

Теория вероятностей и математическая статистика—2

Теоретический и задачный минимумы

ФЭН НИУ ВШЭ

Винер Даниил @danya_vin

Версия от 16 мая 2025 г.

Спасибо Яне Рейтман за эстетик конспект

Содержание

1	Теоретический минимум	3
1.1	Сформулируйте неравенство Крамера-Рао для несмещённых оценок	3
1.2	Дайте определение функции правдоподобия и логарифмической функция правдоподобия	3
1.3	Дайте определение информации Фишера о параметре θ , содержащейся в одном наблюдении	3
1.4	Дайте определение оценки метода моментов параметра θ с использованием первого момента, если $E(X_i) = g(\theta)$ и существует обратная функция g^{-1}	4
1.5	Дайте определение оценки метода максимального правдоподобия параметра θ	4
1.6	Укажите закон распределения выборочного среднего, величины $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, величины $\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}}$, величины $\frac{\hat{\sigma}^2(n-1)}{\sigma^2}$	4
1.7	Укажите формулу доверительного интервала с уровнем доверия $(1 - \alpha)$ для μ при известной дисперсии, для μ при неизвестной дисперсии, для σ^2	4
1.7.1	Для μ при известной дисперсии	4
1.7.2	Для μ при неизвестной дисперсии	5
1.8	Для σ^2	6
1.9	Дайте определение ошибки первого и второго рода, критической области	6
1.10	Укажите формулу доверительного интервала с уровнем доверия $(1 - \alpha)$ для вероятности успеха, построенного по случайной выборке большого размера из распределения Бернулли $Be(p)$	7
2	Задачный минимум	8
2.1	Пусть X_1, \dots, X_n — случайная выборка из распределения с плотностью распределения...	8
2.2	Пусть X_1, \dots, X_n — случайная выборка. Случайные величины X_1, \dots, X_n имеют дискретное распределение, которое задано при помощи таблицы	8
2.3	Пусть X_1, \dots, X_n — случайная выборка из распределения с функцией плотности	9
2.4	Пусть X_1, \dots, X_n — случайная выборка из распределения Бернулли с параметром $p \in (0; 1)$. При помощи метода максимального правдоподобия найдите оценку неизвестного параметра p	9
2.5	Пусть $X = (X_1, \dots, X_n)$ — случайная выборка из распределения с плотностью	10
2.6	[TBD] Стоимость выборочного исследования генеральной совокупности, состоящей из трех страт,..	11
2.7	Пусть X_1, \dots, X_n — случайная выборка из нормального распределения с параметрами μ и $\sigma^2 = 4$	11
2.8	Пусть X_1, \dots, X_n — случайная выборка из нормального распределения с неизвестными параметрами μ и σ^2	11
2.9	Пусть X_1, \dots, X_n — случайная выборка из нормального распределения с неизвестными параметрами μ и σ^2	12
2.10	Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — независимые случайные выборки из нормального распределения с параметрами (μ_X, σ_X^2) и (μ_Y, σ_Y^2) соответственно, причем $\sigma_X^2 = 2$ и $\sigma_Y^2 = 1$	12
2.11	Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — независимые случайные выборки из нормального распределения с параметрами (μ_X, σ_X^2) и (μ_Y, σ_Y^2) соответственно. Известно, что $\sigma_X^2 = \sigma_Y^2$	13
2.12	Пусть X_1, \dots, X_n — случайная выборка из распределения Бернулли с параметром p	14

2.13	Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — независимые случайные выборки из распределения Бернулли с параметрами $p_X \in (0; 1)$ и $p_Y \in (0; 1)$ соответственно	14
2.14	Дядя Вова (Владимир Николаевич) и Скрипач (Гедеван) зарабатывают на Плюке чатлы, чтобы купить гравицапу	14
2.15	Пусть X_1, \dots, X_n — случайная выборка из показательного (экспоненциального) распределения с плотностью распределения	15

1 Теоретический минимум

1.1 Сформулируйте неравенство Крамера-Рао для несмещённых оценок

Определение. Пусть $X = (X_1, \dots, X_n)$ — случайная выборка и $\ell(x_1, \dots, x_n; \theta)$ — логарифмическая функция правдоподобия. Тогда, величина

$$I_n(\theta) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ell(x_1, \dots, x_n; \theta) \right)^2 \right]$$

называется *информацией Фишера* о параметре θ , содержащейся в n наблюдениях случайной выборки X

Теорема. Пусть $\hat{\theta}$ — несмещённая оценка параметра θ , а также выполнены некоторые условия регулярности. Тогда, имеет место неравенство (*Рао-Крамера*):

$$I_n^{-1}(\theta) \leq \mathbb{D} [\hat{\theta}]$$

Определение. Несмещённая оценка $\hat{\theta}$ называется *эффективной* оценкой параметра θ , если для нее неравенство Рао-Крамера обращается в равенство

$$I_n^{-1}(\theta) = \mathbb{D} [\hat{\theta}]$$

1.2 Дайте определение функции правдоподобия и логарифмической функции правдоподобия

Определение. Пусть задана случайная выборка $X = (X_1, \dots, X_n)$, компоненты которой имеют функцию распределения $F(x; \theta)$, зависящую от неизвестного параметра $\theta \in \Theta$

- Если случайный вектор $X = (X_1, \dots, X_n)$ имеет дискретное распределение, то его функцией правдоподобия называется совместная вероятность

$$\mathcal{L}(x_1, \dots, x_n; \theta) := \mathbb{P}_\theta (\{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\}),$$

которая рассматривается как функция от переменной θ при фиксированных значениях переменных x_1, \dots, x_n

- Если случайный вектор $X = (X_1, \dots, X_n)$ имеет абсолютно непрерывное распределение, то его функцией правдоподобия называется совместная плотность

$$\mathcal{L}(x_1, \dots, x_n; \theta) := f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)$$

