

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
ПЕТРА ВЕЛИКОГО

ФИЗИКО-МЕХАНИЧЕСКИЙ ИНСТИТУТ

ВЫСШАЯ ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ВЫЧИСЛИТЕЛЬНОЙ ФИЗИКИ

**Отчёт**  
**по лабораторным работам №1-4**  
**по дисциплине**  
**«Математическая статистика»**

Выполнил студент:  
Желудев К.И.  
группа: 5030102/00101

Проверил:  
к.ф.-м.н., доцент  
Баженов Александр Николаевич

Санкт-Петербург  
2023 г.

# Содержание

<b>1</b>	<b>Постановка задачи</b>	<b>4</b>
<b>2</b>	<b>Теория</b>	<b>5</b>
2.1	Рассматриваемые распределения . . . . .	5
2.2	Гистограмма . . . . .	5
2.2.1	Построение гистограммы . . . . .	5
2.3	Вариационный ряд . . . . .	5
2.4	Выборочные числовые характеристики . . . . .	6
2.4.1	Характеристики положения . . . . .	6
2.4.2	Характеристики рассеяния . . . . .	6
2.5	Боксплот Тьюки . . . . .	7
2.5.1	Построение боксплота . . . . .	7
2.6	Теоретическая вероятность выбросов . . . . .	7
2.7	Эмпирическая функция распределения . . . . .	7
2.7.1	Статистический ряд . . . . .	7
2.7.2	Эмпирическая функция распределения . . . . .	7
2.7.3	Нахождение эмпирической функции распределения . . . . .	8
2.8	Оценки плотности распределения . . . . .	8
2.8.1	Определение . . . . .	8
2.8.2	Ядерные оценки . . . . .	8
<b>3</b>	<b>Реализация</b>	<b>10</b>
<b>4</b>	<b>Результаты</b>	<b>11</b>
4.1	Гистограмма и график плотности распределения . . . . .	11
4.2	Характеристики положения и рассеяния . . . . .	12
4.3	Боксплот Тьюки . . . . .	15
4.4	Доля выбросов . . . . .	17
4.5	Теоретическая вероятность выбросов . . . . .	18
4.6	Эмпирическая функция распределения . . . . .	18
4.7	Ядерные оценки плотности распределения . . . . .	20
<b>5</b>	<b>Обсуждение</b>	<b>25</b>
5.1	Гистограмма и график плотности распределения . . . . .	25
5.2	Характеристики положения и рассеяния . . . . .	25
5.3	Доля и теоретическая вероятность выбросов . . . . .	25
5.4	Эмпирическая функция и ядерные оценки плотности распределения . . . . .	25
	<b>Литература</b>	<b>27</b>

## Список иллюстраций

1	Нормальное распределение . . . . .	11
2	Распределение Коши . . . . .	11
3	Распределение Лапласа . . . . .	11
4	Распределение Пуассона . . . . .	11
5	Равномерное распределение . . . . .	12
6	Боксплот нормального распределения . . . . .	15
7	Боксплот распределения Коши . . . . .	15
8	Боксплот распределения Лапласа . . . . .	16
9	Боксплот распределения Пуассона . . . . .	16
10	Боксплот равномерного распределения . . . . .	17
11	Нормальное распределение . . . . .	18
12	Распределение Коши . . . . .	18
13	Распределение Лапласа . . . . .	19
14	Распределение Пуассона . . . . .	19
15	Равномерное распределение . . . . .	19
16	Нормальное распределение, $n = 20$ . . . . .	20
17	Нормальное распределение, $n = 60$ . . . . .	20
18	Нормальное распределение, $n = 100$ . . . . .	20
19	Распределение Коши, $n = 20$ . . . . .	21
20	Распределение Коши, $n = 60$ . . . . .	21
21	Распределение Коши, $n = 100$ . . . . .	21
22	Распределение Лапласа, $n = 20$ . . . . .	22
23	Распределение Лапласа, $n = 60$ . . . . .	22
24	Распределение Лапласа, $n = 100$ . . . . .	22
25	Распределение Пуассона, $n = 20$ . . . . .	23
26	Распределение Пуассона, $n = 60$ . . . . .	23
27	Распределение Пуассона, $n = 100$ . . . . .	23
28	Равномерное распределение, $n = 20$ . . . . .	24
29	Равномерное распределение, $n = 60$ . . . . .	24
30	Равномерное распределение, $n = 100$ . . . . .	24

## Список таблиц

1	Статистический ряд . . . . .	8
2	Таблица распределения . . . . .	8
3	Характеристики положения и рассеяния нормального распределения . . . . .	12
4	Характеристики положения и рассеяния распределения Коши . . . . .	13
5	Характеристики положения и рассеяния распределения Лапласа . . . . .	13
6	Характеристики положения и рассеяния распределения Пуассона . . . . .	14
7	Характеристики положения и рассеяния равномерного распределения . . . . .	14
8	Доля выбросов . . . . .	17
9	Теоретическая вероятность выбросов . . . . .	18

# 1 Постановка задачи

Для 5 распределений:

- Нормальное распределение  $N(x, 0, 1)$
- Распределение Коши  $C(x, 0, 1)$
- Распределение Лапласа  $L(x, 0, \frac{1}{\sqrt{2}})$
- Распределение Пуассона  $P(k, 10)$
- Равномерное распределение  $U(x, -\sqrt{3}, \sqrt{3})$

1. Сгенерировать выборки размером 10, 50 и 1000 элементов.

Построить на одном рисунке гистограмму и график плотности распределения.

2. Сгенерировать выборки размером 10, 100 и 1000 элементов.

Для каждой выборки вычислить следующие статистические характеристики положения данных:  $\bar{x}$ ,  $med\ x$ ,  $z_R$ ,  $z_Q$ ,  $z_{tr}$ . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

$$E(z) = \bar{z} \quad (1)$$

Вычислить оценку дисперсии по формуле:

$$D(z) = \overline{z^2} - \bar{z}^2 \quad (2)$$

Представить полученные данные в виде таблиц.

3. Сгенерировать выборки размером 20 и 100 элементов.

Построить для них боксплот Тьюки.

Для каждого распределения определить долю выбросов экспериментально (сгенерировав выборку, соответствующую распределению, 1000 раз и вычислив среднюю долю выбросов) и сравнить с результатами, полученными теоретически.

4. Сгенерировать выборки размером 20, 60 и 100 элементов.

Построить на них эмпирические функции распределения и ядерные оценки плотности распределения на отрезке  $[-4; 4]$  для непрерывных распределений и на отрезке  $[6; 14]$  для распределения Пуассона.

## 2 Теория

### 2.1 Рассматриваемые распределения

Плотности распределений:

- Нормальное распределение

$$N(x, \mu, \sigma) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}} \quad (3)$$

- Распределение Коши

$$C(x, x_0, \gamma) = \frac{1}{\pi \cdot \gamma} \frac{1}{1 + \frac{(x - x_0)^2}{\gamma^2}} \quad (4)$$

- Распределение Лапласа

$$L(x, \beta, \alpha) = \frac{\alpha}{2} e^{-\alpha |x - \beta|} \quad (5)$$

- Распределение Пуассона

$$P(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (6)$$

- Равномерное распределение

$$U(x, a, b) = \begin{cases} \frac{1}{b-a} & \text{при } x \in [a, b] \\ 0 & \text{при } x \notin [a, b] \end{cases} \quad (7)$$

### 2.2 Гистограмма

#### 2.2.1 Построение гистограммы

Множество значений, которое может принимать элемент выборки, разбивается на несколько интервалов. Чаще всего эти интервалы берут одинаковыми, но это не является строгим требованием. Эти интервалы откладываются на горизонтальной оси, затем над каждым рисуется прямоугольник. Если все интервалы были одинаковыми, то высота каждого прямоугольника пропорциональна числу элементов выборки, попадающих в соответствующий интервал. Если интервалы разные, то высота прямоугольника выбирается таким образом, чтобы его площадь была пропорциональна числу элементов выборки, которые попали в этот интервал. [1]

### 2.3 Вариационный ряд

Вариационный ряд - последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются. [2, с. 409]

