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

Винер Даниил @danya_vin

Версия от 21 февраля 2025 г.

Содержание

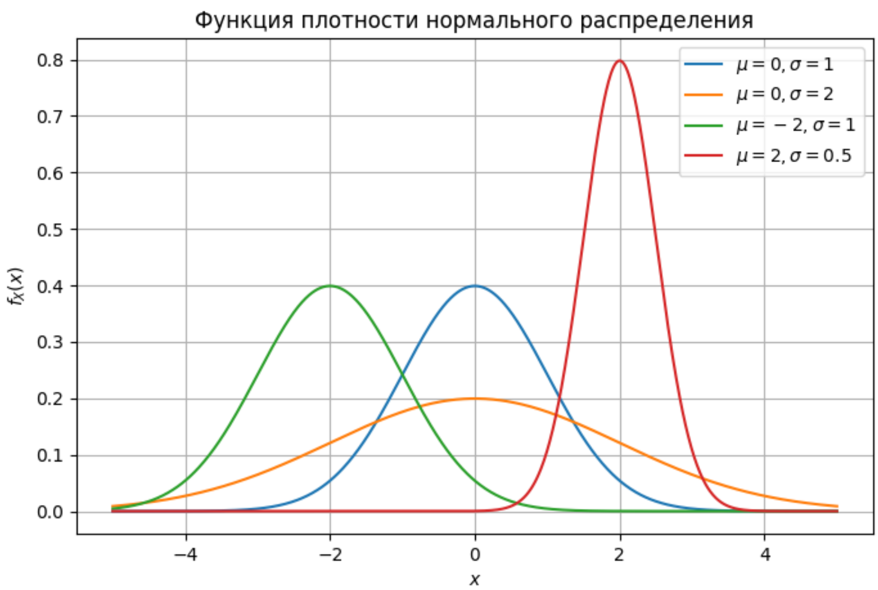
1 Теоретический минимум

1.1 Дайте определение нормально распределённой случайной величины. Для неё укажите диапазон возможных значений, функцию плотности, ожидание, дисперсию. Нарисуйте функцию плотности

Определение. Случайная величина имеет нормальное распределение $X \sim \mathcal{N}(\mu, \sigma^2)$, если функция плотности равна

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

- Случайная величина $X \sim \mathcal{N}(\mu, \sigma^2)$ принимает любые значения $(-\infty, +\infty)$
- $\mathbb{E}[X] = \mu$
- $\mathbb{D}[X] = \sigma^2$



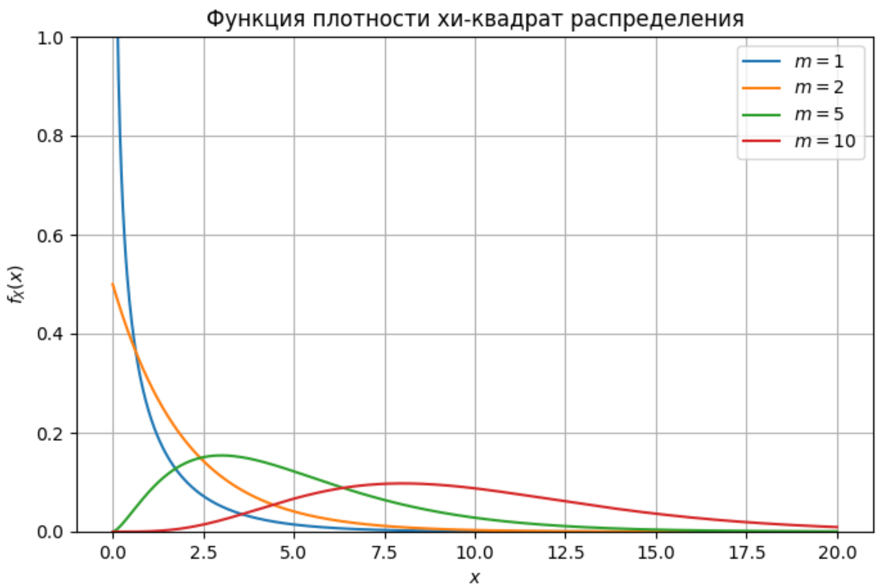
1.2 С помощью нормальных случайных величин дайте определение случайной величины, имеющей хи-квадрат распределение. Для хи-квадрат распределённой случайной величины укажите диапазон возможных значений, математическое ожидание и дисперсию. Нарисуйте функцию плотности при разных степенях свободы

