

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ПЕТРА ВЕЛИКОГО

ФИЗИКО-МЕХАНИЧЕСКИЙ ИНСТИТУТ
КАФЕДРА «ПРИКЛАДНАЯ МАТЕМАТИКА»

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

Выполнил студент:

Абдуллин Данил Якупович
группа: 5030102/00101

Проверил:

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

Санкт-Петербург
2023 г.

Содержание

1	Постановка задачи	4
2	Теория	4
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
3	Реализация	8
4	Результаты	8
4.1	Гистограмма и график плотности распределения	8
4.2	Характеристики положения и рассеяния	11
4.3	Боксплот Тьюки	16
4.4	Доля выбросов	19
4.5	Теоретическая вероятность выбросов	19
4.6	Эмпирическая функция распределения	20
4.7	Ядерные оценки плотности распределения	21
5	Обсуждение	26
5.1	Гистограмма и плотность распределения	26
5.2	Характеристики положения и рассеяния	26
5.3	Доля и теоретическая вероятность выбросов	27
5.4	Эмпирическая функция и ядерные оценки плотности распределения	27
	Литература	28

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

Список таблиц

1	Таблица распределения	7
2	Нормальное распределение	11
3	Распределение Коши	12
4	Распределение Лапласа	13
5	Распределение Пуассона	14
6	Равномерное распределение	15
7	Доля выбросов	19
8	Теоретическая вероятность выбросов	19

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_{tr} . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристическое 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.