## 2.4 Выборочные числовые характеристики

### 2.4.1 Характеристики положения

- Выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

- Выборочная медиана

$$\text{med } x = \begin{cases} x_{(l+1)} & \text{при } n = 2l + 1 \\ \frac{x_{(l)} + x_{(l+1)}}{2} & \text{при } n = 2l \end{cases} \quad (9)$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \quad (10)$$

- Полусумма квартилей

Выборочная квартиль  $z_p$  порядка  $p$  определяется формулой

$$z_p = \begin{cases} x_{([np]+1)} & \text{при } np \text{ дробном,} \\ x_{(np)} & \text{при } np \text{ целом.} \end{cases} \quad (11)$$

- Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \quad (12)$$

- Усечённое среднее

$$z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_{(i)}, \quad r \approx \frac{n}{4} \quad (13)$$

### 2.4.2 Характеристики рассеяния

Выборочная дисперсия

$$D = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (14)$$

## 2.5 Боксплот Тьюки

### 2.5.1 Построение боксплота

Границы ящика – первый и третий квартили, линия в середине ящика – медиана. Концы усов – края статистически значимой выборки (без выбросов). Длина "усов":

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), \quad X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1), \quad (15)$$

где  $X_1$  – нижняя граница уса,  $X_2$  – верхняя граница уса,  $Q_1$  – первый квартиль,  $Q_3$  – третий квартиль.

Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков. [3]

## 2.6 Теоретическая вероятность выбросов

Можно вычислить теоретические первый и третий квартили распределений ( $Q_1^T$  и  $Q_3^T$  соответственно). По формуле (15) можно вычислить теоретические нижнюю и верхнюю границы уса ( $X_1^T$  и  $X_2^T$  соответственно). Выбросами считаются величины  $x$ , такие что:

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases} \quad (16)$$

Теоретическая вероятность выбросов:

- для непрерывных распределений

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)). \quad (17)$$

- для дискретных распределений

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)). \quad (18)$$

Выше  $F(X) = P(x \leq X)$  - функция распределения.

## 2.7 Эмпирическая функция распределения

### 2.7.1 Статистический ряд

Статистическим рядом называется последовательность различных элементов выборки  $z_1, z_2, \dots, z_k$ , расположенных в возрастающем порядке с указанием частот  $n_1, n_2, \dots, n_k$ , с которыми эти элементы содержатся в выборке.

Статистический ряд обычно записывается в виде таблицы.

### 2.7.2 Эмпирическая функция распределения

Эмпирической (выборочной) функцией распределения называется относительная частота события  $X < x$ , полученная по данной выборке:

$$F_n^*(x) = P^*(X < x). \quad (19)$$



$z$	$z_1$	$z_2$	$\dots$	$z_k$
$n$	$n_1$	$n_2$	$\dots$	$n_k$

Таблица 1: Статистический ряд

### 2.7.3 Нахождение эмпирической функции распределения

Для получения относительной частоты  $P^*(X < x)$  просуммируем в статистическом ряде, построенном по данной выборке, все частоты  $n_i$ , для которых элементы  $z_i$  статистического ряда меньше  $x$ . Тогда  $P^*(X < x) = \frac{1}{n} \sum_{z_i < x} n_i$ . Получаем

$$F^*(x) = \frac{1}{n} \sum_{z_i < x} n_i. \quad (20)$$

$F^*(x)$  — функция распределения дискретной случайной величины  $X^*$ , заданной таблицей распределения

$X^*$	$z_1$	$z_2$	$\dots$	$z_k$
$P$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\dots$	$\frac{n_k}{n}$

Таблица 2: Таблица распределения

Эмпирическая функция распределения является оценкой, т. е. приближённым значением, генеральной функции распределения

$$F_n^*(x) \approx F_X(x). \quad (21)$$

## 2.8 Оценки плотности распределения

### 2.8.1 Определение

Оценкой плотности вероятности  $f(x)$  называется функция  $\hat{f}(x)$ , построенная на основе выборки, приближённо равная  $f(x)$

$$\hat{f}(x) \approx f(x). \quad (22)$$

### 2.8.2 Ядерные оценки

Представим оценку в виде суммы с числом слагаемых, равным объёму выборки:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-x_i}{h_n}\right). \quad (23)$$

Здесь функция  $K(u)$ , называемая ядерной (ядром), непрерывна и является плотностью вероятности,  $x_1, \dots, x_n$  — элементы выборки,  $\{h_n\}$  — любая последовательность положительных чисел, обладающая свойствами

$$h_n \xrightarrow{n \rightarrow \infty} 0; \quad \frac{h_n}{n^{-1}} \xrightarrow{n \rightarrow \infty} \infty. \quad (24)$$

Такие оценки называются непрерывными ядерными.

Замечание. Свойство, означающее сближение оценки с оцениваемой величиной при  $n \rightarrow \infty$  в каком-либо смысле, называется состоятельностью оценки.

Если плотность  $f(x)$  кусочно-непрерывная, то ядерная оценка плотности является состоятельной при соблюдении условий, накладываемых на параметр сглаживания  $h_n$ , а также на ядро  $K(u)$ .

Гауссово (нормальное) ядро [4, с. 38]

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}. \quad (25)$$

Правило Сильвермана [4, с. 44]

$$h_n = 1.06 \hat{\sigma} n^{-1/5}, \quad (26)$$

где  $\hat{\sigma}$  - выборочное стандартное отклонение.

### 3 Реализация

Лабораторная работа выполнена с использованием языка программирования Python 3.10 в среде разработки PyCharm Community с использованием библиотек

