - a|$
- **0-1 потери:** $L(\theta, a) = \mathbf{1}_{\theta \neq a}$

Апостериорный риск оценки $\hat{\theta}$:

$$\rho(\hat{\theta}|\mathbf{x}) = \mathbb{E}[L(\theta, \hat{\theta})|\mathbf{X} = \mathbf{x}] = \int L(\theta, \hat{\theta})\pi(\theta|\mathbf{x})\, d\theta$$

Байесовский риск (усреднённый по данным):

$$r(\pi, \hat{\theta}) = \mathbb{E}_{\mathbf{X}}[\rho(\hat{\theta}|\mathbf{X})] = \mathbb{E}_{\theta,\mathbf{X}}[L(\theta, \hat{\theta})]$$

Байесовская оценка — оценка, минимизирующая апостериорный риск:

$$\hat{\theta}^{Bayes} = \arg\min_a \mathbb{E}[L(\theta, a)|\mathbf{X}]$$

14.3 Байесовские оценки для различных функций потерь

(i) Квадратичные потери $L(\theta, a) = (\theta - a)^2$:

$$\boxed{\hat{\theta}^{Bayes} = \mathbb{E}[\theta|\mathbf{X}] = \int \theta \cdot \pi(\theta|\mathbf{x})\, d\theta}$$

— **апостериорное среднее**

(ii) Абсолютные потери $L(\theta, a) = |\theta - a|$:

$$\hat{\theta}^{Bayes} = \text{Med}[\theta|\mathbf{X}]$$

— **апостериорная медиана**

(iii) 0-1 потери:

$$\hat{\theta}^{Bayes} = \text{Mode}[\theta|\mathbf{X}] = \arg\max_{\theta} \pi(\theta|\mathbf{x})$$

— **апостериорная мода** (MAP-оценка)

[redacted]

Минимизируем:

$$\rho(a|\mathbf{x}) = \mathbb{E}[(\theta - a)^2|\mathbf{X} = \mathbf{x}]$$

Дифференцируем по a:

$$\frac{\partial \rho}{\partial a} = -2\,\mathbb{E}[\theta - a|\mathbf{X}] = -2(\mathbb{E}[\theta|\mathbf{X}] - a) = 0$$

Следовательно: $a = \mathbb{E}[\theta|\mathbf{X}]$. ∎

14.4 Примеры байесовских оценок

[redacted]

Пусть $X_1, \ldots, X_n|\mu \sim \mathcal{N}(\mu, \sigma^2)$ (дисперсия σ^2 известна).
Априорное распределение: $\mu \sim \mathcal{N}(\mu_0, \tau^2)$.
Апостериорное распределение:

$$\mu|X_1, \ldots, X_n \sim \mathcal{N}\left(\frac{\frac{\mu_0}{\tau^2} + \frac{n\overline{X}_n}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\right)$$

Байесовская оценка (апостериорное среднее):

$$\boxed{\hat{\mu}^{Bayes} = \frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \cdot \mu_0 + \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \cdot \overline{X}_n = w \cdot \mu_0 + (1 - w) \cdot \overline{X}_n}$$

Это **взвешенное среднее** априорного среднего μ_0 и выборочного среднего $\overline{X}_n$.
При $n \to \infty$: $\hat{\mu}^{Bayes} \to \overline{X}_n$ (влияние априорного распределения исчезает).

[redacted]

Пусть $X_1, \ldots, X_n|p \sim \text{Bernoulli}(p)$.
Априорное: $p \sim \text{Beta}(\alpha, \beta)$.
Апостериорное: $p|X_1, \ldots, X_n \sim \text{Beta}(\alpha + S_n, \beta + n - S_n)$, где $S_n = \sum X_i$.

Байесовская оценка:

$$\hat{p}^{Bayes} = \frac{\alpha + S_n}{\alpha + \beta + n}$$

Это сглаживание Лапласа: добавляем α «успехов» и β «неудач» к данным.

14.5 Свойства байесовских оценок

При $n \to \infty$ и регулярных условиях:

(i) Апостериорное распределение сосредотачивается около истинного θ_0

(ii) Байесовская оценка состоятельна: $\hat{\theta}^{Bayes} \xrightarrow{\mathbb{P}} \theta_0$

(iii) Апостериорное распределение асимптотически нормально:

$$\theta | \mathbf{X} \approx \mathcal{N}\left(\hat{\theta}^{MLE}, \frac{1}{nI(\theta_0)}\right)$$

(iv) $\hat{\theta}^{Bayes} - \hat{\theta}^{MLE} = O_p(n^{-1})$

- Естественное включение априорной информации
- Полное апостериорное распределение (не только точечная оценка)
- Автоматическая регуляризация через априорное распределение
- Когерентный подход к принятию решений

15 Минимаксные оценки

15.1 Мотивация

Минимаксный подход — «пессимистическая» стратегия: минимизируем максимальный возможный риск по всем значениям параметра.

15.2 Функция риска

Функция риска оценки $\hat{\theta}$ при параметре θ:

$$R(\theta, \hat{\theta}) = \mathbb{E}_\theta[L(\theta, \hat{\theta})]$$

Для квадратичных потерь: $R(\theta, \hat{\theta}) = \mathrm{MSE}_\theta(\hat{\theta}) = \mathbb{D}_\theta[\hat{\theta}] + (\mathrm{bias})^2$.

Нельзя минимизировать $R(\theta, \hat{\theta})$ одновременно для всех θ — разные оценки оптимальны при разных θ.

15.3 Минимаксные оценки

Оценка $\hat{\theta}^{minimax}$ называется **минимаксной**, если:

$$\boxed{\sup_{\theta \in \Theta} R(\theta, \hat{\theta}^{minimax}) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} R(\theta, \hat{\theta})}$$

Словами: минимизируем максимальный риск.

Минимаксный риск:

$$R^* = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} R(\theta, \hat{\theta})$$

15.4 Связь с байесовскими оценками

Пусть $\hat{\theta}^{\pi}$ — байесовская оценка относительно априорного распределения π. Если функция риска $R(\theta, \hat{\theta}^{\pi})$ постоянна по θ:

$$R(\theta, \hat{\theta}^{\pi}) = c \quad \text{для всех } \theta \in \Theta$$

то $\hat{\theta}^{\pi}$ — минимаксная оценка, а π — **наименее благоприятное априорное распределение**.

Для любой другой оценки $\tilde{\theta}$:

$$\sup_{\theta} R(\theta, \tilde{\theta}) \geq \int R(\theta, \tilde{\theta}) \pi(\theta)\, d\theta \geq \int R(\theta, \hat{\theta}^{\pi}) \pi(\theta)\, d\theta = c = \sup_{\theta} R(\theta, \hat{\theta}^{\pi})$$

Первое неравенство — $\sup \geq$ среднее. Второе — $\hat{\theta}^{\pi}$ минимизирует байесовский риск. ■

Если существует априорное распределение π^* такое, что:

$$\sup_{\theta} R(\theta, \hat{\theta}^{\pi^*}) = \inf_{\hat{\theta}} \sup_{\theta} R(\theta, \hat{\theta}) = \sup_{\pi} \inf_{\hat{\theta}} r(\pi, \hat{\theta})$$

то $(\hat{\theta}^{\pi^*}, \pi^*)$ образуют **седловую точку** и $\hat{\theta}^{\pi^*}$ — минимаксная оценка.

15.5 Примеры минимаксных оценок

Пусть $X \sim \text{Binomial}(n, p)$, $p \in [0, 1]$.
Для квадратичных потерь $L(p, a) = (p - a)^2$ минимаксная оценка:

$$\boxed{\hat{p}^{minimax} = \frac{X + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}}$$

Это байесовская оценка относительно Beta$(\sqrt{n}/2, \sqrt{n}/2)$.
Сравнение:

- ОМП: $\hat{p}^{MLE} = X/n$
- Минимаксная: сдвигает оценку к $1/2$

Пусть $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, σ^2 известна.
Для квадратичных потерь минимаксная оценка:

$$\hat{\mu}^{minimax} = \overline{X}_n$$

Выборочное среднее минимаксно (и одновременно является ОМП и байесовской оценкой для неинформативного априорного).

Пусть $X \sim \mathcal{N}_d(\theta, I_d)$, $d \geq 3$.
Оценка Стейна:

$$\hat{\theta}^{JS} = \left(1 - \frac{d-2}{\|X\|^2}\right) X$$

Эта оценка **доминирует** обычную оценку X (т.е. имеет меньший риск для всех θ).
Это означает, что X — **недопустимая** оценка при $d \geq 3$ (парадокс Стейна).

15.6 Допустимость

Оценка $\hat{\theta}$ **допустима**, если не существует другой оценки $\tilde{\theta}$, доминирующей её:

$$\nexists \tilde{\theta} : R(\theta, \tilde{\theta}) \leq R(\theta, \hat{\theta}) \ \forall \theta, \quad R(\theta_0, \tilde{\theta}) < R(\theta_0, \hat{\theta}) \text{ для некоторого } \theta_0$$

Единственная байесовская оценка (при строго положительном априорном) допустима.

Единственная минимаксная оценка с постоянным риском допустима.

15.7 Сводка подходов к оценинванию

Подход	**Критерий оптимальности**
ОМП	Максимизация правдоподобия $L(\theta)$
Метод моментов	Приравнивание теоретических и выборочных моментов
Несмещённая эффективная	Минимальная дисперсия среди несмещённых (граница Р-К)
Байесовская	Минимизация апостериорного риска $\mathbb{E}[L(\theta, a)\|\mathbf{X}]$
Минимаксная	Минимизация максимального риска $\sup\limits_{\theta} R(\theta, \hat{\theta})$

16 Доверительные интервалы

16.1 Определение и мотивация

Точечная оценка $\hat{\theta}_n$ даёт одно значение, но не показывает точность оценки. Доверительный интервал даёт диапазон «правдоподобных» значений параметра.

Доверительным интервалом уровня доверия $1-\alpha$ (или уровня значимости α) для параметра θ называется случайный интервал $[\hat{\theta}_L, \hat{\theta}_U]$, где $\hat{\theta}_L = \hat{\theta}_L(X_1, \ldots, X_n)$ и $\hat{\theta}_U = \hat{\theta}_U(X_1, \ldots, X_n)$ — статистики, такие что:

$$\boxed{\mathbb{P}_\theta(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) \geq 1-\alpha \quad \text{для всех } \theta \in \Theta}$$

Типичные значения: $1-\alpha = 0.95$ (95%) или $1-\alpha = 0.99$ (99%).

Важно: После вычисления конкретного интервала $[l, u]$ **нельзя** говорить «θ лежит в $[l, u]$ с вероятностью $1-\alpha$». Параметр θ — фиксированная константа!
Правильная интерпретация: если повторить эксперимент много раз, то примерно $(1-\alpha) \cdot 100\%$ построенных интервалов будут содержать истинное θ.

[redacted]

- **Двусторонний:** $[\hat{\theta}_L, \hat{\theta}_U]$
- **Односторонний снизу:** $[\hat{\theta}_L, +\infty)$, $\mathbb{P}(\theta \geq \hat{\theta}_L) \geq 1 - \alpha$
- **Односторонний сверху:** $(-\infty, \hat{\theta}_U]$, $\mathbb{P}(\theta \leq \hat{\theta}_U) \geq 1 - \alpha$

16.2 Общая схема построения доверительного интервала

Основная идея: Найти **центральную статистику** (pivot) — функцию $G(X_1, \dots, X_n; \theta)$, распределение которой известно и не зависит от θ.

[redacted]

Статистика $G(\mathbf{X}; \theta)$ называется **центральной** (или **опорной**, pivot), если её распределение не зависит от неизвестных параметров.

Алгоритм построения ДИ:
Шаг 1. Найти центральную статистику $G(\mathbf{X}; \theta)$ с известным распределением.
Шаг 2. Найти квантили $g_{\alpha/2}$ и $g_{1-\alpha/2}$ такие, что:

$$\mathbb{P}(g_{\alpha/2} \leq G(\mathbf{X}; \theta) \leq g_{1-\alpha/2}) = 1 - \alpha$$

Шаг 3. Разрешить неравенство относительно θ:

$$g_{\alpha/2} \leq G(\mathbf{X}; \theta) \leq g_{1-\alpha/2} \iff \hat{\theta}_L \leq \theta \leq \hat{\theta}_U$$

[redacted]

Пусть $X_i \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 известна.
Центральная статистика:

$$G = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Распределение G не зависит от μ — это pivot.

16.3 «Универсальный» рецепт

Для асимптотически нормальных оценок существует универсальный подход.

[redacted]

Пусть $\hat{\theta}_n$ — состоятельная и асимптотически нормальная оценка:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$$

Если $\hat{\sigma}_n^2$ — состоятельная оценка $\sigma^2(\theta)$, то асимптотический $(1 - \alpha)$-доверительный интервал:

$$\boxed{\left[\hat{\theta}_n - z_{1-\alpha/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}}, \quad \hat{\theta}_n + z_{1-\alpha/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}}\right]}$$

где $z_{1-\alpha/2}$ — квантиль уровня $1 - \alpha/2$ стандартного нормального распределения.

Частные случаи $z_{1-\alpha/2}$:

$1-\alpha$	90%	95%	99%
$z_{1-\alpha/2}$	1.645	1.96	2.576

16.4 Связь доверительных интервалов и проверки гипотез

Пусть $C_\alpha(\mathbf{X})$ — $(1-\alpha)$-доверительный интервал для θ. Тогда тест:

$$\text{Отвергнуть } H_0 : \theta = \theta_0 \iff \theta_0 \notin C_\alpha(\mathbf{X})$$

имеет уровень значимости α.
Обратно: если тест уровня α не отвергает $H_0 : \theta = \theta_0$, то $\theta_0 \in C_\alpha(\mathbf{X})$.

17 Теорема Фишера. Распределение хи-квадрат

17.1 Распределение хи-квадрат

Пусть $Z_1, \dots, Z_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Тогда:

$$\chi^2_k = Z_1^2 + Z_2^2 + \ldots + Z_k^2 \sim \chi^2(k)$$

имеет **распределение хи-квадрат** с k **степенями свободы**.

Свойства $\chi^2(k)$:

- Плотность: $f(x) = \dfrac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$, $x > 0$
- $\mathbb{E}[\chi^2_k] = k$
- $\mathbb{D}[\chi^2_k] = 2k$
- $\chi^2_{k_1} + \chi^2_{k_2} \sim \chi^2(k_1 + k_2)$ (при независимости)
- При $k \to \infty$: $\dfrac{\chi^2_k - k}{\sqrt{2k}} \xrightarrow{d} \mathcal{N}(0,1)$

$$\chi^2(k) = \Gamma\left(\frac{k}{2}, \frac{1}{2}\right)$$

17.2 Вспомогательные леммы

Пусть $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_n)$ и A — симметричная идемпотентная матрица ($A^2 = A$) ранга r. Тогда:

$$\mathbf{Z}^\top A \mathbf{Z} \sim \chi^2(r)$$

Так как A симметричная и идемпотентная, её собственные значения — только 0 и 1. Ранг r означает, что ровно r собственных значений равны 1.
Существует ортогональная матрица P такая, что $A = P\text{diag}(1, \ldots, 1, 0, \ldots, 0)P^\top$.
Пусть $\mathbf{Y} = P^\top \mathbf{Z}$. Тогда $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, I_n)$ и:

$$\mathbf{Z}^\top A \mathbf{Z} = \mathbf{Y}^\top \text{diag}(1, \ldots, 1, 0, \ldots, 0)\mathbf{Y} = Y_1^2 + \ldots + Y_r^2 \sim \chi^2(r)$$

■

Пусть $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_n)$ и $\|\mathbf{Z}\|^2 = Q_1 + Q_2 + \ldots + Q_k$, где $Q_i = \mathbf{Z}^\top A_i \mathbf{Z}$ — квадратичные формы с рангами r_i и $\sum r_i = n$.
Тогда $Q_1, \ldots, Q_k$ независимы и $Q_i \sim \chi^2(r_i)$.

17.3 Теорема Фишера

Пусть $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Тогда:

(i) $\overline{X}_n \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

(ii) $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^n (X_i - \overline{X}_n)^2 \sim \chi^2(n-1)$

(iii) $\overline{X}_n$ и S^2 **независимы**

Введём стандартизованные величины $Z_i = \dfrac{X_i - \mu}{\sigma} \sim \mathcal{N}(0,1)$, i.i.d.
Запишем:

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^n \left(\frac{X_i - \overline{X}_n}{\sigma} + \frac{\overline{X}_n - \mu}{\sigma}\right)^2$$

Раскрывая:

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \frac{(X_i - \overline{X}_n)^2}{\sigma^2} + n \cdot \frac{(\overline{X}_n - \mu)^2}{\sigma^2}$$

(перекрёстный член равен нулю, так как $\sum(X_i - \overline{X}_n) = 0$).
Обозначим:

$$Q_1 = \frac{(n-1)S^2}{\sigma^2}, \quad Q_2 = \frac{n(\overline{X}_n - \mu)^2}{\sigma^2} = \left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2$$

Имеем:

$$\underbrace{\sum Z_i^2}_{\chi^2(n)} = Q_1 + Q_2$$

$Q_2 \sim \chi^2(1)$ (квадрат стандартной нормальной).
По теореме Кохрана (или прямой проверке): Q_1 и Q_2 независимы и $Q_1 \sim \chi^2(n-1)$.
Независимость Q_1 и Q_2 влечёт независимость S^2 и $\overline{X}_n$. ■

Почему $n-1$ степеней свободы? Потому что $\sum(X_i - \overline{X}_n) = 0$ — одно линейное ограничение, и только $n-1$ отклонений независимы.

18 Распределения Фишера и Стьюдента. ДИ для нормального закона

18.1 Распределение Стьюдента

Пусть $Z \sim \mathcal{N}(0,1)$ и $V \sim \chi^2(k)$ независимы. Тогда:

$$T = \frac{Z}{\sqrt{V/k}} \sim t(k)$$

имеет **распределение Стьюдента** с k степенями свободы.

Свойства $t(k)$:

- Симметрично относительно нуля
- $\mathbb{E}[T] = 0$ при $k > 1$
- $\mathbb{D}[T] = \dfrac{k}{k-2}$ при $k > 2$
- Более тяжёлые хвосты, чем у $\mathcal{N}(0,1)$
- При $k \to \infty$: $t(k) \to \mathcal{N}(0,1)$
- $t(1) = \text{Cauchy}(0,1)$ (распределение Коши)

[redacted]

Пусть $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Тогда:

$$\boxed{T = \frac{\overline{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1)}$$

[redacted]

$$T = \frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{(\overline{X}_n - \mu)/(\sigma/\sqrt{n})}{S/\sigma} = \frac{Z}{\sqrt{V/(n-1)}}$$

где $Z = \dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ и $V = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.
По теореме Фишера Z и V независимы, поэтому $T \sim t(n-1)$. ■

18.2 Распределение Фишера (F-распределение)

[redacted]

Пусть $U \sim \chi^2(m)$ и $V \sim \chi^2(n)$ независимы. Тогда:

$$F = \frac{U/m}{V/n} \sim F(m, n)$$

имеет **распределение Фишера** с (m, n) степенями свободы.