4 Результаты

4.1 Гистограмма и график плотности распределения

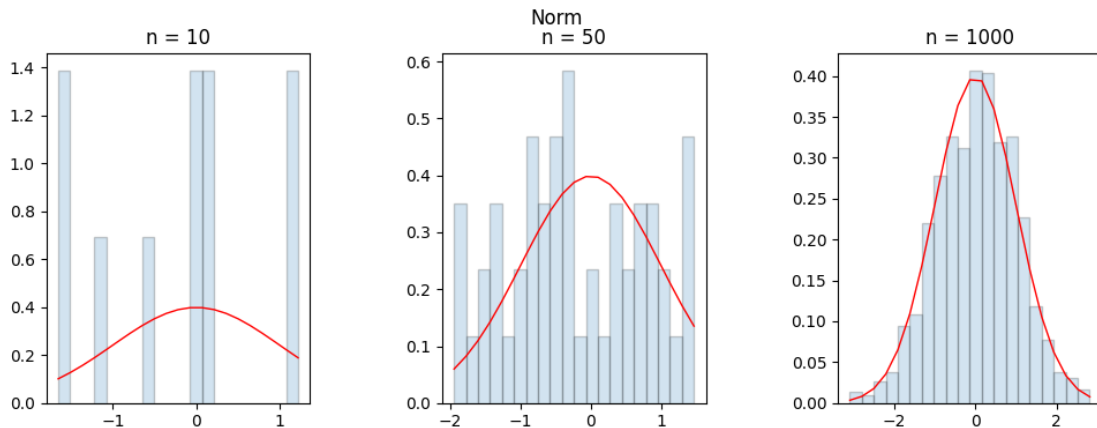


Рис. 1: Нормальное распределение

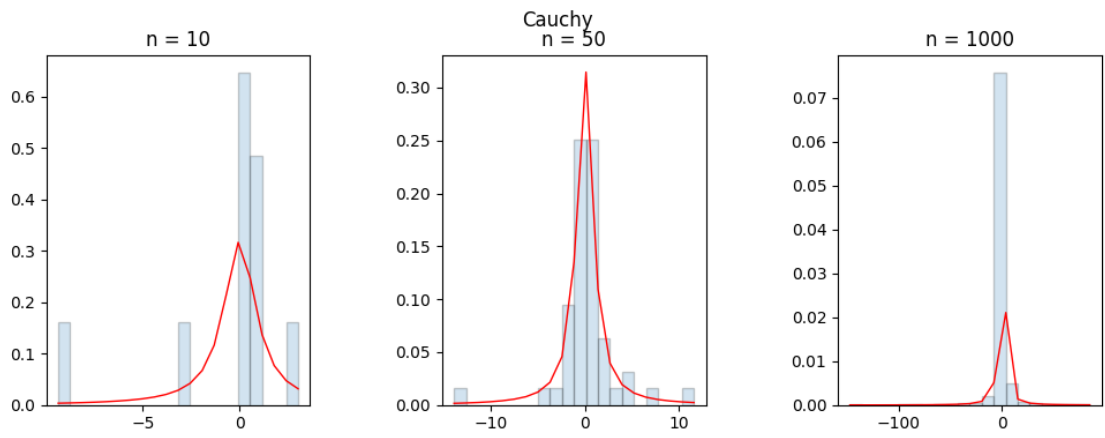


Рис. 2: Распределение Коши

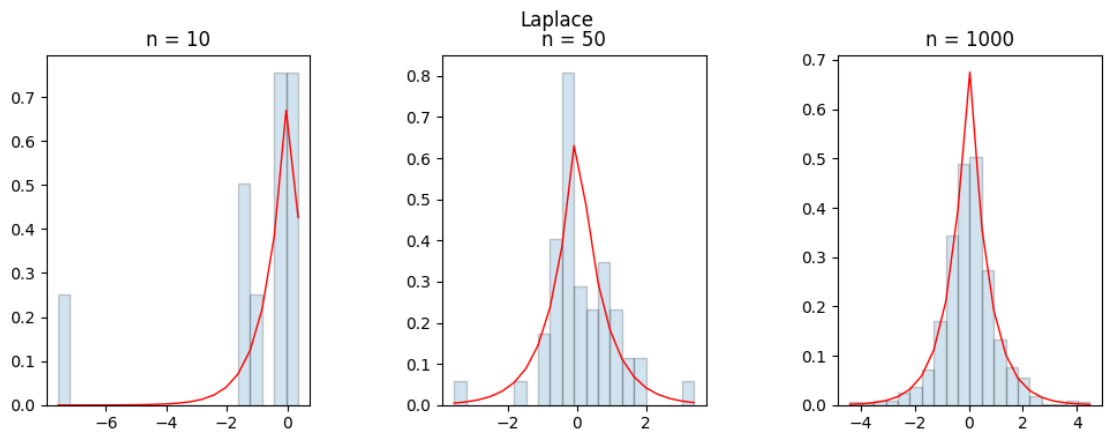


Рис. 3: Распределение Лапласа

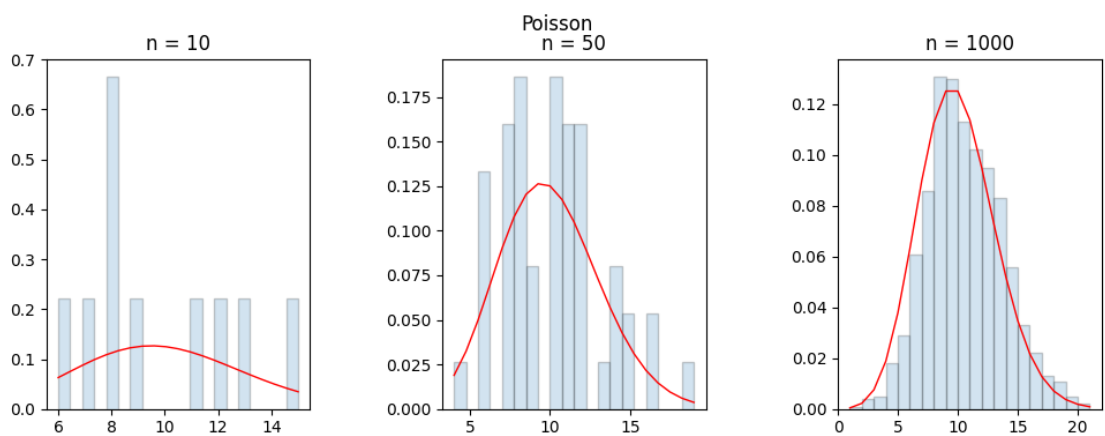


Рис. 4: Распределение Пуассона

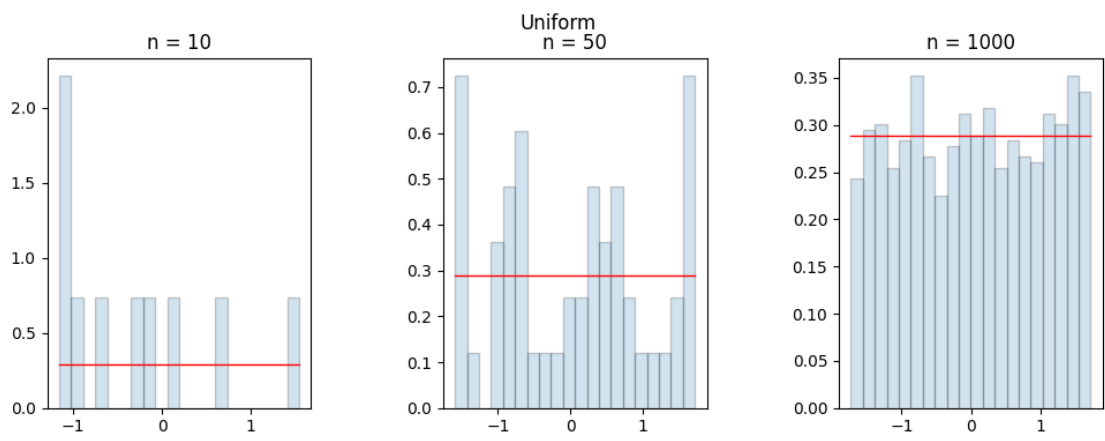


Рис. 5: Равномерное распределение

4.2 Характеристики положения и рассеяния

	\bar{x} (8)	$medx$ (9)	z_R (10)	z_Q (11)	z_{tr} (12)
n =10					
$E(z)$	-0.000283	0.008833	-0.010460	0.302207	0.117178
$D(z)$	0.097321	0.135348	0.179134	0.120981	0.079717
$E(z) - \sqrt{D(z)}$	-0.312246	-0.359064	-0.433703	-0.045616	-0.165163
$E(z) + \sqrt{D(z)}$	0.311681	0.376730	0.412782	0.650031	0.399520
$\hat{E}(z)$	0	0	0	0	0
n =100					
$E(z)$	-0.005917	-0.00619	-0.001552	0.008335	0.007469
$D(z)$	0.009765	0.015500	0.092342	0.011835	0.011044
$E(z) - \sqrt{D(z)}$	-0.104736	-0.130689	-0.305429	-0.100455	-0.097623
$E(z) + \sqrt{D(z)}$	0.092902	0.118309	0.302326	0.117126	0.112562
$\hat{E}(z)$	0	0	0	0	0
n =1000					
$E(z)$	0.001025	0.001256	0.000093	0.002219	0.002152
$D(z)$	0.001030	0.001591	0.059916	0.001250	0.001226
$E(z) - \sqrt{D(z)}$	-0.031067	-0.038628	-0.244685	-0.033133	-0.032862
$E(z) + \sqrt{D(z)}$	0.033116	0.041141	0.244871	0.037572	0.037166
$\hat{E}(z)$	0.0	0.0	0	0.0	0.0

Таблица 2: Нормальное распределение

	\bar{x} (8)	$medx$ (9)	z_R (10)	z_Q (11)
n =10				
$E(z)$	-1.377753	-0.034602	-6.471401	1.045653
$D(z)$	4338.309698	0.372108	106438.412364	3.727669
$E(z) - \sqrt{D(z)}$	-67.243599	-0.644609	-332.720404	-0.885064
$E(z) + \sqrt{D(z)}$	64.488094	0.575405	319.777603	2.976370
$\hat{E}(z)$	–	0	–	–
n =100				
$E(z)$	0.664422	0.008927	33.345240	0.040737
$D(z)$	247.993025	0.024689	598108.806534	0.047824
$E(z) - \sqrt{D(z)}$	-15.083373	-0.148199	-740.029706	-0.177951
$E(z) + \sqrt{D(z)}$	16.412216	0.166053	806.720185	0.259424
$\hat{E}(z)$	–	0	–	0
n =1000				
$E(z)$	0.243245	-0.001115	126.459766	0.004916
$D(z)$	978.427696	0.002458	242881346.020365	0.004922
$E(z) - \sqrt{D(z)}$	-31.036584	-0.050689	-15458.191209	-0.065240
$E(z) + \sqrt{D(z)}$	31.523073	0.048458	15711.110741	0.075072
$\hat{E}(z)$	–	0.0	–	0.0