- matplotlib v. 3.7.1
- numpy v. 1.24.2
- scipy v. 1.10.1
- seaborn v. 0.12.2

## 4 Результаты

### 4.1 Гистограмма и график плотности распределения

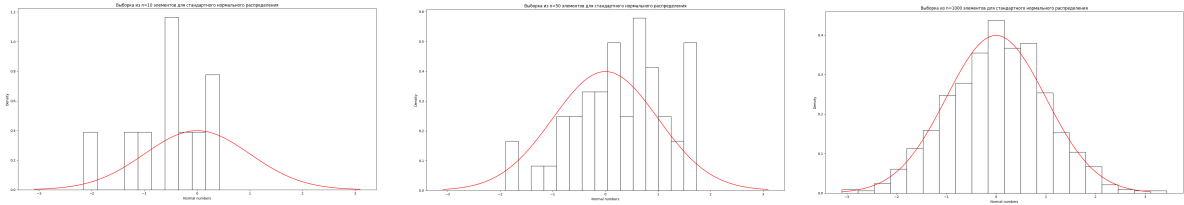


Рис. 1: Нормальное распределение

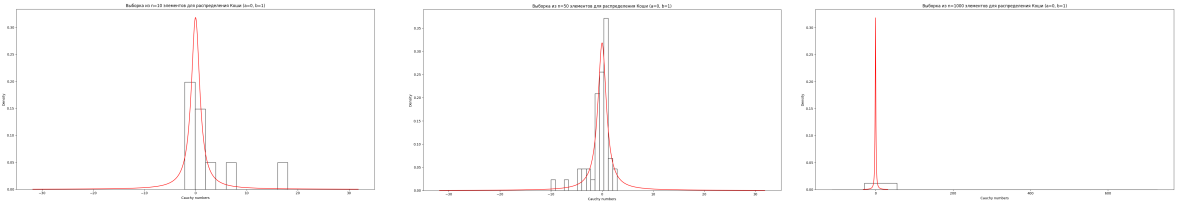


Рис. 2: Распределение Коши

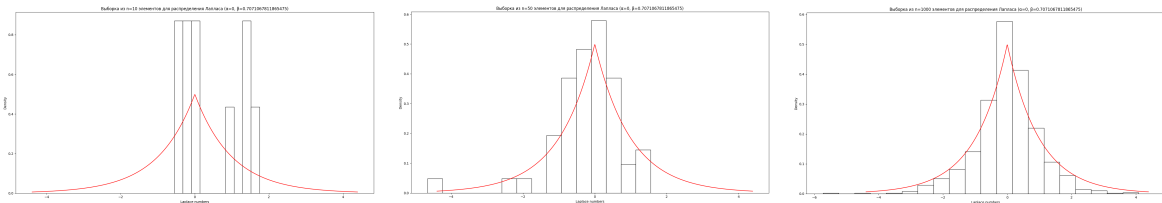


Рис. 3: Распределение Лапласа

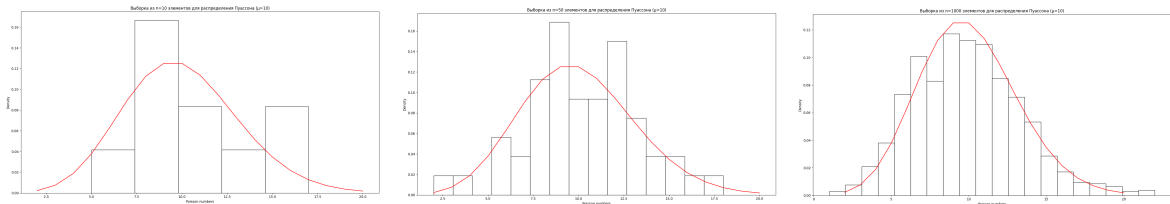


Рис. 4: Распределение Пуассона

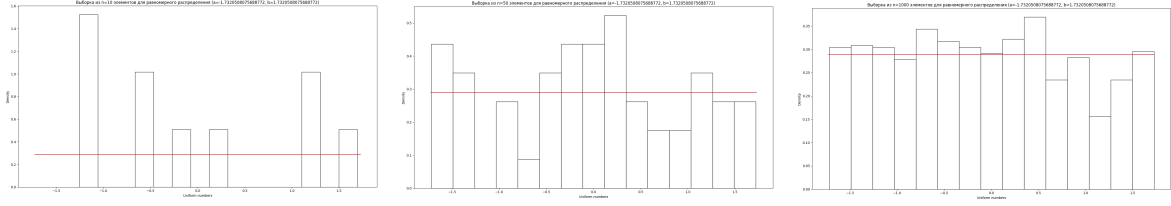


Рис. 5: Равномерное распределение

## 4.2 Характеристики положения и рассеяния

Как было проведено округление:

В оценке  $x = E \pm \sqrt{D}$  вариации подлежит первая цифра после точки.

В данном случае  $= 0.0 \pm 0.1$  где,  $k$  зависит от доверительной вероятности и вида распределения (рассматривается в дальнейшем цикле лабораторных работ).

Округление сделано для  $k = 1$ .

Gauss n = 10	$\bar{x}$	$med x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	-0.001	0.120	-0.008	-0.001	-0.001
$D(z)$	0.107	0.167	0.205	0.123	0.120
$E(z) \pm \sqrt{D(z)}$	[-0.329;0.326]	[-0.289;0.529]	[-0.461;0.446]	[-0.352;0.350]	[-0.347;0.345]
$\hat{E}(z)$	0	0	0	0	0
Gauss n = 100	$\bar{x}$	$med x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	-0.002	0.011	-0.010	-0.015	-0.001
$D(z)$	0.010	0.016	0.096	0.013	0.012
$E(z) \pm \sqrt{D(z)}$	[-0.103;0.098]	[-0.118;0.139]	[-0.320;0.301]	[-0.129;0.098]	[-0.112;0.111]
$\hat{E}(z)$	0	0	0	0	0
Gauss n = 1000	$\bar{x}$	$med x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	0.000	0.001	0.010	-0.001	0.000
$D(z)$	0.001	0.002	0.057	0.001	0.001
$E(z) \pm \sqrt{D(z)}$	[-0.030;0.031]	[-0.038;0.040]	[-0.229;0.249]	[-0.035;0.033]	[-0.034;0.034]
$\hat{E}(z)$	0.0	0.0	0	0.0	0.0

Таблица 3: Характеристики положения и рассеяния нормального распределения

Cauchy n = 10	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	13.257	0.173	66.059	0.041	0.006
$D(z)$	95821.419	0.382	2393086.501	1.240	0.542
$E(z) \pm \sqrt{D(z)}$	[-296.293;322.808]	[-0.445;0.790]	[-1480.901;1613.020]	[-1.072;1.154]	[-0.730;0.742]
$\hat{E}(z)$	-	0	-	-	0
Cauchy n = 100	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	0.934	0.018	43.235	-0.020	0.005
$D(z)$	936.934	0.027	2321693.382	0.055	0.028
$E(z) \pm \sqrt{D(z)}$	[-29.675;31.543]	[-0.147;0.182]	[-1480.475;1566.946]	[-0.254;0.214]	[-0.162;0.173]
$\hat{E}(z)$	-	0	-	0	0
Cauchy n = 1000	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	2.262	0.004	1118.595	0.002	0.003
$D(z)$	954.296	0.002	236247167.367	0.005	0.003
$E(z) \pm \sqrt{D(z)}$	[-28.630;33.153]	[-0.045;0.053]	[-14251.739;16488.929]	[-0.068;0.072]	[-0.048;0.053]
$\hat{E}(z)$	-	0.0	-	0.0	0.0

Таблица 4: Характеристики положения и рассеяния распределения Коши

Laplace n = 10	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	0.009	0.093	0.004	0.011	0.010
$D(z)$	0.093	0.074	0.409	0.092	0.066
$E(z) \pm \sqrt{D(z)}$	[-0.296;0.315]	[-0.179;0.366]	[-0.636;0.644]	[-0.293;0.314]	[-0.247;0.266]
$\hat{E}(z)$	0	0	0	0	0
Laplace n = 100	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	0.003	0.008	0.043	-0.014	0.000
$D(z)$	0.010	0.006	0.381	0.011	0.007
$E(z) \pm \sqrt{D(z)}$	[-0.100;0.105]	[-0.070;0.086]	[-0.574;0.660]	[-0.117;0.088]	[-0.080;0.081]
$\hat{E}(z)$	0	0.0	0	0	0.0
Laplace n = 1000	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	0.000	0.001	0.010	-0.001	-0.000
$D(z)$	0.001	0.001	0.404	0.001	0.001
$E(z) \pm \sqrt{D(z)}$	[-0.031;0.032]	[-0.022;0.023]	[-0.626;0.646]	[-0.033;0.030]	[-0.025;0.025]
$\hat{E}(z)$	0.0	0.0	0	0.0	0.0

Таблица 5: Характеристики положения и рассеяния распределения Лапласа

Poisson n = 10	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	10.017	10.296	10.295	9.938	9.914
$D(z)$	0.994	1.592	1.963	1.214	1.137
$E(z) \pm \sqrt{D(z)}$	[9.020;11.014]	[9.034;11.558]	[8.894;11.696]	[8.836;11.040]	[8.848;10.981]
$\hat{E}(z)$	-	-	-	-	-
Poisson n = 100	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	9.990	9.873	10.945	9.864	9.851
$D(z)$	0.094	0.211	1.062	0.146	0.110
$E(z) \pm \sqrt{D(z)}$	[9.684;10.296]	[9.414;10.332]	[9.915;11.976]	[9.482;10.246]	[9.520;10.182]
$\hat{E}(z)$	10	10	-	10	10
Poisson n = 1000	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	10.003	9.993	11.636	9.992	9.860
$D(z)$	0.010	0.007	0.712	0.004	0.012
$E(z) \pm \sqrt{D(z)}$	[9.901;10.105]	[9.910;10.076]	[10.793;12.480]	[9.925;10.059]	[9.751;9.968]
$\hat{E}(z)$	10	10.0	-	10.0	10