Определение. Случайная величина X имеет χ^2-распределение с m степенями свободы, $H \sim \chi^2(m)$, если X представима в виде

$$X = X_1^2 + \ldots + X_m^2,$$

где $X_1, \ldots, X_m \sim iidN(0,1)$ (independent identically distributed normal — независимые одинаково распределенные нормально распределенные случайные величины)

- Принимает значения $x \in [0; +\infty]$
- $\mathbb{E}[X] = m$
- $\mathbb{D}[X] = 2m$



1.3 С помощью нормальных случайных величин дайте определение случайной величины, имеющей распределения Стьюдента. Для случайной величины, распределённой по Стьюденту, укажите диапазон возможных значений. Нарисуйте функцию плотности распределения Стьюдента при разных степенях свободы на фоне нормальной стандартной функции плотности.

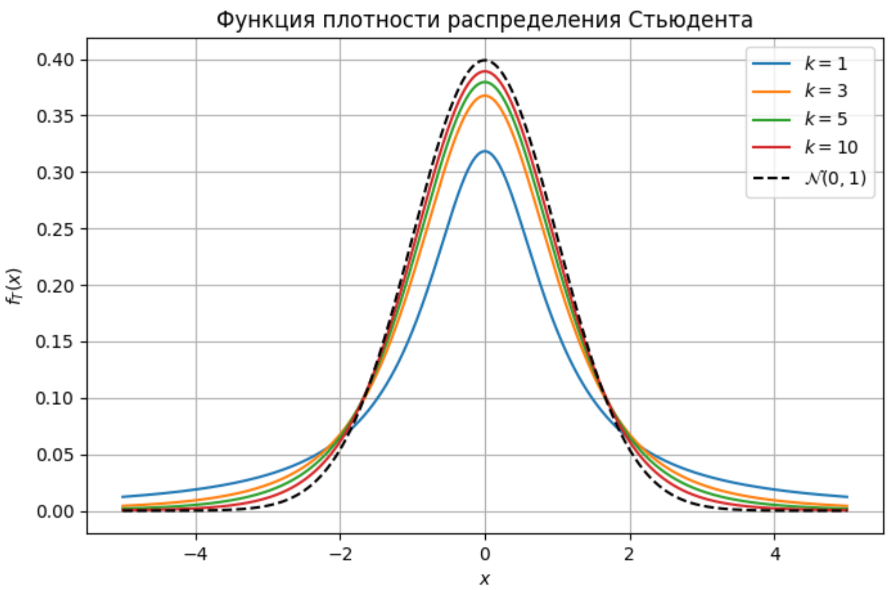
Определение. Пусть $\xi_0, \xi_1, \ldots, \xi_k$ независимы и имеют стандартное нормальное распределение. Распределение случайной величины

$$t_k = \frac{\xi_0}{\sqrt{\frac{\xi_1^2 + \ldots + \xi_k^2}{k}}}$$

называется распределением Стьюдента с k степенями свободы и обозначается T_k.

Распределение Стьюдента совпадает с распределением случайной величины $t_k = \frac{\xi}{\sqrt{\chi_k^2/k}}$, где $\xi \in \mathcal{N}(0,1)$ и $\chi_k^2 \in \mathrm{H}_k$ независимы.

- Принимает значения в диапазоне $(-\infty; +\infty)$
- $\mathbb{E}\left[t_k\right] = 0$ при $k > 1$, а при $k = 1$ матожидание не существует, так как интеграл расходится (распределение Коши)
- $\mathbb{D}\left[t_k\right] = \dfrac{k}{k-2}$, $\exists$ только при $k > 2$