Таблица 3: Распределение Коши

	\bar{x} (8)	$medx$ (9)	z_R (10)	z_Q (11)	z_{tr} (12)
n =10					
$E(z)$	-0.017196	-0.014853	-0.030922	0.275508	0.073415
$D(z)$	0.09144	0.063094	0.392255	0.108368	0.045346
$E(z) - \sqrt{D(z)}$	-0.319586	-0.266039	-0.657225	-0.053685	-0.139531
$E(z) + \sqrt{D(z)}$	0.285195	0.236333	0.59538	0.604701	0.286362
$\hat{E}(z)$	0	0	0	0	0
n =100					
$E(z)$	-0.001597	0.002159	-0.012405	0.013658	0.010094
$D(z)$	0.010055	0.005792	0.451687	0.010103	0.005949
$E(z) - \sqrt{D(z)}$	-0.101872	-0.073948	-0.684482	-0.086856	-0.067036
$E(z) + \sqrt{D(z)}$	0.098678	0.078265	0.659672	0.114172	0.087225
$\hat{E}(z)$	0	0.0	0	0	0.0
n =1000					
$E(z)$	-0.00042	0.000312	0.021462	0.001261	0.000911
$D(z)$	0.001074	0.000508	0.407138	0.001029	0.000628
$E(z) - \sqrt{D(z)}$	-0.033200	-0.02223	-0.616612	-0.030811	-0.024158
$E(z) + \sqrt{D(z)}$	0.032359	0.022855	0.659536	0.033334	0.025979
$\hat{E}(z)$	0.0	0.0	0	0.0	0.0

Таблица 4: Распределение Лапласа

	\bar{x} (8)	$medx$ (9)	z_R (10)	z_Q (11)	z_{tr} (12)
n =10					
$E(z)$	9.979000	9.852500	9.965500	10.044000	9.997167
$D(z)$	0.975799	1.373994	4.683560	5.018564	1.779575
$E(z) - \sqrt{D(z)}$	8.991170	8.680320	7.801340	7.803780	8.663150
$E(z) + \sqrt{D(z)}$	10.966800	11.024600	11.024600	12.284200	11.331100
$\hat{E}(z)$	10_{-1}^{+1}	10_{-1}^{+1}	10_{-2}^{+2}	10_{-2}^{+2}	10_{-1}^{+1}
n =100					
$E(z)$	9.98473	9.853500	9.961500	10.042500	9.986043
$D(z)$	0.534271	0.786788	4.841018	5.073944	0.982735
$E(z) - \sqrt{D(z)}$	9.253790	8.966480	7.761260	7.789950	8.994710
$E(z) + \sqrt{D(z)}$	10.715600	10.740500	12.161700	12.2950	10.977300
$\hat{E}(z)$	10_{-1}^{+1}	10_{-1}^{+1}	10_{-2}^{+2}	10_{-2}^{+2}	10_{-1}^{+1}
n =1000					
$E(z)$	9.990781	9.902000	9.955000	10.057000	9.990350
$D(z)$	0.359408	0.529563	4.769308	5.025751	0.661672
$E(z) - \sqrt{D(z)}$	9.391270	9.174280	7.771120	7.815180	9.176910
$E(z) + \sqrt{D(z)}$	10.590200	10.629700	12.138800	12.298800	10.803700
$\hat{E}(z)$	10_{-1}^{+1}	10_{-1}^{+1}	10_{-2}^{+2}	10_{-2}^{+2}	10_{-1}^{+1}

Таблица 5: Распределение Пуассона

	\bar{x} (8)	$medx$ (9)	z_R (10)	z_Q (11)	z_{tr} (12)
n =10					
$E(z)$	0.001218	-0.002673	0.001898	0.312834	0.134491
$D(z)$	0.100674	0.225633	0.044579	0.125054	0.120186
$E(z) - \sqrt{D(z)}$	-0.316074	-0.477681	-0.209239	-0.040796	-0.212188
$E(z) + \sqrt{D(z)}$	0.318509	0.472336	0.213034	0.666464	0.481170
$\hat{E}(z)$	0	0	0	0	0
n =100					
$E(z)$	-0.000991	-0.004632	0.000108	0.015719	0.016290
$D(z)$	0.010529	0.029954	0.000537	0.015070	0.020221
$E(z) - \sqrt{D(z)}$	-0.103603	-0.177704	-0.023059	-0.107042	-0.125909
$E(z) + \sqrt{D(z)}$	0.101620	0.168441	0.023274	0.138481	0.158489
$\hat{E}(z)$	0	0	0.0	0	0
n =1000					
$E(z)$	0.001090	0.003175	0.000034	0.002122	0.003430
$D(z)$	0.000975	0.003011	0.000006	0.001437	0.001979
$E(z) - \sqrt{D(z)}$	-0.03014	-0.051701	-0.002505	-0.035783	-0.041051
$E(z) + \sqrt{D(z)}$	0.032320	0.058052	0.002437	0.040026	0.047910
$\hat{E}(z)$	0.0	0.0	0.0	0.0	0.0