Свойства $F(m,n)$**:**

- $F > 0$
- $\mathbb{E}[F] = \dfrac{n}{n-2}$ при $n > 2$
- Если $F \sim F(m,n)$, то $1/F \sim F(n,m)$
- Если $T \sim t(n)$, то $T^2 \sim F(1,n)$
- При $m, n \to \infty$: $F(m,n) \to 1$

18.3 Доверительные интервалы для параметров нормального закона

Пусть $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$.

18.3.1 ДИ для μ при известной σ^2

Центральная статистика: $Z = \dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.

$$\mathbb{P}\left(-z_{1-\alpha/2} \le \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le z_{1-\alpha/2}\right) = 1 - \alpha$$

Доверительный интервал:

$$\boxed{\mu \in \left[\overline{X}_n - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad \overline{X}_n + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]}$$

18.3.2 ДИ для μ при неизвестной σ^2

Центральная статистика: $T = \dfrac{\overline{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1)$.

Доверительный интервал:

$$\boxed{\mu \in \left[\overline{X}_n - t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}, \quad \overline{X}_n + t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right]}$$

где $t_{n-1,1-\alpha/2}$ — квантиль уровня $1-\alpha/2$ распределения $t(n-1)$.

18.3.3 ДИ для σ^2 при неизвестном μ

Центральная статистика: $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

$$\mathbb{P}\left(\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}\right) = 1-\alpha$$

Доверительный интервал:

$$\boxed{\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \quad \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]}$$

18.3.4 ДИ для σ^2 при известном μ

Центральная статистика: $\chi^2 = \dfrac{\sum(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$.

Доверительный интервал:

$$\sigma^2 \in \left[\frac{\sum(X_i-\mu)^2}{\chi^2_{n,1-\alpha/2}}, \quad \frac{\sum(X_i-\mu)^2}{\chi^2_{n,\alpha/2}}\right]$$

18.3.5 ДИ для отношения дисперсий двух выборок

Пусть $X_1, \ldots, X_m \sim \mathcal{N}(\mu_1, \sigma_1^2)$ и $Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_2, \sigma_2^2)$ независимы.

Центральная статистика:

$$F = \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} = \frac{S_X^2}{S_Y^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$$

Доверительный интервал для σ_1^2/σ_2^2:

$$\boxed{\frac{\sigma_1^2}{\sigma_2^2} \in \left[\frac{S_X^2}{S_Y^2} \cdot \frac{1}{F_{m-1,n-1,1-\alpha/2}}, \quad \frac{S_X^2}{S_Y^2} \cdot \frac{1}{F_{m-1,n-1,\alpha/2}}\right]}$$

19 Асимптотические доверительные интервалы

19.1 Общая конструкция

Асимптотические ДИ используются, когда:

- Точное распределение статистики неизвестно
- Выборка достаточно велика для применения ЦПТ

Пусть $\hat{\theta}_n$ — асимптотически нормальная оценка:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$$

Если $\hat{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2(\theta)$, то асимптотический $(1-\alpha)$-ДИ:

$$\left[\hat{\theta}_n - z_{1-\alpha/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}}, \quad \hat{\theta}_n + z_{1-\alpha/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}}\right]$$

19.2 Асимптотический ДИ для математического ожидания

Пусть $X_1, \ldots, X_n$ — i.i.d. с $\mathbb{E}[X] = \mu$, $\mathbb{D}[X] = \sigma^2 < \infty$.

По ЦПТ: $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

Оценка дисперсии: $\hat{\sigma}^2 = S^2$.

Асимптотический ДИ:

$$\boxed{\mu \in \left[\overline{X}_n - z_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}, \quad \overline{X}_n + z_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}}\right]}$$

Для нормальной выборки точный ДИ использует t-распределение. Асимптотический ДИ использует z-квантили и даёт хорошее приближение при $n \gtrsim 30$.

19.3 Асимптотический ДИ для дисперсии

Пусть $\mathbb{E}[X^4] < \infty$. По ЦПТ для S_n^2:

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, \mu_4 - \sigma^4)$$

где $\mu_4 = \mathbb{E}[(X-\mu)^4]$.

Оценка асимптотической дисперсии:

$$\widehat{\mu_4 - \sigma^4} = M_4 - S_n^4$$

где $M_4 = \dfrac{1}{n}\sum (X_i - \overline{X}_n)^4$.

Асимптотический ДИ:

$$\boxed{\sigma^2 \in \left[S_n^2 - z_{1-\alpha/2} \cdot \frac{\sqrt{M_4 - S_n^4}}{\sqrt{n}}, \quad S_n^2 + z_{1-\alpha/2} \cdot \frac{\sqrt{M_4 - S_n^4}}{\sqrt{n}}\right]}$$

При $X \sim \mathcal{N}(\mu, \sigma^2)$: $\mu_4 = 3\sigma^4$, поэтому $\mu_4 - \sigma^4 = 2\sigma^4$.
Асимптотический ДИ:

$$\sigma^2 \in \left[S^2 - z_{1-\alpha/2} \cdot \frac{S^2\sqrt{2}}{\sqrt{n}}, \quad S^2 + z_{1-\alpha/2} \cdot \frac{S^2\sqrt{2}}{\sqrt{n}}\right]$$

19.4 Асимптотический ДИ для медианы

Пусть F имеет плотность f, непрерывную в точке медианы $m = F^{-1}(1/2)$.

Выборочная медиана $\hat{m}_n$ асимптотически нормальна:

$$\sqrt{n}(\hat{m}_n - m) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4f(m)^2}\right)$$

Проблема: Асимптотическая дисперсия зависит от $f(m)$ — плотности в точке медианы.
Решения:
1. Ядерная оценка плотности:

$$\hat{f}(\hat{m}_n) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\hat{m}_n - X_i}{h}\right)$$

Асимптотический ДИ:

$$m \in \left[\hat{m}_n - \frac{z_{1-\alpha/2}}{2\hat{f}(\hat{m}_n)\sqrt{n}}, \quad \hat{m}_n + \frac{z_{1-\alpha/2}}{2\hat{f}(\hat{m}_n)\sqrt{n}}\right]$$

2. ДИ на основе порядковых статистик (точный):

Используем тот факт, что $\mathbb{P}(X_{(j)} \le m \le X_{(k)})$ можно вычислить точно через биномиальное распределение.

Для уровня $1-\alpha$ находим j и k такие, что:

$$\sum_{i=j}^{k-1} \binom{n}{i} \left(\frac{1}{2}\right)^n \ge 1 - \alpha$$

Тогда $[X_{(j)}, X_{(k)}]$ — точный ДИ для медианы.

19.5 Асимптотический ДИ для квантиля

Для $\xi_p = F^{-1}(p)$ и выборочного квантиля $\hat{\xi}_p$:

$$\sqrt{n}(\hat{\xi}_p - \xi_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f(\xi_p)^2}\right)$$

Асимптотический ДИ:

$$\xi_p \in \left[\hat{\xi}_p - z_{1-\alpha/2} \cdot \frac{\sqrt{p(1-p)}}{\hat{f}(\hat{\xi}_p)\sqrt{n}}, \quad \hat{\xi}_p + z_{1-\alpha/2} \cdot \frac{\sqrt{p(1-p)}}{\hat{f}(\hat{\xi}_p)\sqrt{n}}\right]$$

19.6 Сравнение точных и асимптотических ДИ

	Точный ДИ	**Асимптотический ДИ**
Покрытие	Ровно $1-\alpha$	Примерно $1-\alpha$
Требования	Знание распределения	Только ЦПТ
Размер выборки	Любой	Достаточно большой
Универсальность	Низкая	Высокая

20 Проверка статистических гипотез

20.1 Постановка задачи

Статистическая гипотеза — утверждение о распределении наблюдений.
Параметрическая гипотеза: $H : \theta \in \Theta_0 \subset \Theta$
Непараметрическая гипотеза: $H : F \in \mathcal{F}_0$ (например, «F — нормальное»)

- H_0 — **нулевая гипотеза** (null hypothesis) — гипотеза, которую проверяем
- H_1 — **альтернативная гипотеза** (alternative hypothesis) — то, что принимаем при отклонении H_0

Примеры:

- $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ — двусторонняя альтернатива
- $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$ — односторонняя альтернатива
- $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ — сравнение двух групп

- **Простая гипотеза:** $H_0 : \theta = \theta_0$ (одна точка)
- **Сложная гипотеза:** $H_0 : \theta \in \Theta_0$ (множество точек)

20.2 Выбор нулевой гипотезы

Принципы выбора H_0:

1. H_0 — гипотеза «статус-кво» (ничего не происходит, нет эффекта)
2. H_0 — та гипотеза, для которой ошибочное отклонение более опасно
3. H_0 — гипотеза, которую хотим опровергнуть (чтобы доказать H_1)

Примеры:

- Клинические испытания: H_0: лекарство не эффективно
- Контроль качества: H_0: изделие соответствует стандарту
- Судебное дело: H_0: подсудимый невиновен

20.3 Общий принцип работы статистического теста

Статистический критерий — правило, по которому на основе выборки принимается решение: отвергнуть H_0 или не отвергать.
Формально: отображение $\phi : \mathcal{X}^n \to \{0, 1\}$

- $\phi(\mathbf{X}) = 1$ — отвергаем H_0
- $\phi(\mathbf{X}) = 0$ — не отвергаем H_0

Тестовая статистика — функция $T = T(X_1, \ldots, X_n)$, на основе которой принимается решение.
Критическая область W — множество значений T, при которых H_0 отвергается:

$$\phi(\mathbf{X}) = 1 \iff T(\mathbf{X}) \in W$$

Алгоритм проверки гипотез:
Шаг 1. Сформулировать H_0 и H_1.
Шаг 2. Выбрать уровень значимости α (обычно 0.05 или 0.01).
Шаг 3. Выбрать тестовую статистику T с известным распределением при H_0.
Шаг 4. Определить критическую область W так, чтобы $\mathbb{P}_{H_0}(T \in W) = \alpha$.
Шаг 5. Вычислить T по выборке и принять решение:

- $T \in W$ — отвергнуть H_0
- $T \notin W$ — не отвергать H_0

20.4 Ошибки I и II рода

	H_0 **верна**	H_0 **неверна**
Не отвергаем H_0	Верное решение	**Ошибка II рода** (β) False Negative
Отвергаем H_0	**Ошибка I рода** (α) False Positive	Верное решение

- **Ошибка I рода** (Type I error, False Positive): отвергнуть верную H_0

$$\alpha = \mathbb{P}(\text{отвергнуть } H_0 | H_0 \text{ верна}) = \mathbb{P}_{H_0}(T \in W)$$

- **Ошибка II рода** (Type II error, False Negative): не отвергнуть неверную H_0

$$\beta = \mathbb{P}(\text{не отвергнуть } H_0 | H_1 \text{ верна}) = \mathbb{P}_{H_1}(T \notin W)$$

Уровень значимости теста — максимальная вероятность ошибки I рода:

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\text{отвергнуть } H_0)$$

Мощность (power) теста — вероятность правильно отвергнуть неверную H_0:

$$\text{Мощность} = 1 - \beta = \mathbb{P}_{H_1}(T \in W)$$

Функция мощности:

$$\pi(\theta) = \mathbb{P}_\theta(\text{отвергнуть } H_0)$$

При фиксированном n нельзя уменьшить обе ошибки одновременно:

- Уменьшение α (более строгий критерий) $\Rightarrow$ увеличение β
- Единственный способ уменьшить обе ошибки — увеличить n

20.5 p-value (достигаемый уровень значимости)

p**-value** (р-значение) — вероятность получить значение тестовой статистики, столь же или более экстремальное, чем наблюдаемое, при условии что H_0 верна:

$$\boxed{p\text{-value} = \mathbb{P}_{H_0}(T \geq T_{\text{obs}})}$$

(для правосторонней альтернативы; для двусторонней — удваиваем хвост).

Интерпретация:

- p-value — «сила доказательства» против H_0
- Малое p-value — данные маловероятны при H_0
- **Правило:** Отвергаем H_0 при p-value $< \alpha$

Типичные пороги:

p-value	Интерпретация
> 0.10	Нет доказательств против H_0
$0.05 - 0.10$	Слабые доказательства
$0.01 - 0.05$	Умеренные доказательства
$0.001 - 0.01$	Сильные доказательства
< 0.001	Очень сильные доказательства

Если H_0 верна, то p-value $\sim \text{Uniform}(0, 1)$.
Следовательно: $\mathbb{P}_{H_0}(p\text{-value} < \alpha) = \alpha$.

p-value **НЕ** является:

- Вероятностью того, что H_0 верна
- Вероятностью того, что результат получен случайно
- Мерой размера эффекта

p-value — это вероятность данных (или более экстремальных) при условии H_0, а не вероятность H_0 при данных.

20.6 Примеры тестов

$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, σ^2 известна.
Тестовая статистика: $Z = \dfrac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$ при H_0.
Критическая область: $W = \{|Z| > z_{1-\alpha/2}\}$.
p-value: $2(1 - \Phi(|Z_{\text{obs}}|))$.

$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, σ^2 неизвестна.
Тестовая статистика: $T = \dfrac{\overline{X}_n - \mu_0}{S/\sqrt{n}} \sim t(n - 1)$ при H_0.
Критическая область: $W = \{|T| > t_{n-1,1-\alpha/2}\}$.

20.7 Связь между понятиями

Уровень значимости $\alpha \xrightarrow{\text{определяет}}$ **Критическая область** $W \xrightarrow{\text{определяет}}$ **Мощность** $1 - \beta$

p-value $\xleftrightarrow{\text{сравнение}}$ Уровень α: отвергаем $H_0 \iff p\text{-value} < \alpha$

21 Статистические тесты на основе доверительных интервалов

21.1 Общий принцип

Пусть $C_{1-\alpha}(\mathbf{X})$ — $(1-\alpha)$-доверительный интервал для θ. Тогда тест:

$$\text{Отвергнуть } H_0 : \theta = \theta_0 \iff \theta_0 \notin C_{1-\alpha}(\mathbf{X})$$

имеет уровень значимости ровно α.

21.2 z-тест для одной выборки

Условия: $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, дисперсия σ^2 **известна**.

Гипотезы:

- $H_0 : \mu = \mu_0$
- H_1: $\mu \neq \mu_0$ (двусторонняя), $\mu > \mu_0$ или $\mu < \mu_0$ (односторонние)

Тестовая статистика:

$$\boxed{Z = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \text{ при } H_0}$$

Критические области:

H_1	Критическая область	p-value
$\mu \neq \mu_0$	$\lvert Z\rvert > z_{1-\alpha/2}$	$2(1-\Phi(\lvert Z\rvert))$
$\mu > \mu_0$	$Z > z_{1-\alpha}$	$1-\Phi(Z)$
$\mu < \mu_0$	$Z < -z_{1-\alpha}$	$\Phi(Z)$

Мощность теста (для двусторонней альтернативы при истинном $\mu = \mu_1$):

$$\pi(\mu_1) = \Phi\left(-z_{1-\alpha/2} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{1-\alpha/2} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

21.3 z-тест для двух независимых выборок

Условия: $X_1, \dots, X_m \sim \mathcal{N}(\mu_1, \sigma_1^2)$ и $Y_1, \dots, Y_n \sim \mathcal{N}(\mu_2, \sigma_2^2)$ независимы, дисперсии известны.

Гипотезы: $H_0 : \mu_1 = \mu_2$ (или $H_0 : \mu_1 - \mu_2 = \delta_0$).

Тестовая статистика:

$$\boxed{Z = \frac{(\overline{X} - \overline{Y}) - \delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim \mathcal{N}(0,1) \text{ при } H_0}$$

21.4 t-тест для одной выборки

Условия: $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, дисперсия σ^2 **неизвестна**.
Гипотезы: $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.
Тестовая статистика:

$$\boxed{T = \frac{\overline{X}_n - \mu_0}{S/\sqrt{n}} \sim t(n-1) \text{ при } H_0}$$

Критическая область: $|T| > t_{n-1,1-\alpha/2}$.
p**-value:** $2 \cdot \mathbb{P}(t_{n-1} > |T_{\text{obs}}|)$.

При больших n ($n \gtrsim 30$) распределение $t(n-1)$ близко к $\mathcal{N}(0,1)$, и t-тест практически совпадает с z-тестом.

21.5 t-тест для двух независимых выборок

Условия: $X_1, \dots, X_m \sim \mathcal{N}(\mu_1, \sigma^2)$ и $Y_1, \dots, Y_n \sim \mathcal{N}(\mu_2, \sigma^2)$, **одинаковые неизвестные дисперсии**.
Гипотезы: $H_0 : \mu_1 = \mu_2$.
Объединённая оценка дисперсии:

$$S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

Тестовая статистика:

$$\boxed{T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2) \text{ при } H_0}$$

21.5.1 t-тест Уэлча (разные дисперсии)

Если дисперсии $\sigma_1^2 \neq \sigma_2^2$, используется **тест Уэлча**:

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}$$

Распределение аппроксимируется $t(\nu)$, где степени свободы по формуле Уэлча—Саттертуэйта:

$$\nu = \frac{\left(\frac{S_X^2}{m} + \frac{S_Y^2}{n}\right)^2}{\frac{(S_X^2/m)^2}{m-1} + \frac{(S_Y^2/n)^2}{n-1}}$$

21.6 Парный t-тест

Условия: Парные наблюдения $(X_1, Y_1), \dots, (X_n, Y_n)$, $D_i = X_i - Y_i$.
Гипотезы: $H_0 : \mathbb{E}[D] = 0$ (нет разницы).
Тестовая статистика:

$$T = \frac{\overline{D}}{S_D/\sqrt{n}} \sim t(n-1) \text{ при } H_0$$

21.7 F-тест для сравнения дисперсий

Условия: $X_1, \dots, X_m \sim \mathcal{N}(\mu_1, \sigma_1^2)$ и $Y_1, \dots, Y_n \sim \mathcal{N}(\mu_2, \sigma_2^2)$ независимы.
Гипотезы: $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$.
Тестовая статистика:

$$\boxed{F = \frac{S_X^2}{S_Y^2} \sim F(m-1, n-1) \text{ при } H_0}$$

Критическая область (двусторонняя):

$$F < F_{m-1,n-1,\alpha/2} \quad \text{или} \quad F > F_{m-1,n-1,1-\alpha/2}$$

F-тест очень чувствителен к отклонению от нормальности. При ненормальных данных лучше использовать тест Левена или Бартлетта.

