Школа анализа данных
Домашнее задание №1 по курсу
«Математическая Статистика в Машинном Обучении»

Кузина Е.М.

27 марта 2020

Задача 1 [1 балл]

Пусть $\boldsymbol{X}^n = \{X_1, X_2, \dots\}$ — независимые одинаково распределенные (н.о.р.) случайные величины с конечными средним $\mu = \mathbb{E}(X_1)$ и дисперсией $\sigma^2 = \mathbb{V}(X_1)$. Покажите, что величины

$$\langle \boldsymbol{X}^n \rangle = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \langle \boldsymbol{X}^n \rangle)^2.$$

являются *несмещенными* и *состоятельными* оценками среднего μ и дисперсии σ^2, т.е. что

- $\mathbb{E}(\langle \boldsymbol{X}^n \rangle) = \mu$ и $\langle \boldsymbol{X}^n \rangle \xrightarrow{\mathsf{P}} \mu$,
- $\mathbb{E}(\hat{S}_n^2) = \sigma^2$ и $\hat{S}_n^2 \xrightarrow{\mathsf{P}} \sigma^2$.

Замечание. *Конкретно в задачах статистики зачастую под $\boldsymbol{X}^n$ понимается выборка независимых значений случайной величины X. В таком случае $\langle \boldsymbol{X}^n \rangle$ и $\hat{S}_n^2$ — оценки среднего и дисперсии по выборке.*

Решение:

В этой задаче будем пользоваться свойствами математического ожидания и дисперсии для независимых случайных величин.

Для доказательства состоятельности будем пользоваться *признаком состоятельности:*

Пусть Θ - статистика, а $\hat{\Theta}$ - её оценка, тогда если $\mathbb{E}(\hat{\Theta} - \Theta)^2 \to 0$ при $n \to \infty$, то оценка является состоятельной. (https://yadi.sk/i/J18gYICvaq9vBA (стр. 7)).

1. Доказательство для оценки среднего $\langle \boldsymbol{X}^n \rangle$
   (a) несмещённость
   $\mathbb{E}(\langle \boldsymbol{X}^n \rangle) = \mathbb{E}(\frac{1}{n}\sum\limits_{i=1}^{n} X_i) = \frac{1}{n}\mathbb{E}(\sum\limits_{i=1}^{n} X_i) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{E}X_i = \frac{n}{n}\mu = \mu$
   (b) состоятельность $\mathbb{E}(\langle \boldsymbol{X}^n \rangle - \mu)^2 = \mathbb{V}(\langle \boldsymbol{X}^n \rangle) = \mathbb{V}(\frac{1}{n}\sum\limits_{i=1}^{n} X_i) = \frac{1}{n^2}\mathbb{V}(\sum\limits_{i=1}^{n} X_i) = \frac{1}{n^2}\sum\limits_{i=1}^{n} \mathbb{V}X_i = \frac{1}{n}\mathbb{V}X_i =$
   $= \frac{\sigma^2}{n} \to 0$ при $n \to \infty$ => по признаку состоятельности оценка состоятельна
2. Доказательство для оценки дисперсии $\hat{S}_n^2$
   (a) несмещённость
   $\sum\limits_{i=1}^{n}(X_i - \langle \boldsymbol{X}^n \rangle)^2 = \sum\limits_{i=1}^{n}((X_i - \mu) - (\langle \boldsymbol{X}^n \rangle - \mu))^2 = \sum\limits_{i=1}^{n}\left((X_i-\mu)^2 - 2(X_i-\mu)(\langle \boldsymbol{X}^n \rangle - \mu) + (\langle \boldsymbol{X}^n \rangle - \mu)^2\right) =$
   $= \sum\limits_{i=1}^{n}(X_i-\mu)^2 - 2(\langle \boldsymbol{X}^n \rangle - \mu)\sum\limits_{i=1}^{n}(X_i - \mu) + n(\langle \boldsymbol{X}^n \rangle - \mu)^2 = \sum\limits_{i=1}^{n}(X_i-\mu)^2 - 2n(\langle \boldsymbol{X}^n \rangle - \mu)^2 + n(\langle \boldsymbol{X}^n \rangle - \mu)^2 =$
   $= \sum\limits_{i=1}^{n}(X_i-\mu)^2 - n(\langle \boldsymbol{X}^n \rangle - \mu)^2$
   $\mathbb{E}(\hat{S}_n^2) = \mathbb{E}(\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \langle \boldsymbol{X}^n \rangle)^2) = \frac{1}{n-1}\mathbb{E}\sum\limits_{i=1}^{n}(X_i - \langle \boldsymbol{X}^n \rangle)^2 = \frac{1}{n-1}(\mathbb{E}\sum\limits_{i=1}^{n}(X_i-\mu)^2 - \mathbb{E}(n(\langle \boldsymbol{X}^n \rangle - \mu)^2)) =$
   $= \frac{1}{n-1}(n\mathbb{E}(X_i-\mu)^2 - n\mathbb{E}(\langle \boldsymbol{X}^n \rangle - \mu)^2) = \frac{n}{n-1}(\mathbb{V}X_i - \mathbb{V}\langle \boldsymbol{X}^n \rangle) = (1) = \frac{n}{n-1}(\sigma^2 - \frac{\sigma^2}{n}) = \sigma^2$
   (b) состоятельность
   $\mathbb{E}(\hat{S}_n^2 - \sigma^2)^2 = \mathbb{V}(\hat{S}_n^2) = \mathbb{V}(\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \langle \boldsymbol{X}^n \rangle)^2) = \frac{n}{(n-1)^2}\mathbb{V}(X_i - \langle \boldsymbol{X}^n \rangle)^2 \to 0$ при $n \to \infty$