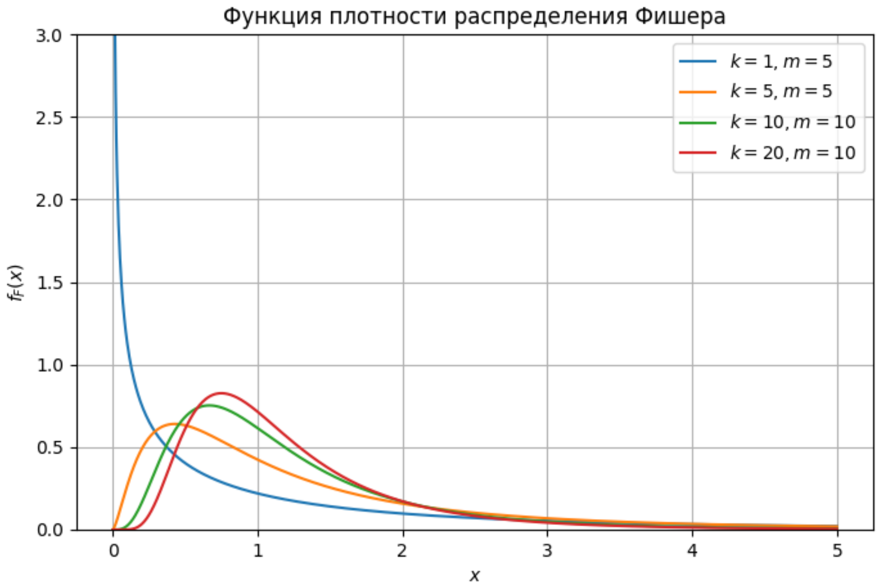
1.4 С помощью нормальных случайных величин дайте определение случайной величины, имеющей распределение Фишера. Для случайной величины, распределённой по Фишеру, укажите диапазон возможных значений. Нарисуйте возможную функцию плотности

Определение. Случайная величина $F \sim F(k, m)$ имеет распределение Фишера с параметрами k и m (степени свободы), если она может быть представлена в виде:

$$F = \frac{X_k/k}{X_m/m}$$

где $X_k \sim \chi^2(k),\ X_m \sim \chi^2(m)$ и $X_k,\ X_m$ — независимы

- F принимает значения из диапазона $[0; +\infty]$



1.5 Дайте определение выборочного среднего и выборочной дисперсии

Определение. Выборочное среднее — $\overline{X} := \dfrac{X_1 + \ldots + X_n}{n}$

Определение. Неисправленная выборочная дисперсия — $s^2 := \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$

Определение. Исправленная выборочная дисперсия — $\widehat{\sigma}^2 := \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$

1.6 Дайте определение выборочного начального и выборочного центрального момента порядка k

Пусть дана выборка $X = (X_1, \ldots, X_n)$

Определение. Выборочным *начальным* моментом порядка k называется число

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

Определение. Выборочным *центральным* моментом порядка k называется число

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

1.7 Дайте определение выборочной функции распределения

Определение. Выборочная (эмпирическая) функция распределения — это

$$\hat{F}_n(x)=\frac{1}{n}\sum_{i=1}^{n}I\{X_i\leqslant x\},$$

где $I\{X_i\leqslant x\}$ — индикаторная функция, равная 1, если $X_i\leqslant x$, и 0 в противном случае

1.8 Выпишите формулу несмещённой оценки дисперсии

Определение. $\widehat{\sigma}^2:=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2$

1.9 Дайте определение несмещённой оценки $\widehat{\theta}$ параметра θ

Определение. Оценка $\hat{\theta}$ неизвестного параметра $\theta\in\Theta$ называется *несмещённой*, если

$$\mathbb{E}\left[\hat{\theta}\right]=\theta\ \forall\theta\in\Theta,$$

где Θ — множество всех параметров θ

1.10 Дайте определение состоятельной последовательности оценок $\widehat{\theta}_n$; Укажите условия на $\mathbb{E}\left[\widehat{\theta}_n\right]$ и $\mathbb{D}\left[\widehat{\theta}_n\right]$, достаточные для состоятельности

Определение. Последовательность оценок $\hat{\theta}_n$ называется *состоятельной оценкой* неизвестного параметра $\theta\in\Theta$, если

$$\forall\theta\in\Theta\ \hat{\theta}_n\xrightarrow[n\to\infty]{\mathbb{P}}\theta,$$

то есть $\hat{\theta}_n$ сходится по вероятности к θ

$$\forall\varepsilon>0:\ \lim_{n\to\infty}\mathbb{P}(|\hat{\theta}_n-\theta|\geqslant\varepsilon)=0$$

Теорема. Если оценка асимптотически несмещенная и её дисперсия стремится к нулю, то такая оценка будет состоятельной. То есть должны выполняться такие условия:

- Асимптотическая несмещённость — $\lim\limits_{n\to\infty}\mathbb{E}\left[\widehat{\theta}_n\right]=\theta$
- Сходимость дисперсии к нулю — $\lim\limits_{n\to\infty}\mathbb{D}\left[\widehat{\theta}_n\right]=0$

1.11 Дайте определение эффективности оценки $\widehat{\theta}$ среди множества оценок $\widehat{\Theta}_n$

Определение. Пусть $\mathcal{K}$ — некоторый класс оценок параметра θ. Оценка $\widehat{\theta}\in\mathcal{K}$ неизвестного параметра θ называется наиболее эффективной в классе $\mathcal{K}$, если для любого конкурента $\widetilde{\theta}\in\mathcal{K}\ \forall\theta\in\Theta$

$$\mathbb{E}\left[(\widehat{\theta}-\theta)^2\right]\leqslant\mathbb{E}\left[(\widetilde{\theta}-\theta)^2\right]$$

2 Задачный минимум

2.1 Для взрослого мужчины рост в сантиметрах, величина X, и вес в килограммах, величина Y...

X — рост, Y — вес, $Z = (X, Y)$, $\mathbb{E}[Z] = \begin{pmatrix} 175 \\ 74 \end{pmatrix}$, $V(Z) = \begin{pmatrix} 49 & 28 \\ 28 & 36 \end{pmatrix}$

а)

Вариант 1. Найдем вероятность противоположного события:

$$\begin{aligned}
\mathbb{P}(\{|X - \mathbb{E}[X]| \leqslant 10\}) &= \mathbb{P}(\{-10 \leqslant X - 175 \leqslant 10\}) \\
&\quad \text{проведем стандартизацию и нормализацию} \\
&= \mathbb{P}\left(\left\{\frac{-10}{7} \leqslant \frac{X - 175}{7} \leqslant \frac{10}{7}\right\}\right) \\
&= \Phi\left(\frac{10}{7}\right) - \Phi\left(-\frac{10}{7}\right) \\
&= 2\Phi\left(\frac{10}{7}\right) - 1 \\
&\approx 0.1531
\end{aligned}$$

Вариант 2. Найдем вероятность непосредственно искомого события

$$\begin{aligned}
\mathbb{P}(\{|X - \mathbb{E}[X]| > 10\}) &= \mathbb{P}\left(\left|\frac{X - 175}{\sqrt{49}}\right| > \frac{10}{\sqrt{49}}\right) \\
&\quad \text{пусть } \left|\frac{X - 175}{\sqrt{49}}\right| = |Z| \\
&= 2 \cdot \left(1 - \mathbb{P}\left(Z \leqslant \frac{10}{\sqrt{49}}\right)\right) \\
&\approx 0.153
\end{aligned}$$

б)

Так как Z имеет многомерное нормальное распределение, то U также имеет нормальное распределение

$$\begin{aligned}
\mathbb{E}[U] &= \mathbb{E}[X] - \mathbb{E}[Y] \\
&= 175 - 74 \\
&= 101 \\
\mathbb{D}[U] &= \mathbb{D}[X] + \mathbb{D}[Y] - 2cov(X, Y) \\
&= 49 + 46 - 2 \cdot 28 \\
&= 29
\end{aligned}$$

Стандартная функция плотности нормального распределения:

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

Тогда, в нашем случае:

$$f_U(u) = \frac{1}{\sqrt{2\pi 29}} \exp\left(-\frac{(u - 101)^2}{2 \cdot 29}\right)$$

в)

$$
\begin{aligned}
\mathbb{P}\left(\{U<90\}\right) &= \mathbb{P}\left(\left\{\frac{U-101}{\sqrt{29}} < \frac{90-101}{\sqrt{29}}\right\}\right)\\
&= \Phi(-2.0426)\\
&\approx 0.0205
\end{aligned}
$$

2.2 Рост в сантиметрах, случайная величина X, и вес в килограммах, случайная величина Y...