21.8 Сводная таблица тестов

Тест	H_0	**Статистика**	**Распределение**
z-тест (1 выб.)	$\mu = \mu_0$	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$\mathcal{N}(0,1)$
t-тест (1 выб.)	$\mu = \mu_0$	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$t(n-1)$
z-тест (2 выб.)	$\mu_1 = \mu_2$	$\dfrac{\overline{X} - \overline{Y}}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}$	$\mathcal{N}(0,1)$
t-тест (2 выб.)	$\mu_1 = \mu_2$	$\dfrac{\overline{X} - \overline{Y}}{S_p\sqrt{1/m + 1/n}}$	$t(m+n-2)$
F-тест	$\sigma_1^2 = \sigma_2^2$	S_X^2/S_Y^2	$F(m-1, n-1)$

22 Критерии Колмогорова—Смирнова

22.1 Критерий согласия Колмогорова (одновыборочный)

Задача: Проверить, имеет ли выборка заданное распределение F_0.
Гипотезы:

- H_0: $F = F_0$ (выборка из F_0)
- H_1: $F \neq F_0$

Статистика Колмогорова:

$$\boxed{D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|}$$

где $F_n(x)$ — эмпирическая функция распределения.

Эквивалентная формула через порядковые статистики:

$$D_n = \max\left(\max_{1\le i\le n}\left(\frac{i}{n} - F_0(X_{(i)})\right), \max_{1\le i\le n}\left(F_0(X_{(i)}) - \frac{i-1}{n}\right)\right)$$

При H_0 распределение D_n не зависит от F_0 (если F_0 непрерывна).
Асимптотически:

$$\sqrt{n}D_n \xrightarrow{d} K$$

где K — **распределение Колмогорова** с функцией распределения:

$$\mathbb{P}(K \le x) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2x^2}, \quad x > 0$$

Преобразуем $U_i = F_0(X_i)$. При H_0: $U_i \sim \text{Uniform}(0,1)$ (интегральное преобразование).
Тогда:

$$D_n = \sup_x |F_n(x) - F_0(x)| = \sup_{u\in[0,1]} |G_n(u) - u|$$

где G_n — ЭФР для $U_1, \dots, U_n \sim \text{Uniform}(0,1)$.
Это распределение не зависит от F_0!
Асимптотика следует из теории эмпирических процессов (процесс $\sqrt{n}(G_n(u) - u)$ сходится к броуновскому мосту). ■

Критерий: Отвергаем H_0 при $D_n > d_{n,\alpha}$, где $d_{n,\alpha}$ — критическое значение.
Приближённое критическое значение:

$$d_{n,\alpha} \approx \sqrt{\frac{-\ln(\alpha/2)}{2n}} \cdot c$$

α	0.10	0.05	0.01
$\sqrt{n}\cdot d_{n,\alpha}$	1.22	1.36	1.63

22.2 Критерий Колмогорова—Смирнова (двухвыборочный)

Задача: Проверить, имеют ли две выборки одинаковое распределение.
Гипотезы:

- H_0: $F_X = F_Y$
- H_1: $F_X \ne F_Y$

Данные: $X_1, \dots, X_m \sim F_X$ и $Y_1, \dots, Y_n \sim F_Y$ — независимые выборки.

$$\boxed{D_{m,n} = \sup_{x\in\mathbb{R}} |F_m(x) - G_n(x)|}$$

где F_m, G_n — ЭФР для выборок X и Y.

При H_0 и $m, n \to \infty$, $m/(m+n) \to \lambda \in (0,1)$:

$$\sqrt{\frac{mn}{m+n}} D_{m,n} \xrightarrow{d} K$$

где K — то же распределение Колмогорова.

Критерий: Отвергаем H_0 при $D_{m,n} > d_{m,n,\alpha}$.

22.3 Односторонние статистики

$$D_n^+ = \sup_x (F_n(x) - F_0(x)), \quad D_n^- = \sup_x (F_0(x) - F_n(x))$$

Используются для односторонних альтернатив:

- $H_1 : F \prec F_0$ (стохастически меньше) — используем D_n^+
- $H_1 : F \succ F_0$ (стохастически больше) — используем D_n^-

22.4 Достоинства и недостатки

Достоинства:

- Непараметрический (не требует предположений о распределении)
- Свободен от распределения при непрерывной F_0
- Чувствителен к различиям в центре распределения

Недостатки:

- Менее мощный для конкретных альтернатив, чем специализированные тесты
- Менее чувствителен к различиям в хвостах
- При оценивании параметров F_0 по данным критические значения неверны (нужна поправка Лиллиефорса)

23 Критерий согласия хи-квадрат Пирсона

23.1 Постановка задачи

Задача: Проверить, согласуются ли данные с заданным распределением.

Разбиваем область значений на k непересекающихся интервалов (ячеек) $A_1, \ldots, A_k$.

Обозначения:

- n_j — наблюдаемое число попаданий в A_j
- $p_j = \mathbb{P}(X \in A_j)$ — теоретическая вероятность при H_0
- $e_j = n \cdot p_j$ — ожидаемое число попаданий

23.2 Критерий для простой гипотезы

Гипотезы: $H_0 : F = F_0$ (полностью задано).

$$\boxed{\chi^2 = \sum_{j=1}^{k} \frac{(n_j - e_j)^2}{e_j} = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j}}$$

При H_0 и $n \to \infty$:

$$\chi^2 \xrightarrow{d} \chi^2(k-1)$$

при условии, что все $e_j = np_j \to \infty$ (практически: $e_j \geq 5$).

Вектор частот $(n_1, \ldots, n_k)$ имеет полиномиальное распределение.
По ЦПТ для мультиномиального распределения:

$$\sqrt{n}\left(\frac{n_j}{n} - p_j\right) \xrightarrow{d} \mathcal{N}(0, p_j(1-p_j))$$

Статистика χ^2 является квадратичной формой от этих нормальных величин. С учётом ограничения $\sum n_j = n$ (одна степень свободы теряется), получаем $\chi^2(k-1)$. ■

Критерий: Отвергаем H_0 при $\chi^2 > \chi^2_{k-1,1-\alpha}$.

23.3 Критерий для сложной гипотезы (с оцениванием параметров)

Гипотезы: $H_0 : F \in \{F_\theta : \theta \in \Theta\}$ — параметрическое семейство.

Процедура:

1. Оценить θ по данным (например, методом максимального правдоподобия): $\hat{\theta}$
2. Вычислить $\hat{p}_j = \mathbb{P}_{\hat{\theta}}(X \in A_j)$
3. Вычислить статистику:

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - n\hat{p}_j)^2}{n\hat{p}_j}$$

Если оценка $\hat{\theta}$ получена методом максимального правдоподобия по сгруппированным данным, то при H_0:

$$\chi^2 \xrightarrow{d} \chi^2(k-1-d)$$

где $d = \dim(\theta)$ — число оцениваемых параметров.

Если параметры оценены по исходным данным (не по группированным), то распределение статистики лежит между $\chi^2(k-1-d)$ и $\chi^2(k-1)$. На практике используют $k-1-d$ как консервативную оценку.

H_0: выборка из $\mathcal{N}(\mu, \sigma^2)$.
Разбиваем на k интервалов, оцениваем $\hat{\mu} = \overline{X}$, $\hat{\sigma}^2 = S^2$.
Степени свободы: $k-1-2 = k-3$.

23.4 Выбор разбиения

Рекомендации:

- Число ячеек: $k \approx \sqrt{n}$ или $k = 1 + \log_2 n$
- Ожидаемые частоты: $e_j \geq 5$ для всех j
- При малых e_j — объединять соседние ячейки
- Равновероятные ячейки ($p_j = 1/k$) часто оптимальны

24 Критерий однородности хи-квадрат

24.1 Постановка задачи

Задача: Проверить, имеют ли несколько выборок одинаковое распределение по категориям.
Данные: r независимых выборок (групп), каждая классифицирована по c категориям.
Таблица сопряжённости:

	Категория 1	Категория 2	$\cdots$	Категория c	Всего
Группа 1	n_{11}	n_{12}	$\cdots$	n_{1c}	$n_{1\cdot}$
Группа 2	n_{21}	n_{22}	$\cdots$	n_{2c}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Группа r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$n_{r\cdot}$
Всего	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot c}$	n

Гипотезы:

- H_0: распределения по категориям одинаковы во всех группах ($p_{1j} = p_{2j} = \ldots = p_{rj}$ для всех j)
- H_1: распределения различаются хотя бы для некоторых групп

24.2 Статистика критерия

Ожидаемые частоты при H_0:

$$e_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$$

Статистика хи-квадрат:

$$\boxed{\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}}$$

[redacted]

При H_0 и $n \to \infty$:

$$\chi^2 \xrightarrow{d} \chi^2((r-1)(c-1))$$

[redacted]

В таблице $r \times c$ всего rc ячеек.
Ограничения:

- r сумм по строкам фиксированы: $-r$ ограничений (но одно следует из $\sum = n$, поэтому $-(r-1)$)
- c сумм по столбцам: $-(c-1)$
- Общая сумма n: уже учтена

Свободных параметров: $rc - (r-1) - (c-1) - 1 = (r-1)(c-1)$. ■

Критерий: Отвергаем H_0 при $\chi^2 > \chi^2_{(r-1)(c-1),1-\alpha}$.

24.3 Случай 2×2

Для таблицы 2×2:

	Категория 1	Категория 2	Всего
Группа 1	a	b	$a+b$
Группа 2	c	d	$c+d$
Всего	$a+c$	$b+d$	n

Упрощённая формула:

$$\boxed{\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}}$$

Степени свободы: $(2-1)(2-1) = 1$.
Поправка Йейтса (для малых выборок):

$$\chi^2_{\text{Yates}} = \frac{n(|ad - bc| - n/2)^2}{(a+b)(c+d)(a+c)(b+d)}$$

[redacted]

При очень малых ожидаемых частотах ($e_{ij} < 5$) лучше использовать точный тест Фишера.

24.4 Точный тест Фишера (для 2×2)

При фиксированных маргинальных суммах вероятность конкретной таблицы:

$$\mathbb{P}(\text{таблица}) = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n! \cdot a! \cdot b! \cdot c! \cdot d!}$$

p-value — сумма вероятностей всех таблиц, не менее экстремальных, чем наблюдаемая.

25 Критерий независимости хи-квадрат

25.1 Постановка задачи

Задача: Проверить независимость двух категориальных признаков.

Данные: Одна выборка объёма n, каждое наблюдение классифицировано по двум признакам:

- Признак A с r уровнями
- Признак B с c уровнями

Таблица сопряжённости: та же структура $r \times c$, но с другой интерпретацией.

Гипотезы:

- H_0: признаки A и B независимы ($p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$)
- H_1: признаки зависимы

25.2 Статистика критерия

При независимости $\mathbb{P}(A = i, B = j) = \mathbb{P}(A = i) \cdot \mathbb{P}(B = j)$.

Оценки маргинальных вероятностей:

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}, \quad \hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}$$

Ожидаемые частоты при H_0:

$$e_{ij} = n \cdot \hat{p}_{i\cdot} \cdot \hat{p}_{\cdot j} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$$

Статистика:

$$\boxed{\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}}$$

При H_0 и $n \to \infty$:

$$\chi^2 \xrightarrow{d} \chi^2((r-1)(c-1))$$

Формулы для статистик однородности и независимости идентичны! Различие — в постановке задачи и интерпретации.

25.3 Случай 2×2

Для таблицы 2×2 (два бинарных признака):

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Степени свободы: 1.

	Болен	Здоров	Всего
Курит	60	40	100
Не курит	30	70	100
Всего	90	110	200

Ожидаемые частоты при независимости:

$$e_{11} = \frac{100 \cdot 90}{200} = 45, \quad e_{12} = \frac{100 \cdot 110}{200} = 55$$

Статистика:

$$\chi^2 = \frac{200 \cdot (60 \cdot 70 - 40 \cdot 30)^2}{100 \cdot 100 \cdot 90 \cdot 110} = \frac{200 \cdot 3000^2}{99000000} \approx 18.18$$

При $\alpha = 0.05$: $\chi^2_{1,0.95} = 3.84$.
$18.18 > 3.84$ — отвергаем H_0, признаки зависимы.

25.4 Меры связи

При отвержении H_0 полезно оценить силу связи.

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(r-1, c-1)}}$$

$V \in [0, 1]$, где $V = 0$ — независимость, $V = 1$ — полная связь.

$$\phi = \sqrt{\frac{\chi^2}{n}} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$\phi \in [-1, 1]$ — аналог корреляции для категориальных данных.

Для таблицы 2×2:

$$OR = \frac{a \cdot d}{b \cdot c}$$

- $OR = 1$ — независимость

- $OR > 1$ — положительная ассоциация
- $OR < 1$ — отрицательная ассоциация

25.5 Сравнение критериев однородности и независимости

	Однородность	**Независимость**
Число выборок	r независимых	Одна
Что фиксировано	Объёмы групп $n_{i\cdot}$	Только n
Гипотеза H_0	$p_{1j} = \ldots = p_{rj}$	$p_{ij} = p_{i\cdot}p_{\cdot j}$
Статистика	Одинаковая	Одинаковая
Распределение	$\chi^2((r-1)(c-1))$	$\chi^2((r-1)(c-1))$

26 Критерии квантилей и знаков

Критерии квантилей и знаков — это непараметрические критерии, не требующие знания вида распределения и основанные на порядковых свойствах данных.

26.1 Критерий знаков (Sign Test)

Критерий знаков предназначен для проверки гипотезы о медиане распределения или для сравнения парных выборок.

26.1.1 Постановка задачи о медиане

Пусть $X_1, \ldots, X_n$ — выборка из непрерывного распределения с медианой Me.

Гипотезы:

- H_0: Me $= m_0$ (медиана равна заданному значению)
- H_1: Me $\neq m_0$ (или односторонняя альтернатива)

26.1.2 Построение статистики

Для каждого наблюдения определим знак:

$$S_i = \text{sign}(X_i - m_0) = \begin{cases} +1, & X_i > m_0 \\ -1, & X_i < m_0 \\ 0, & X_i = m_0 \end{cases}$$

Статистика критерия:

$$\boxed{S^+ = \sum_{i=1}^{n} \mathbf{1}(X_i > m_0) = \#\{i : X_i > m_0\}}$$

— число положительных знаков.

При справедливости H_0 и отбрасывании нулей (пусть их n_0, тогда $n' = n - n_0$):

$$S^+ \sim \text{Bin}(n', 1/2)$$

При H_0: Me $= m_0$, то есть $P(X_i > m_0) = P(X_i < m_0) = 1/2$ по определению медианы (для непрерывного распределения). Каждое $\mathbf{1}(X_i > m_0)$ — независимый бернуллиевский эксперимент с $p = 1/2$. Сумма независимых бернуллиевских величин имеет биномиальное распределение. ■

26.1.3 Критическая область

Двусторонний критерий (уровень α):
Отвергаем H_0, если S^+ слишком мало или слишком велико:

$$S^+ \leq c_1 \quad \text{или} \quad S^+ \geq c_2$$

где c_1, c_2 — квантили Bin$(n', 1/2)$:

$$P(S^+ \leq c_1) = \alpha/2, \quad P(S^+ \geq c_2) = \alpha/2$$

Асимптотическая версия:
При $n' \to \infty$:

$$Z = \frac{S^+ - n'/2}{\sqrt{n'/4}} = \frac{2S^+ - n'}{\sqrt{n'}} \xrightarrow{d} \mathcal{N}(0, 1)$$

26.1.4 Критерий знаков для парных выборок

Пусть $(X_1, Y_1), \ldots, (X_n, Y_n)$ — парные наблюдения.
Гипотезы:

- H_0: распределения X и Y одинаковы (медиана разности равна 0)
- H_1: распределения различаются

Процедура:

1. Вычислить разности $D_i = X_i - Y_i$
2. Подсчитать $S^+ = \#\{i : D_i > 0\}$
3. Применить критерий знаков к разностям с $m_0 = 0$

26.2 Критерий квантилей

Критерий квантилей проверяет гипотезу о том, что заданное значение является квантилью определённого порядка.

26.2.1 Постановка

Гипотезы:

- H_0: $x_p = q_0$ (p-квантиль равна q_0)
- H_1: $x_p \neq q_0$

Статистика:

$$K = \#\{i : X_i \leq q_0\}$$

При H_0:

$$K \sim \text{Bin}(n, p)$$

По определению p-квантили: $P(X \leq x_p) = p$. При H_0: $x_p = q_0$, поэтому $P(X_i \leq q_0) = p$. Каждый индикатор $\mathbf{1}(X_i \leq q_0)$ имеет распределение $\text{Ber}(p)$, и они независимы. ■

Асимптотика:

$$Z = \frac{K - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Проверка гипотезы о медиане ($p = 0.5$) — частный случай критерия квантилей, совпадающий с критерием знаков.

26.3 Знаково-ранговый критерий Уилкоксона

Усиление критерия знаков, учитывающее не только знаки, но и величины отклонений.

Процедура:

1. Вычислить $D_i = X_i - m_0$
2. Упорядочить $|D_i|$ по возрастанию и присвоить ранги R_i
3. Вычислить статистику:

$$\boxed{W^+ = \sum_{i: D_i > 0} R_i}$$

— сумма рангов положительных разностей

При H_0: $\text{Me} = m_0$:

$$\mathbb{E}[W^+] = \frac{n(n+1)}{4}, \quad \text{Var}(W^+) = \frac{n(n+1)(2n+1)}{24}$$

Асимптотика:

$$Z = \frac{W^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \xrightarrow{d} \mathcal{N}(0,1)$$

Критерий Уилкоксона мощнее критерия знаков, так как использует больше информации (величины отклонений), но требует симметричности распределения относительно медианы.

27 Ранговые критерии. Критерий Манна-Уитни-Уилкоксона

Ранговые критерии — непараметрические критерии, основанные на рангах наблюдений. Они устойчивы к выбросам и не требуют предположений о виде распределения.

27.1 Основные понятия

Ранг наблюдения X_i в выборке — это его порядковый номер после упорядочения выборки по возрастанию:

$$R_i = \#\{j : X_j \leq X_i\}$$

Если несколько наблюдений имеют одинаковые значения, им присваивают **средний ранг** — среднее арифметическое рангов, которые они бы заняли.

Выборка: $3, 7, 7, 9, 12$. Ранги: $1, 2.5, 2.5, 4, 5$ (значениям 7 присвоен средний ранг $(2+3)/2 = 2.5$).