Задача 2 [1 балла]

Пусть $\boldsymbol{X}^n = \{X_1, X_2, \dots, X_n\}$ и $\boldsymbol{Y}^m = \{Y_1, Y_2, \dots, Y_m\}$ — две выборки н.о.р. случайных величин объема n и m, полученных из одного и того же распределения. Пусть $\hat{S}_X^2$ и $\hat{S}_Y^2$ — несмещенные оценки дисперсий по выборкам $\boldsymbol{X}^n$

и $\boldsymbol{Y}^m$ соответственно. Выразите несмещенную оценку дисперсии $\hat{S}_{X,Y}$ суммарной выборки через $\hat{S}^2_X$ и $\hat{S}^2_Y$ и средние $\langle \boldsymbol{X}^n \rangle$ и $\langle \boldsymbol{Y}^m \rangle$.

Решение:

Обозначим суммарную выборку через $\langle \boldsymbol{XY}^{n+m} \rangle = \{X_1, \dots, X_n, Y_1, \dots, Y_m\}$

Оценка для матожидания $\boldsymbol{XY}^{n+m} = \frac{1}{n+m} \sum\limits_{i=1}^{n+x} XY_i = \frac{1}{n+m} \left(\sum\limits_{i=1}^{n} X_i + \sum\limits_{i=0}^{m} Y_i \right) = \frac{n\langle \boldsymbol{X}^n \rangle + m \langle \boldsymbol{Y} \rangle^m}{n+m}$

Несмещённая оценка дисперсии $\hat{S}^2_{X,Y} = \frac{1}{n+m-1} \sum\limits_{i=1}^{n+m} (XY_i - \langle \boldsymbol{XY}^{n+m} \rangle)^2 = \frac{1}{n+m-1} \left(\sum\limits_{i=1}^{n} (X_i - \langle \boldsymbol{XY}^{n+m} \rangle)^2 + \sum\limits_{i=1}^{m} (Y_i - \right.$ $\left. - \langle \boldsymbol{XY}^{n+m} \rangle)^2 \right)$

