

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
ПЕТРА ВЕЛИКОГО

ФИЗИКО-МЕХАНИЧЕСКИЙ ИНСТИТУТ  
КАФЕДРА «ПРИКЛАДНАЯ МАТЕМАТИКА»

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

Выполнил студент:

Петрошенко Артём Владимирович  
группа: 5030102/00201

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>4</b>
2.1	Рассматриваемые распределения . . . . .	4
2.2	Гистограмма . . . . .	5
2.2.1	Построение гистограммы . . . . .	5
2.3	Вариационный ряд . . . . .	5
2.4	Выборочные числовые характеристики . . . . .	5
2.4.1	Характеристики положения . . . . .	5
2.4.2	Характеристики рассеяния . . . . .	6
2.5	Боксплот Тьюки . . . . .	6
2.6	Теоретическая вероятность выбросов . . . . .	6
2.7	Эмпирическая функция распределения . . . . .	7
2.7.1	Статистический ряд . . . . .	7
2.7.2	Эмпирическая функция распределения . . . . .	7
2.7.3	Нахождение э. ф. р. . . . .	7
2.8	Оценки плотности вероятности . . . . .	7
2.8.1	Определение . . . . .	7
2.8.2	Ядерные оценки . . . . .	7
<b>3</b>	<b>Реализация</b>	<b>8</b>
<b>4</b>	<b>Результаты</b>	<b>8</b>
4.1	Гистограмма и график плотности распределения . . . . .	8
4.2	Характеристики положения и рассеяния . . . . .	10
4.3	Боксплот Тьюки . . . . .	13
4.4	Доля выбросов . . . . .	16
4.5	Теоретическая вероятность выбросов . . . . .	16
4.6	Эмпирическая функция распределения . . . . .	16
4.7	Ядерные оценки плотности распределения . . . . .	18
<b>5</b>	<b>Обсуждение</b>	<b>23</b>
5.1	Гистограмма и плотность распределения . . . . .	23
5.2	Характеристики положения и рассеяния . . . . .	23
5.3	Доля и теоретическая вероятность выбросов . . . . .	23
5.4	Эмпирическая функция и ядерные оценки плотности распределения . . . . .	24
	<b>Литература</b>	<b>25</b>

## Список иллюстраций

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

## Список таблиц

1	Таблица распределения . . . . .	7
2	Теоретическая вероятность выбросов. . . . .	16

# 1 Постановка задачи

Для 5 распределений

- Нормальное распределение  $N(x, 0, 1)$
- Распределение Коши  $C(x, 0, 1)$
- Распределение Лапласа  $L\left(x, 0, \frac{1}{\sqrt{2}}\right)$
- 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:  $\bar{x}$ ,  $\text{med } x$ ,  $z_R$ ,  $z_Q$ ,  $z_{\text{тг}}$ . 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:

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3)$$

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \quad (4)$$

- Распределение Лапласа

$$L\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \quad (5)$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (6)$$

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при } |x| \leq \sqrt{3} \\ 0 & \text{при } |x| > \sqrt{3} \end{cases} \quad (7)$$

## 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

Вариационным 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 элементов

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

- Полусумма квартилей

Выборочный 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

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

- Усеченное среднее

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

## 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

Границами 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  $Q_3^T$ . По формуле (15) – теоретические нижнюю и верхнюю границы уса  $X_1^T$  и  $X_2^T$ . Выбросы – величины  $x$ :

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

Теоретическая вероятность выбросов:

- для непрерывных распределений

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

- для дискретных распределений

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

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

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

Таблица 1: Таблица распределения

## 2.7 Эмпирическая функция распределения

### 2.7.1 Статистический ряд

Статистический ряд – последовательность различных элементов выборки  $z_1, \dots, z_k$  расположенных в возрастающем порядке с указанием частот  $n_1, \dots, n_k$ , с которыми эти элементы содержатся в выборке. Обычно записывается в виде таблицы.

### 2.7.2 Эмпирическая функция распределения

Эмпирическая (выборочная) функция распределения (э. ф. р.) – относительная частота события  $X < x$ , полученная по данной выборке:

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

### 2.7.3 Нахождение э. ф. р.

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

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

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

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

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

## 2.8 Оценки плотности вероятности

### 2.8.1 Определение

Оценкой плотности вероятности  $f(x)$  называется функция  $\hat{f}(x)$ , построенная на основе выборки, приближенно 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 объему 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,  $\{h_n\}_{n \in \mathbb{N}}$  – 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 [2, с. 421-423].

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 помощью языка программирования Python 3.8 и модулей numpy, scipy.stats (для генерации выборок), matplotlib.pyplot (для рисования графиков) в среде разработки PyCharm. Исходный код программы и отчета загружен на GitHub: [https://github.com/Artem-Petroshenko/MathStatistics/tree/lab\\_1-4](https://github.com/Artem-Petroshenko/MathStatistics/tree/lab_1-4)