Таблица 6: Характеристики положения и рассеяния распределения Пуассона

Uniform n = 10	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	-0.003	0.151	0.001	-0.008	-0.006
$D(z)$	0.097	0.235	0.044	0.132	0.157
$E(z) \pm \sqrt{D(z)}$	[-0.314;0.308]	[-0.333;0.636]	[-0.209;0.212]	[-0.371;0.355]	[-0.402;0.390]
$\hat{E}(z)$	0	0	0	0	0
Uniform n = 100	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	0.001	0.017	0.001	-0.015	-0.001
$D(z)$	0.009	0.027	0.001	0.014	0.018
$E(z) \pm \sqrt{D(z)}$	[-0.096;0.097]	[-0.148;0.182]	[-0.023;0.024]	[-0.133;0.102]	[-0.136;0.134]
$\hat{E}(z)$	0.0	0	0.0	0	0
Uniform n = 1000	$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
$E(z)$	0.002	0.004	0.000	0.000	0.003
$D(z)$	0.001	0.003	0.000	0.002	0.002
$E(z) \pm \sqrt{D(z)}$	[-0.030;0.035]	[-0.052;0.059]	[-0.002;0.003]	[-0.039;0.040]	[-0.043;0.049]
$\hat{E}(z)$	0.0	0.0	0.0	0.0	0.0