В независимом случае:

$$\mathcal{L}(x_1, \dots, x_n; \theta) = f_{X_1}(x_1; \theta) \cdot \dots \cdot f_{X_n}(x_n; \theta)$$

Определение. Логарифмической функцией правдоподобия называется функция

$$l(x_1, \dots, x_n; \theta) := \ln \mathcal{L}(x_1, \dots, x_n; \theta)$$

1.3 Дайте определение информации Фишера о параметре θ , содержащейся в одном наблюдении

Определение. Информацией Фишера за одно наблюдение называется

$$\underbrace{I_1(\theta)}_{i(\theta)} = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ell(x_1; \theta) \right)^2 \right]$$

1.4 Дайте определение оценки метода моментов параметра θ с использованием первого момента, если $E(X_i) = g(\theta)$ и существует обратная функция g^{-1}

$\mu_1 = \hat{\mu}_1$ — моментное тождество, где $\mu_1 = E[X_i] = g(\theta)$, $\hat{\mu}_1 = \bar{X}$, $g(\theta) = \bar{X}$

$g(\theta)$ имеет обратную функцию $\theta = g^{-1}(\bar{X})$

Определение. Оценка метода моментов параметра θ определяется как:

$$\hat{\theta}_{MM} = g^{-1}(\bar{X})$$

1.5 Дайте определение оценки метода максимального правдоподобия параметра θ

Определение. Оценкой $\hat{\theta}_{ML}$ неизвестного параметра $\theta \in \Theta$ по ММП называется точка глобального максимума функции правдоподобия по переменной $\theta \in \Theta$ при фиксированных значениях переменных x_1, \dots, x_n , т.е.

$$\mathcal{L}(x_1, \dots, x_n; \hat{\theta}_{ML}) = \max_{\theta \in \Theta} \mathcal{L}(x_1, \dots, x_n; \theta)$$

1.6 Укажите закон распределения выборочного среднего, величины $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, величины $\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}}$, величины $\frac{\hat{\sigma}^2(n-1)}{\sigma^2}$

Пусть X_1, \dots, X_n — независимые случайные величины, $X_i \sim (\mu, \sigma^2)$

- $\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0; 1)$
- $\frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim t_{n-1}$, $\hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- $\frac{\hat{\sigma}^2(n-1)}{\sigma^2} \sim \chi^2(n-1)$
- $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, $\bar{X} := \frac{X_1 + \dots + X_n}{n}$

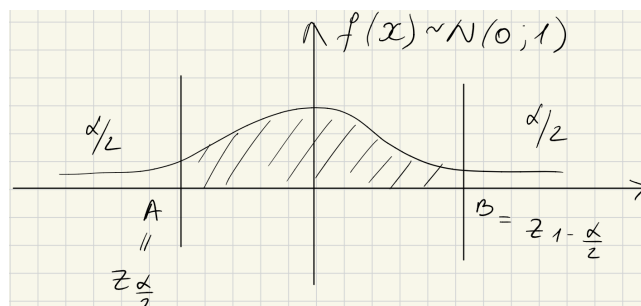
1.7 Укажите формулу доверительного интервала с уровнем доверия $(1 - \alpha)$ для μ при известной дисперсии, для μ при неизвестной дисперсии, для σ^2

Пусть $X = (X_1, \dots, X_n)$ — независимые случайные величины, $X_i \sim (\mu, \sigma^2)$

1.7.1 Для μ при известной дисперсии

Доверительный интервал для μ при известной дисперсии σ^2 уровня доверия $\gamma = 1 - \alpha$

$$T(\text{статистика}) = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0; 1)$$



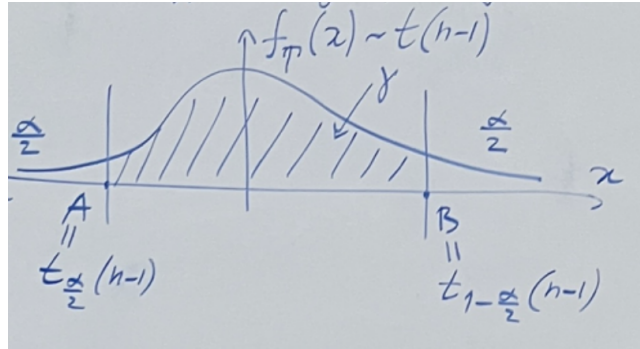
С вероятностью γ выполняется событие:

$$\begin{aligned}
 A \leq T \leq B &\iff A \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq B \\
 &\iff A\sqrt{\frac{\sigma^2}{n}} \leq \bar{X} - \mu \leq B\sqrt{\frac{\sigma^2}{n}} \\
 &\iff -\bar{X} + A\sqrt{\frac{\sigma^2}{n}} \leq -\mu \leq -\bar{X} + B\sqrt{\frac{\sigma^2}{n}} \\
 &\iff \bar{X} - B\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} - A\sqrt{\frac{\sigma^2}{n}} \\
 &\quad -z_{\alpha/2} = z_{1-\alpha/2} \\
 &\iff \bar{X} - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \\
 &\quad -z_{1-\alpha/2} = z_{\alpha/2} \\
 &\iff \underbrace{\bar{X} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}}_{T_1(X)} \leq \mu \leq \underbrace{\bar{X} + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}}_{T_2(X)}
 \end{aligned}$$

1.7.2 Для μ при неизвестной дисперсии

Доверительный интервал для μ при неизвестной дисперсии σ^2 уровня доверия $\gamma = 1 - \alpha$

$$T = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim t_{n-1}, \text{ где } \hat{\sigma}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$