## 4 Результаты

### 4.1 Гистограмма и график плотности распределения

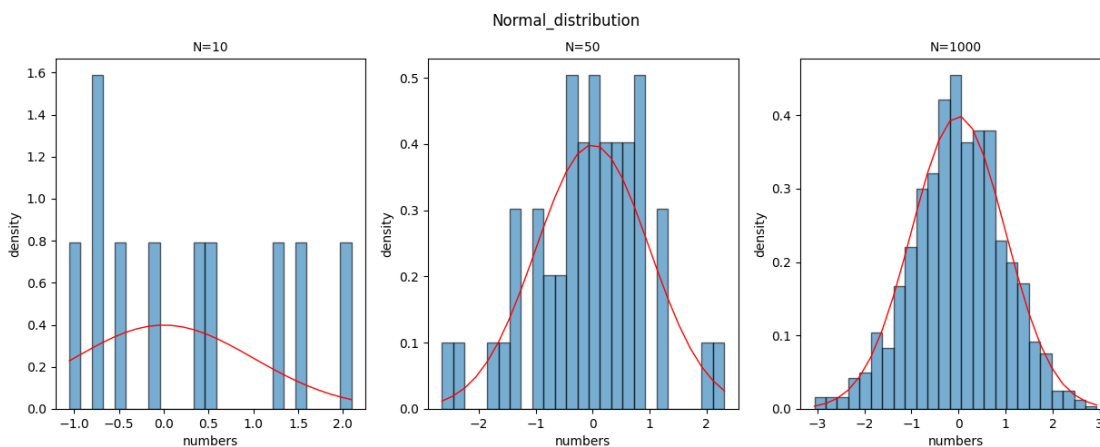


Рис. 1: Нормальное распределение

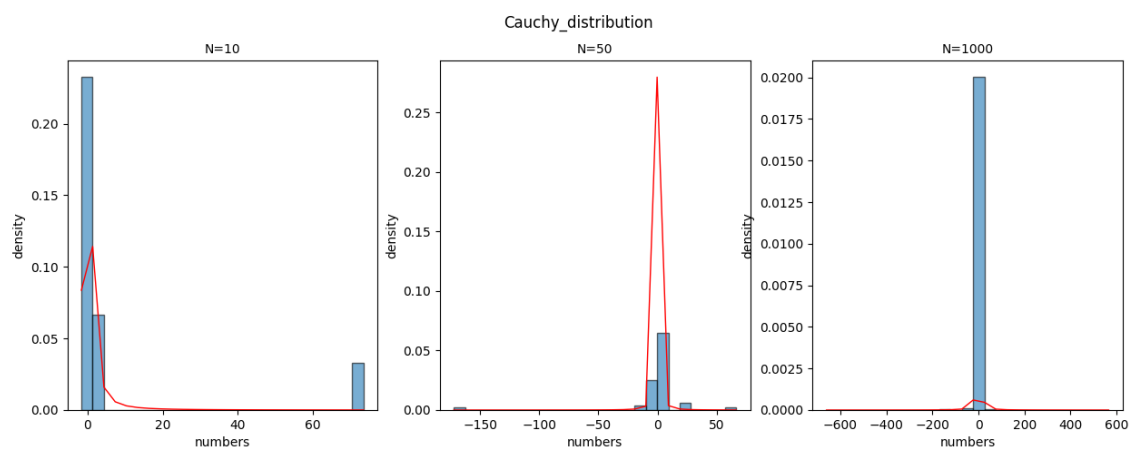


Рис. 2: Распределение Коши

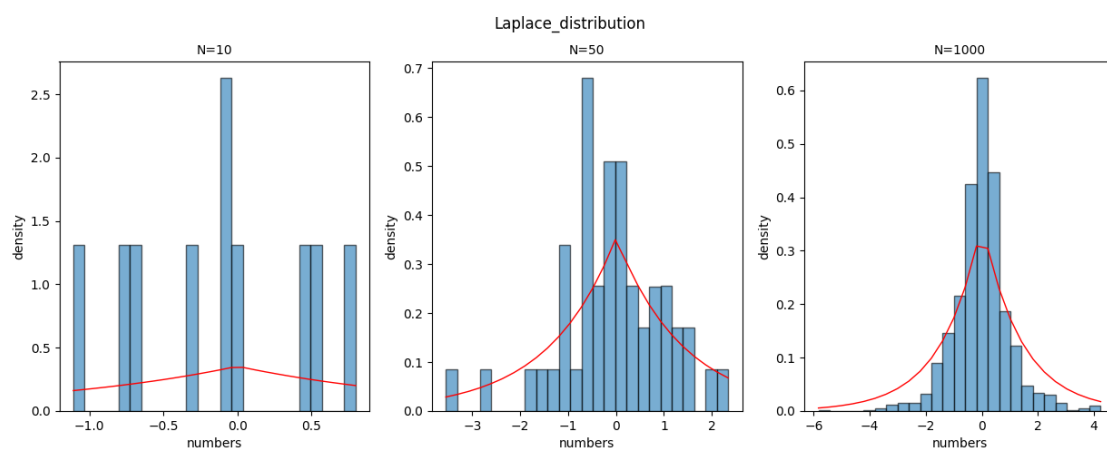


Рис. 3: Распределение Лапласа

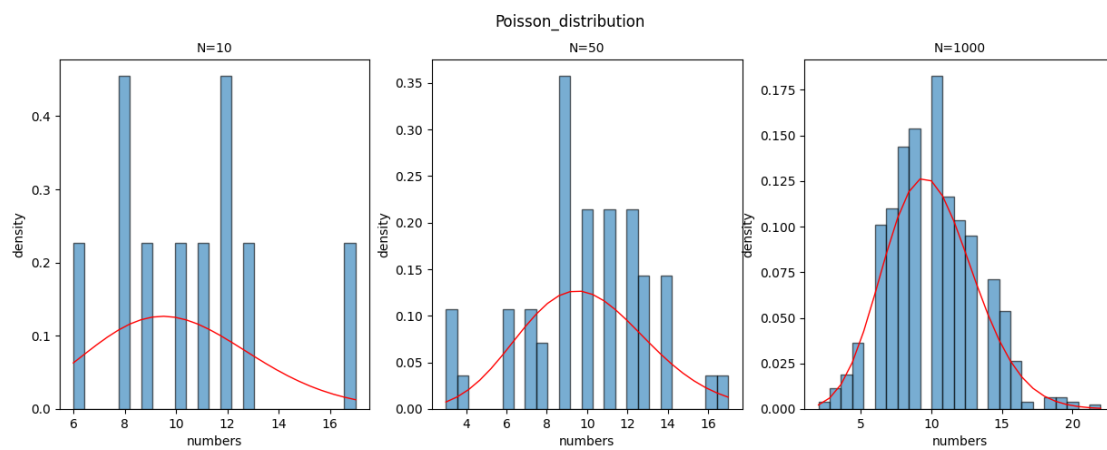


Рис. 4: Распределение Пуассона

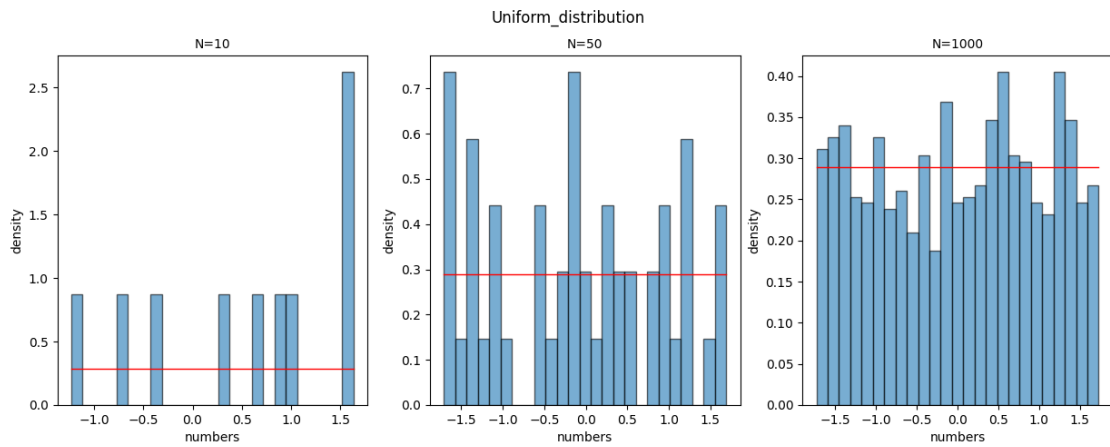


Рис. 5: Равномерное распределение

## 4.2 Характеристики положения и рассеяния

### Нормальное распределение

Размер выборки 10 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	0.0089	-0.0011	0.0015	0.0205	0.0052
$D(z)$	0.0982	0.5233	0.5217	0.4824	0.1596
$E(x) - \sqrt{D(z)}$	-0.3045	-0.7245	-0.7208	-0.6741	-0.3943
$E(x) + \sqrt{D(z)}$	0.3223	0.7223	0.7238	0.7151	0.4047
$\hat{E}$	0	0	0	0	0

Размер выборки 100 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	-0.0072	-0.0448	0.0143	0.023	-0.0153
$D(z)$	0.01	0.5029	0.4897	0.5192	0.02
$E(x) - \sqrt{D(z)}$	-0.1072	-0.754	-0.6855	-0.6976	-0.1567
$E(x) + \sqrt{D(z)}$	0.0928	0.6644	0.7141	0.7436	0.1261
$\hat{E}$	0	0	0	0	0

Размер выборки 1000 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	0.0023	-0.0073	-0.0138	0.0161	0.0018
$D(z)$	0.001	0.5059	0.4994	0.5017	0.0021
$E(x) - \sqrt{D(z)}$	-0.0293	-0.7186	-0.7205	-0.6922	-0.044
$E(x) + \sqrt{D(z)}$	0.0339	0.704	0.6929	0.7244	0.0476
$\hat{E}$	0.0	0	0	0	0.0

### Распределение Коши

Размер выборки 10 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	0.0486	-0.5147	-0.8482	0.0858	-0.1536
$D(z)$	258.6608	2056.7252	1295.6622	90.0341	379.7275
$E(x) - \sqrt{D(z)}$	-16.0343	-45.8658	-36.8435	-9.4028	-19.6402
$E(x) + \sqrt{D(z)}$	16.1315	44.8364	35.1471	9.5744	19.333
$\hat{E}$	-	-	-	-	-

Размер выборки 100 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	0.8162	6.481	-2.007	0.998	1.9626
$D(z)$	879.5897	333231.3781	900.3703	3351.9033	2484.4085
$E(x) - \sqrt{D(z)}$	-28.8417	-570.781	-32.0132	-56.8976	-47.8812
$E(x) + \sqrt{D(z)}$	30.4741	583.743	27.9992	58.8936	51.8064
$\hat{E}$	-	-	-	-	-

Размер выборки 1000 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	0.0298	-0.3433	-0.5171	10.0795	-0.6156
$D(z)$	905.7655	3538.7143	645.3264	52427.1296	432.2051
$E(x) - \sqrt{D(z)}$	-30.0661	-59.8304	-25.9204	-218.8902	-21.4051
$E(x) + \sqrt{D(z)}$	30.1257	59.1438	24.8862	239.0492	20.1739
$\hat{E}$	-	-	-	-	-

### Распределение Лапласа

Размер выборки 10 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	0.0074	-0.0159	-0.0036	0.0184	0.0064
$D(z)$	0.1114	0.4736	0.5552	0.5439	0.178
$E(x) - \sqrt{D(z)}$	-0.3264	-0.7041	-0.7487	-0.7191	-0.4155
$E(x) + \sqrt{D(z)}$	0.3412	0.6723	0.7415	0.7559	0.4283
$\hat{E}$	0	0	0	0	0

Размер выборки 100 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	0.0019	-0.018	0.0032	-0.0114	0.0081
$D(z)$	0.0092	0.4975	0.512	0.4877	0.0191
$E(x) - \sqrt{D(z)}$	-0.094	-0.7233	-0.7123	-0.7098	-0.1301
$E(x) + \sqrt{D(z)}$	0.0978	0.6873	0.7187	0.687	0.1463
$\hat{E}$	0.0	0	0	0	0