Таблица 7: Характеристики положения и рассеяния равномерного распределения

### 4.3 Боксплот Тьюки

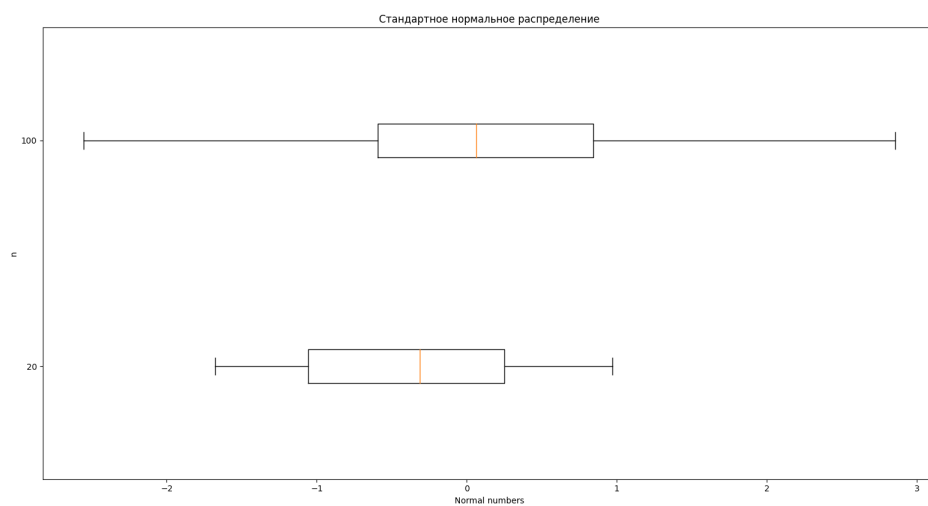


Рис. 6: Боксплот нормального распределения

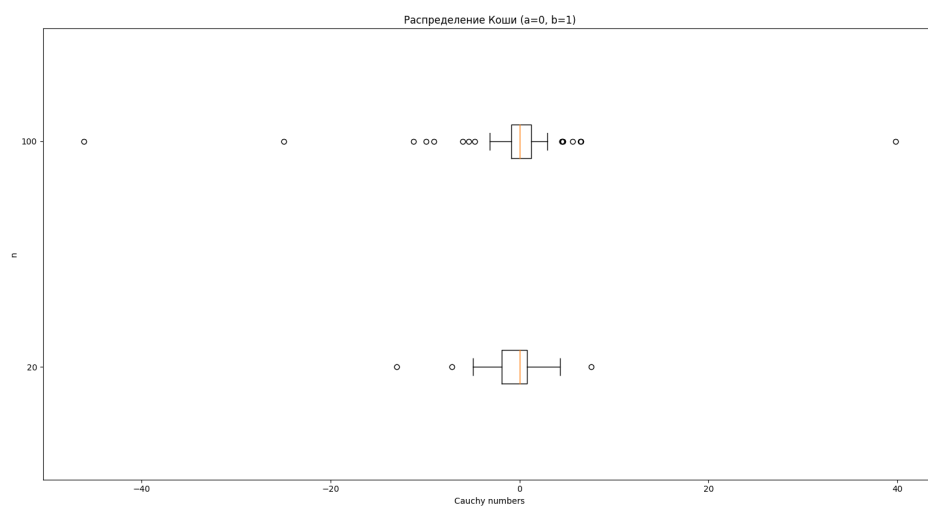


Рис. 7: Боксплот распределения Коши



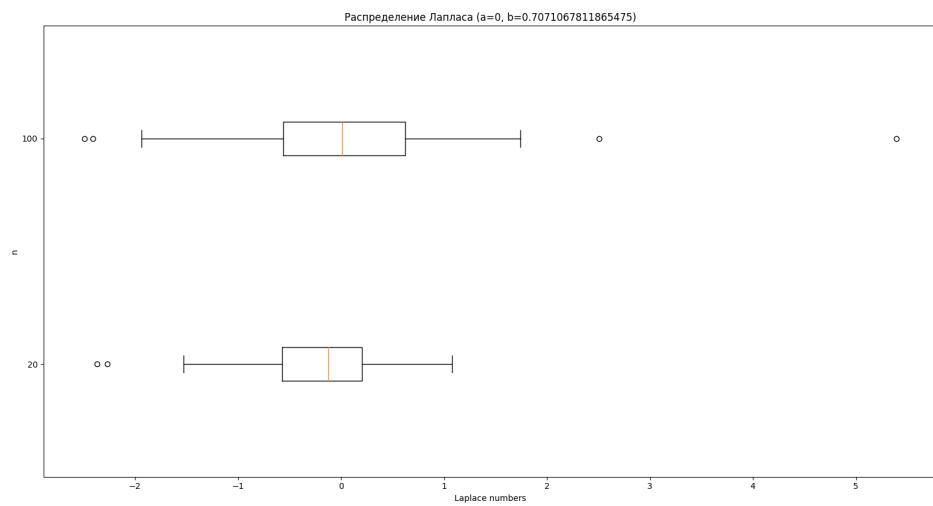


Рис. 8: Боксплот распределения Лапласа

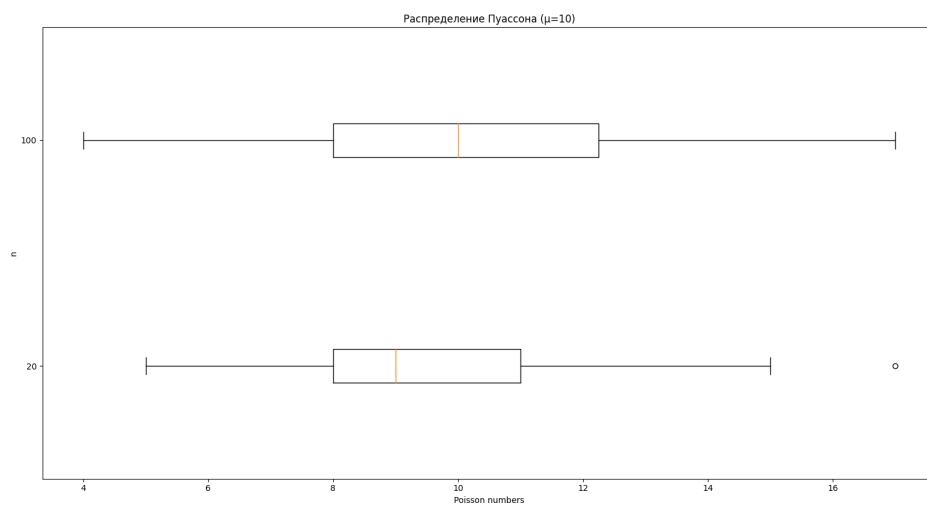


Рис. 9: Боксплот распределения Пуассона

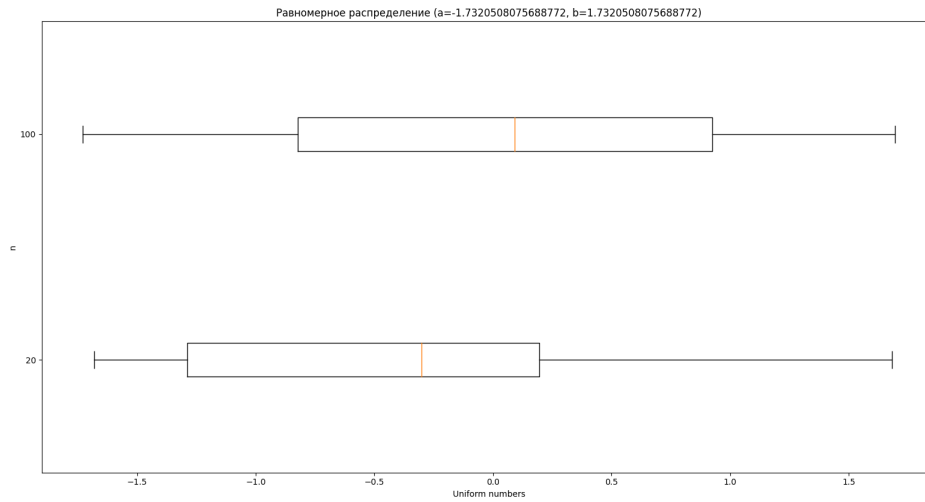


Рис. 10: Боксплот равномерного распределения

#### 4.4 Доля выбросов

Выборка	Доля выбросов
normal n=20	0.01
normal n=100	0.02
cauchy n=20	0.16
cauchy n=100	0.15
laplace n=20	0.07
laplace n=100	0.07
poisson n=20	0.01
poisson n=100	0.02
uniform n=20	0
uniform n=100	0

Таблица 8: Доля выбросов

## 4.5 Теоретическая вероятность выбросов

Распределение	$Q_1^T$	$Q_3^T$	$X_1^T$	$X_2^T$	$P_B^T$
Нормальное распределение	-0.674	0.674	-2.698	2.698	0.007
Распределение Коши	-1	1	-4	4	0.156
Распределение Лапласа	-0.490	0.490	-1.961	1.961	0.063
Распределение Пуассона	8	12	2	18	0.008
Равномерное распределение	-0.866	0.866	-3.464	3.464	0