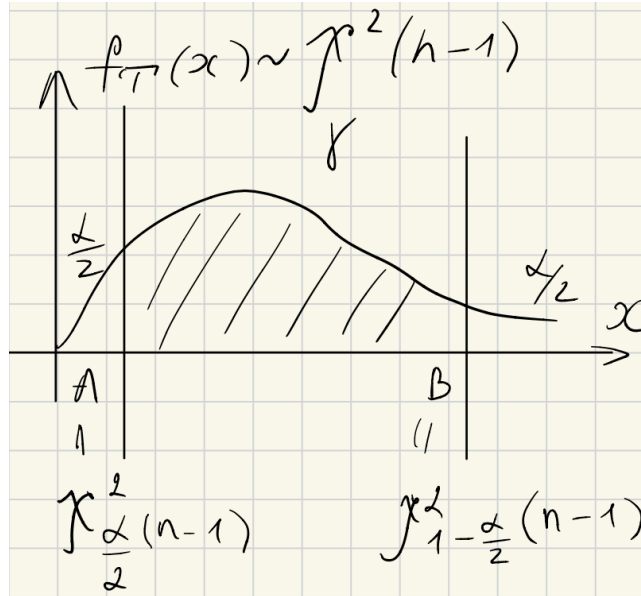
С вероятностью γ выполняется событие:

$$\begin{aligned}
 A \leq T \leq B &\iff A \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \leq B \\
 &\iff A\sqrt{\frac{\hat{\sigma}^2}{n}} \leq \bar{X} - \mu \leq B\sqrt{\frac{\hat{\sigma}^2}{n}} \\
 &\iff -\bar{X} + A\sqrt{\frac{\hat{\sigma}^2}{n}} \leq -\mu \leq -\bar{X} + B\sqrt{\frac{\hat{\sigma}^2}{n}} \\
 &\iff \bar{X} - B\sqrt{\frac{\hat{\sigma}^2}{n}} \leq \mu \leq \bar{X} - A\sqrt{\frac{\hat{\sigma}^2}{n}} \\
 &\iff \bar{X} - t_{n-1, 1-\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{n}} \leq \mu \leq \bar{X} - t_{n-1, \alpha/2}\sqrt{\frac{\hat{\sigma}^2}{n}} \\
 &\iff \underbrace{\bar{X} + t_{n-1, \alpha/2}\sqrt{\frac{\hat{\sigma}^2}{n}}}_{T_1(X)} \leq \mu \leq \underbrace{\bar{X} + t_{n-1, 1-\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{n}}}_{T_2(X)}
 \end{aligned}$$

1.8 Для σ^2

Доверительный интервал для σ^2 при неизвестном математическом ожидании уровня доверия $\gamma = 1 - \alpha$

$$T = \frac{\hat{\sigma}^2(n-1)}{\sigma^2} \sim \chi^2(n-1)$$



С вероятностью γ выполняется событие:

$$\begin{aligned} A \leq T \leq B &\iff A \leq \frac{\hat{\sigma}^2(n-1)}{\sigma^2} \leq B \\ &\iff \begin{cases} A \leq \frac{\hat{\sigma}^2(n-1)}{\sigma^2} \\ \frac{\hat{\sigma}^2(n-1)}{\sigma^2} \leq B \end{cases} \\ &\iff \begin{cases} \sigma^2 \leq \frac{\hat{\sigma}^2(n-1)}{A} \\ \frac{\hat{\sigma}^2(n-1)}{B} \leq \sigma^2 \end{cases} \\ &\iff \frac{\hat{\sigma}^2(n-1)}{B} \leq \sigma^2 \leq \frac{\hat{\sigma}^2(n-1)}{A} \\ &\iff \underbrace{\frac{\hat{\sigma}^2(n-1)}{\chi^2_{1-\alpha/2}(n-1)}}_{T_1(X)} \leq \sigma^2 \leq \underbrace{\frac{\hat{\sigma}^2(n-1)}{\chi^2_{\alpha/2}(n-1)}}_{T_2(X)} \end{aligned}$$

1.9 Дайте определение ошибки первого и второго рода, критической области

Пусть $X = (X_1, \dots, X_n)$ — случайная выборка с плотностью распределения $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$

Пусть $f_0(x_1, \dots, x_n)$ и $f_1(x_1, \dots, x_n)$ — две конкретные функции плотности, при этом

$$f_0(x_1, \dots, x_n) \neq f_1(x_1, \dots, x_n)$$

Мы хотим проверить две гипотезы:

- $H_0 : f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_0(x_1, \dots, x_n)$
- $H_1 : f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_1(x_1, \dots, x_n)$

Рассмотрим множество $S \in \mathbb{R}^n$ — критическую область. Тогда, S -критерием называется правило

- если $(x_1, \dots, x_n) \in S$, то мы принимаем H_1

- если $(x_1, \dots, x_n) \notin S$, то мы принимаем H_0

При этом, возможны такие 4 ситуации:

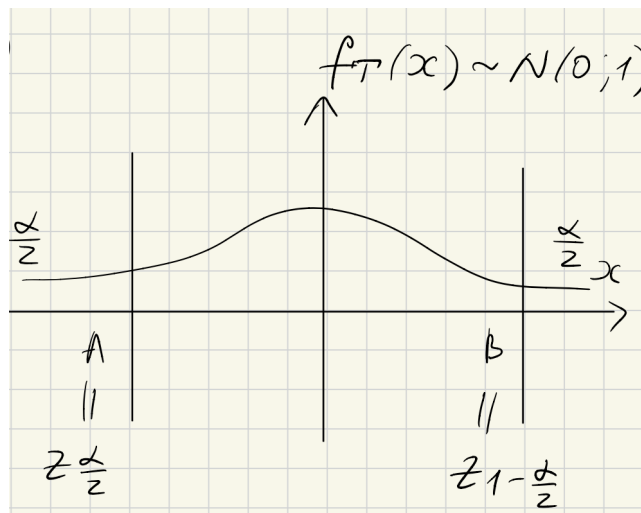
	accepted H_0	accepted H_1
real H_0	ОК	ошибка 1 рода
real H_1	ошибка 2 рода	ОК