Таблица 6: Равномерное распределение

4.3 Боксплот Тьюки

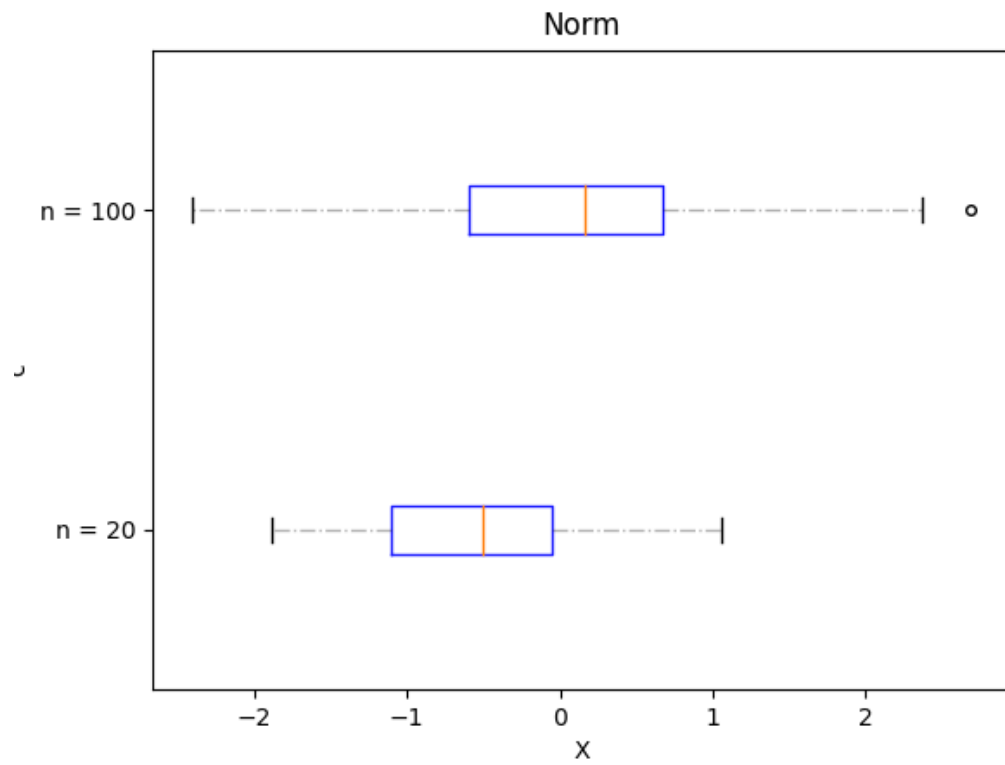


Рис. 6: Нормальное распределение

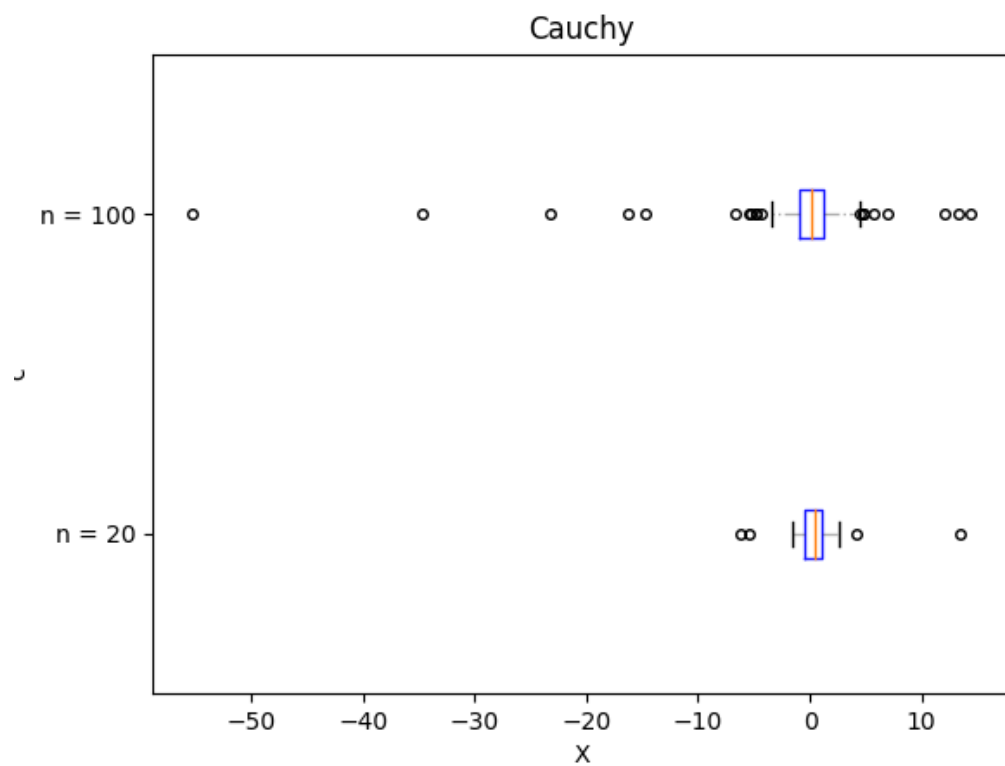


Рис. 7: Распределение Коши

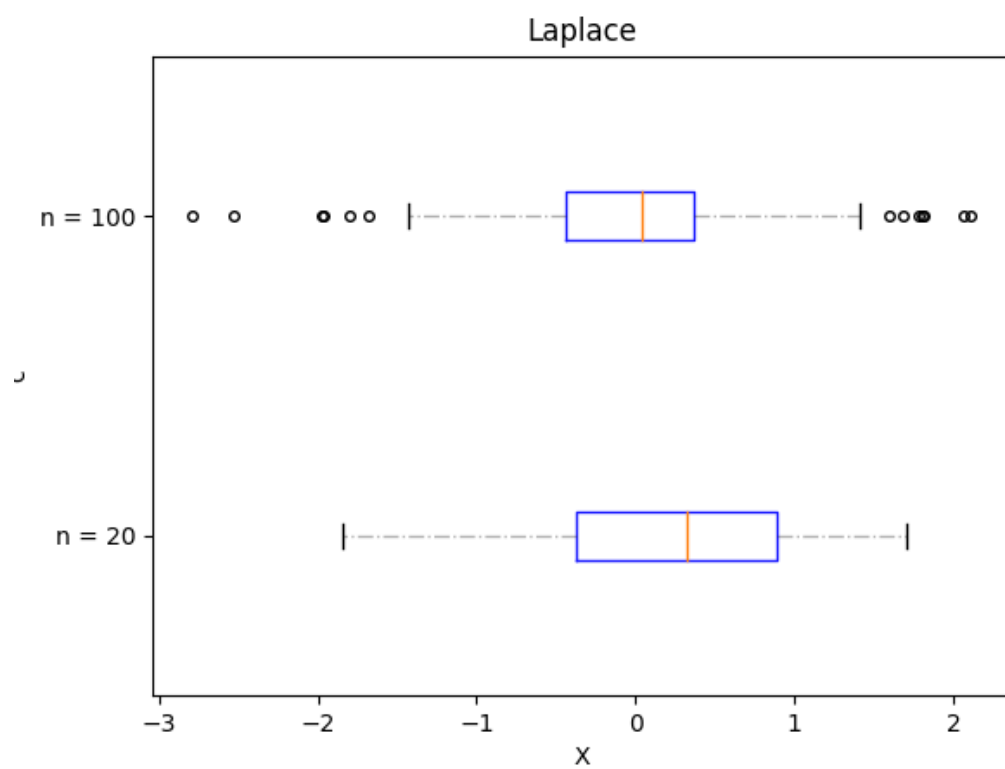


Рис. 8: Распределение Лапласа

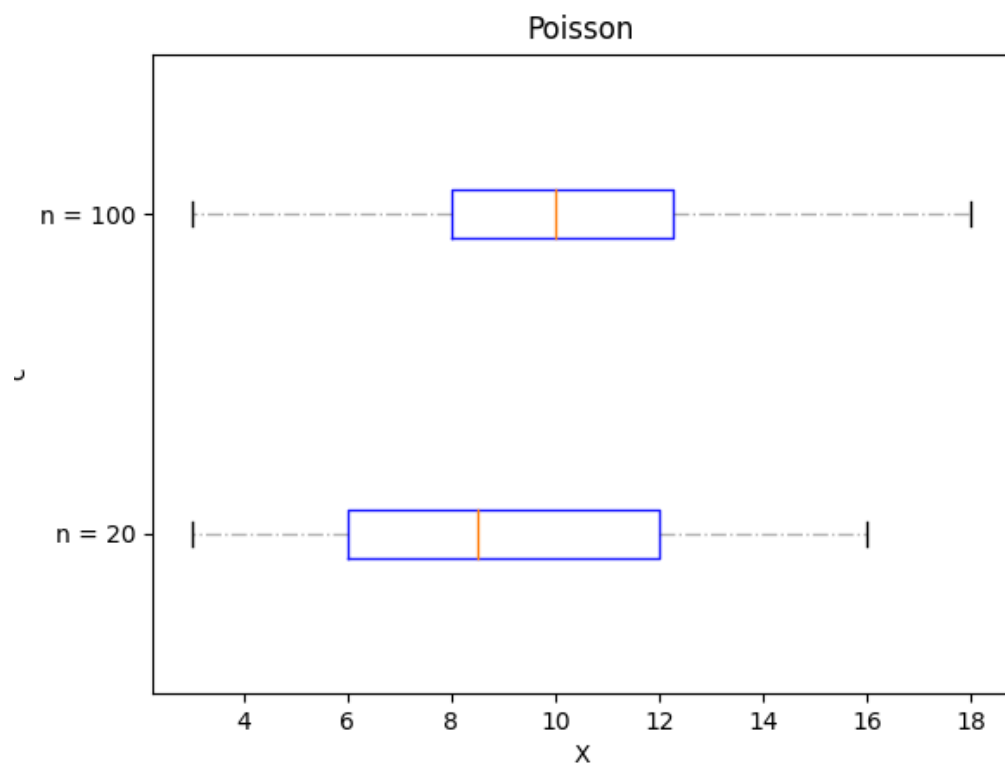