$$\sum_{i=1}^{n} (X_i - \left\langle \boldsymbol{XY}^{n+m} \right\rangle)^2 = \frac{1}{(n+m)^2} \sum_{i=1}^{n} ((n+m)X_i - n \left\langle \boldsymbol{X}^n \right\rangle - m \left\langle \boldsymbol{Y}^m \right\rangle)^2 = \frac{1}{(n+m)^2} \sum_{i=1}^{n} (n(X_i - \langle \boldsymbol{X}^n \rangle) + m(X_i - \langle \boldsymbol{Y}^m \rangle))^2 =$$

$$= \frac{1}{(n+m)^2} \sum_{i=1}^{n} (n^2(X_i - \langle \boldsymbol{X}^n \rangle)^2 + 2nm(X_i - \langle \boldsymbol{X}^n \rangle)(X_i - \langle \boldsymbol{Y}^m \rangle) + m^2(X_i - \langle \boldsymbol{Y}^m \rangle)^2 =$$

$$= \frac{1}{(n+m)^2} \left(n^2(n-1)\hat{S}^2_X + 2nm \sum_{i=1}^{n} (X_i - \langle \boldsymbol{X}^n \rangle)(X_i - \langle \boldsymbol{Y}^m \rangle) + m^2 \sum_{i=1}^{n} (X_i - \langle \boldsymbol{Y}^m \rangle)^2 \right)$$

Аналогично: $\sum\limits_{i=1}^{m} (Y_i - \langle \boldsymbol{XY}^{n+m} \rangle)^2 = \frac{1}{(n+m)^2} \left(n^2 \sum\limits_{i=1}^{m} (Y_i - \langle \boldsymbol{X}^n \rangle)^2 + 2nm \sum\limits_{i=1}^{m} (Y_i - \langle \boldsymbol{X}^n \rangle)(Y_i - \langle \boldsymbol{Y}^m \rangle) + m^2(m-1)\hat{S}^2_Y \right)$

Тогда $\hat{S}^2_{X,Y} = \frac{1}{(n+m-1)(n+m)^2} \left(n^2(n-1)\hat{S}^2_X + 2nm \sum\limits_{i=1}^{n} (X_i - \langle \boldsymbol{X}^n \rangle)(X_i - \langle \boldsymbol{Y}^m \rangle) + m^2 \sum\limits_{i=1}^{m} (X_i - \langle \boldsymbol{Y}^m \rangle)^2 + n^2 \sum\limits_{i=1}^{m} (Y_i - \right.$ $\left. \langle \boldsymbol{X}^n \rangle)^2 + 2nm \sum\limits_{i=1}^{m} (Y_i - \langle \boldsymbol{X}^n \rangle)(Y_i - \langle \boldsymbol{Y}^m \rangle) + m^2(m-1)\hat{S}^2_Y \right) = \frac{1}{(n+m-1)(n+m)^2} \left(n^2(n-1)\hat{S}^2_X + m^2(m-1)\hat{S}^2_Y + (2nm + \right.$ $\left. +m^2) \sum\limits_{i=1}^{n} {X_i}^2 - 2nm(n+m) \left\langle \boldsymbol{X}^n \right\rangle \left\langle \boldsymbol{Y}^m \right\rangle - mn^2 \left\langle \boldsymbol{X}^n \right\rangle^2 + (2nm + n^2) \sum\limits_{i=1}^{m} {Y_i}^2 - nm^2 \left\langle \boldsymbol{Y}^m \right\rangle^2 \right) =$

$$\sum_{i=1}^{m} {Y_i}^2 = \sum_{i=1}^{m} (Y_i - \langle \boldsymbol{Y}^m \rangle + \langle \boldsymbol{Y}^m \rangle)^2 = (m-1)\hat{S}^2_Y + m \left\langle \boldsymbol{Y}_m \right\rangle^2$$

$$\sum_{i=1}^{n} {X_i}^2 = (n-1)\hat{S}^2_X + n \left\langle \boldsymbol{X}_n \right\rangle^2$$

$$= \frac{1}{(n+m-1)(n+m)^2} \left((n+m)^2(n-1)\hat{S}^2_X + (n+m)^2(m-1)\hat{S}^2_Y + nm(n+m)(\langle \boldsymbol{X}^n \rangle - \langle \boldsymbol{Y}^m \rangle)^2 \right)$$

$$= \frac{1}{n+m-1} \left((n-1)\hat{S}^2_X + (m-1)\hat{S}^2_Y + \frac{nm(\langle \boldsymbol{X}^n \rangle - \langle \boldsymbol{Y}^m \rangle)^2}{n+m} \right)$$