27.2 Критерий Манна-Уитни-Уилкоксона

Непараметрический критерий для сравнения двух независимых выборок. Также известен как:

- Критерий суммы рангов Уилкоксона (Wilcoxon rank-sum test)
- U-критерий Манна-Уитни (Mann-Whitney U test)

27.2.1 Постановка задачи

Даны две независимые выборки:

- $X_1, \ldots, X_m$ из распределения F
- $Y_1, \ldots, Y_n$ из распределения G

Гипотезы:

- H_0: $F = G$ (распределения одинаковы)
- H_1: $F \neq G$ (или односторонняя: F стохастически больше/меньше G)

Альтернативная формулировка (модель сдвига): $G(x) = F(x - \Delta)$.

- H_0: $\Delta = 0$
- H_1: $\Delta \neq 0$

27.2.2 Построение статистики Уилкоксона

Процедура:

1. Объединить выборки: $Z_1, \ldots, Z_{m+n}$
2. Упорядочить объединённую выборку и присвоить ранги $R_1, \ldots, R_{m+n}$
3. Вычислить сумму рангов первой выборки:

$$\boxed{W = \sum_{i=1}^{m} R_{X_i}}$$

Если H_0 верна (распределения одинаковы):

$$\mathbb{E}[W] = \frac{m(m+n+1)}{2}$$

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

При H_0 все ранги от 1 до $m+n$ равновероятно распределены между двумя выборками.
Математическое ожидание: Средний ранг в объединённой выборке:

$$\bar{R} = \frac{1 + 2 + \ldots + (m+n)}{m+n} = \frac{m+n+1}{2}$$

Поэтому:

$$\mathbb{E}[W] = m \cdot \bar{R} = \frac{m(m+n+1)}{2}$$

Дисперсия: Дисперсия одного ранга:

$$\mathrm{Var}(R) = \frac{(m+n)^2 - 1}{12}$$

С учётом ковариаций (ранги без повторений):

$$\mathrm{Var}(W) = \frac{mn(m+n+1)}{12}$$

■

27.2.3 Статистика Манна-Уитни

$$\boxed{U = \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{1}(X_i > Y_j)}$$

— число пар (X_i, Y_j), в которых $X_i > Y_j$.

$$U = W - \frac{m(m+1)}{2}$$

или эквивалентно:

$$W = U + \frac{m(m+1)}{2}$$

Сумма рангов первой выборки W складывается из:

- Числа элементов первой выборки, меньших каждого X_i: вклад $\frac{m(m+1)}{2}$ (от самих X)
- Числа элементов второй выборки, меньших каждого X_i: это и есть U

■

$$\mathbb{E}[U] = \frac{mn}{2}, \quad \mathrm{Var}(U) = \frac{mn(m+n+1)}{12}$$

27.2.4 Асимптотическое распределение

При $m, n \to \infty$:

$$Z = \frac{W - \frac{m(m+n+1)}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \xrightarrow{d} \mathcal{N}(0,1)$$

Эквивалентно для U:

$$Z = \frac{U - \frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \xrightarrow{d} \mathcal{N}(0,1)$$

27.2.5 Критическая область

Двусторонний критерий: Отвергаем H_0, если $|Z| > z_{1-\alpha/2}$.
Односторонние критерии:

- H_1: X стохастически больше $Y \Rightarrow$ отвергаем при $Z > z_{1-\alpha}$
- H_1: X стохастически меньше $Y \Rightarrow$ отвергаем при $Z < -z_{1-\alpha}$

27.2.6 Поправка на связи

При наличии связей (одинаковых значений) дисперсия корректируется:

$$\text{Var}(W) = \frac{mn}{12}\left(m+n+1 - \frac{\sum_k t_k(t_k^2-1)}{(m+n)(m+n-1)}\right)$$

где t_k — размер k-й группы связей.

27.3 Интерпретация через вероятности

Статистика $U/(mn)$ является оценкой вероятности $P(X > Y)$:

$$\frac{U}{mn} \approx P(X > Y)$$

При H_0: $P(X > Y) = 1/2$.

Критерий Манна-Уитни проверяет, являются ли два распределения стохастически упорядоченными. Он особенно эффективен для альтернатив сдвига.

27.4 Сравнение с t-критерием

	t-критерий	**Манна-Уитни**
Предположения	Нормальность	Нет
Устойчивость к выбросам	Низкая	Высокая
Мощность (норм. распр.)	Оптимальная	$\approx 95.5\%$ от t
Мощность (тяж. хвосты)	Низкая	Выше

28 Коэффициенты корреляции Пирсона, Спирмена и Кендалла. Статистические тесты

Коэффициенты корреляции измеряют степень связи между двумя переменными. Разные коэффициенты улавливают разные типы зависимости.

28.1 Коэффициент корреляции Пирсона

Коэффициент корреляции Пирсона для случайных величин X и Y:

$$\boxed{\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\cdot\text{Var}(Y)}} = \frac{\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}}$$

1. $\rho \in [-1, 1]$
2. $|\rho| = 1 \Leftrightarrow Y = aX + b$ (линейная зависимость)
3. $\rho = 0$ — отсутствие линейной связи (но возможна нелинейная!)
4. Инвариантен к линейным преобразованиям: $\rho(aX+b, cY+d) = \text{sign}(ac)\cdot\rho(X,Y)$

$$\boxed{r = \frac{\sum_{i=1}^n (X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^n (X_i-\bar{X})^2\cdot\sum_{i=1}^n (Y_i-\bar{Y})^2}}}$$

или эквивалентно:

$$r = \frac{n\sum X_iY_i - \sum X_i\sum Y_i}{\sqrt{[n\sum X_i^2 - (\sum X_i)^2][n\sum Y_i^2 - (\sum Y_i)^2]}}$$

28.1.1 Тест на значимость корреляции Пирсона

Гипотезы:

- H_0: $\rho = 0$ (нет линейной связи)
- H_1: $\rho \neq 0$ (есть линейная связь)

При $(X,Y) \sim \mathcal{N}_2$ и H_0: $\rho = 0$:

$$\boxed{t = r\sqrt{\frac{n-2}{1-r^2}} \sim t(n-2)}$$

При нормальном распределении и $\rho = 0$:

$$r\sqrt{\frac{n-2}{1-r^2}} = \frac{r/\sqrt{1-r^2}}{\sqrt{1/(n-2)}}$$

Числитель асимптотически нормален, знаменатель связан с χ^2-распределением остатков регрессии. ■

Критерий: отвергаем H_0, если $|t| > t_{1-\alpha/2}(n-2)$.

28.1.2 Преобразование Фишера

Для проверки H_0: $\rho = \rho_0 \neq 0$ используется z-преобразование Фишера:

$$\boxed{z = \frac{1}{2}\ln\frac{1+r}{1-r} = \operatorname{arctanh}(r)}$$

При больших n:

$$z \approx \mathcal{N}\left(\frac{1}{2}\ln\frac{1+\rho}{1-\rho}, \frac{1}{n-3}\right)$$

28.2 Коэффициент ранговой корреляции Спирмена

Коэффициент Спирмена — это коэффициент Пирсона, вычисленный для рангов:

$$\boxed{r_S = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \cdot \sum_{i=1}^n (S_i - \bar{S})^2}}}$$

где R_i — ранг X_i, S_i — ранг Y_i.

При отсутствии связей:

$$\boxed{r_S = 1 - \frac{6\sum_{i=1}^n d_i^2}{n(n^2-1)}}$$

где $d_i = R_i - S_i$ — разность рангов.

Поскольку $\bar{R} = \bar{S} = (n+1)/2$ и $\sum(R_i - \bar{R})^2 = \sum(S_i - \bar{S})^2 = \dfrac{n(n^2-1)}{12}$:

$$\begin{aligned} r_S &= \frac{\sum(R_i - \bar{R})(S_i - \bar{S})}{\frac{n(n^2-1)}{12}} \\ &= \frac{\sum R_i S_i - n\bar{R}\bar{S}}{\frac{n(n^2-1)}{12}} \end{aligned}$$

Используя $\sum d_i^2 = \sum R_i^2 + \sum S_i^2 - 2\sum R_i S_i$ и $\sum R_i^2 = \sum S_i^2 = \dfrac{n(n+1)(2n+1)}{6}$:

$$\sum R_i S_i = \frac{n(n+1)(2n+1)}{6} - \frac{\sum d_i^2}{2}$$

Подстановка даёт искомую формулу. ■

28.2.1 Свойства коэффициента Спирмена

- $r_S \in [-1, 1]$
- $|r_S| = 1 \Leftrightarrow$ строгая монотонная зависимость
- Инвариантен к монотонным преобразованиям
- Устойчив к выбросам
- Измеряет монотонную (не обязательно линейную) связь

28.2.2 Тест на значимость Спирмена

Гипотезы:

- H_0: $\rho_S = 0$ (нет монотонной связи)
- H_1: $\rho_S \neq 0$

При $n \to \infty$ и H_0:

$$\sqrt{n-1} \cdot r_S \xrightarrow{d} \mathcal{N}(0,1)$$

или эквивалентно:

$$t = r_S \sqrt{\frac{n-2}{1-r_S^2}} \approx t(n-2)$$

Для малых n используются точные таблицы критических значений.

28.3 Коэффициент ранговой корреляции Кендалла

Пара наблюдений (X_i, Y_i) и (X_j, Y_j) называется:

- **Согласованной**, если $(X_i - X_j)(Y_i - Y_j) > 0$
- **Несогласованной**, если $(X_i - X_j)(Y_i - Y_j) < 0$
- **Связанной**, если $X_i = X_j$ или $Y_i = Y_j$

$$\boxed{\tau = \frac{C - D}{\binom{n}{2}} = \frac{C - D}{\frac{n(n-1)}{2}}}$$

где C — число согласованных пар, D — число несогласованных пар.

Альтернативные формы:

$$\tau = \frac{2(C - D)}{n(n-1)}$$

$$\tau = \frac{\sum_{i<j} \text{sign}(X_j - X_i) \cdot \text{sign}(Y_j - Y_i)}{\binom{n}{2}}$$

28.3.1 Свойства коэффициента Кендалла

1. $\tau \in [-1, 1]$
2. $\tau = 1 \Leftrightarrow$ строго возрастающая монотонная зависимость ($D = 0$)
3. $\tau = -1 \Leftrightarrow$ строго убывающая монотонная зависимость ($C = 0$)
4. $\tau = 0$ — нет монотонной связи

$$\tau = P(\text{согласованная пара}) - P(\text{несогласованная пара})$$

28.3.2 Связь с коэффициентом Спирмена

Для любой выборки:

$$-1 \leq \frac{3\tau - r_S}{2} \leq \tau \leq r_S \leq 1 \quad (\text{при } r_S \geq 0)$$

Асимптотически:

$$r_S \approx \frac{3}{2}\tau \quad (\text{грубое приближение})$$

28.3.3 Тест на значимость Кендалла

При H_0: независимость X и Y:

$$\mathbb{E}[\tau] = 0, \quad \mathrm{Var}(\tau) = \frac{2(2n+5)}{9n(n-1)}$$

При $n \to \infty$:

$$Z = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} = \tau \cdot \frac{3\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} \xrightarrow{d} \mathcal{N}(0,1)$$

28.3.4 Модификации при наличии связей

При наличии связей:

$$\tau_b = \frac{C - D}{\sqrt{(C + D + T_X)(C + D + T_Y)}}$$

где T_X — число пар, связанных по X, T_Y — по Y.

28.4 Сравнение коэффициентов корреляции

	Пирсон	**Спирмен**	**Кендалл**
Тип связи	Линейная	Монотонная	Монотонная
Использует	Значения	Ранги	Пары
Чувствительность	К выбросам	Устойчив	Устойчив
Предположения	Нормальность	Нет	Нет
Интерпретация	Доля дисперсии	Ранговая связь	Вероятности пар

29 Критерий инверсий

Критерий инверсий — непараметрический критерий для проверки гипотезы о случайности последовательности наблюдений. Особенно полезен для обнаружения трендов.

29.1 Определение инверсии

Инверсией в последовательности $X_1, X_2, \ldots, X_n$ называется пара индексов (i, j) такая, что $i < j$, но $X_i > X_j$.

$$\boxed{I_n = \sum_{i<j} \mathbf{1}(X_i > X_j) = \#\{(i,j) : i < j, X_i > X_j\}}$$

Последовательность: $3, 1, 4, 2$.

- $(3, 1)$: $3 > 1$ — инверсия
- $(3, 2)$: $3 > 2$ — инверсия
- $(4, 2)$: $4 > 2$ — инверсия

Итого: $I_4 = 3$.

29.2 Связь с другими статистиками

Число инверсий связано с числом несогласованных пар:

$$I_n = D \quad \text{(число несогласованных пар)}$$

Коэффициент Кендалла:

$$\tau = \frac{\binom{n}{2} - 2I_n}{\binom{n}{2}} = 1 - \frac{4I_n}{n(n-1)}$$

Для последовательности рангов статистика U Манна-Уитни может быть выражена через инверсии.

29.3 Распределение числа инверсий

Пусть $X_1, \ldots, X_n$ — случайная перестановка чисел $1, \ldots, n$ (все перестановки равновероятны). Тогда:

$$\mathbb{E}[I_n] = \frac{n(n-1)}{4}$$

$$\text{Var}(I_n) = \frac{n(n-1)(2n+5)}{72}$$

Математическое ожидание:
Для каждой пары (i, j), $i < j$, определим индикатор $A_{ij} = \mathbf{1}(X_i > X_j)$. При случайном порядке: $P(X_i > X_j) = 1/2$.

$$\mathbb{E}[I_n] = \mathbb{E}\left[\sum_{i<j} A_{ij}\right] = \sum_{i<j} P(X_i > X_j) = \binom{n}{2} \cdot \frac{1}{2} = \frac{n(n-1)}{4}$$

Дисперсия:
$\text{Var}(I_n) = \sum_{i<j} \text{Var}(A_{ij}) + 2 \sum_{(i,j)\neq(k,l)} \text{Cov}(A_{ij}, A_{kl})$
$\text{Var}(A_{ij}) = 1/4$.
Ковариации ненулевые только для пар с общим индексом. После вычислений получаем указанную формулу. ■

При $n \to \infty$:

$$Z = \frac{I_n - \frac{n(n-1)}{4}}{\sqrt{\frac{n(n-1)(2n+5)}{72}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

29.4 Критерий инверсий для проверки случайности

29.4.1 Постановка

Гипотезы:

- H_0: наблюдения независимы и одинаково распределены (случайный порядок)
- H_1: есть тренд или зависимость между последовательными наблюдениями

29.4.2 Обнаружение трендов

- **Возрастающий тренд**: малое число инверсий (I_n мало)
- **Убывающий тренд**: большое число инверсий (I_n велико)

Критическая область:
Двусторонний критерий (любой тренд):

$$|Z| > z_{1-\alpha/2}$$

Односторонний критерий (возрастающий тренд):

$$Z < -z_{1-\alpha}$$

Односторонний критерий (убывающий тренд):

$$Z > z_{1-\alpha}$$

29.4.3 Алгоритм вычисления

Наивный алгоритм: $O(n^2)$ — перебор всех пар.

Эффективный алгоритм: $O(n \log n)$ через модифицированную сортировку слиянием (merge sort).

29.5 Приложения критерия инверсий

1. **Проверка случайности временных рядов**
2. **Обнаружение трендов** в данных
3. **Контроль качества**: обнаружение дрейфа процесса
4. **Проверка гипотезы об отсутствии автокорреляции**

Критерий инверсий тесно связан с ранговыми критериями. Фактически, статистика инверсий эквивалентна статистике Кендалла с заменой $Y_i = i$ (порядковые номера).

30 Модификация критерия хи-квадрат для проверки гипотезы о согласованности данных с моделью простейшей выборки

Модификация критерия хи-квадрат позволяет проверить, являются ли данные простой выборкой (i.i.d. наблюдениями) из некоторого распределения, с учётом оцениваемых параметров.

30.1 Модель простейшей выборки

Простейшая выборка — это набор независимых одинаково распределённых (i.i.d.) случайных величин $X_1, \ldots, X_n$ из некоторого распределения F.

Проверка согласованности данных с моделью простейшей выборки включает:

1. Проверку соответствия данных заданному семейству распределений
2. Учёт того, что параметры распределения оценены по данным

30.2 Стандартный критерий хи-квадрат (напоминание)

Разбиваем область значений на k интервалов $\Delta_1, \ldots, \Delta_k$.

Статистика:

$$\chi^2 = \sum_{j=1}^{k} \frac{(\nu_j - np_j)^2}{np_j}$$

где ν_j — наблюдаемое число попаданий в Δ_j, $p_j = P(X \in \Delta_j)$.

При известных p_j и H_0: $\chi^2 \xrightarrow{d} \chi^2(k-1)$.

30.3 Проблема оценивания параметров

Если параметры распределения $\theta = (\theta_1, \dots, \theta_r)$ оценены по тем же данным, то при H_0:

$$\chi^2 = \sum_{j=1}^{k} \frac{(\nu_j - np_j(\hat{\theta}))^2}{np_j(\hat{\theta})} \xrightarrow{d} \chi^2(k-1-r')$$

где $r' \le r$ — **эффективное число оценённых параметров**.

Важно: $r' = r$ только если параметры оценены методом минимума хи-квадрат по сгруппированным данным. При оценивании методом максимального правдоподобия по исходным данным $r' < r$ в общем случае.

30.4 Модификация Фишера

Если параметры θ оценены методом максимального правдоподобия по **сгруппированным данным** (то есть максимизируется мультиномиальное правдоподобие $\prod p_j(\theta)^{\nu_j}$), то:

$$\chi^2 \xrightarrow{d} \chi^2(k-1-r)$$

где r — число оценённых параметров.

30.5 Модификация для параметров, оценённых по исходным данным

30.5.1 Проблема

При оценивании параметров методом максимального правдоподобия по **исходным (не сгруппированным)** данным распределение статистики χ^2 лежит между $\chi^2(k-1-r)$ и $\chi^2(k-1)$.

30.5.2 Подход Черноффа-Лемана

Пусть $\hat{\theta}$ — МПО по исходным данным. Тогда распределение χ^2-статистики:

$$\chi^2 \xrightarrow{d} \sum_{i=1}^{k-1} \lambda_i Z_i^2$$

где Z_i — независимые $\mathcal{N}(0,1)$, а $\lambda_i \in [0,1]$ зависят от способа разбиения и распределения.

Практические рекомендации:

1. Использовать консервативную оценку с $k-1$ степенями свободы (завышает p-value)
2. Использовать $k-1-r$ степеней свободы (занижает p-value)
3. Использовать специальные поправки или бутстреп

30.6 Рекомендации по выбору разбиения

Для минимизации ошибок II рода рекомендуется:

$$k \approx 2n^{2/5}$$

(правило Манна-Уолда).