Теорема. Если $(X,Y) \sim N\left(\begin{pmatrix}\mu x\\ \mu y\end{pmatrix}, \begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_y\sigma_x & \sigma_y^2\end{pmatrix}\right)$, где $\rho = \dfrac{cov(X,Y)}{\sigma_X\sigma_Y}$ — коэффициент корреляции, то

$$
\begin{aligned}
(X|Y=y) &\sim N\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y),\ \sigma_x^2\left(1-\rho^2\right)\right)\\
(Y|X=x) &\sim N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x),\ \sigma_y^2\left(1-\rho^2\right)\right)
\end{aligned}
$$

а)

$$
\rho = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{28}{\sqrt{49\cdot 36}} = \frac{2}{3}
$$

$$
\begin{aligned}
(Y|X=170) &\sim N\left(\underbrace{74 + \frac{2}{3}\cdot\frac{6}{7}\cdot(170-175)}_{\mathbb{E}[Y|X=170]=71.14};\ \underbrace{36\left(1-\frac{4}{9}\right)}_{\mathbb{D}[Y|X=170]}\right)\\
(Y|X=170) &\sim N\left(71.1; 20\right)
\end{aligned}
$$

б)

$$
f_{Y|X=170}(a) = \frac{1}{\sqrt{2\pi 20}}\exp\left(-\frac{(a-71.14)^2}{2\cdot 20}\right)
$$

в)

$$
\begin{aligned}
\mathbb{P}\left(\{Y>90|X=170\}\right) &= \mathbb{P}\left(\left\{\frac{Y-71.14}{\sqrt{20}} > \frac{90-71.14}{\sqrt{20}}\right\}\right)\\
&= 1-\mathbb{P}\left(\left\{\frac{Y-71.14}{\sqrt{20}} \leqslant \frac{90-71.14}{\sqrt{20}}\right\}\right)
\end{aligned}
$$

2.3 Для реализации случайной выборки $x = (1, 0, -1, 1)$ найдите...

a) выборочное среднее

$$
\overline{X} = \frac{X_1+\ldots+X_n}{n} = \frac{1+0-1+1}{4} = \frac{1}{4}
$$

b) неисправленную выборочную дисперсию

$$
\begin{aligned}
s^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\\
&= \frac{1}{4}\left(\left(1-\frac{1}{4}\right)^2 + \left(0-\frac{1}{4}\right)^2 + \left(-1-\frac{1}{4}\right)^2 + \left(1-\frac{1}{4}\right)^2\right)\\
&\approx 0.68
\end{aligned}
$$

c) исправленную выборочную дисперсию

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}s^2$$
$$\approx 0.91$$

d) выборочный второй начальный момент

Выборочный k-й начальный момент $= \dfrac{1}{n}\sum_{i=1}^{n}(X_i)^k$

В нашем случае:

$$\frac{1}{4}\left(1^2 + 0^2 + (-1)^2 + 1^2\right) = \frac{3}{4}$$

e) выборочный третий центральный момент

Выборочный k-й центральный момент $= \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k$

В нашем случае:

$$\frac{1}{4}\left(\left(1-\frac{1}{4}\right)^3 + \left(0-\frac{1}{4}\right)^3 + \left(-1-\frac{1}{4}\right)^3 + \left(1-\frac{1}{4}\right)^3\right) = -\frac{9}{32} \approx -0.28$$

2.4 Для реализации случайной выборки $x = (1, 0, -1, 1)$ найдите...

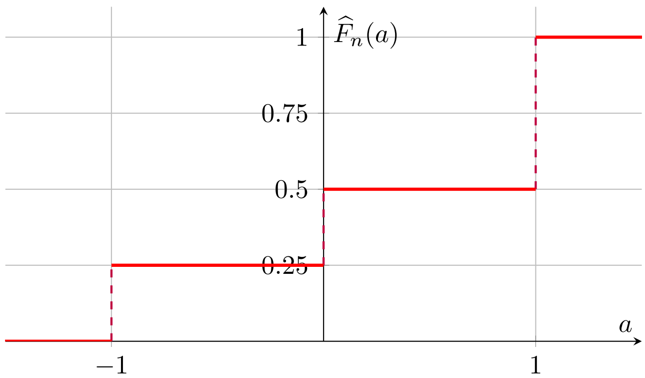
a) вариационный ряд

$$X = sort(list(X_1, \ldots, X_n))$$
$$= (-1, 0, 1, 1)$$

b) первый член вариационного ряда — -1

c) последний член вариационного ряда — 1

d) график выборочной функции распределения



2.5 Пусть $X_1, \ldots, X_n$ — случайная выборка из дискретного распределения, заданного с помощью таблицы

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

Рассмотрите оценку $\widehat{\theta} = \dfrac{\overline{X} + 2}{5}$

a)

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

b) Является ли оценка $\widehat{\theta}$ несмещённой оценкой неизвестного параметра θ?