Таблица 9: Теоретическая вероятность выбросов

## 4.6 Эмпирическая функция распределения

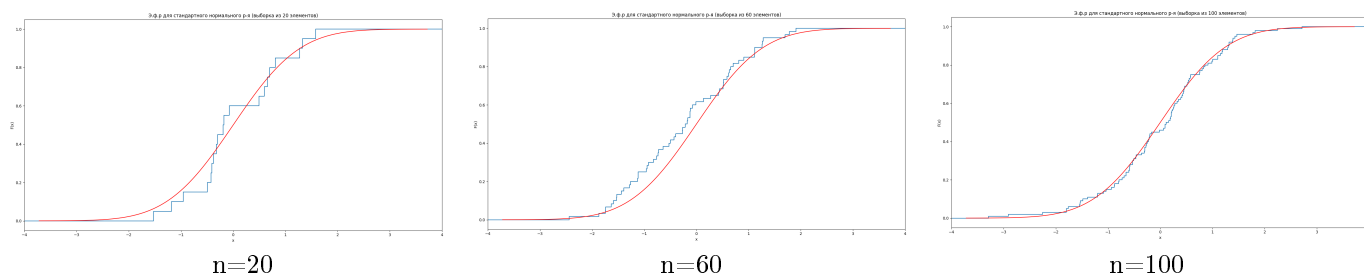


Рис. 11: Нормальное распределение

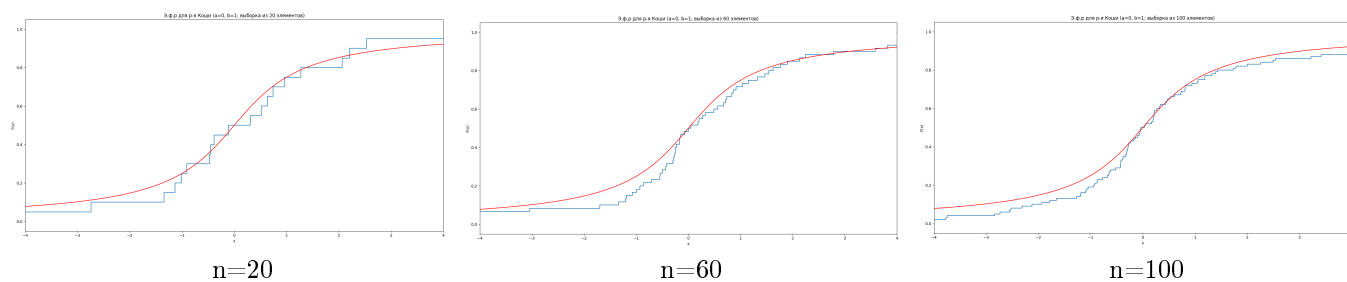


Рис. 12: Распределение Коши

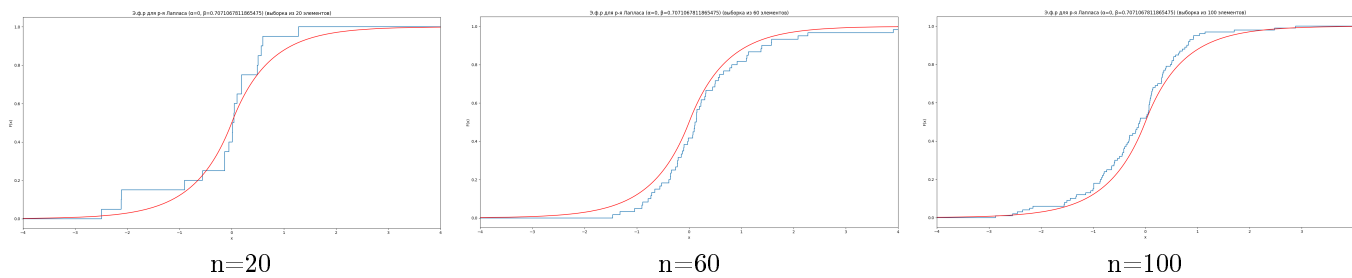


Рис. 13: Распределение Лапласа

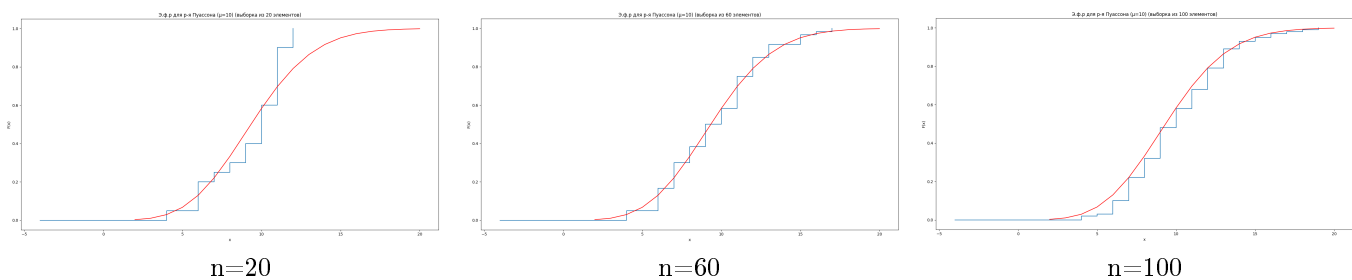


Рис. 14: Распределение Пуассона

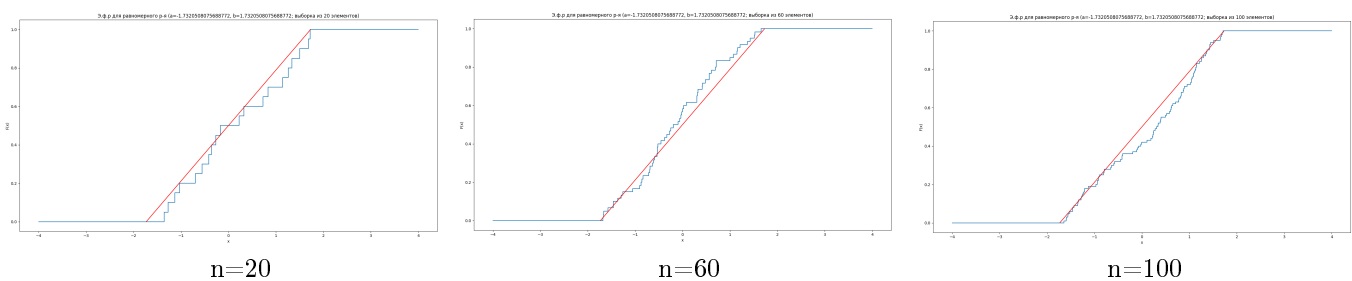


Рис. 15: Равномерное распределение

## 4.7 Ядерные оценки плотности распределения

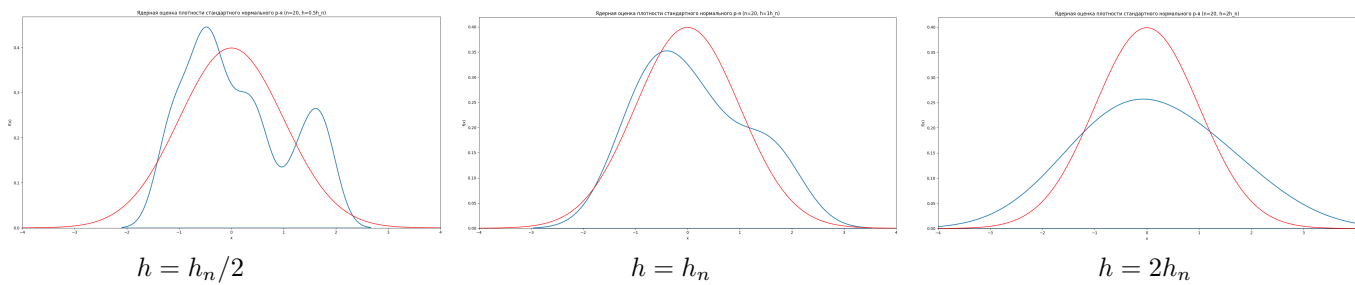


Рис. 16: Нормальное распределение,  $n = 20$

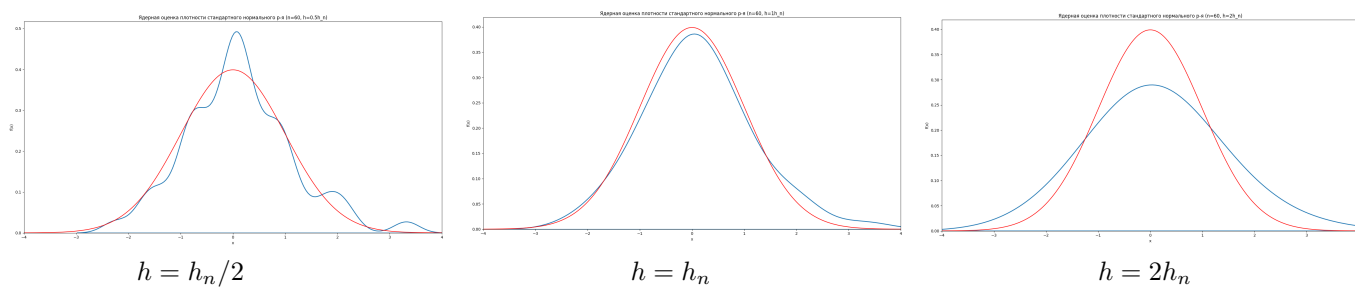


Рис. 17: Нормальное распределение,  $n = 60$

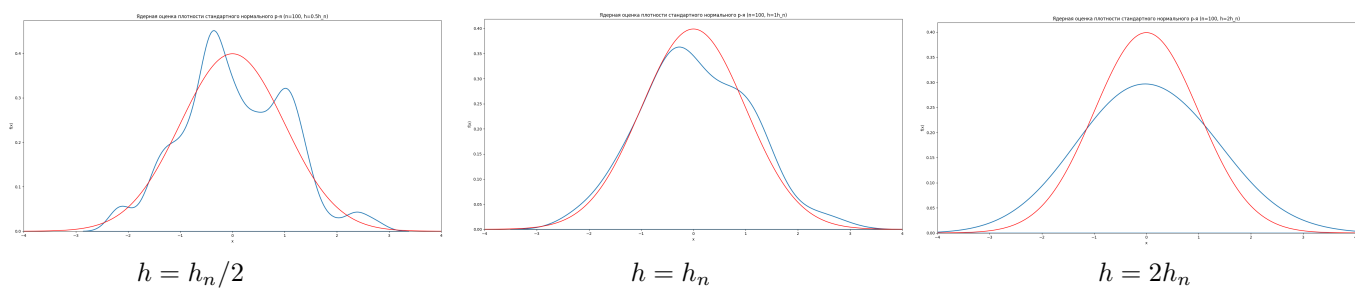


Рис. 18: Нормальное распределение,  $n = 100$

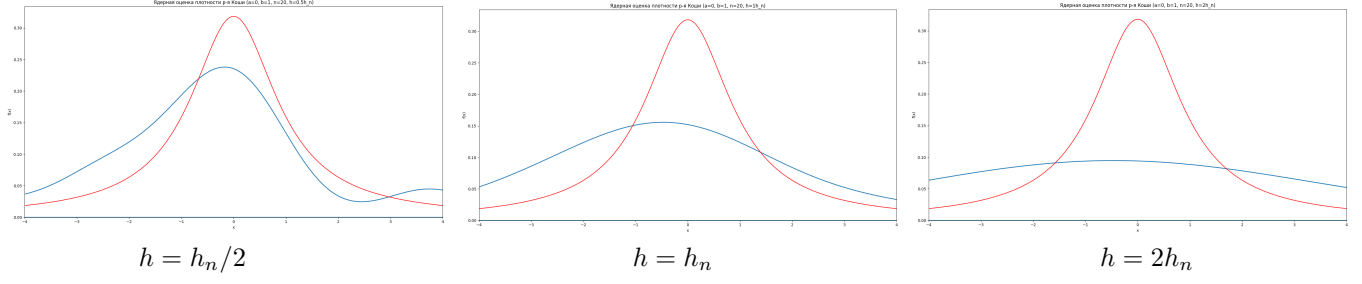


Рис. 19: Распределение Коши,  $n = 20$

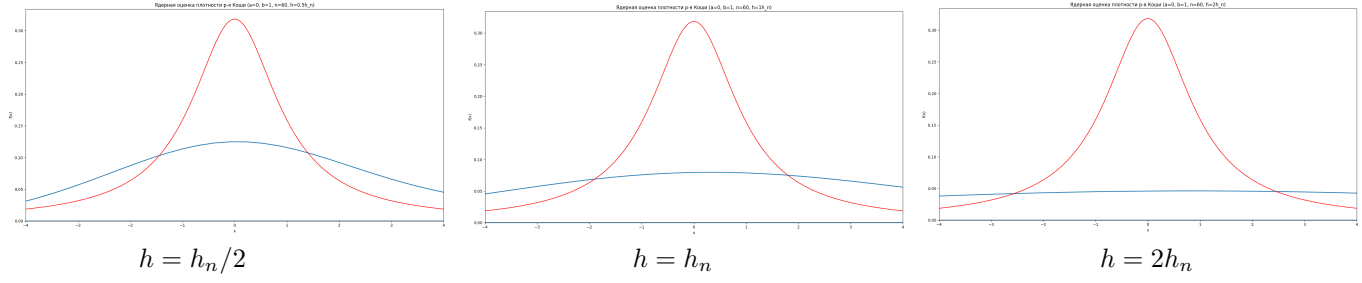