Рис. 9: Распределение Пуассона

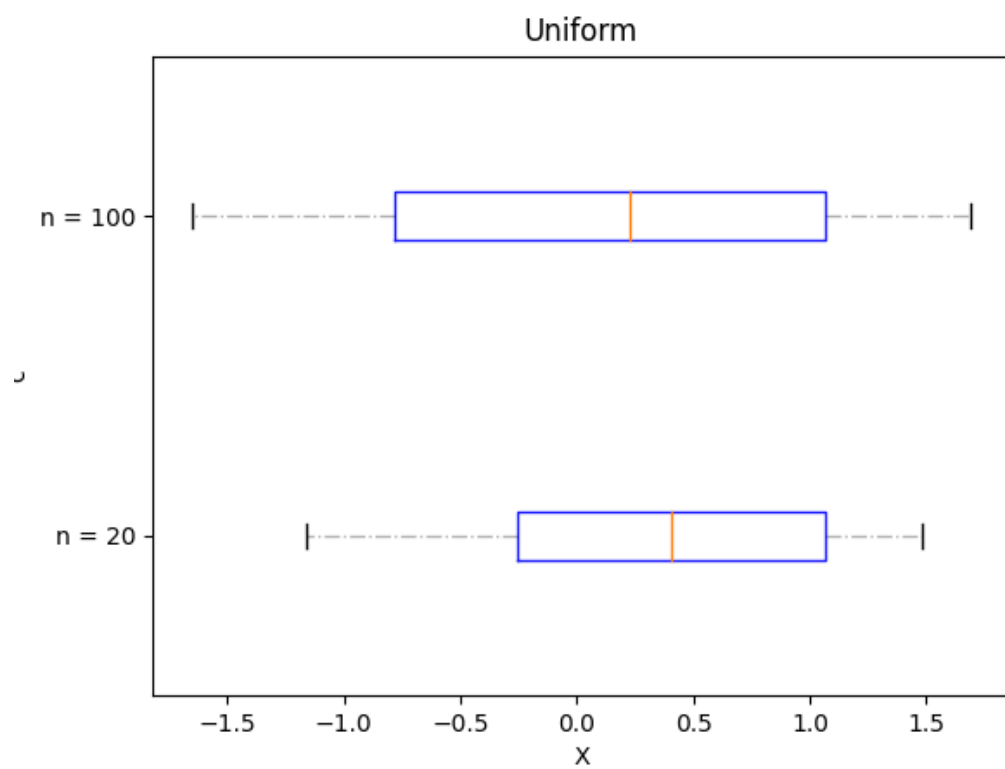


Рис. 10: Равномерное распределение

4.4 Доля выбросов

Выборка	Доля выбросов
Normal, 20	0.02
Normal, 100	0.01
Cauchy, 20	0.13
Cauchy, 100	0.16
Laplace, 20	0.06
Laplace, 100	0.06
Poisson, 20	0.01
Poisson, 100	0.01
Uniform, 20	0.00
Uniform, 100	0.00