Практические правила:

- $np_j \geq 5$ для всех j (классическое правило)
- $np_j \geq 1$ и не более 20% интервалов с $np_j < 5$ (ослабленное правило)
- Равновероятные интервалы: $p_j = 1/k$ — часто оптимальны

30.7 Критерий хи-квадрат с равновероятными интервалами

Процедура:

1. Оценить параметры $\hat{\theta}$ по данным
2. Выбрать k и построить интервалы Δ_j так, что $p_j(\hat{\theta}) = 1/k$
3. Подсчитать ν_j — число наблюдений в каждом интервале
4. Вычислить статистику:
$$\chi^2 = \frac{k}{n}\sum_{j=1}^{k}(\nu_j - n/k)^2 = \frac{k}{n}\sum_{j=1}^{k}\nu_j^2 - n$$
5. Сравнить с $\chi^2(k-1-r)$ или $\chi^2(k-1)$

30.8 Специфика для некоторых распределений

30.8.1 Нормальное распределение

- При оценке μ и σ^2 по исходным данным (МПО)
- Рекомендуется использовать $k-3$ степени свободы
- Альтернатива: критерии Шапиро-Уилка, Андерсона-Дарлинга (мощнее)

30.8.2 Экспоненциальное распределение

- При оценке λ по исходным данным
- Использовать $k-2$ степени свободы

30.8.3 Пуассоновское распределение

- При оценке $\lambda = \bar{X}$
- Объединять категории с малыми ожидаемыми частотами
- Использовать $k-2$ степени свободы (после объединения)

30.9 Альтернативные подходы

30.9.1 Критерий минимума хи-квадрат

Оценить параметры, минимизируя статистику хи-квадрат:

$$\hat{\theta}_{MCS} = \arg\min_{\theta} \sum_{j=1}^{k} \frac{(\nu_j - np_j(\theta))^2}{np_j(\theta)}$$

При такой оценке гарантируется $\chi^2 \xrightarrow{d} \chi^2(k-1-r)$.

30.9.2 Бутстреп-калибровка

1. Оценить параметры $\hat{\theta}$ по данным
2. Генерировать B бутстреп-выборок из $F_{\hat{\theta}}$
3. Для каждой выборки оценить параметры и вычислить χ^2
4. Использовать эмпирическое распределение χ^2 для определения критических значений

30.10 Проверка независимости наблюдений

Для проверки того, что данные представляют простую выборку (i.i.d.), можно дополнительно использовать:

1. **Критерий серий** — проверка на автокорреляцию
2. **Критерий инверсий** — проверка на тренд
3. **Критерий Дарбина-Уотсона** — проверка автокорреляции первого порядка

30.11 Комбинированная процедура проверки

Полная проверка модели простейшей выборки:

1. **Проверка вида распределения:** критерий хи-квадрат, Колмогорова-Смирнова, Андерсона-Дарлинга
2. **Проверка независимости:** критерий серий, инверсий, автокорреляционный анализ
3. **Проверка однородности:** разбить выборку на части и проверить критерием однородности

При множественной проверке необходима поправка на множественные сравнения (Бонферрони, Холма и др.) для контроля вероятности ошибки I рода.

30.12 Сводная таблица степеней свободы

Ситуация	**Степени свободы**
Параметры известны	$k-1$
Параметры оценены по сгруппированным данным (МПО)	$k-1-r$
Параметры оценены методом мин. хи-квадрат	$k-1-r$
Параметры оценены по исходным данным (МПО)	между $k-1-r$ и $k-1$

31 Модель линейной регрессии. Предположения. Оценка наименьших квадратов

Линейная регрессия — фундаментальный метод статистического анализа, позволяющий моделировать зависимость отклика от предикторов.

31.1 Модель линейной регрессии

Линейная модель связывает наблюдения $Y_1, \ldots, Y_n$ с объясняющими переменными (регрессорами):

$$\boxed{Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \ldots, n}$$

или в матричной форме:

$$\boxed{\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}}$$

где:

- $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ — вектор откликов ($n \times 1$)
- $\mathbf{X}$ — матрица плана (дизайна) размера $n \times (p+1)$
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ — вектор параметров ($(p+1) \times 1$)
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ — вектор ошибок ($n \times 1$)

Матрица плана:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

Первый столбец из единиц соответствует свободному члену β_0.

31.2 Минимальные предположения

Минимальные (слабые) предположения о модели линейной регрессии:

1. **Линейность:** $\mathbb{E}[Y_i|\mathbf{x}_i] = \mathbf{x}_i^\top \boldsymbol{\beta}$
2. **Полный ранг:** $\text{rank}(\mathbf{X}) = p + 1$ (матрица **X** полного столбцового ранга)
3. **Экзогенность:** $\mathbb{E}[\varepsilon_i|\mathbf{X}] = 0$ (или слабее: $\mathbb{E}[\varepsilon_i] = 0$)

При минимальных предположениях можно построить оценку наименьших квадратов и показать её несмещённость, но нельзя строить доверительные интервалы и проводить тесты.

31.3 Обычные (классические) предположения

Обычные предположения включают минимальные и дополнительно:

(GM1) **Линейность:** $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

(GM2) **Полный ранг:** $\text{rank}(\mathbf{X}) = p + 1 \leq n$

(GM3) **Строгая экзогенность:** $\mathbb{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$

(GM4) **Сферические ошибки:** $\text{Var}(\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2 \mathbf{I}_n$

- Гомоскедастичность: $\text{Var}(\varepsilon_i) = \sigma^2$ (одинаковая дисперсия)
- Некоррелированность: $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$

Нормальная линейная модель дополнительно требует:

(GM5) **Нормальность:** $\boldsymbol{\varepsilon}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$

Эквивалентно: $\varepsilon_1, \ldots, \varepsilon_n$ — i.i.d. $\mathcal{N}(0, \sigma^2)$.

31.4 Оценка наименьших квадратов (МНК/OLS)

Оценка наименьших квадратов (МНК, OLS — Ordinary Least Squares) минимизирует сумму квадратов остатков:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2$$

При условии полного ранга $\mathbf{X}$:

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}}$$

Минимизируем $Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.
Раскрываем:

$$Q(\boldsymbol{\beta}) = \mathbf{Y}^\top \mathbf{Y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{Y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}$$

Градиент:

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}$$

Приравниваем к нулю:

$$\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = \mathbf{X}^\top \mathbf{Y}$$

Это **нормальные уравнения**. При $\text{rank}(\mathbf{X}) = p + 1$ матрица $\mathbf{X}^\top \mathbf{X}$ обратима:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

Гессиан $\dfrac{\partial^2 Q}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2\mathbf{X}^\top \mathbf{X}$ положительно определён, значит это минимум. ■

31.5 Геометрическая интерпретация

Матрица проектирования (hat matrix):

$$\boxed{\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top}$$

Проецирует $\mathbf{Y}$ на пространство столбцов $\mathbf{X}$.

1. $\mathbf{H}$ — симметричная: $\mathbf{H}^\top = \mathbf{H}$
2. $\mathbf{H}$ — идемпотентная: $\mathbf{H}^2 = \mathbf{H}$
3. $\text{rank}(\mathbf{H}) = \text{tr}(\mathbf{H}) = p + 1$

4. Собственные значения: 0 или 1

Предсказанные значения:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{HY}$$

Остатки:

$$\hat{\varepsilon} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{MY}$$

где $\mathbf{M} = \mathbf{I} - \mathbf{H}$ — также проекционная матрица.

31.6 Свойства МНК-оценки

[redacted]

При условиях GM1-GM3:

$$\mathbb{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = \boldsymbol{\beta}$$

[redacted]

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} \\ &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{X}\boldsymbol{\beta} + \varepsilon) \\ &= \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\varepsilon\end{aligned}$$

Следовательно:

$$\mathbb{E}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbb{E}[\varepsilon|\mathbf{X}] = \boldsymbol{\beta}$$

■

[redacted]

При условиях GM1-GM4:

$$\boxed{\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}}$$

[redacted]

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= \mathrm{Var}\left((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\varepsilon|\mathbf{X}\right) \\ &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathrm{Var}(\varepsilon|\mathbf{X})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\end{aligned}$$

■

31.7 Простая линейная регрессия

Для случая одного регрессора: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xY}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$$

32 Теорема Гаусса-Маркова

Теорема Гаусса-Маркова — фундаментальный результат, обосновывающий оптимальность МНК-оценки в классе линейных несмещённых оценок.

32.1 Формулировка теоремы

При выполнении условий GM1-GM4 (без предположения нормальности) оценка наименьших квадратов $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ является **наилучшей линейной несмещённой оценкой** (BLUE — Best Linear Unbiased Estimator).
То есть для любой линейной несмещённой оценки $\tilde{\boldsymbol{\beta}} = \mathbf{CY}$:

$$\mathrm{Var}(\mathbf{a}^\top \tilde{\boldsymbol{\beta}}) \geq \mathrm{Var}(\mathbf{a}^\top \hat{\boldsymbol{\beta}}) \quad \forall \mathbf{a} \in \mathbb{R}^{p+1}$$

или эквивалентно:

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) - \mathrm{Var}(\hat{\boldsymbol{\beta}}) \text{ — неотрицательно определённая матрица}$$

32.2 Доказательство теоремы Гаусса-Маркова

Пусть $\tilde{\boldsymbol{\beta}} = \mathbf{CY}$ — произвольная линейная оценка.
Шаг 1: Условие несмещённости.
Для несмещённости: $\mathbb{E}[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ для любого $\boldsymbol{\beta}$.

$$\mathbb{E}[\mathbf{CY}] = \mathbf{C}\mathbb{E}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \mathbf{CX}\boldsymbol{\beta}$$

Для равенства $\boldsymbol{\beta}$ при всех $\boldsymbol{\beta}$ необходимо:

$$\mathbf{CX} = \mathbf{I}_{p+1}$$

Шаг 2: Представление линейной оценки.
Запишем $\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top + \mathbf{D}$ для некоторой матрицы $\mathbf{D}$.
Из условия несмещённости:

$$\mathbf{CX} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} + \mathbf{DX} = \mathbf{I} + \mathbf{DX} = \mathbf{I}$$

Следовательно: $\mathbf{DX} = \mathbf{0}$.
Шаг 3: Вычисление дисперсии.

$$\begin{aligned}\mathrm{Var}(\tilde{\boldsymbol{\beta}}) &= \mathrm{Var}(\mathbf{CY}) = \sigma^2\mathbf{CC}^\top \\ &= \sigma^2\left[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{D}\right]\left[\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} + \mathbf{D}^\top\right]\end{aligned}$$

Раскрываем:

$$\begin{aligned}\mathbf{CC}^\top = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{D}^\top \\ + \mathbf{DX}(\mathbf{X}^\top\mathbf{X})^{-1} + \mathbf{DD}^\top\end{aligned}$$

Так как $\mathbf{DX} = \mathbf{0}$, то $\mathbf{X}^\top\mathbf{D}^\top = (\mathbf{DX})^\top = \mathbf{0}$:

$$\mathbf{CC}^\top = (\mathbf{X}^\top\mathbf{X})^{-1} + \mathbf{DD}^\top$$

Шаг 4: Сравнение дисперсий.

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} + \sigma^2\mathbf{DD}^\top = \mathrm{Var}(\hat{\boldsymbol{\beta}}) + \sigma^2\mathbf{DD}^\top$$

Матрица $\mathbf{DD}^\top$ неотрицательно определена, поэтому:

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) \geq \mathrm{Var}(\hat{\boldsymbol{\beta}})$$

в смысле положительной определённости разности.
Равенство достигается только при $\mathbf{D} = \mathbf{0}$, то есть $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$. ■

32.3 Следствия теоремы Гаусса-Маркова

Для любой линейной комбинации $\mathbf{a}^\top\boldsymbol{\beta}$ оценка $\mathbf{a}^\top\hat{\boldsymbol{\beta}}$ имеет минимальную дисперсию среди всех линейных несмещённых оценок.

Предсказание $\hat{Y}(\mathbf{x}_0) = \mathbf{x}_0^\top\hat{\boldsymbol{\beta}}$ имеет минимальную дисперсию среди всех линейных несмещённых предсказаний.

Теорема Гаусса-Маркова **не утверждает**, что МНК-оценка:

- лучшая среди всех несмещённых (нелинейные могут быть лучше)
- имеет минимальный среднеквадратичный риск (смещённые могут быть лучше)
- асимптотически эффективна (это отдельное свойство)

32.4 Эффективность при нормальности

При выполнении GM1-GM5 (включая нормальность) МНК-оценка является:

1. **BLUE** (по теореме Гаусса-Маркова)
2. **BUE** — наилучшей несмещённой (не только линейной) — достигает границы Рао-Крамера
3. **МПО** (оценкой максимального правдоподобия)

33 Оценка остаточной дисперсии

Оценка остаточной дисперсии σ^2 необходима для построения доверительных интервалов и проверки гипотез.

33.1 Остаточная сумма квадратов

Остаточная сумма квадратов (Residual Sum of Squares):

$$\boxed{\mathrm{RSS} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \|\hat{\boldsymbol{\varepsilon}}\|^2 = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 = \mathbf{Y}^\top \mathbf{M} \mathbf{Y}}$$

где $\mathbf{M} = \mathbf{I} - \mathbf{H}$.

33.2 Несмещённая оценка дисперсии

При условиях GM1-GM4:

$$\boxed{s^2 = \frac{\mathrm{RSS}}{n-p-1} = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n-p-1}}$$

является несмещённой оценкой σ^2.

Шаг 1: Представление RSS.

$$\mathrm{RSS} = \mathbf{Y}^\top \mathbf{M} \mathbf{Y} = (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})^\top \mathbf{M} (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

Так как $\mathbf{M}\mathbf{X} = (\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{X} - \mathbf{X} = \mathbf{0}$:

$$\mathrm{RSS} = \boldsymbol{\varepsilon}^\top \mathbf{M} \boldsymbol{\varepsilon}$$

Шаг 2: Математическое ожидание.
Используем свойство: $\mathbb{E}[\boldsymbol{\varepsilon}^\top \mathbf{A} \boldsymbol{\varepsilon}] = \mathrm{tr}(\mathbf{A}\mathrm{Var}(\boldsymbol{\varepsilon})) = \sigma^2 \mathrm{tr}(\mathbf{A})$ при $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$.

$$\mathbb{E}[\mathrm{RSS}] = \mathbb{E}[\boldsymbol{\varepsilon}^\top \mathbf{M} \boldsymbol{\varepsilon}] = \sigma^2 \mathrm{tr}(\mathbf{M})$$

Шаг 3: След матрицы M.

$$\text{tr}(\mathbf{M}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{H}) = n - (p+1)$$

Шаг 4: Вывод.

$$\mathbb{E}[\text{RSS}] = \sigma^2(n-p-1)$$

Следовательно:

$$\mathbb{E}[s^2] = \mathbb{E}\left[\frac{\text{RSS}}{n-p-1}\right] = \sigma^2$$

■

33.3 Распределение RSS при нормальности

При GM1-GM5:

$$\boxed{\frac{\text{RSS}}{\sigma^2} \sim \chi^2(n-p-1)}$$

$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$.
$\text{RSS} = \boldsymbol{\varepsilon}^\top \mathbf{M}\boldsymbol{\varepsilon}$, где **M** — симметричная идемпотентная матрица.
По теореме о квадратичных формах: если $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ и **A** идемпотентна с рангом r, то $\mathbf{Z}^\top\mathbf{A}\mathbf{Z} \sim \chi^2(r)$.
$\frac{\boldsymbol{\varepsilon}}{\sigma} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\text{rank}(\mathbf{M}) = n-p-1$.

$$\frac{\text{RSS}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}^\top}{\sigma}\mathbf{M}\frac{\boldsymbol{\varepsilon}}{\sigma} \sim \chi^2(n-p-1)$$

■

При GM1-GM5:

1. $\dfrac{(n-p-1)s^2}{\sigma^2} \sim \chi^2(n-p-1)$
2. $\mathbb{E}[s^2] = \sigma^2$
3. $\text{Var}(s^2) = \dfrac{2\sigma^4}{n-p-1}$

33.4 Независимость от оценки коэффициентов

При GM1-GM5:

$$\hat{\boldsymbol{\beta}} \text{ и } s^2 \text{ независимы}$$

$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}$ — линейная функция от $\boldsymbol{\varepsilon}$.
$\text{RSS} = \boldsymbol{\varepsilon}^\top\mathbf{M}\boldsymbol{\varepsilon}$, где $\hat{\boldsymbol{\varepsilon}} = \mathbf{M}\boldsymbol{\varepsilon}$.
Ковариация:

$$\text{Cov}\left((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}, \mathbf{M}\boldsymbol{\varepsilon}\right) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{M} = \mathbf{0}$$

так как $\mathbf{X}^\top\mathbf{M} = \mathbf{X}^\top(\mathbf{I} - \mathbf{H}) = \mathbf{X}^\top - \mathbf{X}^\top = \mathbf{0}$.
При нормальности некоррелированность влечёт независимость. ■

34 Условная оценка наименьших квадратов

Условная оценка наименьших квадратов (Restricted Least Squares, RLS) используется при наличии ограничений на параметры.

34.1 Постановка задачи

Рассмотрим модель $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ с линейными ограничениями:

$$\boxed{\mathbf{R}\boldsymbol{\beta} = \mathbf{r}}$$

где $\mathbf{R}$ — матрица размера $q \times (p+1)$, $\mathbf{r}$ — вектор размера $q \times 1$.
Предполагаем: $\text{rank}(\mathbf{R}) = q \leq p + 1$.

Примеры ограничений:

- $\beta_1 = 0$ (исключение переменной): $\mathbf{R} = (0, 1, 0, \ldots, 0)$, $r = 0$
- $\beta_1 = \beta_2$ (равенство коэффициентов): $\mathbf{R} = (0, 1, -1, 0, \ldots)$, $r = 0$
- $\beta_1 + \beta_2 = 1$ (нормировка): $\mathbf{R} = (0, 1, 1, 0, \ldots)$, $r = 1$

34.2 Метод множителей Лагранжа

$$\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 \quad \text{при условии} \quad \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

Лагранжиан:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}^\top (\mathbf{R}\boldsymbol{\beta} - \mathbf{r})$$

Условия первого порядка:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top \mathbf{Y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} + 2\mathbf{R}^\top \boldsymbol{\lambda} = \mathbf{0}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} = 2(\mathbf{R}\boldsymbol{\beta} - \mathbf{r}) = \mathbf{0}$$

Условная оценка наименьших квадратов:

$$\boxed{\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \left[\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}$$

где $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ — безусловная МНК-оценка.