1.10 Укажите формулу доверительного интервала с уровнем доверия $(1 - \alpha)$ для вероятности успеха, построенного по случайной выборке большого размера из распределения Бернулли $Be(p)$

Дана выборка X_1, \dots, X_n — независимые случайные величины, $X_i \sim Be(p)$, $p \in (0; 1)$

Доверительный интервал для p уровня доверия $\gamma = 1 - \alpha$

$$T = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \sim \mathcal{N}(0; 1)$$



С вероятностью γ выполняется событие:

$$\begin{aligned}
 A \leq p \leq B &\iff A \leq \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \leq B \\
 &\iff A \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \bar{X} - p \leq B \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \\
 &\iff -\bar{X} + A \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq -p \leq -\bar{X} + B \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \\
 &\iff \bar{X} - A \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} - B \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \\
 &\iff \bar{X} - z_{\alpha/2} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X} - z_{1-\alpha/2} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}
 \end{aligned}$$

2 Задачный минимум

2.1 Пусть X_1, \dots, X_n — случайная выборка из распределения с плотностью распределения...

$$f(x, \theta) = \begin{cases} \frac{6x(\theta-x)}{\theta^3} & \text{при } x \in [0; \theta] \\ 0 & \text{при } x \notin [0; \theta] \end{cases}$$

где $\theta > 0$ — неизвестный параметр распределения. Используя центральный момент второго порядка, при помощи метода моментов найдите оценку для неизвестного параметра θ .

Составим моментное тождество $\nu_2 = \hat{\nu}_2$. Найдем ν_2 :

$$\begin{aligned} \nu_2 &= \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] = \mathbb{D}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 \\ &= \frac{3\theta^2}{10} - \frac{\theta^2}{4} \\ &= \frac{\theta^2}{20} \end{aligned}$$

При этом, $\hat{\nu}_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, тогда моментное тождество $\nu_2 = \hat{\nu}_2$ принимает вид

$$\frac{\theta^2}{20} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \implies \hat{\theta}_{MM} = \sqrt{\frac{20}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

2.2 Пусть X_1, \dots, X_n — случайная выборка. Случайные величины X_1, \dots, X_n имеют дискретное распределение, которое задано при помощи таблицы

x	-3	0	2
$\mathbb{P}(X_i = x)$	$2/3 - \theta$	$1/3$	θ

Используя второй начальный момент, при помощи метода моментов найдите оценку неизвестного параметра θ . Для реализации случайной выборки $x = (0, 0, -3, 0, 2)$ найдите числовое значение найденной оценки параметра θ .

По определению метода моментов:

$$\mathbb{E}[X_i^k] = \mu_k = \hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k,$$

то есть теоретический момент равен выборочному. Найдем $\mu_2 = \mathbb{E}[X_i^2]$:

$$\begin{aligned} \mathbb{E}[X_i^2] &= (-3)^2 \cdot \left(\frac{2}{3} - \theta\right) + 0^2 \cdot \frac{1}{3} + 2^2 \cdot \theta \\ &= 6 - 5\theta \end{aligned}$$

Тогда,

$$\begin{aligned} \mu_2 &= \hat{\mu}_2 \\ 6 - 5\theta &= \frac{1}{n} \sum_{i=1}^n X_i^k \implies \theta = \frac{1}{5} \left(6 - \frac{1}{n} \sum_{i=1}^n X_i^k\right) \end{aligned}$$

Вычислим $\hat{\mu}_2$:

$$\begin{aligned} \hat{\mu}_2 &= \frac{1}{n} \sum_{i=1}^n X_i^k \\ &= \frac{1}{5} (0^2 + 0^2 + (-3)^2 + 0^2 + 2^2) \\ &= 2.6 \end{aligned}$$

Тогда, численно $\theta = \frac{1}{5} \left(6 - \frac{1}{n} \sum_{i=1}^n X_i^k\right) = \frac{1}{5} (6 - 2.6) = \frac{1}{5} \cdot 3.4 = 0.68$

2.3 Пусть X_1, \dots, X_n — случайная выборка из распределения с функцией плотности

$$f(x, \theta) = \begin{cases} \frac{2x}{\theta} e^{-\frac{x^2}{\theta}} & \text{при } x > 0 \\ 0 & \text{при } x \leq 0 \end{cases}$$

где $\theta > 0$. При помощи метода максимального правдоподобия найдите оценку неизвестного параметра θ

Функция правдоподобия имеет вид

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n f_{X_i}(x_i; \theta) \\ &= \frac{2^n}{\theta^n} \prod_{i=1}^n x_i \cdot \exp\left(-\frac{x_i^2}{\theta}\right) \end{aligned}$$

Логарифмическая функция правдоподобия:

$$\begin{aligned} \ell(x_1, \dots, x_n; \theta) &= n \ln \frac{2}{\theta} + \sum_{i=1}^n \ln \left(x_i \cdot \exp\left(-\frac{x_i^2}{\theta}\right) \right) \\ &= n \ln \frac{2}{\theta} + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln \exp\left(-\frac{x_i^2}{\theta}\right) \\ &= n \ln \frac{2}{\theta} + \sum_{i=1}^n \ln x_i - \frac{1}{\theta} \sum_{i=1}^n x_i^2 \\ &= n \ln 2 - n \ln \theta + \sum_{i=1}^n \ln x_i - \frac{1}{\theta} \sum_{i=1}^n x_i^2 \end{aligned}$$

Дифференцируем по θ , приравняем к 0 и максимизируем

$$\begin{aligned} \frac{\partial \ell}{\partial \theta} &= -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i^2 = 0 \\ \frac{1}{\theta^2} \sum_{i=1}^n x_i^2 &= \frac{n}{\theta} \\ \sum_{i=1}^n x_i^2 &= n\theta \implies \hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^n x_i^2 \end{aligned}$$