Размер выборки 1000 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	0.0002	-0.0362	0.0015	0.0297	0.0005
$D(z)$	0.001	0.4932	0.512	0.5045	0.0021
$E(x) - \sqrt{D(z)}$	-0.0314	-0.7385	-0.714	-0.6806	-0.0453
$E(x) + \sqrt{D(z)}$	0.0318	0.6661	0.717	0.74	0.0463
$\hat{E}$	0.0	0	0	0	0.0

### Распределение Пуассона

Размер выборки 10 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	9.94	9.9285	10.1405	9.8495	9.8907
$D(z)$	0.9661	4.8991	5.074	5.2646	1.7006
$E(x) - \sqrt{D(z)}$	8.9571	7.7151	7.8879	7.555	8.5866
$E(x) + \sqrt{D(z)}$	10.9229	12.1419	12.3931	12.144	11.1948
$\hat{E}$	-	-	-	-	-

Размер выборки 100 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	10.0069	10.0975	10.0605	10.126	10.0177
$D(z)$	0.1039	4.9797	5.2561	4.9971	0.2036
$E(x) - \sqrt{D(z)}$	9.6846	7.866	7.7679	7.8906	9.5665
$E(x) + \sqrt{D(z)}$	10.3292	12.329	12.3531	12.3614	10.4689
$\hat{E}$	-	-	-	-	-

Размер выборки 1000 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	10.0002	10.0	10.042	9.9615	10.0044
$D(z)$	0.0091	4.6535	5.0587	5.1613	0.0203
$E(x) - \sqrt{D(z)}$	9.9048	7.8428	7.7928	7.6897	9.8619
$E(x) + \sqrt{D(z)}$	10.0956	12.1572	12.2912	12.2333	10.1469
$\hat{E}$	-	-	-	-	-

### Равномерное распределение

Размер выборки 10 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_tr$
$E(z)$	-0.003	0.0173	0.0014	-0.0396	-0.0108
$D(z)$	0.0993	0.4596	0.4807	0.4957	0.1684
$E(x) - \sqrt{D(z)}$	-0.3181	-0.6606	-0.6919	-0.7437	-0.4212
$E(x) + \sqrt{D(z)}$	0.3121	0.6952	0.6947	0.6645	0.3996
$\hat{E}$	0	0	0	0	0

Размер выборки 100 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_t r$
$E(z)$	0.0007	0.0123	-0.0033	-0.0225	-0.0009
$D(z)$	0.01	0.5037	0.4761	0.4768	0.0202
$E(x) - \sqrt{D(z)}$	-0.0993	-0.6974	-0.6933	-0.713	-0.143
$E(x) + \sqrt{D(z)}$	0.1007	0.722	0.6867	0.668	0.1412
$\hat{E}$	0	0	0	0	0

Размер выборки 1000 элементов

	$\bar{x}$	median	$z_r$	$z_Q$	$z_t r$
$E(z)$	0.0006	-0.0271	0.0048	0.0121	0.0002
$D(z)$	0.0009	0.5208	0.4766	0.4862	0.0018
$E(x) - \sqrt{D(z)}$	-0.0294	-0.7488	-0.6856	-0.6852	-0.0422
$E(x) + \sqrt{D(z)}$	0.0306	0.6946	0.6952	0.7094	0.0426
$\hat{E}$	0.0	0	0	0	0.0

