Теория вероятностей и математическая статистика—2
Теоретический и задачный минимумы
ФЭН НИУ ВШЭ

Винер Даниил @danya_vin

Версия от 14 мая 2025 г.

Содержание

1 Теоретический минимум

1.1 Сформулируйте неравенство Крамера-Рао для несмещённых оценок

Пусть $\widehat{\theta}$ — несмещенная оценка параметра θ, а также выполняются все условия гладкости и регулярности, тогда для несмещённых оценок верно:

$$\mathbb{D}\left[\widehat{\theta}\right] \geqslant \frac{1}{I_n(\theta)},$$

где $I_n(\theta) = n \cdot I(\theta)$

1.2 Дайте определение функции правдоподобия и логарифмической функция правдоподобия

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

- Для абсолютно непрерывных величин: $\mathcal{L}\left(x_1, \ldots, x_n; \theta\right) = f\left(x_1, \ldots, x_n; \theta\right) = \prod\limits_{i=1}^{n} f(x_i;\theta)$
- Для дискретных величин: $\mathcal{L}\left(x_1, \ldots, x_n; \theta\right) = \mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n; \theta\right) = \prod\limits_{i=1}^{n} \mathbb{P}\left(x_i;\theta\right)$

Определение. Логарифмической функцией правдоподобия называется функция

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

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

Имеется выборка с неизвестным параметром — $X_1, \ldots, X_n \sim F(x;\theta)$

Определение. Информацией Фишера называется

$$I(\theta; X) = \mathbb{E}\left[\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\right)^2\right],$$

где $\mathcal{L}$ — функция правдоподобия

Примечание. Определние применимо для регулярного случая, то есть область значений X не зависит от θ

Определение. Равносильное определение информации Фишера:

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

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

Оценка ММ параметра θ определяется как:

$$\hat{\theta}_{MM} = g^{-1}\left(\overline{X}\right),$$

где $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ — выборочное среднее, $g(\theta) = \mathbb{E}(X_i)$, а g^{-1} — обратная функция к g, существующая по условию[1]

[1] стр.23 учебника Черновой — тоже самое, только в общем виде

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 $x_1, \ldots, x_n$, т.е.

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

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

Пусть $X_1, \ldots, X_n$ — независимые нормальные случайные величины с параметрами μ и σ^2, $\overline{X} := \frac{X_1+\ldots+X_n}{n}$ - выборочное среднее, а $\widehat{\sigma^2} := \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ исправленная выборочная дисперсия. Тогда

- $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
- $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0; 1)$
- $\frac{\overline{X}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$
- $\frac{\hat{\sigma}^2(n-1)}{\sigma^2} \sim \chi^2(n-1)$

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

Дана выборка $X_1, \ldots X_n \sim N\left(\mu, \sigma^2\right)$

- Если известна σ^2:
$$\left(\overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right),$$
где $z_{\alpha/2}$ — квантиль стандартного нормального распределения
- Если σ^2 неизвестна:
$$\left(\overline{X} - t_{\alpha/2,\, n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}},\ \overline{X} + t_{\alpha/2,\, n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right),$$
где $t_{\alpha/2,\, n-1}$ — квантиль распределения Стьюдента с $n-1$ степенями свободы
- Для σ^2:
$$\left(\frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2}},\ \frac{(n-1)\hat{\sigma}^2}{\chi^2_{\alpha/2}}\right),$$
где $\chi^2_{\alpha/2}$, $\chi^2_{1-\alpha/2}$ — квантили хи-квадрат распределения с $n-1$ степенями свободы.

1.8 Дайте определение ошибки первого и второго рода, критической области

Определение. Есть выборка $X_1, \ldots, X_n$, а множество значений $\mathcal{X} \in \mathbb{R}^n$

$$\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$$
$$\mathcal{X}_0 \cap \mathcal{X}_1 = 0$$

$\mathcal{X}_1$ — критическая область, где H_0 отвергается, а в $\mathcal{X}_0$ — не отвергается