2.4 Пусть X_1, \dots, X_n — случайная выборка из распределения Бернулли с параметром $p \in (0; 1)$. При помощи метода максимального правдоподобия найдите оценку неизвестного параметра p

Функцию вероятности для величины из распределения Бернулли можно записать так:

$$\mathbb{P}(X_i, p) = p^{X_i} (1-p)^{1-X_i},$$

где $X_i \in \{0, 1\}$

Тогда функция правдоподобия имеет вид

$$\begin{aligned} \mathcal{L}(X_1, \dots, X_n; p) &= \prod_{i=1}^n \mathbb{P}(X_i, p) \\ &= p^{\sum X_i} \cdot (1-p)^{n - \sum X_i} \end{aligned}$$

Тогда Логарифмическая функция правдоподобия:

$$\ell(X_1, \dots, X_n; p) = \left(\sum_{i=1}^n X_i \right) \ln p + \left(n - \sum_{i=1}^n X_i \right) \ln(1-p)$$

Дифференцируем по p , приравняем к 0 и максимизируем

$$\frac{\partial \ell}{\partial p} = \frac{1}{p} \left(\sum_{i=1}^n X_i \right) - \frac{1}{1-p} \left(n - \sum_{i=1}^n X_i \right) = 0$$

Найдем оценку параметра:

$$\begin{aligned} \frac{1}{p} \left(\sum_{i=1}^n X_i \right) &= \frac{1}{1-p} \left(n - \sum_{i=1}^n X_i \right) \\ \sum_{i=1}^n X_i (1-p) &= \left(n - \sum_{i=1}^n X_i \right) p \\ \sum_{i=1}^n X_i - p \sum_{i=1}^n X_i &= np - p \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i &= np \implies p = \frac{1}{n} \sum_{i=1}^n X_i \end{aligned}$$

2.5 Пусть $X = (X_1, \dots, X_n)$ — случайная выборка из распределения с плотностью

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{при } x \geq 0 \\ 0 & \text{при } x < 0 \end{cases}$$

где $\theta > 0$ — неизвестный параметр. Является ли оценка $\hat{\theta} = \bar{X}$ эффективной?

Плотность указывает, что у нас экспоненциальное распределение с параметром $\lambda = \frac{1}{\theta}$. Проверим несмещенность оценки

$$\begin{aligned} \mathbb{E} [\hat{\theta}] &= \mathbb{E} [\bar{X}] \\ &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] \\ &= \mathbb{E} [X_i] \\ &= \theta \end{aligned}$$

Значит, оценка несмещенная

Теперь найдем информацию Фишера для одного наблюдения, которая определяется как

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

Составим функцию правдоподобия

$$\ln f(x, \theta) = -\ln \theta - \frac{x}{\theta} \implies \frac{\partial}{\partial \theta} \ln f(x, \theta) = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

Тогда,

$$\begin{aligned} I(\theta) &= \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ln f(x, \theta) \right)^2 \right] \\ &= \mathbb{E} \left[\left(-\frac{1}{\theta} + \frac{x}{\theta^2} \right)^2 \right] \\ &= \mathbb{E} \left[\frac{1}{\theta^2} - \frac{2x}{\theta^3} + \frac{x^2}{\theta^4} \right] \\ &= \frac{1}{\theta^2} - \frac{2}{\theta^3} \mathbb{E} [X] + \frac{1}{\theta^4} \mathbb{E} [X^2] \end{aligned}$$

Вычислим $\mathbb{E} [X^2] = \mathbb{D} [X] + (\mathbb{E} [X])^2 = \left(\frac{1}{\theta} \right)^{-2} + \theta^2 = 2\theta^2$, тогда

$$I(\theta) = \frac{1}{\theta^2} - \frac{2}{\theta^3} \theta + \frac{1}{\theta^4} 2\theta^2 = \frac{1}{\theta^2}$$

Используем неравенство Рао-Крамера:

$$\begin{aligned}\mathbb{D}[\hat{\theta}] &\geq \frac{1}{n \cdot I(\theta)} \\ \mathbb{D}[\bar{X}] &\geq \frac{1}{n \cdot \frac{1}{\theta^2}} \\ \frac{1}{n} \mathbb{D}[X_i] &\geq \frac{\theta^2}{n} \\ \frac{\theta^2}{n} &\geq \frac{\theta^2}{n}\end{aligned}$$

Неравенство выполняется, значит оценка эффективна

2.6 [TBD] Стоимость выборочного исследования генеральной совокупности, состоящей из трех страт,..

определяется по формуле $TC = c_1 n_1 + c_2 n_2 + c_3 n_3$, где c_i — цена одного наблюдения в i -й страте, а n_i — число наблюдений, которые приходятся на i -ю страту. Найдите n_1, n_2 и n_3 , при которых дисперсия стратифицированного среднего достигает наименьшего значения, если бюджет исследования 8000 и имеется следующая информация:

Страта	1	2	3
Среднее значение	30	40	50
Стандартная ошибка	5	10	20
Вес	25%	25%	50%
Цена наблюдения	1	5	10

2.7 Пусть X_1, \dots, X_n — случайная выборка из нормального распределения с параметрами μ и $\sigma^2 = 4$

Используя реализацию случайной выборки,

$$x_1 = -1.11, x_2 = -6.10, x_3 = 2.42$$

постройте 90%-й доверительный интервал для неизвестного параметра μ

Уровень доверия — $0.9 = 1 - \alpha \implies \alpha = 0.1$ — уровень значимости. Для выборки из нормального распределения с **известной** дисперсией доверительный интервал имеет вид:

$$\bar{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

где $z_{1-\alpha/2}$ — квантиль уровня $1 - \alpha/2$ для стандартного нормального распределения (из таблички)