Ответ: $\frac{1}{n+m-1} \left((n-1)\hat{S}^2_X + (m-1)\hat{S}^2_Y + \frac{nm(\langle \boldsymbol{X}^n \rangle - \langle \boldsymbol{Y}^m \rangle)^2}{n+m} \right)$

Задача 3 [2 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \dots, X_n\} \sim \mathrm{Exp}(\lambda)$, $\hat{\lambda} = 1/\left\langle \boldsymbol{X}^n \right\rangle$. Найдите bias, se, MSE этой оценки. Является ли оценка смещенной? Состоятельной?

Задача 4 [2 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \dots, X_n\} \sim \mathcal{N}(0, \sigma^2)$. Пусть для оценки параметра σ нормального распределения используется выборочное линейное отклонение $\hat{\sigma} = \langle |\boldsymbol{X}^n| \rangle = n^{-1} \sum_{i=1}^{n} |X_i|$. Найдите bias, se, MSE оценки $\hat{\sigma}$. Является ли оценка несмещенной? Если «нет», то постройте исправленную оценку. Найдите se исправленной оценки. Является ли исправленная оценка $\hat{\sigma}$ состоятельной?

Решение:

$\mathbb{E}\hat{\sigma} = \mathbb{E}|X_i| = \int\limits_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}} dx = \frac{2\sigma}{\sqrt{2\pi}}$

$bias(\hat{\sigma}) = \mathbb{E}\hat{\sigma} - \sigma = (\frac{2}{\sqrt{2\pi}} - 1)\sigma$

$se(\hat{\sigma}) = \sqrt{\mathbb{V}(\hat{\sigma})} = \sqrt{\frac{1}{n}\mathbb{V}|X_i|} = \sqrt{\frac{1}{n}\int\limits_{-\infty}^{+\infty}(|x| - \frac{2\sigma}{\sqrt{2\pi}})^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}} dx} = \sigma\sqrt{\frac{1}{n}(1 - \frac{2}{\pi})}$

$MSE = bias^2(\hat{\sigma}) + \mathbb{V}(\hat{\sigma}) = (\frac{2}{\sqrt{2\pi}} - 1)^2\sigma^2 + \frac{1}{n}(1 - \frac{2}{\pi})\sigma^2$

Оценка является смещенной, так как $\mathbb{E}\hat{\sigma} \neq \sigma$. Исправленная оценка: $\frac{\sqrt{2\pi}}{2}\hat{\sigma} = \frac{\sqrt{2\pi}}{2n}\sum\limits_{i=1}^{n}|X_i|$.

$se(\frac{\sqrt{2\pi}}{2}\hat{\sigma}) = \sqrt{\mathbb{V}(\frac{\sqrt{2\pi}}{2}\hat{\sigma})} = \sqrt{\frac{\pi}{2}\mathbb{V}(\hat{\sigma})} = \sigma\sqrt{\frac{\pi}{2n}(1 - \frac{2}{\pi})}$

Исправленная оценка является состоятельной, так как $\mathbb{V}(\frac{\sqrt{2\pi}}{2}\hat{\sigma}) = \frac{\pi}{2n}(1 - \frac{2}{\pi})\sigma^2 \to 0$ при $n \to \infty$

Задача 5 [3 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \ldots, X_n\} \sim \mathcal{N}(\mu, \sigma^2)$, $\theta = e^\mu$ и $\hat{\theta} = e^{\langle \boldsymbol{X}_n \rangle}$. Найдите аналитически плотность распределения $p_{\hat{\theta}}(x)$ оценки $\hat{\theta} = e^{\langle \boldsymbol{X}^n \rangle}$, математическое ожидание $\mathbb{E}(\hat{\theta})$, и дисперсию $\mathbb{V}(\hat{\theta})$, а также bias, se, MSE оценки $\hat{\theta}$. Является ли оценка $\hat{\theta}$ смещенной? Состоятельной?