Да, так как $\mathbb{E}\left[\widehat{\theta}\right] = \theta\ \forall\theta\in\Theta$ по доказанному в пункте **a)**

2.6 Пусть $X_1, \ldots, X_n$ — 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 и $\hat{\theta} = \overline{X}$

a) Является ли оценка $\widehat{\theta} = \overline{X}$ несмещённой оценкой неизвестного параметра θ?

$$\begin{aligned}
\mathbb{E}\left[\widehat{\theta}\right] &= \mathbb{E}\left[\overline{X}\right] \\
&= \mathbb{E}\left[X_i\right] \\
&= \int\limits_0^\theta x\cdot f(x)\mathrm{d}x \\
&= \frac{6}{\theta^3}\int\limits_0^\theta x^2(\theta-x)\mathrm{d}x \\
&= \frac{6}{\theta^3}\left(\theta\int\limits_0^\theta x^2\mathrm{d}x - \int\limits_0^\theta x^3\mathrm{d}x\right) \\
&= \frac{6}{\theta^3}\left(\theta\cdot\frac{\theta^3}{3}-\frac{\theta^4}{4}\right) \\
&= \frac{\theta}{2} \\
&\neq \theta
\end{aligned}$$

Значит, *не является несмещенной*

b) Подберите константу c так, чтобы оценка $\hat{\theta} = c\overline{X}$ оказалась несмещенной оценкой неизвестного параметра θ

$$\begin{aligned}
\mathbb{E}\left[c\overline{X}\right] = \theta &\Longrightarrow c\cdot\mathbb{E}\left[\overline{X}\right] = \theta \\
&\Longrightarrow c\cdot\frac{1}{2}\cdot\theta = \theta \\
&\Longrightarrow c = 2
\end{aligned}$$

2.7 Пусть X_1, X_2, X_3 — случайная выборка из распределения Бернулли с неизвестным параметром $p \in (0,1)$. Какие из следующих ниже оценкой являются несмещенными? Среди перечисленных ниже оценок найдите наиболее эффективную оценку...

Примечание. Так как выборка случайная, то

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mathbb{E}[X_3] = p$$
$$\mathbb{D}[X_1] = \mathbb{D}[X_2] = \mathbb{D}[X_3] = p(1-p)$$

- $\hat{p}_1 = \dfrac{X_1 + X_3}{2}$

  $\mathbb{E}[\hat{p}_1] = \dfrac{1}{2}\mathbb{E}[X_1 + X_3] = \dfrac{1}{2} \cdot 2\mathbb{E}[X_1] = p \Longrightarrow$ *несмещенная*

  $\mathbb{D}[\hat{p}_1] = \dfrac{1}{4}X_1 + \dfrac{1}{4}X_3 + 2 \cdot \dfrac{1}{2} \cdot \dfrac{1}{2} \underbrace{cov(X_1, X_3)}_{=0,\text{т.к.независ.}} = \dfrac{1}{4} \cdot \mathbb{D}[X_1] = \dfrac{1}{2}p(1-p)$

- $\hat{p}_2 = \dfrac{1}{4}X_1 + \dfrac{1}{2}X_2 + \dfrac{1}{4}X_3$

  $\mathbb{E}[\hat{p}_2] = \mathbb{E}\left[\dfrac{1}{4}X_1 + \dfrac{1}{2}X_2 + \dfrac{1}{4}X_3\right] = \dfrac{1}{4}\mathbb{E}[X_1] + \dfrac{1}{2}\mathbb{E}[X_2] + \dfrac{1}{4}\mathbb{E}[X_3] = p \Longrightarrow$ *несмещенная*

  $\mathbb{D}[\hat{p}_2] = \left(\dfrac{1}{16} + \dfrac{1}{4} + \dfrac{1}{16}\right)\mathbb{D}[X_1] = \dfrac{6}{16}\mathbb{D}[X_1]$