Примечание. $\bar{X} = \frac{-1.11 + (-6.10) + 2.42}{3} = -\frac{4.79}{3}$

Тогда, в нашем случае:

$$\begin{aligned}\bar{X} - z_{0.95} \cdot \frac{2}{\sqrt{3}} &< \mu < \bar{X} + z_{0.95} \cdot \frac{2}{\sqrt{3}} \\ -\frac{4.79}{3} - 1.65 \cdot \frac{2}{\sqrt{3}} &< \mu < -\frac{4.79}{3} + 1.65 \cdot \frac{2}{\sqrt{3}} \\ -3.50192 &< \mu < 0.30859\end{aligned}$$

2.8 Пусть X_1, \dots, X_n — случайная выборка из нормального распределения с неизвестными параметрами μ и σ^2

Используя реализацию случайной выборки,

$$x_1 = -1.11, x_2 = -6.10, x_3 = 2.42$$

постройте 90%-й доверительный интервал для неизвестного параметра μ

Уровень доверия — $0.9 = 1 - \alpha \implies \alpha = 0.1$ — уровень значимости. Для выборки из нормального распределения с **неизвестной** дисперсией доверительный интервал имеет вид:

$$\bar{X} - t_{n-1, 1-\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + t_{n-1, 1-\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

Найдем несмещенную выборочную дисперсию

$$\bar{X} = \frac{-1.11 + (-6.10) + 2.42}{3} = -\frac{4.79}{3}$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \approx 18.32523$$

Тогда, для нашей выборки:

$$-\frac{4.79}{3} - t_{2, 0.95} \cdot \frac{\sqrt{\hat{\sigma}^2}}{\sqrt{3}} < \mu < -\frac{4.79}{3} + t_{2, 0.95} \cdot \frac{\sqrt{\hat{\sigma}^2}}{\sqrt{3}}$$

$$-\frac{4.79}{3} - 2.92 \cdot \frac{4.2808}{\sqrt{3}} < \mu < -\frac{4.79}{3} + 2.92 \cdot \frac{4.2808}{\sqrt{3}}$$

$$-8.81351 < \mu < 5.62017$$

2.9 Пусть X_1, \dots, X_n — случайная выборка из нормального распределения с неизвестными параметрами μ и σ^2

Используя реализацию случайной выборки,

$$x_1 = 1.07, \quad x_2 = 3.66, \quad x_3 = -4.51$$

постройте 80%-й доверительный интервал для неизвестного параметра σ^2

Уровень доверия — $0.8 = 1 - \alpha \implies \alpha = 0.2$ — уровень значимости. Для выборки из нормального распределения доверительный интервал имеет вид:

$$\frac{(n-1)\hat{\sigma}^2}{\chi_{n-1, 1-\alpha/2}^2} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi_{n-1, \alpha/2}^2},$$

где $\chi_{\alpha/2}^2, \chi_{1-\alpha/2}^2$ — квантили хи-квадрат распределения с $n-1$ степенями свободы.

Вычислим несмещенную выборочную дисперсию:

$$\bar{X} = \frac{1.07 + 3.66 - 4.51}{3} = \frac{11}{150}$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \approx 17.43223$$

Тогда, для нашей выборки

$$\frac{2 \cdot 17.43223}{\chi_{2, 0.9}^2} < \sigma^2 < \frac{2 \cdot 17.43223}{\chi_{2, 0.1}^2}$$

$$\frac{2 \cdot 17.43223}{4.605} < \sigma^2 < \frac{2 \cdot 17.43223}{0.211}$$

$$7.571 < \sigma^2 < 165.23441$$

2.10 Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — независимые случайные выборки из нормального распределения с параметрами (μ_X, σ_X^2) и (μ_Y, σ_Y^2) соответственно, причем $\sigma_X^2 = 2$ и $\sigma_Y^2 = 1$

Используя реализации случайных выборок

$$x_1 = -1.11, \quad x_2 = -6.10, \quad x_3 = 2.42$$

$$y_1 = -2.29, \quad y_2 = -2.91$$

постройте 95%-й доверительный интервал для разности математических ожиданий $\mu_X - \mu_Y$

Уровень доверия $- 0.95 = 1 - \alpha \implies \alpha = 0.05$ — уровень значимости. Для выборки из нормального распределения с известными дисперсиями доверительный интервал имеет вид:

$$(\bar{X} - \bar{Y}) - z_{1-\alpha/2} \sqrt{\frac{\hat{\sigma}_X^2}{n} + \frac{\hat{\sigma}_Y^2}{m}} < \mu_X - \mu_Y < (\bar{X} - \bar{Y}) + z_{1-\alpha/2} \sqrt{\frac{\hat{\sigma}_X^2}{n} + \frac{\hat{\sigma}_Y^2}{m}}$$

Вычислим выборочные средние:

$$\begin{aligned}\bar{X} &= \frac{-1.11 - 6.10 + 2.42}{3} = -\frac{4.79}{3} \\ \bar{Y} &= \frac{-2.29 - 2.91}{2} = -2.6\end{aligned}$$

Тогда, для наших выборок:

$$\begin{aligned}\left(-\frac{4.79}{3} + 2.6\right) - z_{0.975} \sqrt{\frac{2}{3} + \frac{1}{2}} &< \mu_X - \mu_Y < \left(-\frac{4.79}{3} + 2.6\right) + z_{0.975} \sqrt{\frac{2}{3} + \frac{1}{2}} \\ \left(-\frac{4.79}{3} + 2.6\right) - 1.96 \cdot \sqrt{\frac{2}{3} + \frac{1}{2}} &< \mu_X - \mu_Y < \left(-\frac{4.79}{3} + 2.6\right) + 1.96 \cdot \sqrt{\frac{2}{3} + \frac{1}{2}} \\ -1.11371 &< \mu_X - \mu_Y < 3.12038\end{aligned}$$

2.11 Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — независимые случайные выборки из нормального распределения с параметрами (μ_X, σ_X^2) и (μ_Y, σ_Y^2) соответственно. Известно, что $\sigma_X^2 = \sigma_Y^2$