Решение:

$X_i \sim \mathcal{N}(\mu, \sigma^2). X_i$ независимы, поэтому $\sum\limits_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$.

Пусть $\sum\limits_{i=1}^{n} X_i = x$. Тогда:

$\mathbb{E}(\hat{\theta}) = \int\limits_{-\infty}^{+\infty} e^{\frac{x}{n}} \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{(x-n\mu)^2}{2n\sigma^2}} dx = e^{\mu + \frac{\sigma^2}{2n}}$

$\mathbb{V}(\hat{\theta}) = \mathbb{E}(\hat{\theta}^2) - (\mathbb{E}\hat{\theta})^2 = \int\limits_{-\infty}^{+\infty} e^{\frac{2x}{n}} \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{(x-n\mu)^2}{2n\sigma^2}} dx - e^{2\mu + \frac{\sigma^2}{n}} = e^{2\mu + \frac{2\sigma^2}{n}} - e^{2\mu + \frac{\sigma^2}{n}} = e^{2\mu + \frac{\sigma^2}{n}}\left(e^{\frac{\sigma^2}{n}} - 1\right)$

$bias(\hat{\theta}) = \mathbb{E}\hat{\theta} - \theta = e^{\mu + \frac{\sigma^2}{2n}} - e^\mu$

$se(\hat{\theta}) = \sqrt{\mathbb{V}\hat{\theta}} = \sqrt{e^{2\mu + \frac{\sigma^2}{n}}\left(e^{\frac{\sigma^2}{n}} - 1\right)}$

$MSE = bias^2(\hat{\theta}) + \mathbb{V}(\hat{\theta}) = e^{2\mu}\left(e^{\frac{2\sigma^2}{n}} - 2e^{\frac{\sigma^2}{2n}} + 1\right)$

Оценка является смещенной, тк $\mathbb{E}\hat{\theta} \neq \theta$, но состоятельной ($\mathbb{V}\hat{\theta} \to 0$, при $n \to \infty$)

Плотность распределения $p_{\hat{\theta}}(x)$:

$F_{\hat{\theta}}(x) = P(\hat{\theta} < x) = P(e^{\langle \boldsymbol{X}_n \rangle} < x) = P(\sum\limits_{i=1}^{n} X_i < n\ln(x)) = F_{\sum\limits_{i=1}^{n} X_i}(n\ln(x))$

$p_{\hat{\theta}}(x) = \frac{\partial F_{\sum\limits_{i=1}^{n} X_i}(n\ln(x))}{\partial x} = p_{\sum\limits_{i=1}^{n} X_i}(n\ln(x)) * \frac{n}{x} = \frac{n}{x\sqrt{2\pi n\sigma^2}} e^{-\frac{(n\ln(x) - n\mu)^2}{2n\sigma^2}}$

Задача 6 [2 балла]

Пусть $\hat{F}_n(x)$ — эмпирическая функция распределения. Пусть x, $y \in \mathbb{R}$. Найдите ковариацию $\text{Cov}(\hat{F}_n(x), \hat{F}_n(y))$.

Решение: $\text{Cov}(\hat{F}_n(x), \hat{F}_n(y)) = \mathbb{E}\left[(\hat{F}_n(x) - \mathbb{E}\hat{F}_n(x))(\hat{F}_n(y) - \mathbb{E}\hat{F}_n(y))\right] = \mathbb{E}\left(\hat{F}_n(x)\hat{F}_n(y)\right) - \mathbb{E}\hat{F}_n(x)\mathbb{E}\hat{F}_n(y) =$

$= \mathbb{E}\left(\hat{F}_n(x)\hat{F}_n(y)\right) - F(x)F(y)$

Здесь было использовано утверждение $\mathbb{E}\hat{F}_n(x) = \mathbb{E}\frac{1}{n}\sum\limits_{i=1}^{n} I(X_i \leq x) = \frac{n}{n}\mathbb{E}I(X_i \leq x) = F(x)$, где $I(X_i \leq x)$ - индикаторная случайная величина, имеющая распределение Бернулли