Рис. 20: Распределение Коши,  $n = 60$

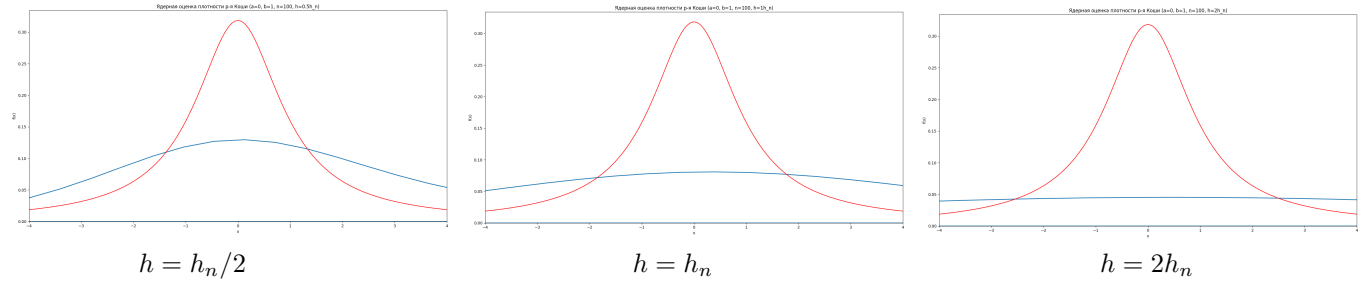


Рис. 21: Распределение Коши,  $n = 100$

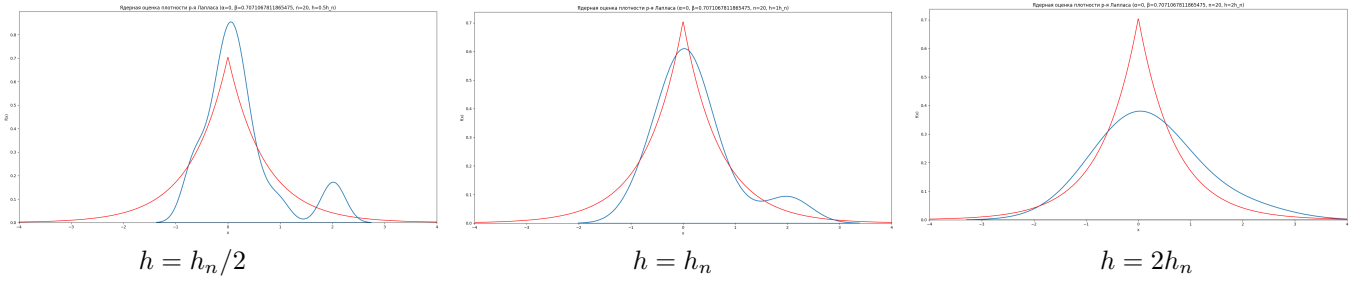


Рис. 22: Распределение Лапласа,  $n = 20$

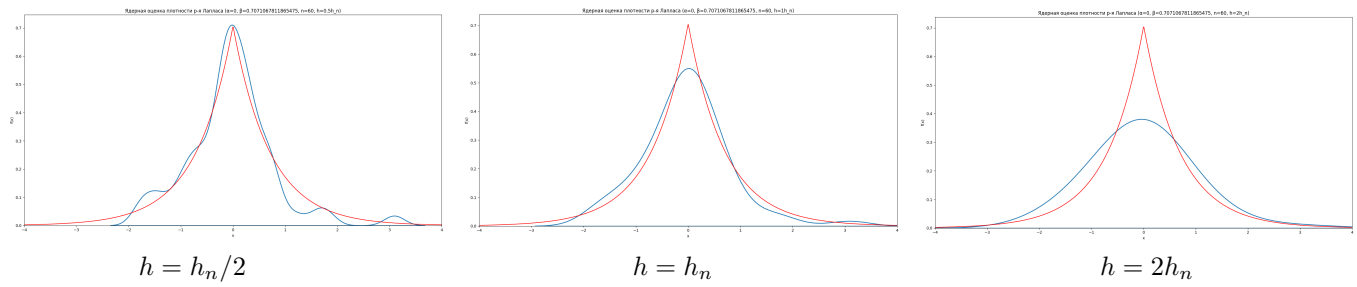


Рис. 23: Распределение Лапласа,  $n = 60$

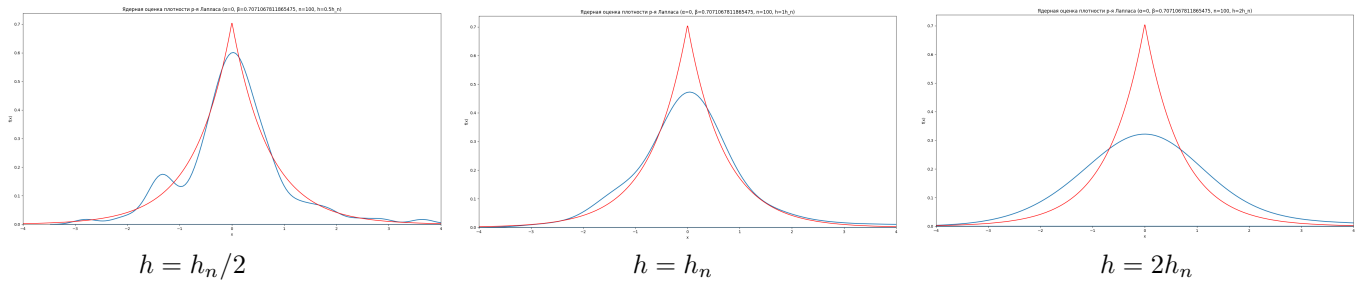
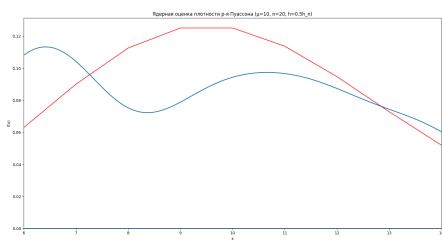
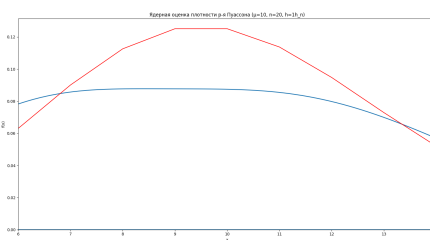


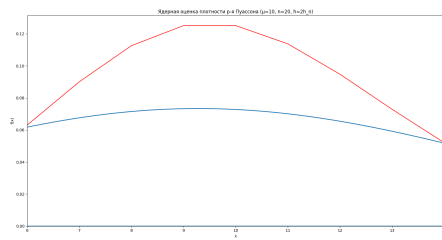
Рис. 24: Распределение Лапласа,  $n = 100$



$$h = h_n/2$$

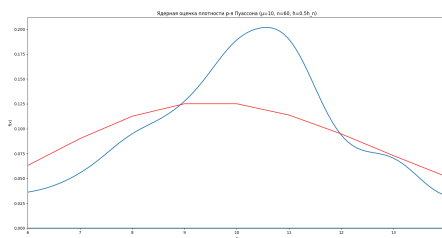


$$h = h_n$$

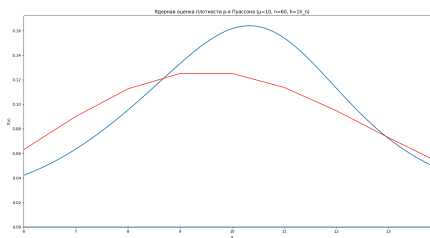


$$h = 2h_n$$

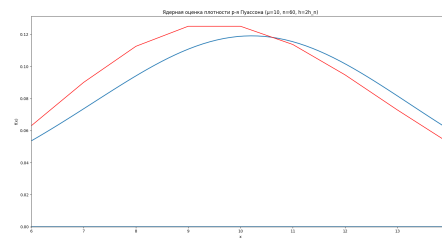
Рис. 25: Распределение Пуассона,  $n = 20$



$$h = h_n/2$$

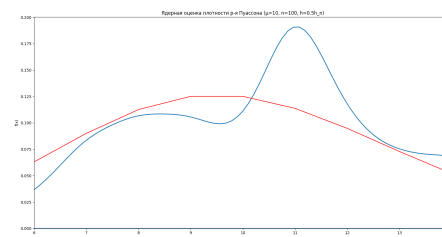


$$h = h_n$$

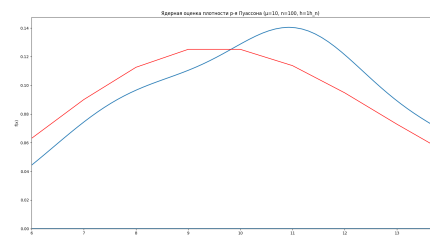


$$h = 2h_n$$

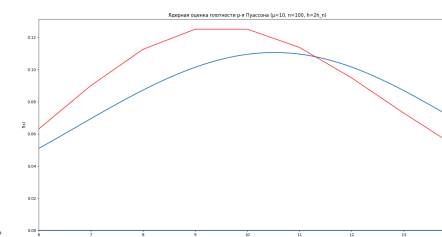
Рис. 26: Распределение Пуассона,  $n = 60$