Таблица 7: Доля выбросов

4.5 Теоретическая вероятность выбросов

Распределение	$P_{\text{в}}^{\text{T}}(16)$
Normal	0.007
Cauchy	0.156
Laplace	0.063
Poisson	0.008
Uniform	0

Таблица 8: Теоретическая вероятность выбросов

4.6 Эмпирическая функция распределения

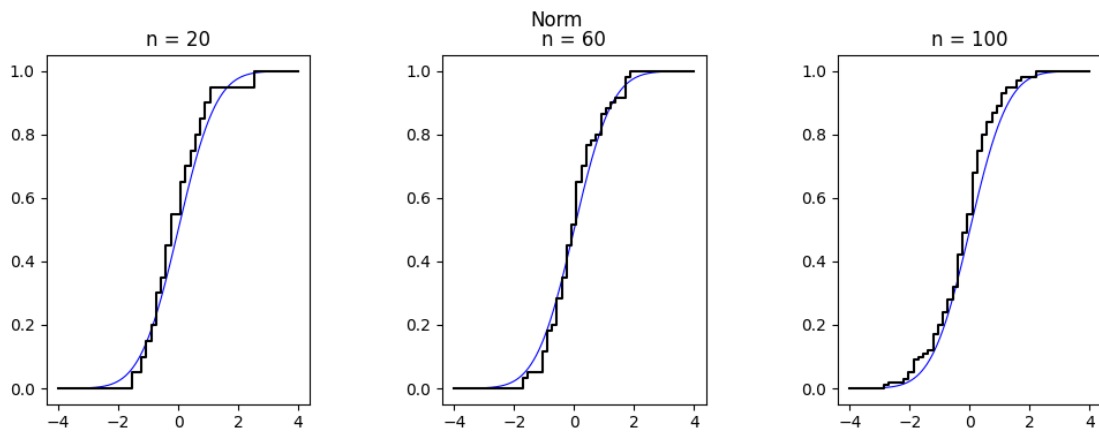


Рис. 11: Нормальное распределение

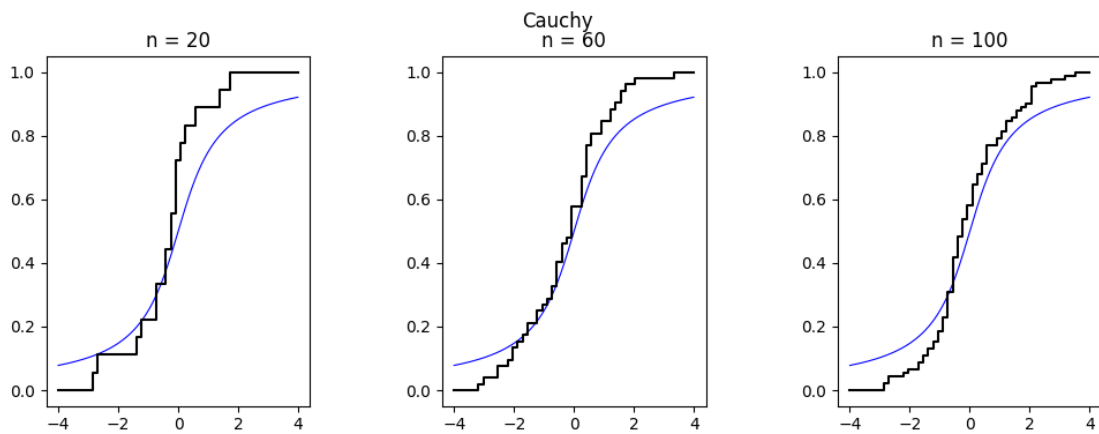