Рассмотрим два случая:

1. $x = y$
   Тогда $\hat{F}_n(x)$ и $\hat{F}_n(y)$ зависимы, получаем $\text{Cov}(\hat{F}_n(x), \hat{F}_n(y)) = \mathbb{E}(\hat{F}_n(x))^2 - F^2(x)$
   $\mathbb{E}(\hat{F}_n(x))^2 = \mathbb{V}\hat{F}_n(x) + \mathbb{E}^2\hat{F}_n(x) = \mathbb{V}\frac{1}{n}\sum\limits_{i=1}^{n} I(X_i \leq x) + F^2(x) = \frac{1}{n}\mathbb{V}I(X_i \leq x) + F^2(x) = \frac{F(x)(1-F(x))}{n} + F^2(x)$
   В итоге: $\text{Cov}(\hat{F}_n(x), \hat{F}_n(y)) = \frac{F(x)(1-F(x))}{n} + F^2(x) - F^2(x) = \frac{F(x)(1-F(x))}{n}$
2. $x \neq y$
   Тогда $\hat{F}_n(x)$ и $\hat{F}_n(y)$ не зависимы, получаем $\text{Cov}(\hat{F}_n(x), \hat{F}_n(y)) = \mathbb{E}\hat{F}_n(x)\mathbb{E}\hat{F}_n(y) - F(x)F(y) = F(x)F(y) - F(x)F(y) = 0$

Ответ:

$\text{Cov}(\hat{F}_n(x), \hat{F}_n(y)) = \begin{cases} 0, x = y \\ \frac{F(x)(1-F(x))}{n}, x \neq y \end{cases}$

Задача 7 [2 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \dots, X_n\} \sim F(x)$, и пусть $\hat{F}_n(x)$ — эмпирическая функция распределения. Для фиксированных числе $a, b \in \mathbb{R}$, таких что $a < b$ определим статистический функционал $T(F) = F(b) - F(a)$. Пусть $\hat{\theta} = \hat{F}_n(b) - \hat{F}_n(a)$. Найдите оценку $\hat{\text{se}}$ стандартного отклонения и $(1-\alpha)$-доверительный интервал.

Решение: $I(X_i \leq x)$-индикаторная величина, имеющая распределение Бернулли

$$\mathbb{E}\hat{\theta} = \mathbb{E}\hat{F}_n(b) - \mathbb{E}\hat{F}_n(a) = \frac{n}{n}(\mathbb{E}I(X_i \leq b) - \mathbb{E}I(X_i \leq a)) = F(b) - F(a)$$

$$\mathbb{V}\hat{\theta} = \mathbb{V}\hat{F}_n(b) + \mathbb{V}\hat{F}_n(a) = \frac{1}{n}(\mathbb{V}I(X_i \leq b) + \mathbb{V}I(X_i \leq a)) = \frac{1}{n}\Big(F(b)(1-F(b)) + F(a)(1-F(a))\Big)$$

$$se(\hat{\theta}) = \sqrt{\mathbb{V}\hat{\theta}}, \text{ поэтому } \hat{\text{se}} = \sqrt{\frac{1}{n}\Big(\hat{F}_n(b)(1-\hat{F}_n(b)) + \hat{F}_n(a)(1-\hat{F}_n(a))\Big)}$$