Из первого условия:

$$\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = \mathbf{X}^\top \mathbf{Y} - \mathbf{R}^\top \boldsymbol{\lambda}$$

$$\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \boldsymbol{\lambda} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \boldsymbol{\lambda}$$

Подставляем в ограничение $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$:

$$\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \boldsymbol{\lambda} = \mathbf{r}$$

$$\boldsymbol{\lambda} = \left[\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

Подставляем обратно:

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \left[\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

■

34.3 Свойства условной оценки

Если $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ выполнено (ограничения верны):

$$\mathbb{E}[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta}$$

$\mathbb{E}[\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}] = \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ при истинных ограничениях.
Следовательно:

$$\mathbb{E}[\tilde{\boldsymbol{\beta}}] = \mathbb{E}[\hat{\boldsymbol{\beta}}] - \mathbf{0} = \boldsymbol{\beta}$$

■

При истинности ограничений:

$$\operatorname{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2 \left[(\mathbf{X}^\top\mathbf{X})^{-1} - (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}^\top\mathbf{W}^{-1}\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\right]$$

где $\mathbf{W} = \mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}^\top$.

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) - \operatorname{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}^\top\mathbf{W}^{-1}\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1} \geq \mathbf{0}$$

Условная оценка имеет меньшую (или равную) дисперсию при истинных ограничениях.

34.4 Остаточная сумма квадратов

$$\mathrm{RSS}_R = \mathrm{RSS} + (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^\top\mathbf{W}^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})$$

где RSS — остаточная сумма квадратов без ограничений, RSS_R — с ограничениями.
Всегда: $\mathrm{RSS}_R \geq \mathrm{RSS}$.

34.5 F-тест для проверки ограничений

Для проверки H_0: $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ против H_1: $\mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$:

$$\boxed{F = \frac{(\mathrm{RSS}_R - \mathrm{RSS})/q}{\mathrm{RSS}/(n-p-1)} \sim F(q, n-p-1)}$$

при H_0 и GM1-GM5.

Альтернативная форма:

$$F = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^\top[\mathbf{R}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{R}^\top]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})/q}{s^2}$$

35 Основная теорема о линейной регрессии. Следствия. t-тест

Основная теорема объединяет результаты о распределениях в линейной регрессии и даёт основу для статистического вывода.

35.1 Основная теорема

При выполнении условий GM1-GM5 (нормальная линейная модель):

1. $\hat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\right)$
2. $\dfrac{\text{RSS}}{\sigma^2} = \dfrac{(n-p-1)s^2}{\sigma^2} \sim \chi^2(n-p-1)$
3. $\hat{\boldsymbol{\beta}}$ и s^2 независимы

Пункт 1: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ — линейная функция от $\mathbf{Y}$.
При $\mathbf{Y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$:

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}\left((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \cdot \mathbf{X}\boldsymbol{\beta}, (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\right)$$
$$= \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\right)$$

Пункты 2-3: Доказаны ранее (раздел об оценке остаточной дисперсии). ■

35.2 Следствия основной теоремы

Для каждого $j = 0, 1, \ldots, p$:
$$\hat{\beta}_j \sim \mathcal{N}\left(\beta_j, \sigma^2[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}\right)$$

$$\boxed{\frac{\hat{\beta}_j - \beta_j}{s \cdot \sqrt{[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}}} \sim t(n-p-1)}$$

Обозначим $c_{jj} = [(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}$.

$$\frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{c_{jj}}} \sim \mathcal{N}(0,1)$$
$$\frac{(n-p-1)s^2}{\sigma^2} \sim \chi^2(n-p-1)$$

$\hat{\beta}_j$ и s^2 независимы. По определению t-распределения:

$$\frac{\frac{\hat{\beta}_j-\beta_j}{\sigma\sqrt{c_{jj}}}}{\sqrt{\frac{(n-p-1)s^2}{\sigma^2(n-p-1)}}} = \frac{\hat{\beta}_j - \beta_j}{s\sqrt{c_{jj}}} \sim t(n-p-1)$$

■

$$\mathrm{SE}(\hat{\beta}_j) = s \cdot \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{jj}}$$

35.3 t-тест для коэффициентов регрессии

35.3.1 Тест на значимость коэффициента

Гипотезы:

- H_0: $\beta_j = 0$ (переменная x_j не влияет на Y)
- H_1: $\beta_j \neq 0$ (переменная x_j значима)

Статистика:

$$\boxed{t_j = \frac{\hat{\beta}_j}{\mathrm{SE}(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{s\sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{jj}}} \sim t(n-p-1)}$$

при H_0.

Критическая область:

- Двусторонний тест: отвергаем H_0, если $|t_j| > t_{1-\alpha/2}(n-p-1)$
- Односторонний тест (H_1: $\beta_j > 0$): отвергаем, если $t_j > t_{1-\alpha}(n-p-1)$

35.3.2 Тест на заданное значение

Гипотезы:

- H_0: $\beta_j = \beta_j^0$
- H_1: $\beta_j \neq \beta_j^0$

Статистика:

$$t = \frac{\hat{\beta}_j - \beta_j^0}{\mathrm{SE}(\hat{\beta}_j)} \sim t(n-p-1)$$

35.4 Доверительный интервал для коэффициента

$(1-\alpha)$-доверительный интервал для β_j:

$$\boxed{\hat{\beta}_j \pm t_{1-\alpha/2}(n-p-1) \cdot \mathrm{SE}(\hat{\beta}_j)}$$

$$P\left(-t_{1-\alpha/2} < \frac{\hat{\beta}_j - \beta_j}{\mathrm{SE}(\hat{\beta}_j)} < t_{1-\alpha/2}\right) = 1 - \alpha$$

Преобразуя:

$$P\left(\hat{\beta}_j - t_{1-\alpha/2} \cdot \text{SE}(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + t_{1-\alpha/2} \cdot \text{SE}(\hat{\beta}_j)\right) = 1 - \alpha$$

■

35.5 Простая линейная регрессия: частные случаи

Для модели $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$:

$$\text{SE}(\hat{\beta}_1) = \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{s}{\sqrt{S_{xx}}}$$

$$\text{SE}(\hat{\beta}_0) = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

t-статистика для наклона:

$$t = \frac{\hat{\beta}_1}{s/\sqrt{S_{xx}}} = \frac{\hat{\beta}_1\sqrt{S_{xx}}}{s}$$

35.6 F-тест для общей значимости регрессии

Гипотезы:

- H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ (все коэффициенты, кроме свободного члена, равны нулю)
- H_1: хотя бы один $\beta_j \neq 0$

$$\boxed{F = \frac{(\text{TSS} - \text{RSS})/p}{\text{RSS}/(n-p-1)} = \frac{\text{ESS}/p}{\text{RSS}/(n-p-1)} \sim F(p, n-p-1)}$$

где:

- $\text{TSS} = \sum(Y_i - \bar{Y})^2$ — полная сумма квадратов
- $\text{ESS} = \sum(\hat{Y}_i - \bar{Y})^2$ — объяснённая сумма квадратов
- $\text{RSS} = \sum(Y_i - \hat{Y}_i)^2$ — остаточная сумма квадратов

$\text{TSS} = \text{ESS} + \text{RSS}$ (разложение дисперсии).

35.7 Коэффициент детерминации

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

— доля объяснённой вариации.

1. $0 \leq R^2 \leq 1$
2. $R^2 = r^2_{Y\hat{Y}}$ — квадрат корреляции между Y и $\hat{Y}$
3. В простой регрессии: $R^2 = r^2_{XY}$

$$\bar{R}^2 = 1 - \frac{\text{RSS}/(n-p-1)}{\text{TSS}/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Учитывает число параметров; может быть отрицательным.

Связь с F-статистикой:

$$F = \frac{R^2/p}{(1-R^2)/(n-p-1)}$$

35.8 Сводная таблица распределений

Статистика	**Распределение при H_0**
$\hat{\boldsymbol{\beta}}$	$\mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1})$
$\dfrac{\hat{\beta}_j - \beta_j}{\text{SE}(\hat{\beta}_j)}$	$t(n-p-1)$
$\dfrac{(n-p-1)s^2}{\sigma^2}$	$\chi^2(n-p-1)$
$F = \dfrac{\text{ESS}/p}{\text{RSS}/(n-p-1)}$	$F(p, n-p-1)$

36 F-тест. Коэффициент детерминации

F-тест — универсальный инструмент для сравнения вложенных моделей и проверки гипотез о группах параметров. *Коэффициент детерминации* измеряет качество подгонки модели.

36.1 F-тест: общая теория

Модель M_0 (ограниченная) называется **вложенной** в модель M_1 (полную), если M_0 получается из M_1 наложением линейных ограничений на параметры.

Полная модель: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$
Ограниченная модель ($\beta_2 = \beta_3 = 0$): $Y = \beta_0 + \beta_1 x_1 + \varepsilon$

36.2 F-статистика для сравнения моделей

Пусть RSS_R — остаточная сумма квадратов ограниченной модели (с q ограничениями), RSS — полной модели. При H_0 (ограничения верны):

$$\boxed{F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(n-p-1)} \sim F(q, n-p-1)}$$

При H_0:

- $\dfrac{\text{RSS}}{\sigma^2} \sim \chi^2(n-p-1)$
- $\dfrac{\text{RSS}_R - \text{RSS}}{\sigma^2} \sim \chi^2(q)$ (разность независимых χ^2)
- Эти величины независимы

По определению F-распределения:

$$F = \frac{(\text{RSS}_R - \text{RSS})/q}{\text{RSS}/(n-p-1)} = \frac{\chi^2(q)/q}{\chi^2(n-p-1)/(n-p-1)} \sim F(q, n-p-1)$$

■

Критерий: Отвергаем H_0, если $F > F_{1-\alpha}(q, n-p-1)$.

36.3 Частные случаи F-теста

36.3.1 Тест на общую значимость регрессии

H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$
Ограниченная модель: $Y = \beta_0 + \varepsilon \Rightarrow \text{RSS}_R = \text{TSS}$

$$F = \frac{(\text{TSS} - \text{RSS})/p}{\text{RSS}/(n-p-1)} = \frac{\text{ESS}/p}{\text{RSS}/(n-p-1)} \sim F(p, n-p-1)$$

36.3.2 Тест на значимость группы переменных

H_0: $\beta_{k+1} = \beta_{k+2} = \ldots = \beta_p = 0$
Ограниченная модель включает только $x_1, \ldots, x_k$.

$$F = \frac{(\text{RSS}_k - \text{RSS}_{full})/(p-k)}{\text{RSS}_{full}/(n-p-1)} \sim F(p-k, n-p-1)$$

36.3.3 Тест на один коэффициент

H_0: $\beta_j = 0$ (частный случай $q = 1$)

$$F = \frac{(\text{RSS}_{-j} - \text{RSS})/1}{\text{RSS}/(n-p-1)} = t_j^2$$

где t_j — t-статистика для β_j.

$F(1, \nu) = t^2(\nu)$, то есть F-тест с одним ограничением эквивалентен двустороннему t-тесту.

36.4 Коэффициент детерминации R^2

$$\boxed{\text{TSS} = \text{ESS} + \text{RSS}}$$

где:

- $\text{TSS} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$ — полная сумма квадратов (Total)
- $\text{ESS} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$ — объяснённая сумма (Explained/Regression)
- $\text{RSS} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ — остаточная сумма (Residual)

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

36.4.1 Интерпретация R^2

- R^2 — доля вариации Y, объяснённая моделью
- $R^2 = 0$: модель не объясняет вариацию (предсказание = $\bar{Y}$)
- $R^2 = 1$: идеальная подгонка (все точки на линии регрессии)

$$F = \frac{R^2/p}{(1-R^2)/(n-p-1)} = \frac{R^2(n-p-1)}{(1-R^2)p}$$

36.4.2 Проблемы с R^2

1. **Монотонность:** R^2 не убывает при добавлении переменных
2. **Переобучение:** высокий R^2 не гарантирует хорошие предсказания
3. **Несравнимость:** R^2 нельзя сравнивать для разных зависимых переменных

36.5 Скорректированный R^2

$$\boxed{\bar{R}^2 = 1 - \frac{\text{RSS}/(n-p-1)}{\text{TSS}/(n-1)} = 1 - \frac{n-1}{n-p-1}(1-R^2)}$$

Свойства:

- Штрафует за добавление незначимых переменных
- Может быть отрицательным
- $\bar{R}^2 < R^2$ всегда (при $p \geq 1$)
- $\bar{R}^2$ увеличивается при добавлении переменной $\Leftrightarrow |t_j| > 1$

36.6 Другие меры качества

AIC (Akaike): $\text{AIC} = n\ln(\text{RSS}/n) + 2(p+1)$
BIC (Bayesian): $\text{BIC} = n\ln(\text{RSS}/n) + (p+1)\ln n$
Меньшие значения предпочтительнее.

37 Однофакторный дисперсионный анализ

Дисперсионный анализ (ANOVA — Analysis of Variance) — метод сравнения средних нескольких групп путём анализа дисперсий.

37.1 Постановка задачи

Имеется k групп (уровней фактора). В i-й группе n_i наблюдений:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i$$

или эквивалентно:

$$\boxed{Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}}$$

где:

- μ — общее среднее
- α_i — эффект i-го уровня фактора
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ — независимые ошибки

Ограничение идентифицируемости: $\sum_{i=1}^{k} n_i \alpha_i = 0$ или $\alpha_1 = 0$ (базовая группа).

Гипотезы:

- H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$ (или $\alpha_1 = \ldots = \alpha_k = 0$)
- H_1: хотя бы два средних различаются

37.2 Обозначения

- $N = \sum_{i=1}^{k} n_i$ — общий объём выборки
- $\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ — среднее i-й группы
- $\bar{Y}_{\cdot\cdot} = \frac{1}{N} \sum_{i,j} Y_{ij}$ — общее среднее

37.3 Разложение суммы квадратов

$$\boxed{\underbrace{\sum_{i,j}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{k} n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}_{\text{SSB}} + \underbrace{\sum_{i,j}(Y_{ij} - \bar{Y}_{i\cdot})^2}_{\text{SSW}}}$$

где:

- SST — полная сумма квадратов (Total)
- SSB — межгрупповая сумма (Between groups)
- SSW — внутригрупповая сумма (Within groups)

$$Y_{ij} - \bar{Y}_{\cdot\cdot} = (Y_{ij} - \bar{Y}_{i\cdot}) + (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})$$

Возводим в квадрат и суммируем. Перекрёстные члены обращаются в ноль:

$$\sum_j (Y_{ij} - \bar{Y}_{i\cdot})(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) = (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) \underbrace{\sum_j (Y_{ij} - \bar{Y}_{i\cdot})}_{=0} = 0$$

■

37.4 Степени свободы

Источник	Сумма квадратов	Степени свободы
Между группами	SSB	$k-1$
Внутри групп	SSW	$N-k$
Полная	SST	$N-1$

37.5 Средние квадраты и F-статистика

$$\text{MSB} = \frac{\text{SSB}}{k-1}, \quad \text{MSW} = \frac{\text{SSW}}{N-k}$$

При любой гипотезе:

$$\mathbb{E}[\text{MSW}] = \sigma^2$$

При H_0:

$$\mathbb{E}[\text{MSB}] = \sigma^2$$

При H_1:

$$\mathbb{E}[\text{MSB}] = \sigma^2 + \frac{\sum_i n_i \alpha_i^2}{k-1} > \sigma^2$$

При H_0:

$$\boxed{F = \frac{\text{MSB}}{\text{MSW}} = \frac{\text{SSB}/(k-1)}{\text{SSW}/(N-k)} \sim F(k-1, N-k)}$$

При H_0:

- $\frac{\text{SSB}}{\sigma^2} \sim \chi^2(k-1)$

- $\dfrac{\text{SSW}}{\sigma^2} \sim \chi^2(N-k)$
- SSB и SSW независимы

По определению F-распределения получаем результат. ■

Критерий: Отвергаем H_0, если $F > F_{1-\alpha}(k-1, N-k)$.

37.6 Таблица ANOVA

Источник	**SS**	**df**	**MS**	**F**
Между группами	SSB	$k-1$	MSB	$\dfrac{\text{MSB}}{\text{MSW}}$
Внутри групп	SSW	$N-k$	MSW	
Итого	SST	$N-1$		

37.7 Связь с регрессией

ANOVA эквивалентна регрессии с фиктивными переменными:

$$Y_{ij} = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \ldots + \beta_{k-1} D_{k-1} + \varepsilon_{ij}$$

где $D_i = 1$ для i-й группы, 0 иначе (первая группа — базовая).

F-тест ANOVA = F-тест на H_0: $\beta_1 = \ldots = \beta_{k-1} = 0$.

37.8 Множественные сравнения

После отвержения H_0 необходимо определить, какие группы различаются.

При m сравнениях использовать уровень α/m для каждого.

Критическое значение для разности средних:

$$|\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}| > q_\alpha(k, N-k) \cdot \sqrt{\frac{\text{MSW}}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

где q_α — квантиль стьюдентизированного размаха.

38 Двухфакторный дисперсионный анализ

Двухфакторный дисперсионный анализ исследует влияние двух категориальных факторов и их взаимодействие.

38.1 Модель без взаимодействия

Фактор A имеет a уровней, фактор B — b уровней:

$$\boxed{Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}}$$

где:

- μ — общее среднее
- α_i — эффект i-го уровня фактора A ($i = 1, \ldots, a$)
- β_j — эффект j-го уровня фактора B ($j = 1, \ldots, b$)
- $k = 1, \ldots, n_{ij}$ — номер наблюдения в ячейке (i, j)

Ограничения: $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$.

38.2 Модель с взаимодействием

$$\boxed{Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}}$$

где $(\alpha\beta)_{ij}$ — эффект взаимодействия факторов A и B.
Ограничения: $\sum_i (\alpha\beta)_{ij} = 0$ для всех j, $\sum_j (\alpha\beta)_{ij} = 0$ для всех i.

Взаимодействие означает, что эффект одного фактора зависит от уровня другого.