Рис. 12: Распределение Коши

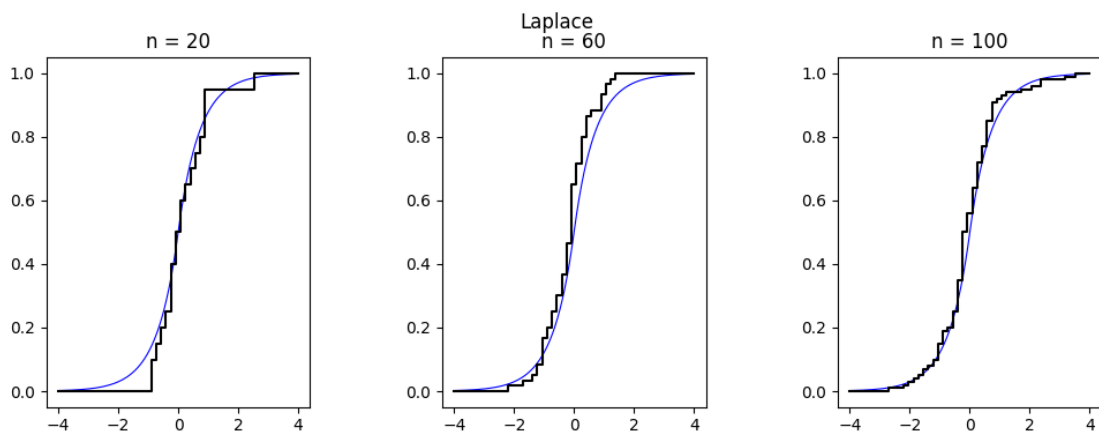


Рис. 13: Распределение Лапласа

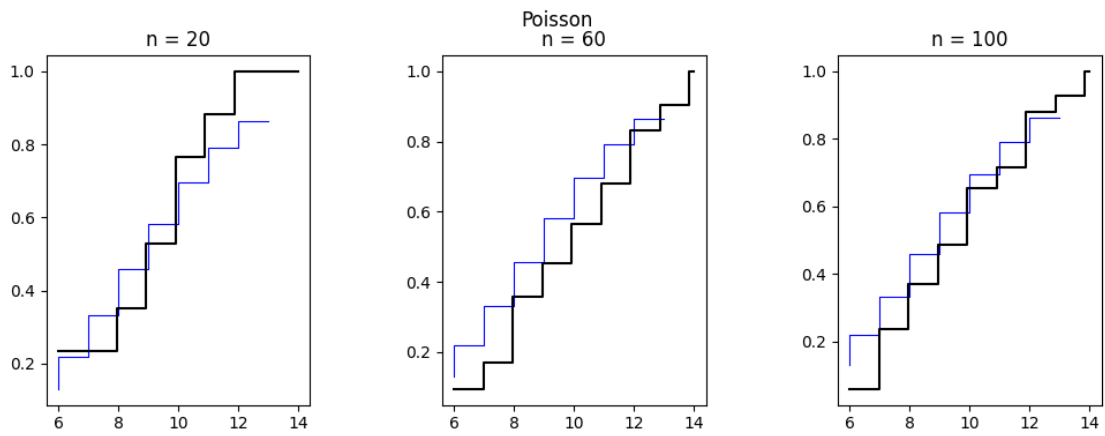


Рис. 14: Распределение Пуассона

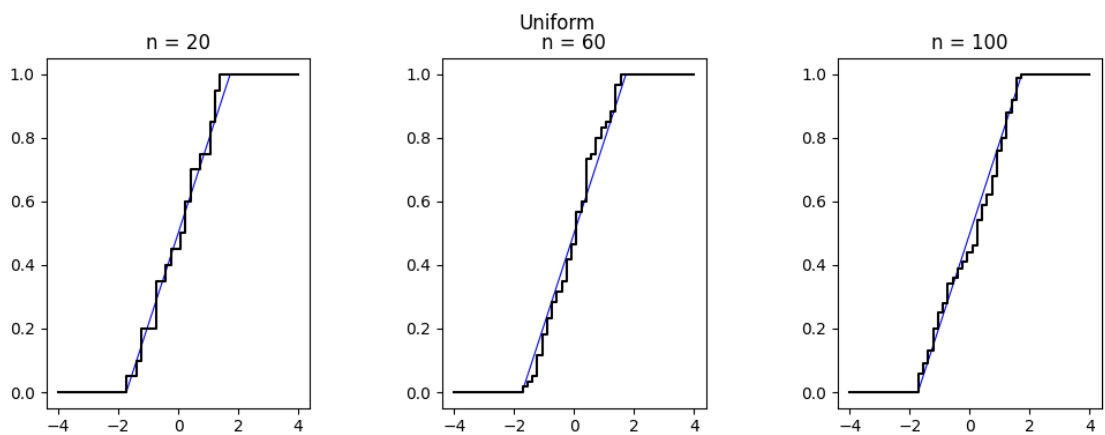


Рис. 15: Равномерное распределение

4.7 Ядерные оценки плотности распределения

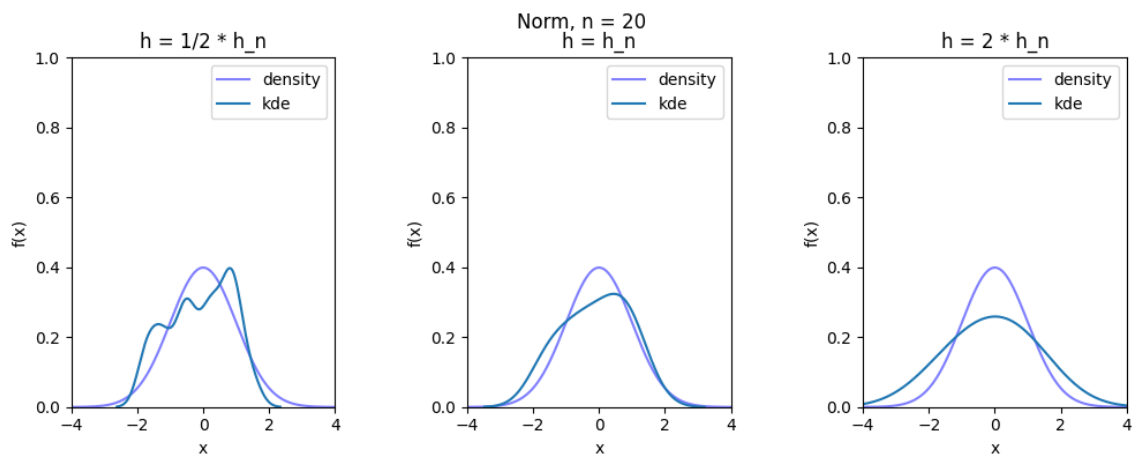