- $\hat{p}_3 = \dfrac{1}{3}X_1 + \dfrac{1}{3}X_2 + \dfrac{1}{3}X_3$

  $\mathbb{E}[\hat{p}_3] = \left(\dfrac{1}{3} + \dfrac{1}{3} + \dfrac{1}{3}\right)\mathbb{E}[X_1] = p \Longrightarrow$ *несмещенная*

  $\mathbb{D}[\hat{p}_3] = \left(\dfrac{1}{9} + \dfrac{1}{9} + \dfrac{1}{9}\right)\mathbb{D}[X_1] = \dfrac{1}{3}p(1-p)$

Рассмотрим дисперсии оценок, для них верно следующее

$$\mathbb{D}[\hat{p}_1] > \mathbb{D}[\hat{p}_2] > \mathbb{D}[\hat{p}_3]$$

Значит, оценка $\hat{p}_1$ наиболее эффективна, так как ее дисперсия наименьшая из данных

2.8 Пусть $X_1, \ldots, X_n$— случайная выборка из распределения с плотностью...

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

где $\theta > 0$ — неизвестный параметр. Является ли оценка $\widehat{\theta}_n = \frac{X_1 + \ldots + X_n}{n+1}$ состоятельной?

Примечание. По функции плотности распределения видно, что эта выборка из эскпоненциального распределения с параметром $\lambda = \frac{1}{\theta}$, и так как выборка случайная, то

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mathbb{E}[X_3] = \theta$$
$$\mathbb{D}[X_1] = \mathbb{D}[X_2] = \mathbb{D}[X_3] = \theta^2$$

Теорема. Если оценка *асимптотически* несмещенная и её дисперсия стремится к нулю, то такая оценка будет состоятельной

$$\left.\begin{aligned} &\mathbb{E}\left[\frac{X_1 + \ldots + X_n}{n+1}\right] = \frac{n}{n+1}\mathbb{E}[X_1] = \frac{n}{n+1}\theta \underset{n\to\infty}{\longrightarrow} \theta \\ &\mathbb{D}\left[\frac{X_1 + \ldots + X_n}{n+1}\right] = \frac{1}{(n+1)^2}\mathbb{D}[X_1 + \ldots + X_n] = \frac{n}{(n+1)^2}\theta^2 \underset{n\to\infty}{\longrightarrow} 0 \end{aligned}\right\} \Longrightarrow \text{оценка состоятельна}$$

2.9 Пусть $X_1, \ldots, X_n$ — 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. Является ли оценка $\hat{\theta}_n = \frac{2n+1}{n}\overline{X_n}$ состоятельной оценкой неизвестного параметра θ

$$\begin{aligned} \mathbb{E}\left[\widehat{\theta}\right] &= \frac{2n+1}{n}\mathbb{E}\left[\overline{X}\right] \\ &= \frac{2n+1}{n}\mathbb{E}\left[X_i\right] \\ &= \frac{2n+1}{n}\int\limits_0^\theta x \cdot f(x)\mathrm{d}x \\ &= \frac{2n+1}{n} \cdot \frac{1}{2} \cdot \theta \underset{n\to\infty}{\longrightarrow} \theta \end{aligned}$$

Найдем $\mathbb{D}\left[X\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}\left[X\right])^2$

$$\begin{aligned} \mathbb{D}\left[X\right] &= \mathbb{E}\left[X^2\right] - (\mathbb{E}\left[X\right])^2 \\ &= \int\limits_0^\theta x^2 f(x)\mathrm{d}x - \left(\frac{\theta}{2}\right)^2 \\ &= \frac{6}{\theta^3}\int\limits_0^\theta x^3(\theta - x)\mathrm{d}x - \left(\frac{\theta}{2}\right)^2 \\ &= \frac{6}{\theta^3} \cdot \frac{\theta^5}{20} - \left(\frac{\theta}{2}\right)^2 \\ &= \frac{\theta^2}{20} \end{aligned}$$

Тогда, $\mathbb{D}\left[\frac{2n+1}{n}\overline{X_n}\right] = \frac{(2n+1)^2}{n^2 \cdot n}\mathbb{D}\left[X\right] = \frac{(2n+1)^2}{n^3} \cdot \frac{\theta^2}{20} \underset{n\to\infty}{\longrightarrow} 0$

Значит, *оценка состоятельна*