38.3 Разложение суммы квадратов

Для сбалансированного плана ($n_{ij} = n$ для всех i, j):

$$\boxed{\text{SST} = \text{SS}_A + \text{SS}_B + \text{SS}_{AB} + \text{SSE}}$$

где:

- $\text{SS}_A = bn \sum_i (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$ — сумма квадратов фактора A
- $\text{SS}_B = an \sum_j (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$ — сумма квадратов фактора B
- $\text{SS}_{AB} = n \sum_{i,j} (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$ — взаимодействие

- $\text{SSE} = \sum_{i,j,k}(Y_{ijk} - \bar{Y}_{ij\cdot})^2$ — остаточная (ошибка)

38.4 Степени свободы

Источник	**SS**	**df**
Фактор A	SS_A	$a-1$
Фактор B	SS_B	$b-1$
Взаимодействие AB	SS_{AB}	$(a-1)(b-1)$
Ошибка	SSE	$ab(n-1)$
Итого	SST	$abn-1$

38.5 F-тесты

При соответствующих H_0:

$$F_A = \frac{\text{SS}_A/(a-1)}{\text{SSE}/(ab(n-1))} \sim F(a-1, ab(n-1))$$

$$F_B = \frac{\text{SS}_B/(b-1)}{\text{SSE}/(ab(n-1))} \sim F(b-1, ab(n-1))$$

$$F_{AB} = \frac{\text{SS}_{AB}/((a-1)(b-1))}{\text{SSE}/(ab(n-1))} \sim F((a-1)(b-1), ab(n-1))$$

38.6 Таблица двухфакторного ANOVA

Источник	**SS**	**df**	**MS**	**F**
Фактор A	SS_A	$a-1$	MS_A	F_A
Фактор B	SS_B	$b-1$	MS_B	F_B
A × B	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	F_{AB}
Ошибка	SSE	$ab(n-1)$	MSE	
Итого	SST	$abn-1$		

38.7 Интерпретация взаимодействия

- Если F_{AB} значим: эффект A зависит от уровня B (и наоборот)
- Графически: непараллельные линии на графике взаимодействия
- При значимом взаимодействии главные эффекты интерпретировать осторожно

38.8 Случай без повторений ($n = 1$)

При одном наблюдении на ячейку:

- Невозможно оценить взаимодействие отдельно от ошибки
- Используется аддитивная модель
- $\text{SSE} = \text{SS}_{AB}$ служит оценкой ошибки
- df ошибки = $(a-1)(b-1)$

39 Ковариационный анализ

Ковариационный анализ (ANCOVA — Analysis of Covariance) объединяет ANOVA и регрессию, позволяя учитывать влияние непрерывных ковариат при сравнении групп.

39.1 Постановка задачи

$$\boxed{Y_{ij} = \mu + \alpha_i + \beta(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij}}$$

где:

- μ — общее среднее
- α_i — эффект i-й группы (фактор)
- X_{ij} — ковариата (непрерывная переменная)
- β — коэффициент регрессии (общий для всех групп)
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$

39.2 Цели ANCOVA

1. **Контроль** влияния мешающей переменной (ковариаты)
2. **Повышение мощности** за счёт уменьшения остаточной дисперсии
3. **Корректировка** групповых средних на различия в ковариате

Сравнение эффективности методов обучения (группы) с учётом начального уровня знаний (ковариата).

39.3 Предположения ANCOVA

1. Линейная связь между Y и X
2. **Гомогенность наклонов:** β одинаков для всех групп
3. Независимость ковариаты от фактора (в идеале)
4. Нормальность и гомоскедастичность ошибок

39.4 Оценивание параметров

Модель является частным случаем линейной регрессии. МНК-оценки:

$$\hat{\beta} = \frac{\sum_{i,j}(X_{ij} - \bar{X}_{i\cdot})(Y_{ij} - \bar{Y}_{i\cdot})}{\sum_{i,j}(X_{ij} - \bar{X}_{i\cdot})^2} = \frac{SP_W}{SS_{X,W}}$$

Скорректированные групповые средние:

$$\bar{Y}_{i\cdot}^{adj} = \bar{Y}_{i\cdot} - \hat{\beta}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})$$

39.5 Разложение сумм квадратов

$$\mathrm{SST}_{adj} = \mathrm{SS}_{groups,adj} + \mathrm{SSE}_{adj}$$

Скорректированная остаточная сумма:

$$\mathrm{SSE}_{adj} = \mathrm{SSE}_Y - \frac{(SP_W)^2}{SS_{X,W}}$$

где SSE_Y — остаточная сумма квадратов ANOVA без ковариаты.

39.6 F-тест ANCOVA

Для проверки H_0: $\alpha_1 = \ldots = \alpha_k = 0$ (нет различий между группами с учётом ковариаты):

$$\boxed{F = \frac{\mathrm{SS}_{groups,adj}/(k-1)}{\mathrm{SSE}_{adj}/(N-k-1)} \sim F(k-1, N-k-1)}$$

Обратите внимание на изменение степеней свободы: $N - k - 1$ вместо $N - k$ (потеря одной степени свободы на ковариату).

39.7 Проверка гомогенности наклонов

Расширенная модель с разными наклонами:

$$Y_{ij} = \mu + \alpha_i + \beta_i(X_{ij} - \bar{X}_{\cdot\cdot}) + \varepsilon_{ij}$$

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k$ (гомогенность)
F-тест сравнивает модели с общим и разными наклонами.

39.8 Таблица ANCOVA

Источник	**SS**	**df**
Группы (скорр.)	$\mathrm{SS}_{groups,adj}$	$k-1$
Ковариата	SS_{cov}	1
Ошибка	SSE_{adj}	$N-k-1$
Итого	SST	$N-1$

39.9 Преимущества ANCOVA

1. Уменьшение MSE повышает мощность
2. Статистический контроль мешающих переменных
3. Более точные оценки групповых различий

40 Логистическая регрессия. Регрессия Пуассона

Обобщённые линейные модели (GLM) расширяют классическую регрессию на случаи, когда отклик не является нормально распределённым.

40.1 Обобщённые линейные модели: структура

1. **Случайный компонент:** $Y \sim$ распределение из экспоненциального семейства
2. **Систематический компонент:** $\eta = \mathbf{x}^\top \boldsymbol{\beta}$ — линейный предиктор
3. **Функция связи:** $g(\mu) = \eta$, где $\mu = \mathbb{E}[Y]$

40.2 Логистическая регрессия

Для бинарного отклика $Y \in \{0, 1\}$:

$$Y_i \sim \text{Bernoulli}(p_i)$$

$$\boxed{\text{logit}(p_i) = \ln \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} = \mathbf{x}_i^\top \boldsymbol{\beta}}$$

40.2.1 Связь с вероятностью

Обратная функция связи:

$$\boxed{p_i = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}} = \frac{1}{1 + e^{-\mathbf{x}_i^\top \boldsymbol{\beta}}}}$$

Это **логистическая (сигмоидная) функция**.

40.2.2 Интерпретация коэффициентов

$$\frac{p_i}{1 - p_i} = e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$$

При увеличении x_j на единицу (при прочих равных):

$$\text{Odds Ratio} = e^{\beta_j}$$

Шансы успеха умножаются на e^{β_j}.

$\beta_j = 0.5 \Rightarrow \text{OR} = e^{0.5} \approx 1.65$: шансы увеличиваются на 65%.
$\beta_j = -0.3 \Rightarrow \text{OR} = e^{-0.3} \approx 0.74$: шансы уменьшаются на 26%.

40.2.3 Оценивание: метод максимального правдоподобия

Функция правдоподобия:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} p_i^{Y_i}(1-p_i)^{1-Y_i}$$

Логарифм правдоподобия:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[Y_i \mathbf{x}_i^\top \boldsymbol{\beta} - \ln(1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}) \right]$$

Максимизируется численно (метод Ньютона-Рафсона / IRLS).

40.2.4 Асимптотические свойства МПО

При $n \to \infty$:

$$\hat{\boldsymbol{\beta}} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{\beta}, \mathbf{I}^{-1}(\boldsymbol{\beta})\right)$$

где $\mathbf{I}(\boldsymbol{\beta}) = \mathbf{X}^\top \mathbf{W} \mathbf{X}$ — информация Фишера, $\mathbf{W} = \text{diag}(p_i(1-p_i))$.

40.2.5 Тесты в логистической регрессии

Тест Вальда:

$$z_j = \frac{\hat{\beta}_j}{\text{SE}(\hat{\beta}_j)} \approx \mathcal{N}(0,1)$$

Тест отношения правдоподобия:

$$\Lambda = -2(\ell_0 - \ell_1) \xrightarrow{d} \chi^2(q)$$

где q — число ограничений.

Девианс:

$$D = -2\ell(\hat{\boldsymbol{\beta}}) + 2\ell_{saturated}$$

40.2.6 Меры качества

$$R^2_{McF} = 1 - \frac{\ell(\hat{\boldsymbol{\beta}})}{\ell_0}$$

где ℓ_0 — логарифм правдоподобия модели только с константой.

40.3 Регрессия Пуассона

Для счётного отклика $Y \in \{0, 1, 2, \ldots\}$:

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\boxed{\ln(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} = \mathbf{x}_i^\top \boldsymbol{\beta}}$$

Функция связи — логарифм.

40.3.1 Связь с интенсивностью

$$\lambda_i = \mathbb{E}[Y_i] = e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$$

40.3.2 Интерпретация коэффициентов

При увеличении x_j на единицу:

$$\frac{\lambda_{new}}{\lambda_{old}} = e^{\beta_j}$$

Ожидаемое число событий умножается на e^{β_j}.

$\beta_j = 0.2 \Rightarrow e^{0.2} \approx 1.22$: ожидаемое число увеличивается на 22%.

40.3.3 Оценивание МПО

Функция правдоподобия:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\lambda_i^{Y_i} e^{-\lambda_i}}{Y_i!}$$

Логарифм правдоподобия:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[Y_i \mathbf{x}_i^\top \boldsymbol{\beta} - e^{\mathbf{x}_i^\top \boldsymbol{\beta}} - \ln(Y_i!) \right]$$

40.3.4 Сверхдисперсия

В модели Пуассона: $\mathbb{E}[Y] = \text{Var}(Y) = \lambda$.
Сверхдисперсия — ситуация, когда $\text{Var}(Y) > \mathbb{E}[Y]$.

Решения:

- Квази-пуассоновская регрессия: $\text{Var}(Y) = \phi\lambda$
- Отрицательная биномиальная регрессия

40.3.5 Смещение (Offset)

Для моделирования интенсивности с учётом экспозиции t_i:

$$\ln(\lambda_i) = \ln(t_i) + \mathbf{x}_i^\top \boldsymbol{\beta}$$

где $\ln(t_i)$ — **смещение** (offset), известная величина.

40.4 Сравнение GLM

	Линейная	**Логистическая**	**Пуассона**
Отклик	Непрерывный	Бинарный	Счётный
Распределение	Нормальное	Бернулли	Пуассон
Функция связи	Тождественная	Логит	Логарифм
$\mathbb{E}[Y]$	$\mathbf{x}^\top \boldsymbol{\beta}$	$\dfrac{e^\eta}{1+e^\eta}$	e^η
Интерпретация β	Разность	$\ln(\text{OR})$	$\ln(\text{RR})$
Оценивание	МНК	МПО	МПО

41 Критерий отношения правдоподобий для простых гипотез. Лемма Неймана-Пирсона

Лемма Неймана-Пирсона — фундаментальный результат теории проверки гипотез, определяющий наиболее мощный критерий для простых гипотез.

41.1 Постановка задачи

Гипотеза называется **простой**, если она полностью определяет распределение:

- H_0: $X \sim f_0(x)$ (или P_0)
- H_1: $X \sim f_1(x)$ (или P_1)

Для наблюдений $X_1, \ldots, X_n$ **отношение правдоподобия**:

$$\boxed{\Lambda(X_1, \ldots, X_n) = \frac{L_1(X_1, \ldots, X_n)}{L_0(X_1, \ldots, X_n)} = \frac{\prod_{i=1}^n f_1(X_i)}{\prod_{i=1}^n f_0(X_i)}}$$

41.2 Непрерывный случай

Пусть $f_0(x)$ и $f_1(x)$ — плотности при H_0 и H_1 соответственно.

Критерий отношения правдоподобия (LRT — Likelihood Ratio Test):
Отвергаем H_0, если $\Lambda(\mathbf{X}) > c$, где c выбирается из условия $P_{H_0}(\Lambda > c) = \alpha$.

41.3 Лемма Неймана-Пирсона

Пусть H_0: $X \sim f_0$ и H_1: $X \sim f_1$ — простые гипотезы. Для заданного уровня значимости $\alpha \in (0, 1)$ рассмотрим критерий:

$$\boxed{\phi(x) = \begin{cases} 1, & \text{если } \Lambda(x) > c \\ \gamma, & \text{если } \Lambda(x) = c \\ 0, & \text{если } \Lambda(x) < c \end{cases}}$$

где константы $c \geq 0$ и $\gamma \in [0, 1]$ выбираются так, что:

$$\mathbb{E}_{H_0}[\phi(X)] = \alpha$$

Тогда:

1. Критерий ϕ имеет уровень α

2. Критерий ϕ является **наиболее мощным** (НМ) среди всех критериев уровня α

То есть для любого другого критерия ψ с $\mathbb{E}_{H_0}[\psi] \leq \alpha$:

$$\mathbb{E}_{H_1}[\phi] \geq \mathbb{E}_{H_1}[\psi]$$

(мощность ϕ не меньше мощности ψ).

Обозначим мощности: $\beta_\phi = \mathbb{E}_{H_1}[\phi]$, $\beta_\psi = \mathbb{E}_{H_1}[\psi]$.
Покажем, что $\beta_\phi \geq \beta_\psi$.
Шаг 1: Из определения ϕ:

- При $\Lambda(x) > c$: $\phi(x) = 1 \geq \psi(x)$, поэтому $(\phi - \psi)(f_1 - cf_0) \geq 0$

- При $\Lambda(x) < c$: $\phi(x) = 0 \le \psi(x)$, и $f_1 < cf_0$, поэтому $(\phi - \psi)(f_1 - cf_0) \ge 0$
- При $\Lambda(x) = c$: $f_1 = cf_0$, поэтому $(\phi - \psi)(f_1 - cf_0) = 0$

Шаг 2: Интегрируем:

$$\int (\phi(x) - \psi(x))(f_1(x) - cf_0(x))\, dx \ge 0$$

Шаг 3: Раскрываем:

$$\int \phi f_1\, dx - \int \psi f_1\, dx - c\int \phi f_0\, dx + c\int \psi f_0\, dx \ge 0$$
$$\beta_\phi - \beta_\psi - c\alpha + c\mathbb{E}_{H_0}[\psi] \ge 0$$

Шаг 4: Так как $\mathbb{E}_{H_0}[\psi] \le \alpha$:

$$\beta_\phi - \beta_\psi \ge c(\alpha - \mathbb{E}_{H_0}[\psi]) \ge 0$$

Следовательно: $\beta_\phi \ge \beta_\psi$. ■

41.4 Следствия леммы Неймана-Пирсона

Если критерий ψ имеет ту же мощность, что и ϕ, то $\psi = \phi$ почти всюду относительно $P_0 + P_1$.

Рандомизация ($\gamma \in (0,1)$) нужна только если $P_{H_0}(\Lambda = c) > 0$, что происходит в дискретном случае.

Мощность НМ-критерия не убывает по уровню α.

41.5 Примеры применения

H_0: $\mu = \mu_0$ vs H_1: $\mu = \mu_1 > \mu_0$.
$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 известна.

$$\Lambda = \frac{\prod \frac{1}{\sigma}\varphi\left(\frac{X_i - \mu_1}{\sigma}\right)}{\prod \frac{1}{\sigma}\varphi\left(\frac{X_i - \mu_0}{\sigma}\right)} = \exp\left\{\frac{(\mu_1 - \mu_0)}{\sigma^2}\sum X_i - \frac{n(\mu_1^2 - \mu_0^2)}{2\sigma^2}\right\}$$

$\Lambda > c$ эквивалентно $\bar{X} > c'$.
НМ-критерий: отвергать H_0, если $\bar{X} > \mu_0 + z_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}$.

H_0: $\lambda = \lambda_0$ vs H_1: $\lambda = \lambda_1 < \lambda_0$.
$\Lambda > c$ эквивалентно $\sum X_i > c'$.
НМ-критерий основан на статистике $\sum X_i$.

42 Критерий отношения правдоподобий для простых гипотез в дискретном случае

В дискретном случае необходимо учитывать особенности: отношение правдоподобия может принимать одно и то же значение с положительной вероятностью, что требует рандомизации.

42.1 Особенности дискретного случая

Пусть X принимает значения в счётном множестве $\mathcal{X}$:

- H_0: $P_0(X = x) = p_0(x)$
- H_1: $P_1(X = x) = p_1(x)$

Отношение правдоподобия:

$$\Lambda(x) = \frac{p_1(x)}{p_0(x)}$$

Проблема: Функция $\alpha(c) = P_{H_0}(\Lambda > c)$ — ступенчатая, поэтому для заданного α может не существовать c такого, что $P_{H_0}(\Lambda > c) = \alpha$ точно.

42.2 Рандомизированный критерий

$$\phi(x) = \begin{cases} 1, & \text{если } \Lambda(x) > c \\ \gamma, & \text{если } \Lambda(x) = c \\ 0, & \text{если } \Lambda(x) < c \end{cases}$$

где γ выбирается так, что:

$$P_{H_0}(\Lambda > c) + \gamma \cdot P_{H_0}(\Lambda = c) = \alpha$$

Интерпретация: При $\Lambda(x) = c$ отвергаем H_0 с вероятностью γ (бросаем монету).

42.3 Алгоритм построения НМ-критерия в дискретном случае

1. Вычислить $\Lambda(x)$ для всех $x \in \mathcal{X}$
2. Упорядочить значения: $\Lambda(x_{(1)}) > \Lambda(x_{(2)}) > \dots$
3. Найти c и γ:

- c — наибольшее значение такое, что $P_{H_0}(\Lambda > c) \leq \alpha$
- $\gamma = \dfrac{\alpha - P_{H_0}(\Lambda > c)}{P_{H_0}(\Lambda = c)}$

42.4 Пример: биномиальное распределение

$X \sim \text{Bin}(n, p)$.
H_0: $p = p_0$ vs H_1: $p = p_1 > p_0$.

$$\Lambda(k) = \frac{\binom{n}{k} p_1^k (1-p_1)^{n-k}}{\binom{n}{k} p_0^k (1-p_0)^{n-k}} = \left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)^k \cdot \left(\frac{1-p_1}{1-p_0}\right)^n$$

При $p_1 > p_0$: $\Lambda(k)$ возрастает по k.
$\Lambda > c$ эквивалентно $X > c'$.
НМ-критерий: отвергать H_0, если $X \geq k^*$, где k^* — наименьшее k такое, что $P_{H_0}(X \geq k) \leq \alpha$.
Рандомизация: при $X = k^* - 1$ отвергать с вероятностью γ.

42.5 Пример: распределение Пуассона

$X \sim \text{Poisson}(\lambda)$.
H_0: $\lambda = \lambda_0$ vs H_1: $\lambda = \lambda_1 > \lambda_0$.

$$\Lambda(k) = \frac{e^{-\lambda_1} \lambda_1^k / k!}{e^{-\lambda_0} \lambda_0^k / k!} = e^{\lambda_0 - \lambda_1} \left(\frac{\lambda_1}{\lambda_0}\right)^k$$

При $\lambda_1 > \lambda_0$: $\Lambda(k)$ возрастает по k.
НМ-критерий: отвергать при больших значениях X.