Используя реализации случайных выборок

$$\begin{aligned}x_1 &= 1.53, & x_2 &= 2.83, & x_3 &= -1.25 \\ y_1 &= -0.8, & y_2 &= 0.06\end{aligned}$$

постройте 95%-й доверительный интервал для разности математических ожиданий $\mu_X - \mu_Y$

Уровень доверия $- 0.95 = 1 - \alpha \implies \alpha = 0.05$ — уровень значимости. Для выборки из нормального распределения с неизвестными, но равными дисперсиями доверительный интервал имеет вид:

$$\bar{X} - \bar{Y} - t_{n+m-2; 1-\alpha/2} \widehat{\sigma}_0 \sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_X - \mu_Y < \bar{X} - \bar{Y} + t_{n+m-2; 1-\alpha/2} \widehat{\sigma}_0 \sqrt{\frac{1}{n} + \frac{1}{m}},$$

где

$$\widehat{\sigma}_0^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n + m - 2}$$

Выпишем альтернативный вариант доверительного интервала и работать будем с этим вариантом:

$$(\bar{X} - \bar{Y}) \mp t_{n+m-2; 1-\alpha/2} \sqrt{\left(\frac{n+m}{nm}\right) \left(\frac{(n-1)\hat{\sigma}_X^2 + (m-1)\hat{\sigma}_Y^2}{n+m-2}\right)}$$

Найдем выборочные средние:

$$\begin{aligned}\bar{X} &= \frac{1.53 + 2.83 - 1.25}{3} = \frac{3.11}{3} \\ \bar{Y} &= \frac{-0.8 + 0.06}{2} = -0.37\end{aligned}$$

Вычислили выборочные дисперсии:

$$\begin{aligned}\hat{\sigma}_X^2 &= \frac{(1.53 - \frac{3.11}{3})^2 + (2.83 - \frac{3.11}{3})^2 + (-1.25 - \frac{3.11}{3})^2}{2} \approx 4.34413 \\ \hat{\sigma}_Y^2 &= \frac{(-0.8 + 0.37)^2 + (0.06 + 0.37)^2}{1} = 0.3698\end{aligned}$$

Тогда, доверительный интервал имеет вид:

$$\begin{aligned} \left(\frac{3.11}{3} + 0.37\right) - t_{3,0.975} \sqrt{\frac{5}{6} \cdot \frac{2 \cdot 4.34413 + 1 \cdot 0.3698}{3}} < \mu_X - \mu_Y < \left(\frac{3.11}{3} + 0.37\right) + t_{3,0.975} \sqrt{\frac{5}{6} \cdot \frac{2 \cdot 4.34413 + 1 \cdot 0.3698}{3}} \\ \frac{21.1}{15} - 3.182 \cdot \sqrt{\frac{5}{6} \cdot \frac{2 \cdot 4.34413 + 1 \cdot 0.3698}{3}} < \mu_X - \mu_Y < \frac{21.1}{15} + 3.182 \cdot \sqrt{\frac{5}{6} \cdot \frac{2 \cdot 4.34413 + 1 \cdot 0.3698}{3}} \\ -3.64072 < \mu_X - \mu_Y < 6.45405 \end{aligned}$$

2.12 Пусть X_1, \dots, X_n — случайная выборка из распределения Бернулли с параметром p

Используя реализацию случайной выборки X_1, \dots, X_n , в которой 55 нулей и 45 единиц, постройте приближенный 95%-й доверительный интервал для неизвестного параметра p

Уровень доверия — $0.95 = 1 - \alpha \implies \alpha = 0.05$ — уровень значимости. Для выборки из распределения Бернулли с неизвестным параметром p доверительный интервал имеет вид:

$$\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Так как у нас 55 нулей и 45 единиц, то $n = 100$, тогда $\hat{p} = \bar{X} = \frac{45}{100}$ и доверительный интервал такой:

$$\begin{aligned} 0.45 - z_{0.975} \cdot \sqrt{\frac{0.45 \cdot 0.55}{100}} < p < 0.45 + z_{0.975} \cdot \sqrt{\frac{0.45 \cdot 0.55}{100}} \\ 0.45 - 1.96 \cdot \frac{3\sqrt{11}}{200} < p < 0.45 + 1.96 \cdot \frac{3\sqrt{11}}{200} \\ 0.352491 < p < 0.547509 \end{aligned}$$

2.13 Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — независимые случайные выборки из распределения Бернулли с параметрами $p_X \in (0; 1)$ и $p_Y \in (0; 1)$ соответственно

Известно, что $n = 100, \bar{x}_n = 0.6, m = 200, \bar{y}_m = 0.4$. Постройте приближенный 95%-й доверительный интервал для отношения разности вероятностей успеха $p_X - p_Y$

Уровень доверия — $0.95 = 1 - \alpha \implies \alpha = 0.05$ — уровень значимости. Для выборки из распределения Бернулли с неизвестными параметрами p_X, p_Y границы доверительного интервала имеют вид:

$$(\hat{p}_X - \hat{p}_Y) \mp z_{1-\alpha/2} \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{m}}$$

Отметим, что $\hat{p}_X = \bar{x}_n, \hat{p}_Y = \bar{y}_m$, тогда доверительный интервал:

$$\begin{aligned} (0.6 - 0.2) - z_{0.975} \sqrt{\frac{0.6 \cdot 0.4}{100} + \frac{0.4 \cdot 0.6}{200}} < p_X - p_Y < (0.6 - 0.2) + z_{0.975} \sqrt{\frac{0.6 \cdot 0.4}{100} + \frac{0.4 \cdot 0.6}{200}} \\ 0.2 - 1.96 \cdot 0.06 < p_X - p_Y < 0.2 + 1.96 \cdot 0.06 \\ 0.0824 < p_X - p_Y < 0.3176 \end{aligned}$$