$$h = h_n/2$$



$$h = h_n$$



$$h = 2h_n$$

Рис. 27: Распределение Пуассона,  $n = 100$



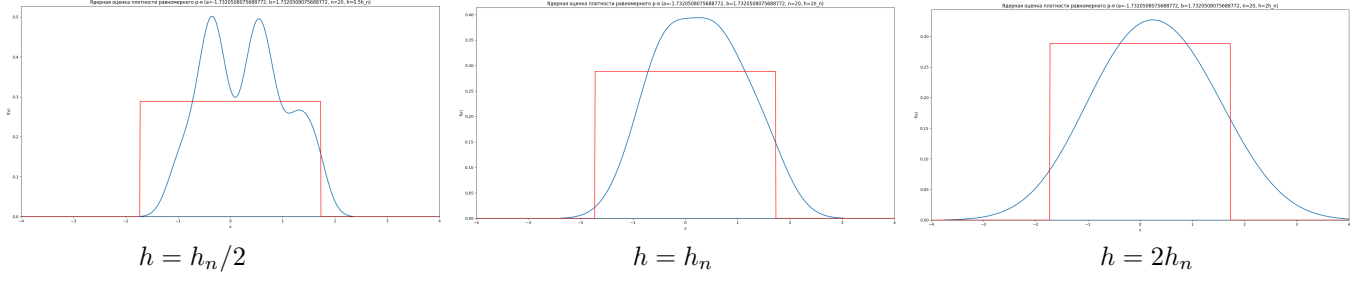


Рис. 28: Равномерное распределение,  $n = 20$

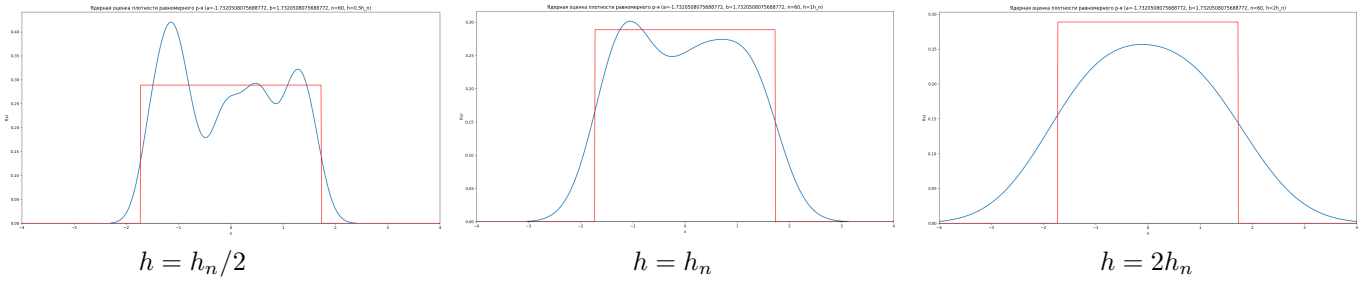


Рис. 29: Равномерное распределение,  $n = 60$

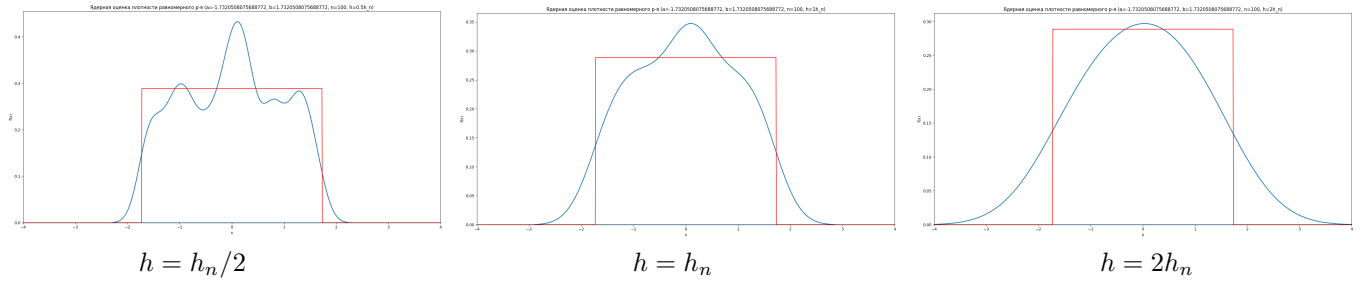


Рис. 30: Равномерное распределение,  $n = 100$

## 5 Обсуждение

### 5.1 Гистограмма и график плотности распределения

По итогам проделанной работы можем сказать, что чем больше выборка для каждого из распределений, тем ближе ее гистограмма к графику плотности вероятности того закона, по которому распределены величины сгенерированной выборки. Это наглядно показывает гистограммы большой выборки ( $n = 10$ ) стандартного нормального распределения.

Чем меньше выборка, тем хуже по ней определяется характер распределения величины. Действительно, гистограммы по маленькой выборке ( $n = 10$ ) различных распределений очень похожи друг на друга (например нормальное распределение и распределение Коши).

### 5.2 Характеристики положения и рассеяния

Проанализировав данные из таблиц, можно судить о том, что дисперсия характеристик рассеяния для распределения Коши является некой аномалией: значения слишком большие даже при увеличении размера выборки - понятно, что это результат выбросов, которые мы могли наблюдать в результатах предыдущего задания.

### 5.3 Доля и теоретическая вероятность выбросов

По данным, приведенным в таблице, можно сказать, что чем больше выборка, тем ближе доля выбросов будет к теоретической оценке. Снова доля выбросов для распределения Коши значительно выше, чем для остальных распределений. Равномерное распределение же в точности повторяет теоретическую оценку - выбросов мы не получали.

Боксплоты Тьюки действительно позволяют более наглядно и с меньшими усилиями оценивать важные характеристики распределений. Так, исходя из полученных рисунков, наглядно видно то, что мы довольно трудоёмко анализировали в предыдущих частях.

### 5.4 Эмпирическая функция и ядерные оценки плотности распределения

Можем наблюдать на иллюстрациях с э. ф. р., что ступенчатая эмпирическая функция распределения тем лучше приближает функцию распределения реальной выборки, чем мощнее эта выборка. Заметим так же, что для распределения Пуассона и равномерного распределения отклонение функций друг от друга наибольшее.

Рисунки, посвященные ядерным оценкам, иллюстрируют сближение ядерной оценки и функции плотности вероятности для всех  $h$  с ростом размера выборки. Для распределения Пуассона наиболее ярко видно, как сглаживает отклонения увеличение параметра сглаживания  $h$ .

В зависимости от особенностей распределений для их описания лучше подходят разные параметры  $h$  в ядерной оценке: для равномерного распределения и распределения Пуассона лучше подойдет параметр  $h = 2h_n$ , для распределения Лапласа  $h = h_n/2$ , а для нормального и Коши  $h = h_n$ . Такие значения дают вид ядерной оценки наиболее близкий к плотности, характерной данным распределениям.

Также можно увидеть, что чем больше коэффициент при параметре сглаживания  $h_n$ , тем меньше изменений знака производной у аппроксимирующей функции, вплоть до того, что при  $h = 2h_n$  функция становится унимодальной на рассматриваемом промежутке. Также видно, что при  $h = 2h_n$  по полученным приближениям становится сложно сказать плотность вероятности какого распределения они должны повторять, так как они очень похожи между собой.

## Литература

- [1] Histogram. URL: <https://en.wikipedia.org/wiki/Histogram>
- [2] Вероятностные разделы математики. Учебник для бакалавров технических направлений. //Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.
- [3] Box plot. URL: [https://en.wikipedia.org/wiki/Box\\_plot](https://en.wikipedia.org/wiki/Box_plot)
- [4] Анатольев, Станислав (2009) «Непараметрическая регрессия», Квантиль, №7, стр. 37-52.