Доверительный интервал:

$$\frac{\hat{\theta}-\theta}{\hat{\text{se}}} \sim N(0,1) \text{ по ЦПТ}$$

$$\left|\frac{\hat{\theta}-\theta}{\hat{\text{se}}}\right| \leq t_{\alpha/2}$$

$$-t_{\alpha/2} \leq \frac{\hat{\theta}-\theta}{\hat{\text{se}}} \leq t_{\alpha/2}$$

$$\hat{\theta} - t_{\alpha/2}\hat{\text{se}} \leq \theta \leq \hat{\theta} + t_{\alpha/2}\hat{\text{se}}$$

$$\hat{F}_n(b) - \hat{F}_n(a) - t_{\alpha/2}\sqrt{\frac{1}{n}\Big(\hat{F}_n(b)(1-\hat{F}_n(b)) + \hat{F}_n(a)(1-\hat{F}_n(a))\Big)} \leq \theta \leq \hat{F}_n(b) - \hat{F}_n(a) + t_{\alpha/2}\sqrt{\frac{1}{n}\Big(\hat{F}_n(b)(1-\hat{F}_n(b)) + \hat{F}_n(a)(1-\hat{F}_n(a))\Big)}$$

Получили доверительный интервал для θ

Задача 8 [2 балла]

Скачайте данные о качестве красных вин. Постройте график для $\hat{F}(x; \boldsymbol{x}^n)$ для уровня кислотности (pH). Для каждой точки x постройте:

- 95%-ый доверительный интервал на основе неравенства Дворецкого-Кифера-Вольфовица.
- Асимптотический нормальный 95%-ый доверительный интервал для значения $F(x)$.

По значениям уровня кислотности $\boldsymbol{x}^n$ подсчитайте оценку $T(\boldsymbol{x}^n)$ для функционала $T(F) = F(3.5) - F(3.4)$ и найдите оцените аналитически стандартное отклонение $\hat{\text{se}}$ оценки $T(\boldsymbol{x}^n)$. Постройте асимптотический нормальный 95%-ый доверительный интервал для $T(F)$.

Решение в приложенном ноутбуке

Задача 9 [2 балла]

В процессе очистки питьевой воды выпадает значительный осадок. Для его уменьшения можно воздействовать на разные факторы, в т.ч. на количество микроорганизмов в жидкости, способствующих окислению органики. В группу из 261 очистительных установок был добавлен реагент, подавляющих активность микроорганизмов, а состав остальных 119 остался без изменений. Пусть θ — разность в средних значениях количества твердых частиц в этих двух группах установок. Оценить по данным `WaterTreatment` величину θ, оценить стандартную ошибку оценки, построить 95% и 99% доверительные интервалы. Какие выводы можно сделать на основе полученных результатов?

Решение в приложенном ноутбуке

Задача 10 [2 балла]

Провести моделирование, чтобы сравнить различные типы доверительных интервалов, построенных с помощью бутстрепа. Пусть $n = 50$, $T(F) = \int (x-\mu)^3 \, dF(x)/\sigma^3$ — коэффициент асимметрии, где F — логнормальное распределение. Постройте 95% доверительные интервалы для $T(F)$ (под F понимается распределение элементов выборки $X_1, \dots, X_n$) по данным $\boldsymbol{X}^n = \{X_1, \dots, X_n\}$, используя три подхода на основе бутстрепа.

Замечание. *Выборку из логнормального распределения можно сгенерировать из нормального, сначала сгенерировав выборку н.о.р. величин* $\boldsymbol{Y}^n = \{Y_1, \dots, Y_n\} \sim \mathcal{N}(0,1)$*, после чего положив* $X_i = e^{Y_i}$*,* $i = 1, 2, \dots, n$*.*

Решение в приложенном ноутбуке

Задача 11 [2 балла]

Пусть $\boldsymbol{X}^n = \{X_1, \ldots, X_n\} \sim \mathcal{N}(\mu, 1)$, $\theta = e^{\mu}$ и $\hat{\theta} = e^{\langle \boldsymbol{X}^n \rangle}$. Сгенерируйте выборку $\boldsymbol{X}^n$ из $n = 100$ наблюдений для $\mu = 10$. Нарисуйте гистограмму значений $\{\hat{\theta}_i^*\}_{i=1}^B$ бутстрепных оценок. Эта гистограмма является оценкой распределения $p_{\hat{\theta}}(x)$. Сравните ее с настоящим распределением $p_{\hat{\theta}}(x)$. Используя бутстреп, подсчитайте величину se и постройте тремя способами 95% доверительный интервал для θ.