2.14 Дядя Вова (Владимир Николаевич) и Скрипач (Гедеван) зарабатывают на Плюке чатлы, чтобы купить гравицапу

Число заработанных за i -й день чатлов имеет распределение Пуассона с неизвестным параметром λ . Заработки в различные дни независимы. За прошедшие 100 дней они заработали 250 чатлов. С помощью метода максимального правдоподобия постройте приближенный 95%-й доверительный интервал для неизвестного параметра λ

Дана выборка (X_1, \dots, X_n) , $n = 100, X_i \sim Pois(\lambda)$

Функция вероятности для распределения Пуассона имеет вид

$$\mathbb{P}(X_i = x_i) = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

Тогда, функция правдоподобия:

$$\mathcal{L}(X_1, \dots, X_n; \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{\sum x_i} \cdot \left(\prod_{i=1}^n \frac{1}{x_i!} \right)$$

Логарифмическая функция правдоподобия:

$$\ell(X_1, \dots, X_n; \lambda) = -n\lambda + \left(\sum x_i \right) \ln \lambda - \sum \ln(x_i!)$$

Дифференцируем и максимизируем:

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= -n + \frac{\sum x_i}{\lambda} = 0 \\ \lambda &= \frac{1}{n} \sum x_i \\ \lambda &= \bar{X} \\ \lambda &= \frac{250}{100} \Rightarrow \lambda_{ML} = 2.5 \end{aligned}$$

Для выборки из некоторого распределения с параметром λ , который оценивается ММП доверительный интервал имеет вид:

$$\lambda_{ML} - z_{1-\alpha/2} \frac{1}{\sqrt{I_n(\lambda_{ML})}} < \lambda < \lambda_{ML} + z_{1-\alpha/2} \frac{1}{\sqrt{I_n(\lambda_{ML})}}$$

Информация Фишера определяется так:

$$I(\lambda) = -\mathbb{E} \left[\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2} \right]$$

При этом, $I_n(\lambda_{ML}) = n \cdot I_1(\lambda_{ML})$. Для одного наблюдения x_i из выборки:

$$\begin{aligned} \mathcal{L}(\lambda) &= \mathbb{P}(x_i) = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \\ \ell(x_i, \lambda) &= -\lambda + x \ln \lambda \end{aligned}$$

Тогда,

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= -1 + \frac{x}{\lambda} \\ \frac{\partial^2 \ell}{\partial \lambda^2} &= -\frac{x}{\lambda^2} \end{aligned}$$

Значит, информация Фишера:

$$I_1(\lambda) = -\mathbb{E} \left[-\frac{x}{\lambda^2} \right] = \frac{\mathbb{E}[x_i]}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} \Rightarrow I_n(\lambda) = \frac{n}{\lambda}$$

И наконец, искомый доверительный интервал:

$$\begin{aligned} 2.5 - z_{0.975} \frac{1}{\sqrt{I_n(2.5)}} &< \lambda < 2.5 + z_{0.975} \frac{1}{\sqrt{I_n(2.5)}} \\ 2.5 - 1.96 \cdot \frac{1}{\sqrt{\frac{100}{2.5}}} &< \lambda < 2.5 + 1.96 \cdot \frac{1}{\sqrt{\frac{100}{2.5}}} \\ 2.1901 &< \lambda < 2.8099 \end{aligned}$$

2.15 Пусть X_1, \dots, X_n — случайная выборка из показательного (экспоненциального) распределения с плотностью распределения

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

где $\lambda > 0$ — неизвестный параметр распределения. Известно, что $n = 100$ и $\bar{x}_n = 0.52$. С помощью метода максимального правдоподобия постройте приближенный 95%-й доверительный интервал для параметра λ

Функция правдоподобия:

$$\mathcal{L}(X_1, \dots, X_n; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$$

Логарифмическая функция правдоподобия:

$$\ell(X_1, \dots, X_n; \lambda) = \ln \mathcal{L} = n \ln \lambda - \lambda \sum x_i$$

Дифференцируем и максимизируем:

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i = 0 \implies \hat{\lambda}_{ML} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}} = \frac{1}{0.52} \approx 1.9231$$

Для выборки из некоторого распределения с параметром λ , который оценивается ММП доверительный интервал имеет вид:

$$\lambda_{ML} - z_{1-\alpha/2} \frac{1}{\sqrt{I_n(\lambda_{ML})}} < \lambda < \lambda_{ML} + z_{1-\alpha/2} \frac{1}{\sqrt{I_n(\lambda_{ML})}}$$

Информация Фишера определяется так:

$$I(\lambda) = -\mathbb{E} \left[\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2} \right]$$

При этом, $I_n(\lambda_{ML}) = n \cdot I_1(\lambda_{ML})$. Для одного наблюдения x_i из выборки:

$$\begin{aligned} \mathcal{L}(\lambda) &= \mathbb{P}(x_i) = \lambda e^{-\lambda x} \\ \ell(x_i, \lambda) &= \ln \lambda - \lambda x \\ \frac{\partial^2 \ell}{\partial \lambda^2} &= -\frac{1}{\lambda^2} \end{aligned}$$

Значит, информация Фишера:

$$I_1(\lambda) = \frac{1}{\lambda^2} \implies I_n(\lambda) = \frac{n}{\lambda^2}$$

И наконец, искомый доверительный интервал:

$$\begin{aligned} 1.9231 - z_{0.975} \frac{1}{\sqrt{\frac{100}{1.9231^2}}} &< \lambda < 1.9231 + z_{0.975} \frac{1}{\sqrt{\frac{100}{1.9231^2}}} \\ 1.9231 - 1.96 \cdot 0.19231 &< \lambda < 1.9231 + 1.96 \cdot 0.19231 \\ 1.5461 &< \lambda < 2.3 \end{aligned}$$