### 4.3 Боксплот Тьюки

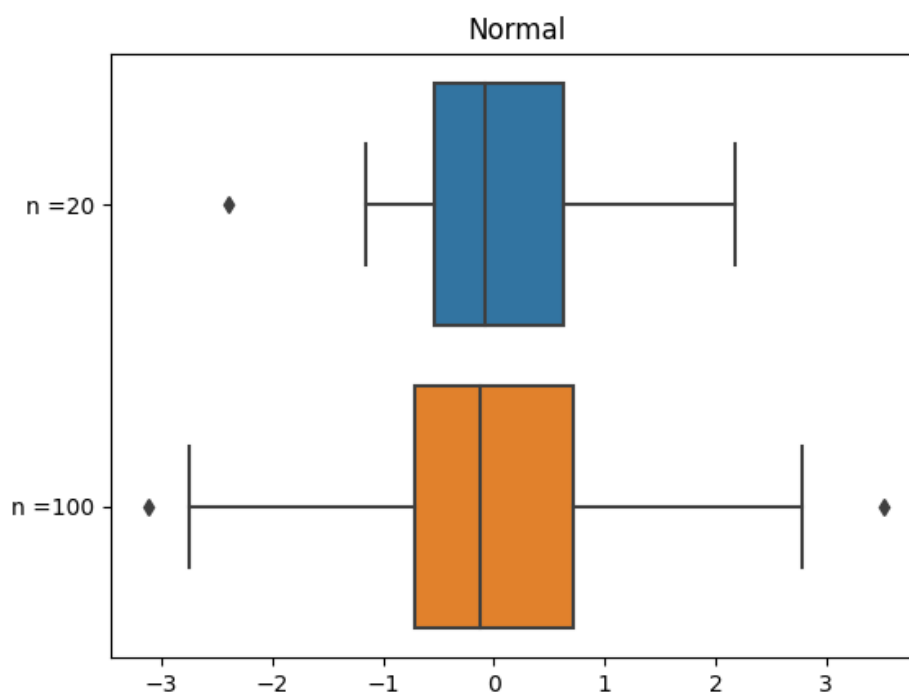


Рис. 6: Нормальное распределение

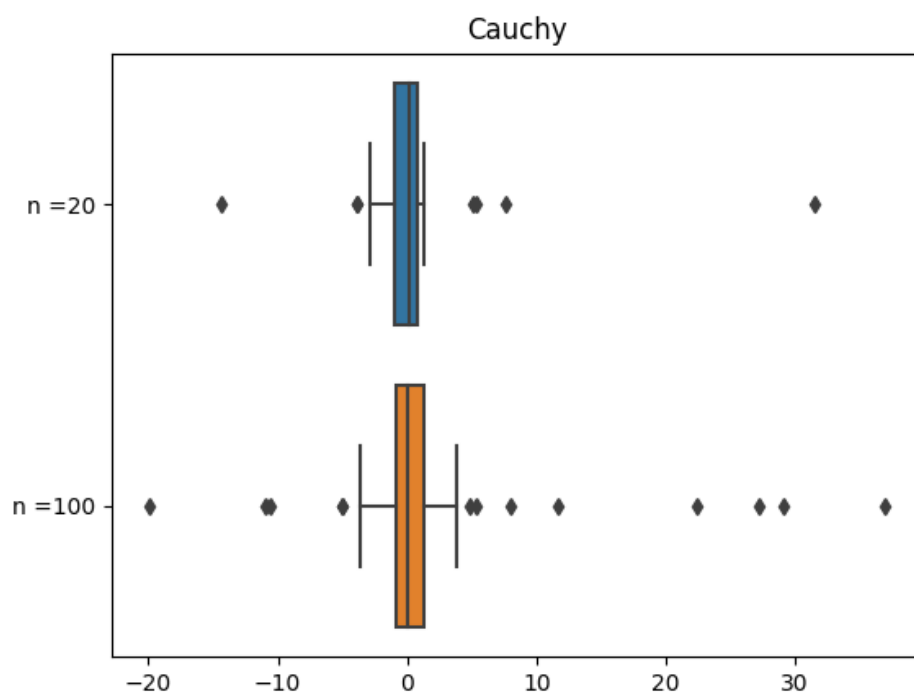


Рис. 7: Распределение Коши

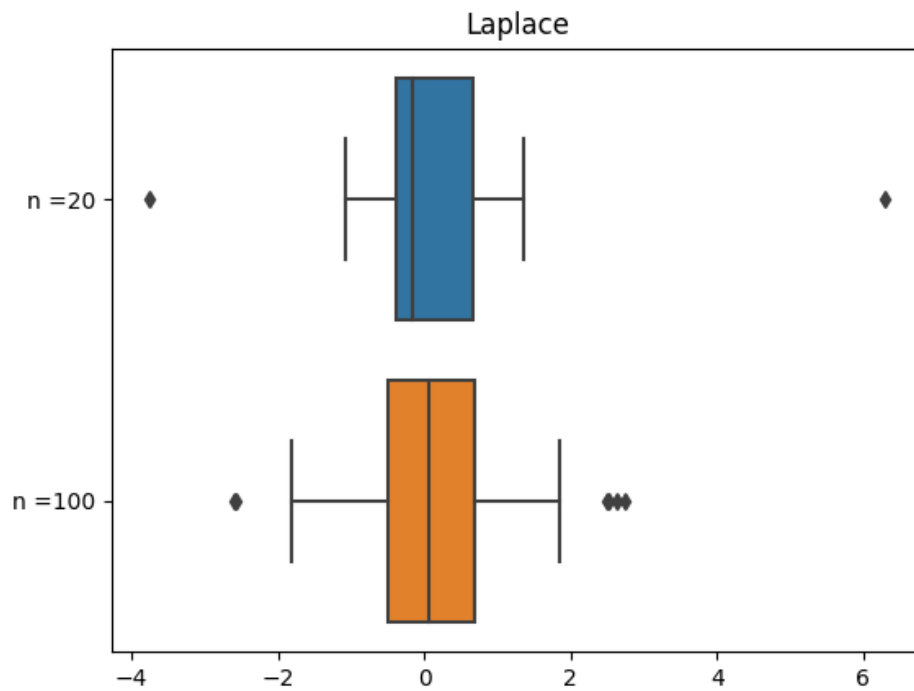


Рис. 8: Распределение Лапласа

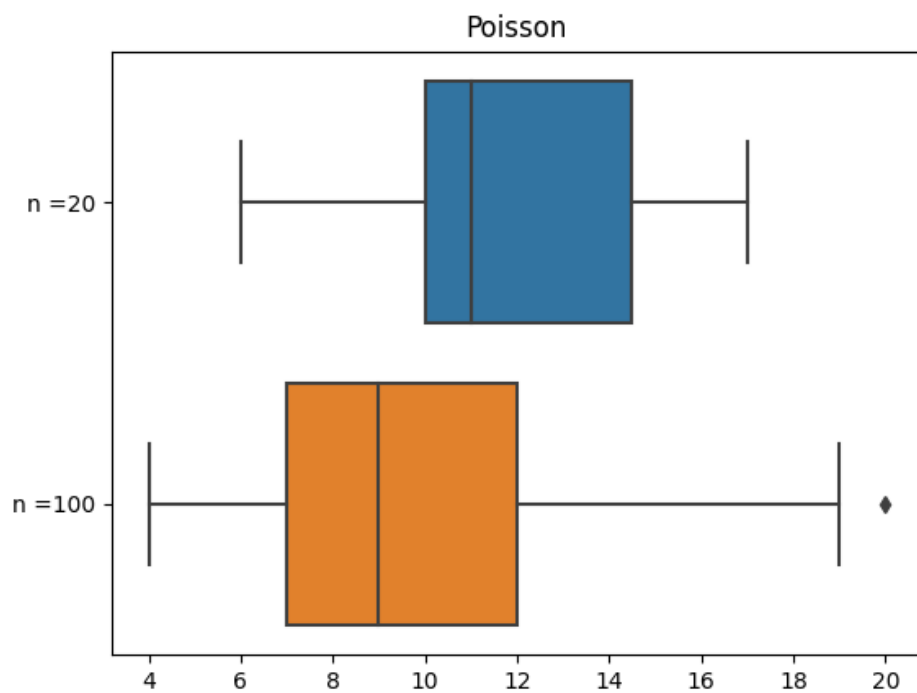


Рис. 9: Распределение Пуассона

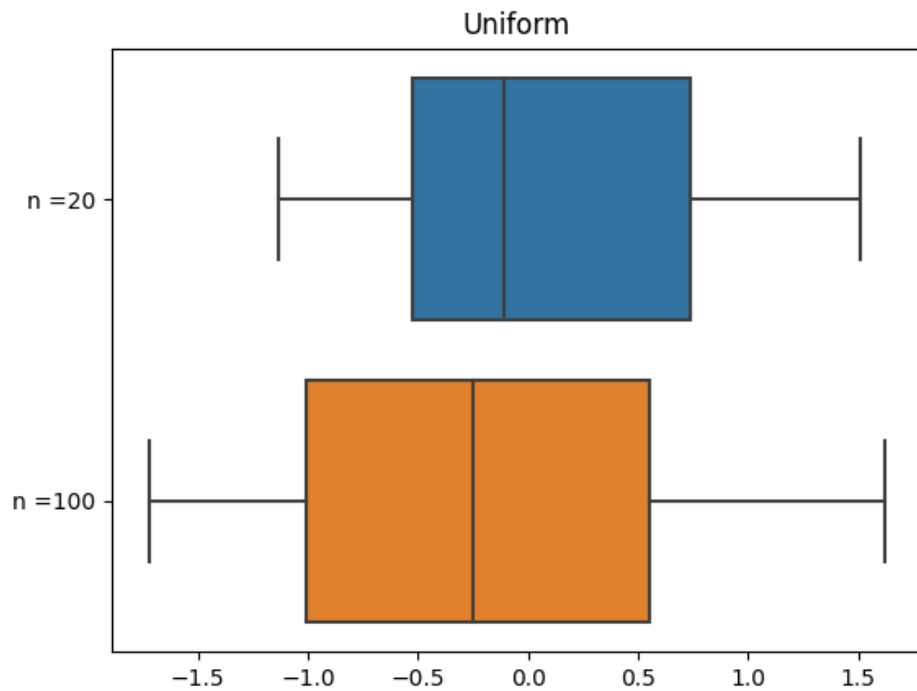


Рис. 10: Равномерное распределение



#### 4.4 Доля выбросов

Sample	Share of emissions
Normal n = 20	0.108
Normal n = 100	0.097
Cauchy n = 20	0.217
Cauchy n = 100	0.224
Laplace n = 20	0.16
Laplace n = 100	0.157
Poisson n = 20	0.111
Poisson n = 100	0.09
Uniform n = 20	0.052
Uniform n = 100	0.024

#### 4.5 Теоретическая вероятность выбросов

Распределение	$P_B^T$
Нормальное распределение	0.007
Распределение Коши	0.156
Распределение Лапласа	0.063
Распределение Пуассона	0.008
Равномерное распределение	0

Таблица 2: Теоретическая вероятность выбросов.