Определение. Ошибка первого рода — вероятность отвергнуть H_0, когда она на самом деле верна:

$$\mathbb{P}(X \in \mathcal{X}_1 \mid H_0 \text{ верна})$$

Определение. Ошибка второго рода — вероятность не отвергнуть H_0, когда на самом деле верна H_1:

$$\mathbb{P}(X \in \mathcal{X}_0 \mid H_1 \text{ верна})$$

Определение. Говорят, что произошла ошибка i-го рода критерия δ, если критерий отверг верную гипотезу H_i. Вероятностью ошибки i-го рода критерия $\boldsymbol{\delta}$ называется число

$$\alpha_i(\boldsymbol{\delta}) = \mathrm{P}_{H_i}\left(\boldsymbol{\delta}(\vec{X}) \neq H_i\right)$$

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

При больших n почти всегда имеет место интервал:

$$\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Тогда, для выборки из распределения Бернулли $\mathrm{Bin}(1, p)$:

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

где

- $\hat{p} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ — выборочная доля успехов,
- $z_{\alpha/2}$ — квантиль стандартного нормального распределения

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 $\nu_2 = \widehat{\nu}_2$. Найдем ν_2:

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

При этом, $\widehat{\nu}_2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$, тогда моментное тождество $\nu_2 = \widehat{\nu}_2$ принимает вид

$$\frac{\theta^2}{20} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \Longrightarrow \widehat{\theta}_{MM} = \sqrt{\frac{20}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

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:

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

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

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

Тогда,

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

Вычислим $\widehat{\mu}_2$:

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

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

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

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

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

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

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:

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

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:

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

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

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

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

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

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

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

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

Тогда,

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

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

$$\mathbb{D}\left[\widehat{\theta}\right] \geqslant \frac{1}{n \cdot I(\theta)}$$
$$\mathbb{D}\left[\overline{X}\right] \geqslant \frac{1}{n \cdot \frac{1}{\theta^2}}$$
$$\frac{1}{n}\mathbb{D}\left[X_i\right] \geqslant \frac{\theta^2}{n}$$
$$\frac{\theta^2}{n} \geqslant \frac{\theta^2}{n}$$

Неравенство выполняется, значит оценка эффективна

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

определяется по формуле $TC = c_1n_1 + c_2n_2 + c_3n_3$, 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 $X_1, \ldots, X_n$ — 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 интревал имеет вид:

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

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

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

Тогда, в нашем случае:

$$\overline{X} - z_{0.95} \cdot \frac{2}{\sqrt{3}} < \mu < \overline{X} + z_{0.95} \cdot \frac{2}{\sqrt{3}}$$
$$-\frac{4.79}{3} - 1.65 \cdot \frac{2}{\sqrt{3}} < \mu < -\frac{4.79}{3} + 1.65 \cdot \frac{2}{\sqrt{3}}$$
$$-3.50192 < \mu < 0.30859$$

2.8 Пусть $X_1, \ldots, X_n$ — 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 интревал имеет вид:

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

Найдем несмещенную выборочную дисперсию

$$\overline{X} = \frac{-1.11 + (-6.10) + 2.42}{3} = -\frac{4.79}{3}$$
$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \approx 18.32523$$

Тогда, для нашей выборки:

$$-\frac{4.79}{3} - t_{2,0.95}\cdot\frac{\sqrt{\widehat{\sigma}^2}}{\sqrt{3}} < \mu < -\frac{4.79}{3} + t_{2,0.95}\cdot\frac{\sqrt{\widehat{\sigma}^2}}{\sqrt{3}}$$
$$-\frac{4.79}{3} - 2.92\cdot\frac{4.2808}{\sqrt{3}} < \mu < -\frac{4.79}{3} + 2.92\cdot\frac{4.2808}{\sqrt{3}}$$
$$-8.81351 < \mu < 5.62017$$

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

Используя реализацию случайной выборки,

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

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 интревал имеет вид:

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

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

Вычислим несмещенную выборочную дисперсию:

$$\overline{X} = \frac{1.07 + 3.66 - 4.51}{3} = \frac{11}{150}$$
$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \approx 17.43223$$

Тогда, для нашей выборки

$$\frac{2\cdot 17.43223}{\chi^2_{2,0.9}} < \sigma^2 < \frac{2\cdot 17.43223}{\chi^2_{2,0.1}}$$
$$\frac{2\cdot 17.43223}{4.605} < \sigma^2 < \frac{2\cdot 17.43223}{0.211}$$
$$7.571 < \sigma^2 < 165.23441$$