42.6 Практические замечания

- **Рандомизация на практике:** часто избегают, используя ближайший достижимый уровень
- **p-value:** минимальный уровень, при котором H_0 отвергается
- **Консервативный критерий:** не использовать рандомизацию, уровень $\leq \alpha$

43 Общий критерий отношения максимальных правдоподобий для сложных гипотез

Для сложных гипотез (когда параметр не полностью определён) используется обобщённый критерий отношения правдоподобий (GLRT).

43.1 Сложные гипотезы

- H_0: $\theta \in \Theta_0$ (подмножество параметрического пространства)
- H_1: $\theta \in \Theta_1 = \Theta \setminus \Theta_0$

H_0: $\mu \leq \mu_0$ vs H_1: $\mu > \mu_0$ — обе гипотезы сложные.
H_0: $\mu = \mu_0$ vs H_1: $\mu \neq \mu_0$ — H_0 простая, H_1 сложная.

43.2 Обобщённый критерий отношения правдоподобий (GLRT)

$$\boxed{\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta; X)}{\sup_{\theta \in \Theta} L(\theta; X)} = \frac{L(\hat{\theta}_0; X)}{L(\hat{\theta}; X)}}$$

где $\hat{\theta}_0$ — МПО при H_0, $\hat{\theta}$ — безусловный МПО.

Свойства:

- $0 \leq \Lambda \leq 1$
- Малые значения Λ свидетельствуют против H_0

Отвергаем H_0, если $\Lambda \leq c$, где c выбирается из условия:

$$\sup_{\theta \in \Theta_0} P_\theta(\Lambda \leq c) = \alpha$$

43.3 Асимптотическое распределение

При регулярных условиях, если H_0 верна и Θ_0 задаётся r независимыми ограничениями:

$$\boxed{-2 \ln \Lambda \xrightarrow{d} \chi^2(r)}$$

при $n \to \infty$.

При H_0: $\hat{\theta} - \hat{\theta}_0 = O_p(n^{-1/2})$.

Разложение в ряд Тейлора логарифма правдоподобия:

$$\ln L(\hat{\theta}) - \ln L(\hat{\theta}_0) \approx \frac{1}{2}(\hat{\theta} - \hat{\theta}_0)^\top I(\theta_0)(\hat{\theta} - \hat{\theta}_0)$$

Это квадратичная форма от асимптотически нормального вектора с r ограничениями, что даёт $\chi^2(r)$. ■

Критерий: Отвергаем H_0, если $-2\ln\Lambda > \chi^2_{1-\alpha}(r)$.

43.4 Примеры GLRT

43.4.1 Проверка значения среднего

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, оба параметра неизвестны.

H_0: $\mu = \mu_0$ vs H_1: $\mu \neq \mu_0$.

GLRT-статистика эквивалентна t-статистике:

$$-2\ln\Lambda = n\ln\left(1 + \frac{t^2}{n-1}\right)$$

где $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$.

При $n \to \infty$: $-2\ln\Lambda \xrightarrow{d} \chi^2(1)$.

43.4.2 Проверка равенства дисперсий

$X_1, \ldots, X_m \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y_1, \ldots, Y_n \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

H_0: $\sigma_1^2 = \sigma_2^2$ vs H_1: $\sigma_1^2 \neq \sigma_2^2$.

GLRT приводит к F-статистике: $F = s_1^2/s_2^2$.

43.4.3 Проверка в таблицах сопряжённости

H_0: независимость признаков.

$-2\ln\Lambda = 2\sum_{ij} O_{ij}\ln\dfrac{O_{ij}}{E_{ij}}$ — статистика G^2 (G-test).

При H_0: $G^2 \xrightarrow{d} \chi^2((r-1)(c-1))$.

43.5 Сравнение GLRT с другими критериями

Критерий	Статистика	Асимптотика
GLRT	$-2\ln\Lambda$	$\chi^2(r)$
Вальда	$(\hat{\theta} - \theta_0)^\top \hat{I}(\hat{\theta} - \theta_0)$	$\chi^2(r)$
Множ. Лагранжа (Score)	$S(\theta_0)^\top I^{-1}(\theta_0) S(\theta_0)$	$\chi^2(r)$

Все три асимптотически эквивалентны при H_0.

44 Последовательный анализ Вальда. Алгоритм. Теорема о конечном числе шагов

Последовательный анализ — метод, при котором решение о принятии/отвержении гипотезы принимается по мере поступления данных, а не после сбора фиксированной выборки.

44.1 Мотивация

- В классических тестах объём выборки n фиксирован заранее
- Последовательные тесты позволяют принять решение раньше
- В среднем требуется меньше наблюдений для достижения тех же ошибок

44.2 Постановка задачи

Для простых гипотез:

- H_0: $X \sim f_0$
- H_1: $X \sim f_1$

После каждого наблюдения X_n принимается одно из решений:

1. Принять H_0
2. Принять H_1
3. Продолжить наблюдения

44.3 Последовательный критерий отношения вероятностей (SPRT)

После n наблюдений вычисляется отношение правдоподобия:

$$\Lambda_n = \frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} = \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)}$$

Правило решения:

$$\boxed{\begin{cases} \text{Принять } H_0, & \text{если } \Lambda_n \le A \\ \text{Принять } H_1, & \text{если } \Lambda_n \ge B \\ \text{Продолжить}, & \text{если } A < \Lambda_n < B \end{cases}}$$

где $0 < A < 1 < B$ — граничные константы.

44.4 Логарифмическая форма

Удобнее работать с логарифмом:

$$S_n = \ln \Lambda_n = \sum_{i=1}^{n} \ln \frac{f_1(X_i)}{f_0(X_i)} = \sum_{i=1}^{n} Z_i$$

где $Z_i = \ln \dfrac{f_1(X_i)}{f_0(X_i)}$.

Правило:

$$\begin{cases} \text{Принять } H_0, & \text{если } S_n \le a = \ln A \\ \text{Принять } H_1, & \text{если } S_n \ge b = \ln B \\ \text{Продолжить}, & \text{если } a < S_n < b \end{cases}$$

44.5 Графическая интерпретация

S_n — случайное блуждание. Тест останавливается при первом выходе за границы $[a, b]$:

- Выход через нижнюю границу $a \Rightarrow$ принять H_0
- Выход через верхнюю границу $b \Rightarrow$ принять H_1

44.6 Теорема о конечном числе шагов

При условии, что $P_0 \neq P_1$ (распределения различны), последовательный критерий отношения правдоподобностей завершается с вероятностью 1:

$$P(\text{тест завершится за конечное число шагов}) = 1$$

независимо от того, какая гипотеза верна.

Случай 1: $\mathbb{E}[Z_i] \neq 0$ при истинной гипотезе.
По закону больших чисел: $S_n/n \to \mathbb{E}[Z_i] \neq 0$.
Следовательно, $|S_n| \to \infty$, и блуждание выйдет за границы.

Случай 2: Если $\mathbb{E}_{H_0}[Z_i] < 0$ (всегда верно при H_0, так как $\mathbb{E}_{H_0}[\ln \frac{f_1}{f_0}] = -D_{KL}(f_0 \| f_1) < 0$), то $S_n \to -\infty$ при H_0.
Аналогично, $\mathbb{E}_{H_1}[Z_i] = D_{KL}(f_1 \| f_0) > 0$, и $S_n \to +\infty$ при H_1.
В обоих случаях тест завершится. ■

Ключевое условие: $P_0 \neq P_1$, то есть гипотезы различимы. Если $f_0 = f_1$, то $Z_i = 0$, и тест никогда не завершится.

45 Последовательный анализ Вальда. Оценка констант и вероятностей ошибок. Среднее число итераций

45.1 Вероятности ошибок

- $\alpha = P_{H_0}(\text{принять } H_1)$ — ошибка I рода
- $\beta = P_{H_1}(\text{принять } H_0)$ — ошибка II рода

45.2 Теорема об оценке граничных констант

Для заданных вероятностей ошибок α и β граничные константы приближённо равны:

$$\boxed{A \approx \frac{\beta}{1-\alpha}, \quad B \approx \frac{1-\beta}{\alpha}}$$

или в логарифмической форме:

$$\boxed{a = \ln A \approx \ln\frac{\beta}{1-\alpha}, \quad b = \ln B \approx \ln\frac{1-\beta}{\alpha}}$$

Рассмотрим событие $\{\text{принять } H_1\} = \{\Lambda_N \geq B\}$ при остановке на шаге N.
При H_0:

$$\alpha = P_{H_0}(\Lambda_N \geq B) = \mathbb{E}_{H_0}[\mathbf{1}(\Lambda_N \geq B)]$$

При H_1:

$$1-\beta = P_{H_1}(\Lambda_N \geq B) = \mathbb{E}_{H_1}[\mathbf{1}(\Lambda_N \geq B)]$$

Используя замену меры: $dP_1 = \Lambda \cdot dP_0$:

$$1-\beta = \mathbb{E}_{H_0}[\Lambda_N \cdot \mathbf{1}(\Lambda_N \geq B)]$$

Приближённо (при выходе через верхнюю границу $\Lambda_N \approx B$):

$$1-\beta \approx B \cdot \alpha \Rightarrow B \approx \frac{1-\beta}{\alpha}$$

Аналогично для нижней границы (при выходе $\Lambda_N \approx A$):

$$\beta \approx (1-\alpha) \cdot A \Rightarrow A \approx \frac{\beta}{1-\alpha}$$

■

Приближения точны, когда блуждание выходит ровно на границу. На практике возможен “перескок” границы, и реальные ошибки могут отличаться. Формулы Вальда дают оценку сверху для α и β одновременно.

45.3 Уточнённые оценки ошибок

Для любого SPRT с границами A и B:

$$\alpha + \beta \leq 1$$

$$\alpha \leq \frac{1}{B}, \quad \beta \leq A$$

Более точно:

$$\frac{\alpha}{1-\beta} \leq \frac{1}{B}, \quad \frac{\beta}{1-\alpha} \leq A$$

45.4 Среднее число итераций (ASN)

$$\mathbb{E}[N] = \text{среднее число наблюдений до остановки}$$

Пусть $\mu = \mathbb{E}[Z_i]$ при истинной гипотезе, $\mu \neq 0$. Тогда:

$$\boxed{\mathbb{E}[N] \approx \frac{\mathbb{E}[S_N]}{\mu}}$$

где S_N — значение логарифма отношения правдоподобия при остановке.

При H_0 ($\mu_0 = \mathbb{E}_{H_0}[Z_i] < 0$):

$$\mathbb{E}_{H_0}[S_N] \approx (1-\alpha) \cdot a + \alpha \cdot b = (1-\alpha) \ln \frac{\beta}{1-\alpha} + \alpha \ln \frac{1-\beta}{\alpha}$$

При H_1 ($\mu_1 = \mathbb{E}_{H_1}[Z_i] > 0$):

$$\mathbb{E}_{H_1}[S_N] \approx \beta \cdot a + (1-\beta) \cdot b = \beta \ln \frac{\beta}{1-\alpha} + (1-\beta) \ln \frac{1-\beta}{\alpha}$$

$$\boxed{\mathbb{E}_{H_0}[N] \approx \frac{(1-\alpha) \ln \frac{\beta}{1-\alpha} + \alpha \ln \frac{1-\beta}{\alpha}}{\mathbb{E}_{H_0}[Z]}}$$

$$\boxed{\mathbb{E}_{H_1}[N] \approx \frac{\beta \ln \frac{\beta}{1-\alpha} + (1-\beta)\ln \frac{1-\beta}{\alpha}}{\mathbb{E}_{H_1}[Z]}}$$

45.5 Сравнение с фиксированным объёмом выборки

SPRT с заданными α и β требует в среднем примерно вдвое меньше наблюдений, чем наилучший фиксированный тест с теми же ошибками:

$$\mathbb{E}[N_{SPRT}] \approx \frac{1}{2} n_{fixed}$$

Точнее, экономия составляет от 30% до 70% в зависимости от параметров.

45.6 Пример: нормальное распределение

$X_i \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 известна.

H_0: $\mu = \mu_0$ vs H_1: $\mu = \mu_1 > \mu_0$.

$$Z_i = \ln \frac{f_1(X_i)}{f_0(X_i)} = \frac{(\mu_1 - \mu_0)}{\sigma^2}\left(X_i - \frac{\mu_0 + \mu_1}{2}\right)$$

$$\mathbb{E}_{H_0}[Z_i] = -\frac{(\mu_1-\mu_0)^2}{2\sigma^2}, \quad \mathbb{E}_{H_1}[Z_i] = \frac{(\mu_1-\mu_0)^2}{2\sigma^2}$$

46 Введение в байесовскую статистику. Credible intervals. Проверка гипотез

Байесовская статистика — альтернативный подход к статистическому выводу, основанный на теореме Байеса и субъективной интерпретации вероятности.

46.1 Основные принципы

- Параметр θ — случайная величина с **априорным распределением** $\pi(\theta)$
- Данные X имеют условное распределение $f(x|\theta)$ — **правдоподобие**
- После наблюдения данных обновляем знания до **апостериорного распределения** $\pi(\theta|x)$

$$\boxed{\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta')\pi(\theta')\,d\theta'} \propto f(x|\theta)\pi(\theta)}$$

Апостериорное $\propto$ Правдоподобие $\times$ Априорное

46.2 Сравнение с частотным подходом

	Частотный	**Байесовский**
Параметр θ	Фиксированный	Случайный
Вероятность	Частотная	Субъективная
Результат	Точечная/интервальная оценка	Апостериорное распределение
Предварительное знание	Не учитывается	Априорное распределение

46.3 Априорные распределения

- **Информативное:** отражает существенные предварительные знания
- **Неинформативное (диффузное):** минимальное влияние на апостериорное
- **Сопряжённое:** апостериорное принадлежит тому же семейству, что и априорное

- Бернулли + Бета $\to$ Бета
- Пуассон + Гамма $\to$ Гамма
- Нормальное (известная σ^2) + Нормальное $\to$ Нормальное
- Нормальное (известное μ) + Обратное гамма $\to$ Обратное гамма

46.4 Байесовские точечные оценки

- **Апостериорное среднее:** $\hat{\theta}_{PM} = \mathbb{E}[\theta|X]$
- **Апостериорная медиана:** $\mathrm{Med}[\theta|X]$
- **Апостериорная мода (MAP):** $\hat{\theta}_{MAP} = \arg\max_{\theta} \pi(\theta|X)$

- Апостериорное среднее минимизирует квадратичный риск
- Апостериорная медиана минимизирует абсолютный риск
- MAP максимизирует апостериорную плотность

46.5 Credible Intervals (доверительные интервалы Байеса)

$(1-\alpha)$-**байесовский доверительный интервал** (credible interval) для θ — это интервал $[L, U]$ такой, что:

$$\boxed{P(\theta \in [L, U] | X) = 1 - \alpha}$$

Интерпретация: Вероятность того, что θ лежит в интервале, равна $1-\alpha$ *при условии наблюдённых данных*.

В отличие от частотного доверительного интервала, байесовский интервал имеет прямую вероятностную интерпретацию относительно параметра.

46.5.1 Типы credible intervals

Равнохвостый интервал:

$$P(\theta < L|X) = P(\theta > U|X) = \frac{\alpha}{2}$$

Highest Posterior Density (HPD) интервал: кратчайший интервал с заданной апостериорной вероятностью.
C — HPD интервал, если:

1. $P(\theta \in C|X) = 1 - \alpha$
2. Для любого $\theta_1 \in C$ и $\theta_2 \notin C$: $\pi(\theta_1|X) \geq \pi(\theta_2|X)$

$X_1, \ldots, X_n|\mu \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 известна.
Априорное: $\mu \sim \mathcal{N}(\mu_0, \tau^2)$.

Апостериорное: $\mu|X \sim \mathcal{N}(\mu_n, \sigma_n^2)$, где:

$$\mu_n = \frac{\frac{n}{\sigma^2}\bar{X} + \frac{1}{\tau^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \quad \sigma_n^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

95% credible interval: $\mu_n \pm 1.96\sigma_n$.

46.6 Байесовская проверка гипотез

46.6.1 Апостериорные вероятности гипотез

Для гипотез H_0 и H_1 с априорными вероятностями $P(H_0)$ и $P(H_1) = 1 - P(H_0)$:

$$\boxed{P(H_0|X) = \frac{P(X|H_0)P(H_0)}{P(X|H_0)P(H_0) + P(X|H_1)P(H_1)}}$$

46.6.2 Байес-фактор

$$\boxed{BF_{01} = \frac{P(X|H_0)}{P(X|H_1)} = \frac{\int f(X|\theta)\pi_0(\theta)\,d\theta}{\int f(X|\theta)\pi_1(\theta)\,d\theta}}$$

— отношение маргинальных правдоподобий.

Связь с апостериорными шансами:

$$\frac{P(H_0|X)}{P(H_1|X)} = BF_{01} \cdot \frac{P(H_0)}{P(H_1)}$$

Апостериорные шансы = Байес-фактор × Априорные шансы

46.6.3 Интерпретация Байес-фактора

BF_{01}	**Сила доказательства в пользу** H_0
> 100	Решающее
$30 - 100$	Очень сильное
$10 - 30$	Сильное
$3 - 10$	Существенное
$1 - 3$	Слабое
$1/3 - 1$	Слабое в пользу H_1
$< 1/3$	Существенное и более в пользу H_1

46.6.4 Точечные нулевые гипотезы

H_0: $\theta = \theta_0$ vs H_1: $\theta \neq \theta_0$.
Требуется присвоить H_0 ненулевую априорную вероятность $\pi_0 = P(\theta = \theta_0)$.

Априорное распределение:

$$\pi(\theta) = \pi_0 \delta_{\theta_0}(\theta) + (1 - \pi_0) g(\theta)$$

где $g(\theta)$ — непрерывная плотность при H_1.

Апостериорная вероятность:

$$P(H_0|X) = \frac{\pi_0 f(X|\theta_0)}{\pi_0 f(X|\theta_0) + (1 - \pi_0) m_1(X)}$$

где $m_1(X) = \displaystyle\int f(X|\theta) g(\theta)\, d\theta$ — маргинальное правдоподобие при H_1.

46.7 Сравнение байесовского и частотного подходов к проверке гипотез

	Частотный	**Байесовский**
Результат	p-value	$P(H_0|X)$, Байес-фактор
Интерпретация	$P(\text{данные}|H_0)$	$P(H_0|\text{данные})$
Априорное	Не нужно	Необходимо
H_0 vs H_1	Асимметрия	Симметрия

Малое p-value не обязательно означает высокую апостериорную вероятность H_1. Это называется **парадоксом Линдли-Джеффриса**: при больших n p-value может быть малым, а Байес-фактор — в пользу H_0.