#### 4.6 Эмпирическая функция распределения

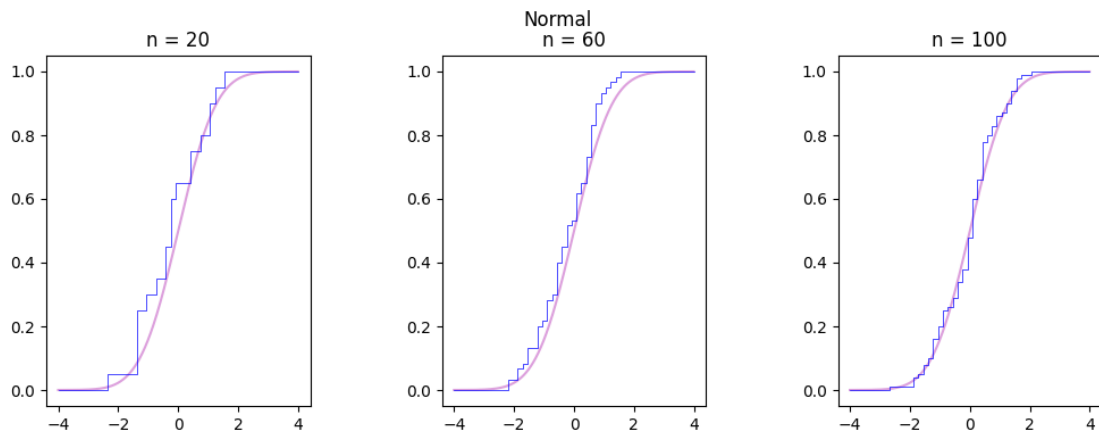


Рис. 11: Нормальное распределение

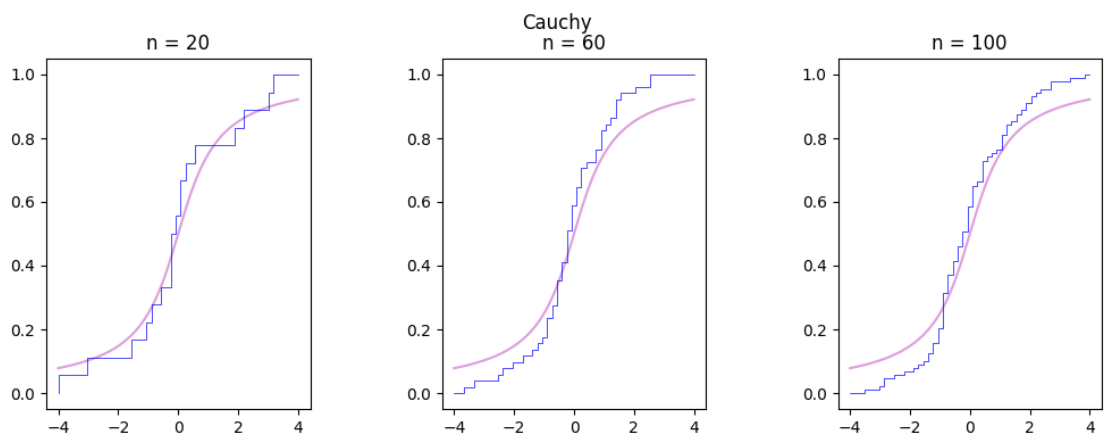


Рис. 12: Распределение Коши

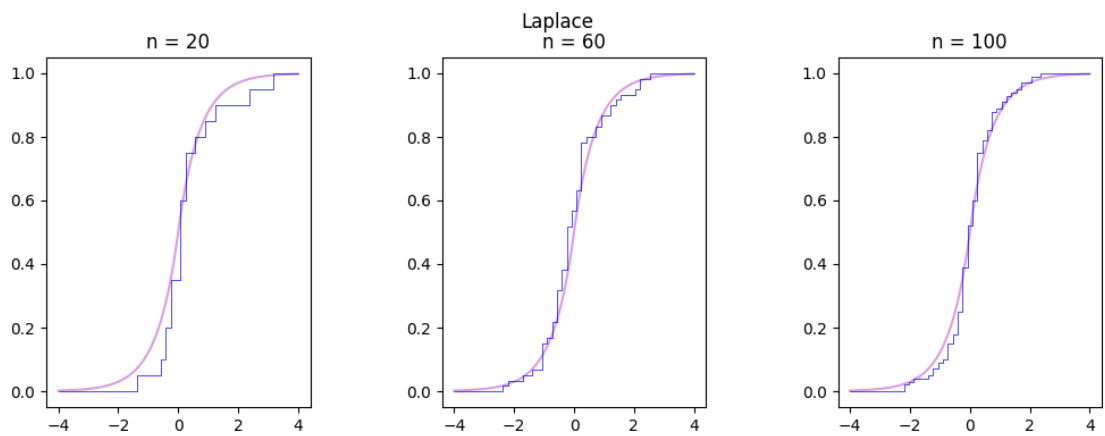


Рис. 13: Распределение Лапласа

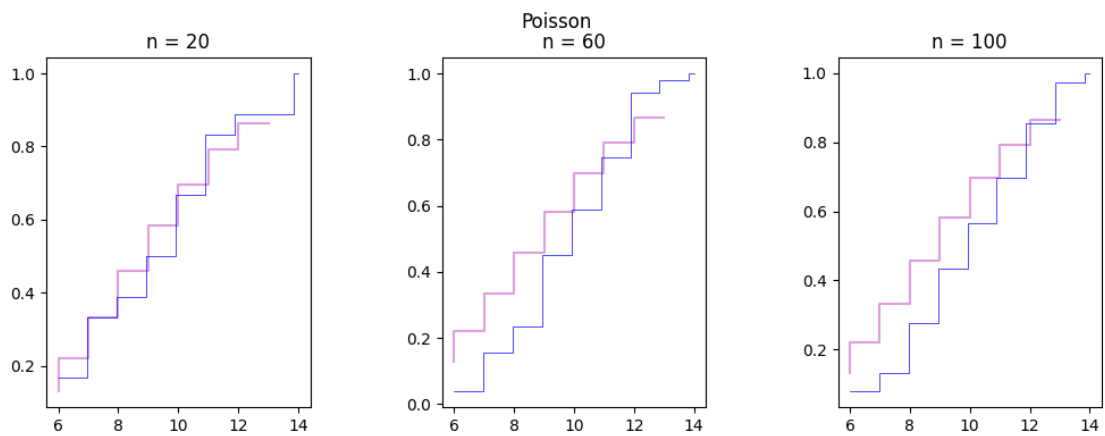


Рис. 14: Распределение Пуассона

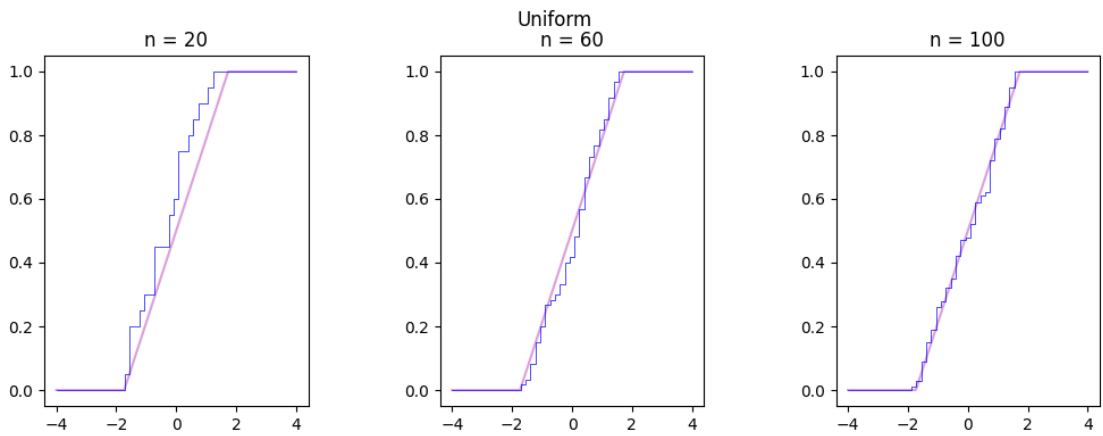


Рис. 15: Равномерное распределение

## 4.7 Ядерные оценки плотности распределения

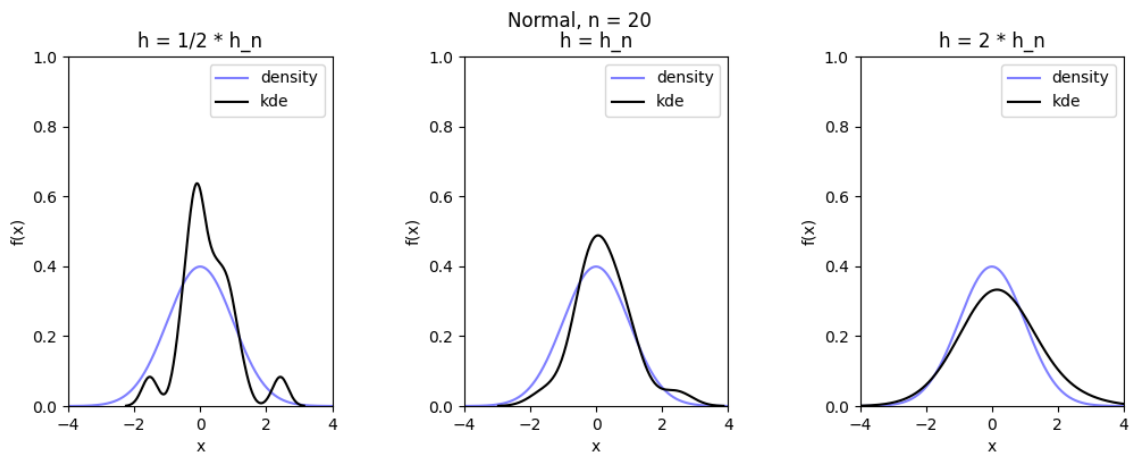


Рис. 16: Нормальное распределение,  $n = 20$

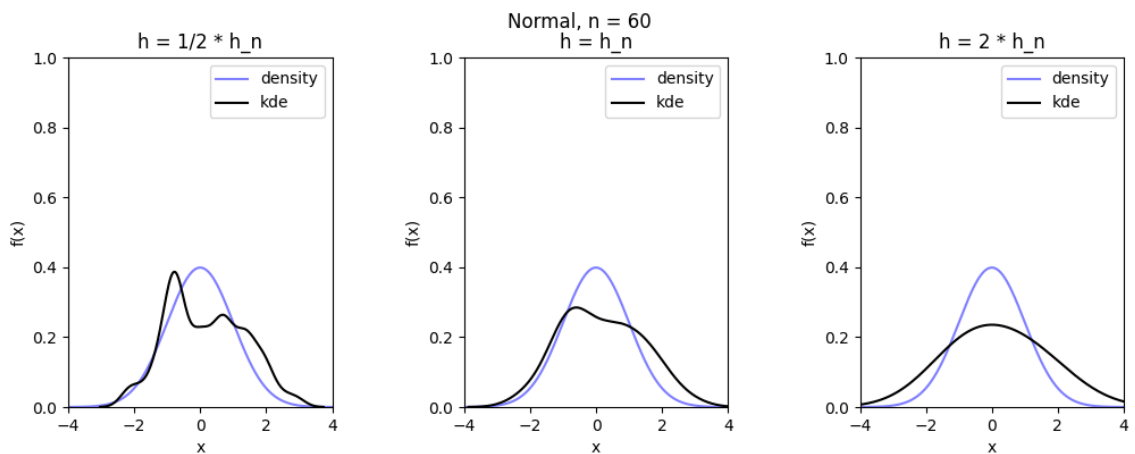


Рис. 17: Нормальное распределение,  $n = 60$

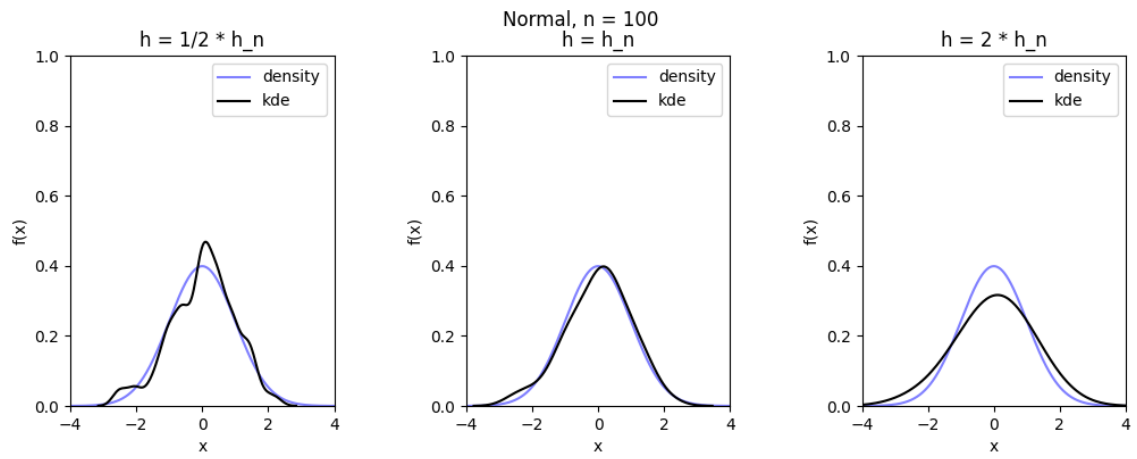


Рис. 18: Нормальное распределение,  $n = 100$

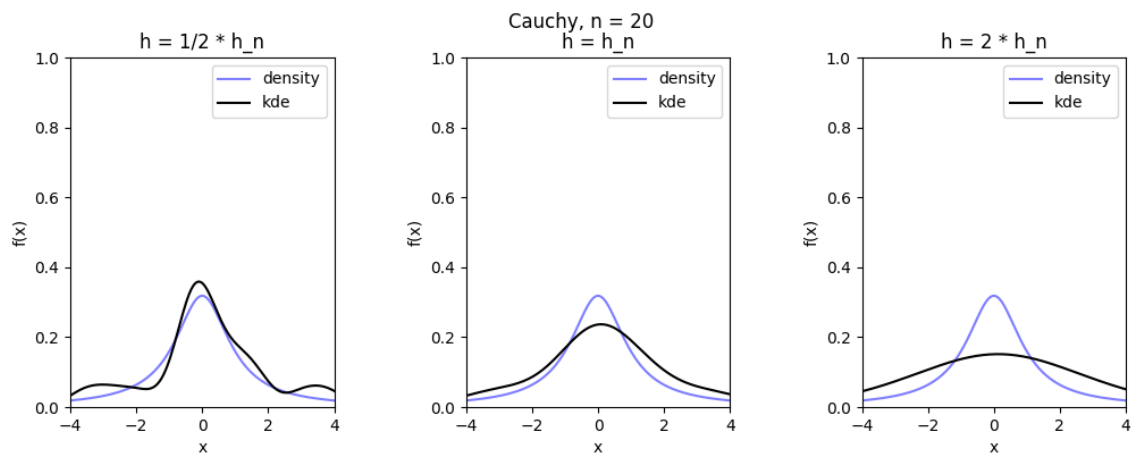


Рис. 19: Распределение Коши,  $n = 20$

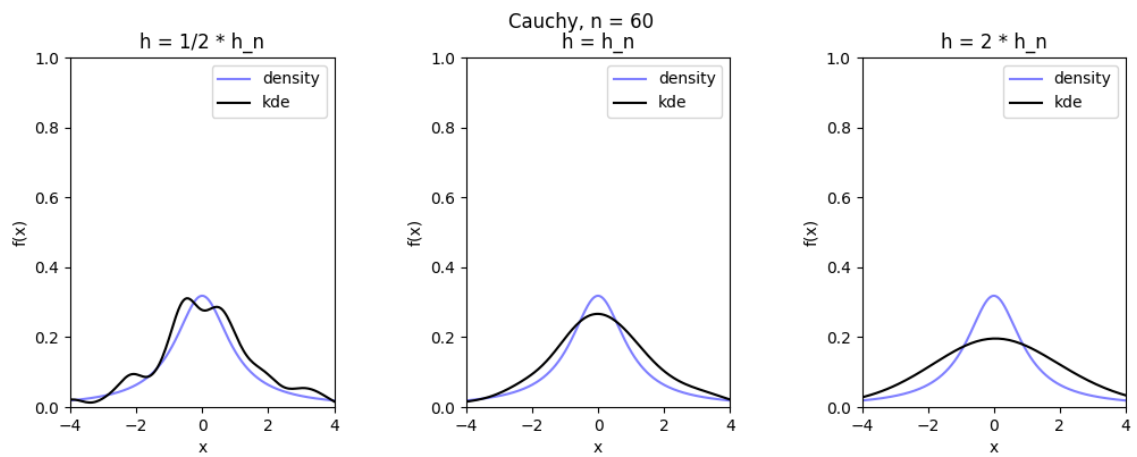


Рис. 20: Распределение Коши,  $n = 60$

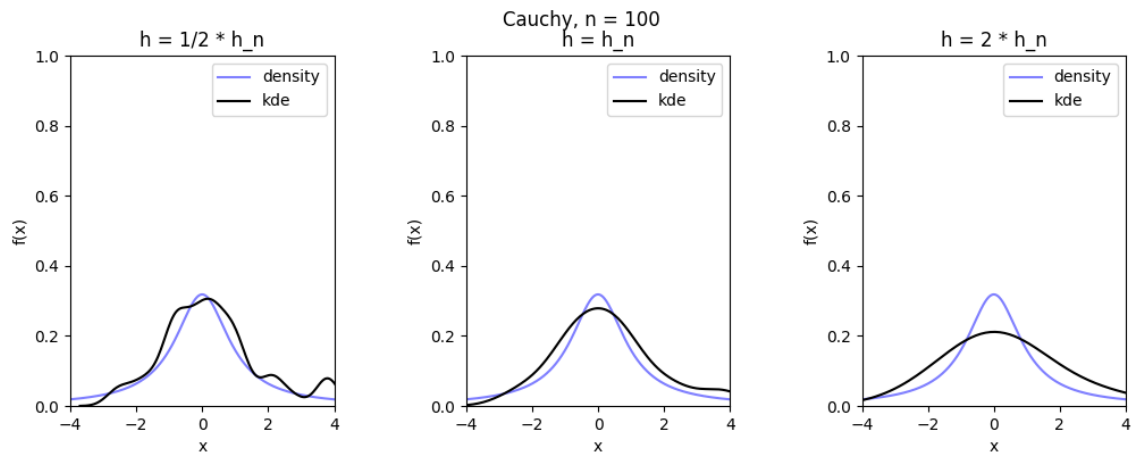


Рис. 21: Распределение Коши,  $n = 100$

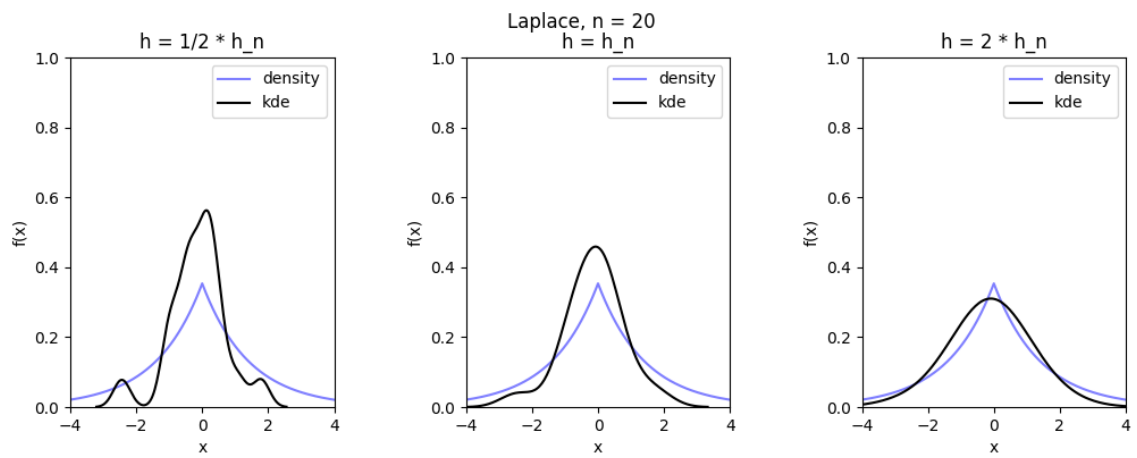


Рис. 22: Распределение Лапласа,  $n = 20$

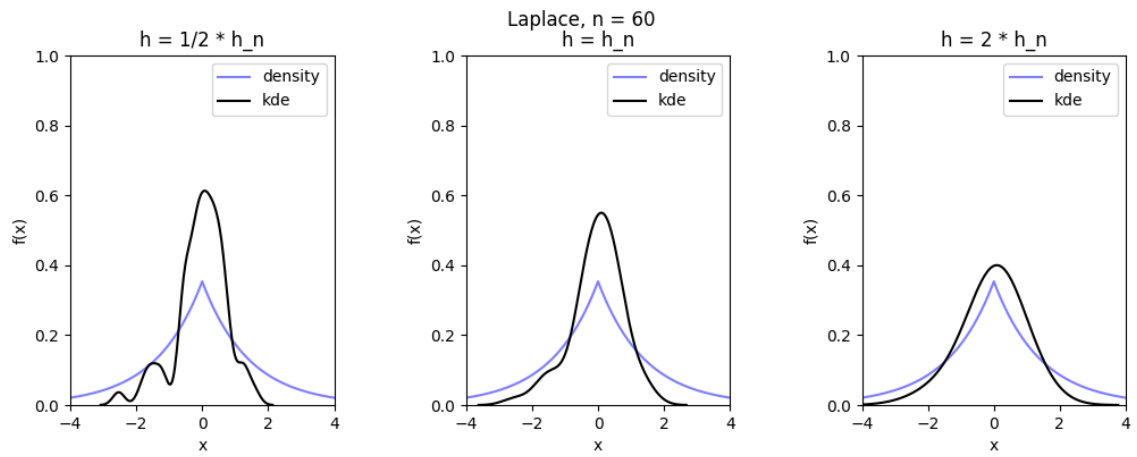