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

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

Вычислим выборочные средние:

$$\overline{X} = \frac{-1.11 - 6.10 + 2.42}{3} = -\frac{4.79}{3}$$
$$\overline{Y} = \frac{-2.29 - 2.91}{2} = -2.6$$

Тогда, для наших выборок:

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

2.11 Пусть $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$ — 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

$$x_1 = 1.53, \quad x_2 = 2.83, \quad x_3 = -1.25$$
$$y_1 = -0.8, \quad y_2 = 0.06$$

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

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

$$\overline{X} - \overline{Y} - t_{n+m-2;1-\alpha/2}\widehat{\sigma_0}\sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + t_{n+m-2;1-\alpha/2}\widehat{\sigma_0}\sqrt{\frac{1}{n} + \frac{1}{m}},$$

где

$$\widehat{\sigma_0^2} = \frac{\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2 + \sum\limits_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2}{n + m - 2}$$

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

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

Найдем выборочные средние:

$$\overline{X} = \frac{1.53 + 2.83 - 1.25}{3} = \frac{3.11}{3}$$
$$\overline{Y} = \frac{-0.8 + 0.06}{2} = -0.37$$

Вычилим выборочные дисперсии:

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

Тогда, доверительный интервал имеет вид:

$$\left(\frac{3.11}{3} + 0.37\right) - t_{3,0.975}\sqrt{\frac{5}{6} \cdot \frac{2 \cdot 4.34413 + 1 \cdot 0.3698}{3}} < \mu_X - \mu_Y < \left(\frac{3.11}{3} + 0.37\right) + t_{3,0.975}\sqrt{\frac{5}{6} \cdot \frac{2 \cdot 4.34413 + 1 \cdot 0.3698}{3}}$$

$$\frac{21.1}{15} - 3.182 \cdot \sqrt{\frac{5}{6} \cdot \frac{2 \cdot 4.34413 + 1 \cdot 0.3698}{3}} < \mu_X - \mu_Y < \frac{21.1}{15} + 3.182 \cdot \sqrt{\frac{5}{6} \cdot \frac{2 \cdot 4.34413 + 1 \cdot 0.3698}{3}}$$

$$-3.64072 < \mu_X - \mu_Y < 6.45405$$

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

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

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

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

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

$$0.45 - z_{0.975} \cdot \sqrt{\frac{0.45 \cdot 0.55}{100}} < p < 0.45 + z_{0.975} \cdot \sqrt{\frac{0.45 \cdot 0.55}{100}}$$

$$0.45 - 1.96 \cdot \frac{3\sqrt{11}}{200} < p < 0.45 + 1.96 \cdot \frac{3\sqrt{11}}{200}$$

$$0.352491 < p < 0.547509$$

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

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

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

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

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

$$(0.6 - 0.2) - z_{0.975}\sqrt{\frac{0.6 \cdot 0.4}{100} + \frac{0.4 \cdot 0.6}{200}} < p_X - p_Y < (0.6 - 0.2) + z_{0.975}\sqrt{\frac{0.6 \cdot 0.4}{100} + \frac{0.4 \cdot 0.6}{200}}$$

$$0.2 - 1.96 \cdot 0.06 < p_X - p_Y < 0.2 + 1.96 \cdot 0.06$$

$$0.0824 < p_X - p_Y < 0.3176$$

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 $(X_1, \ldots, X_n)$, $n = 100$, $X_i \sim Pois(\lambda)$

Функция вероятности для распределения Пуассона имеет вид

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

Тогда, функция правдоподобия:

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

Логарифмическая функция правдоподобия:

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

Дифференцируем и максимизируем:

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

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

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

Информация Фишера определяется так:

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

При этом, $I_n\left(\lambda_{ML}\right) = n \cdot I_1(\lambda_{ML})$. 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, информаицця Фишера:

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

И наконец, искомый доверительный интервал:

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

2.15 Пусть $X_1, \ldots, X_n$ — 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 $\overline{x}_n = 0.52$. 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:

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i = 0 \Longrightarrow \widehat{\lambda}_{ML} = \frac{n}{\sum x_i} = \frac{1}{\overline{x}} = \frac{1}{0.52} \approx 1.9231$$

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

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

Информация Фишера определяется так:

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

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

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

Значит, информаиця Фишера:

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

И наконец, искомый доверительный интервал:

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