Рис. 23: Распределение Лапласа,  $n = 60$

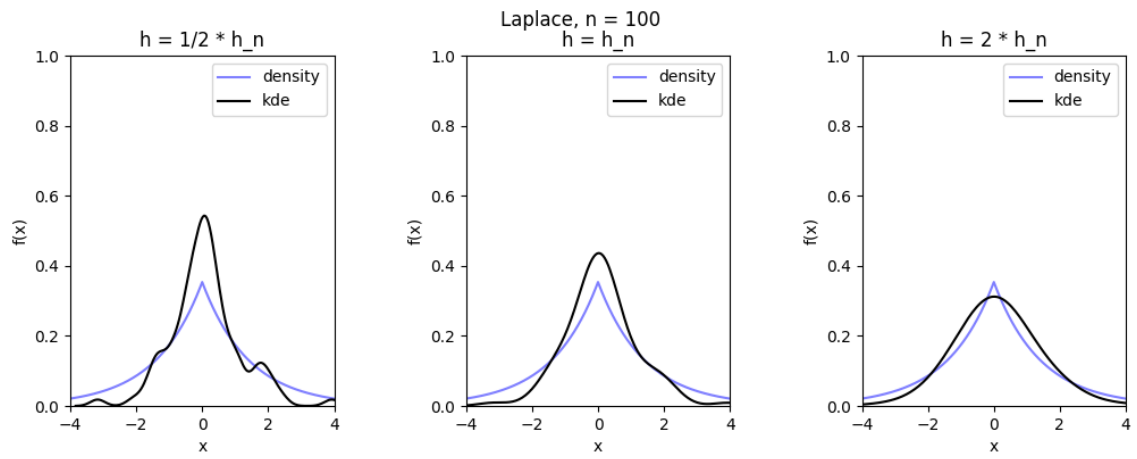


Рис. 24: Распределение Лапласа,  $n = 100$

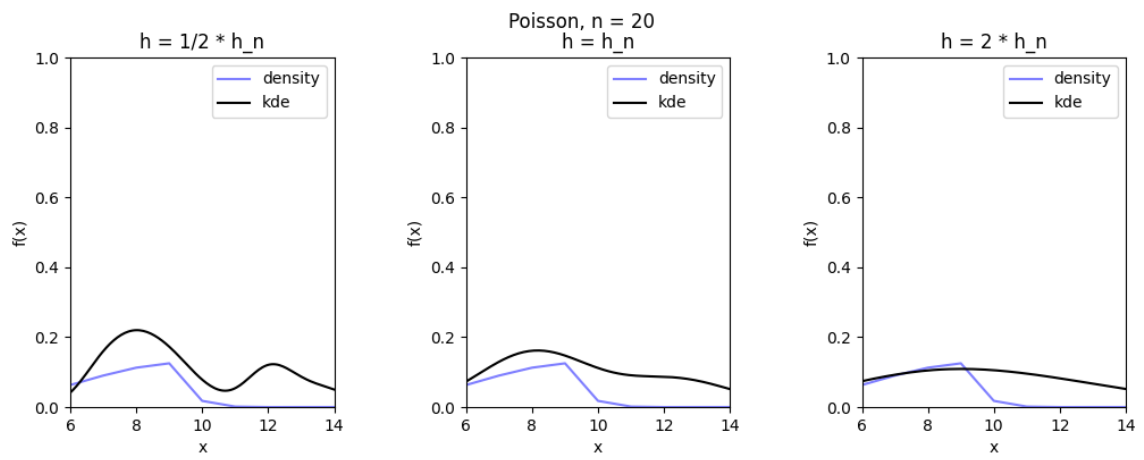


Рис. 25: Распределение Пуассона,  $n = 20$

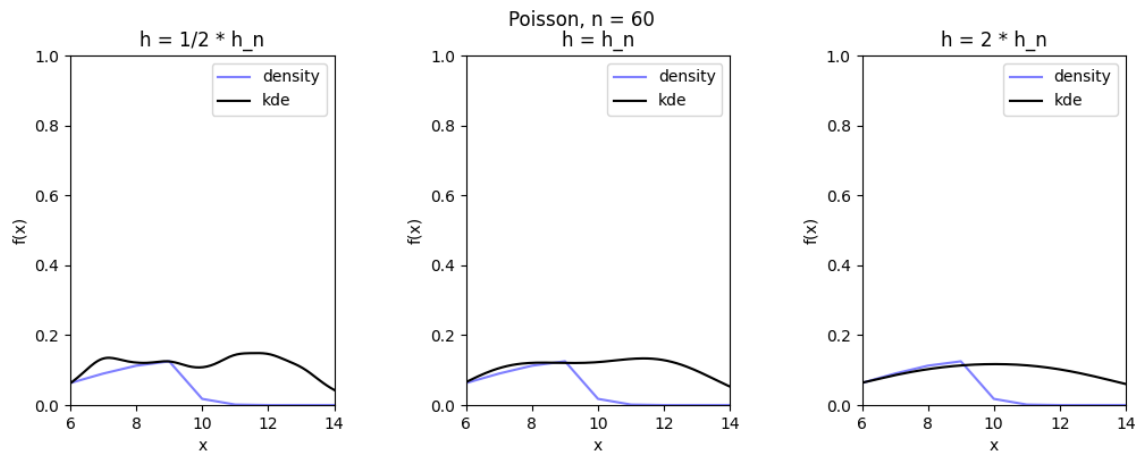


Рис. 26: Распределение Пуассона,  $n = 60$

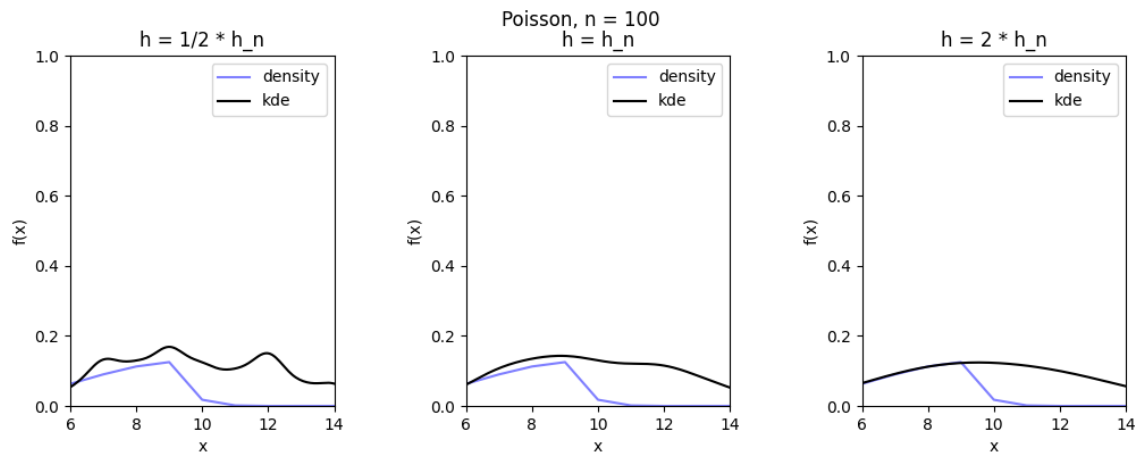


Рис. 27: Распределение Пуассона,  $n = 100$

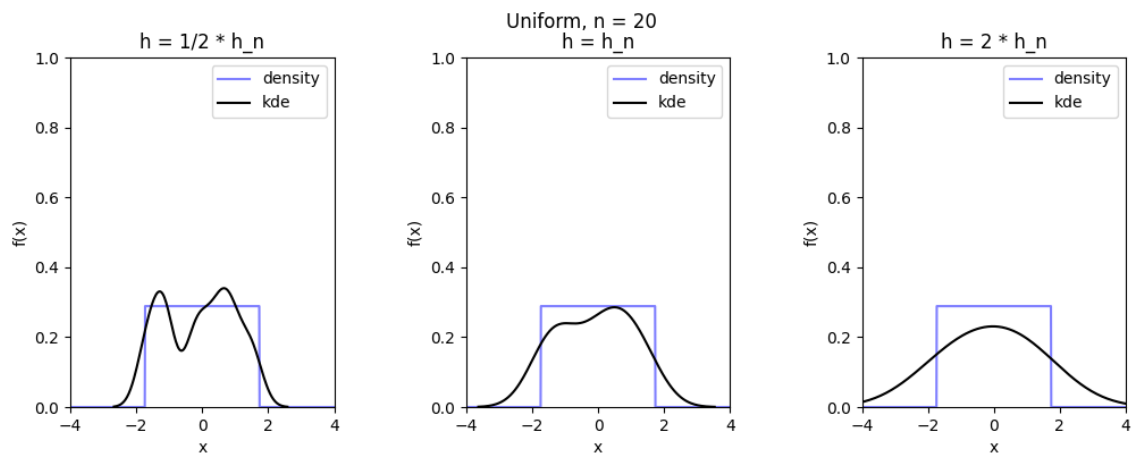


Рис. 28: Равномерное распределение,  $n = 20$

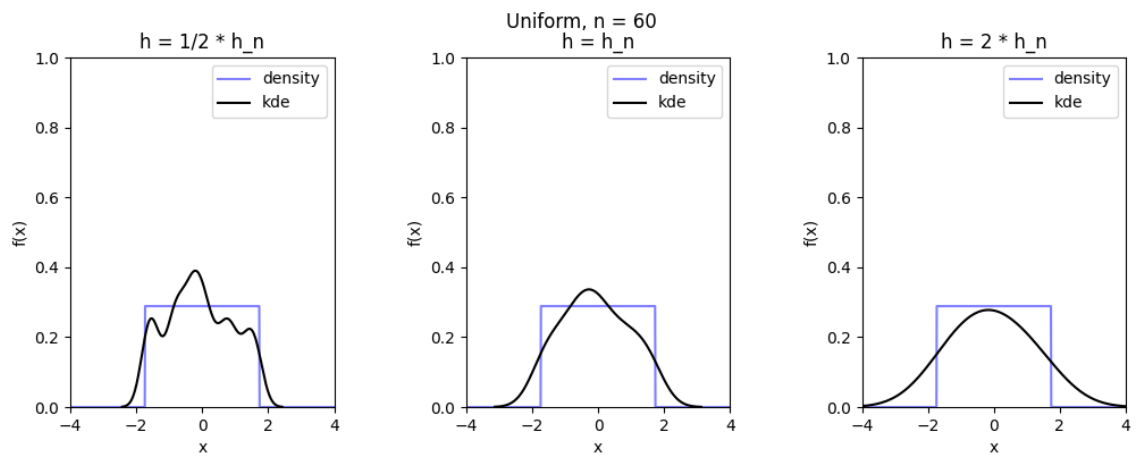


Рис. 29: Равномерное распределение,  $n = 60$

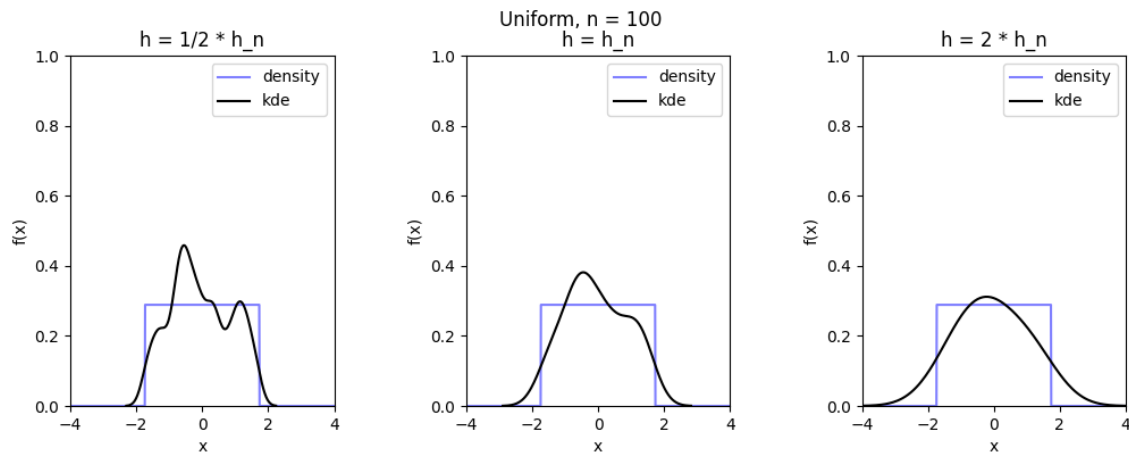


Рис. 30: Равномерное распределение,  $n = 100$

## 5 Обсуждение

### 5.1 Гистограмма и плотность распределения

Из построенных графиков видно, что чем больше мощность выборки для каждого из распределений, тем лучше гистограмма приближает плотность распределения, из которого сгенерирована выборка. Отметим, что чем меньше выборка, тем менее она показательна и тем менее точно она приближает плотность распределения – так, например, при  $n = 10$  гистограммы нормального и равномерного распределений выглядят достаточно похоже, несмотря на принципиальные отличия плотностей этих распределений.

Также отметим, что максимумы гистограмм и плотностей даже для больших выборок у большинства распределений не совпали. Кроме того, для распределения Пуассона и  $n = 1000$  гистограмма неплохо приближает плотность распределения, но при этом в выборке вообще не оказалось числа 9 (при параметре распределения Пуассона  $\lambda = 10$ )

### 5.2 Характеристики положения и рассеяния

Исходя из данных, приведенных в таблицах, видим, что для большинства распределений при увеличении мощности выборки выборочные характеристики положения и рассеяния приближаются к теоретическим. Однако, для распределения Коши данные довольно хаотичны, а дисперсии выборочных характеристик крайне велики – это следствие большого числа выбросов и не существования у распределения Коши моментов. Тем не менее, медиана существует и равна 0, поэтому и в таблице данные для медианы приближаются к теоретическим, а дисперсия выборочной медианы уменьшается при увеличении мощности выборки.

### 5.3 Доля и теоретическая вероятность выбросов

По данным, приведенным в таблицах, можем сделать вывод, что чем больше мощность выборки, тем ближе доля выбросов к теоретической вероятности выбросов. Отметим, что для распределения Коши доля выбросов значительно выше, чем для остальных распределений.

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

На графиках э. ф. р. можем наблюдать, что ступенчатая эмпирическая функция распределения тем лучше приближает теоретическую, чем больше мощность выборки. Отметим, что для распределения Пуассона и равномерного распределения отклонение эмпирической функции от теоретической наибольшее.

Графики ядерных оценок плотности распределения показывают, как с ростом мощности выборки ядерная оценка плотности распределения приближается к теоретической. Отметим, что это приближение наименее выражено для  $h = 2h_n$ , поскольку этот параметр оказывается велик и ядерные оценки для разных распределений становятся мало отличимы друг от друга.

Отметим, что чем больше параметр  $h$ , тем меньше у ядерной оценки плотности распределения число изменений знака производной – так, при  $h = 2h_n$  все ядерные оценки унимодальны.

## 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.