Рис. 16: Нормальное распределение, $n = 20$

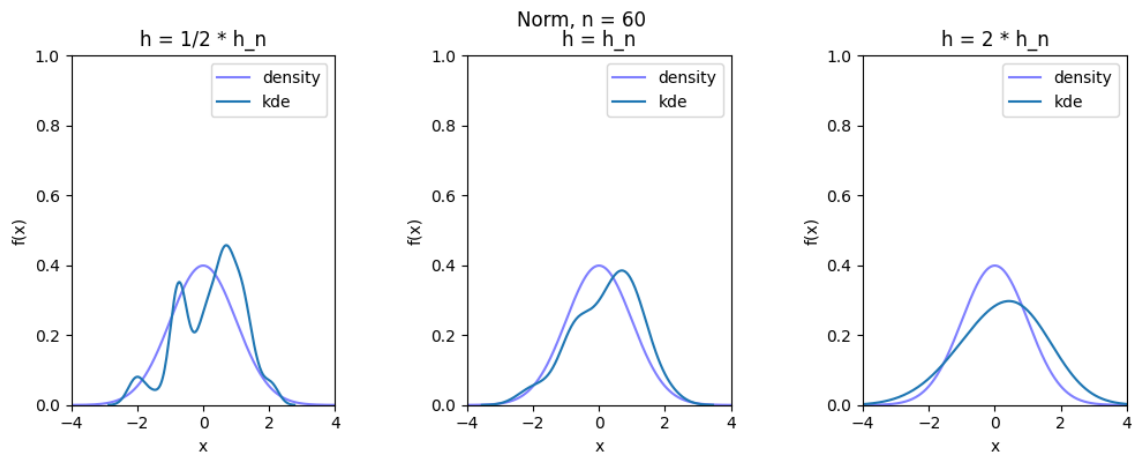


Рис. 17: Нормальное распределение, $n = 60$

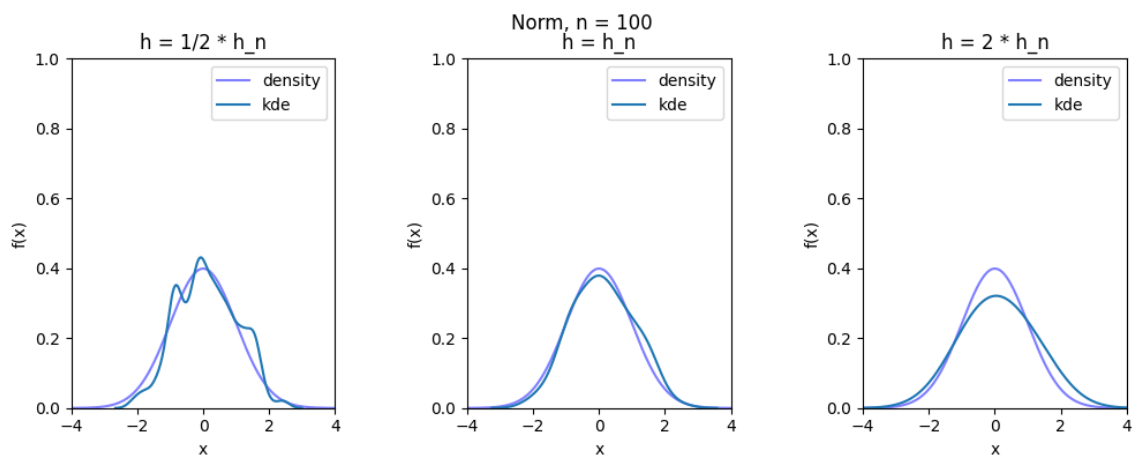


Рис. 18: Нормальное распределение, $n = 100$

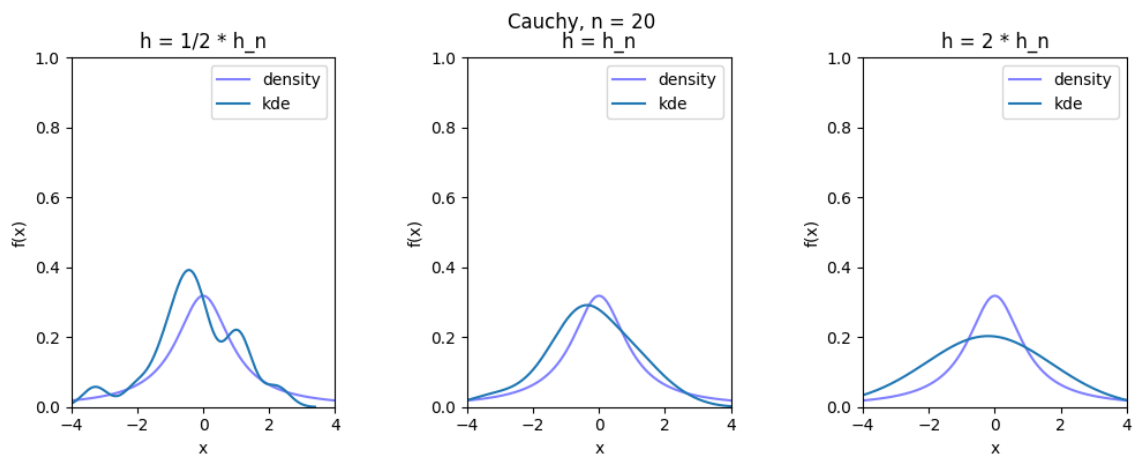


Рис. 19: Распределение Коши, $n = 20$

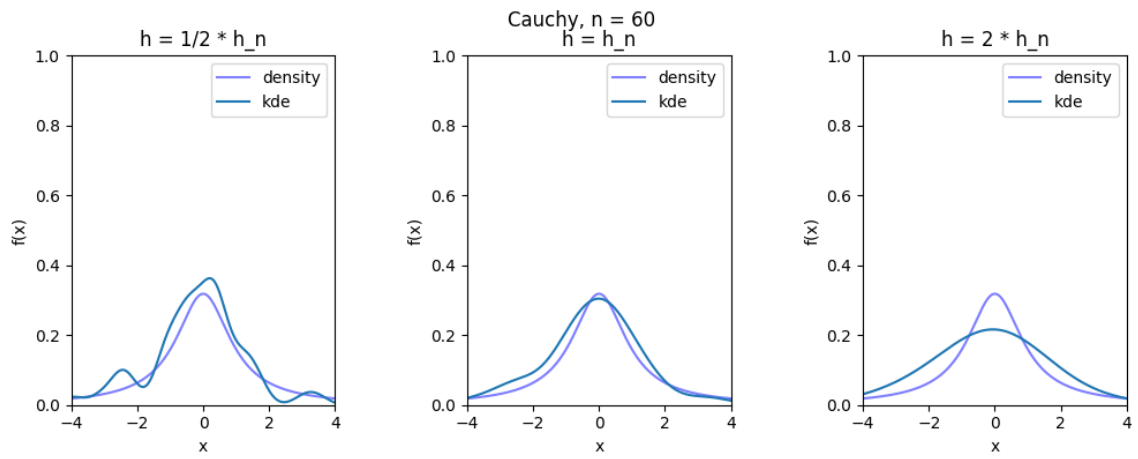


Рис. 20: Распределение Коши, $n = 60$

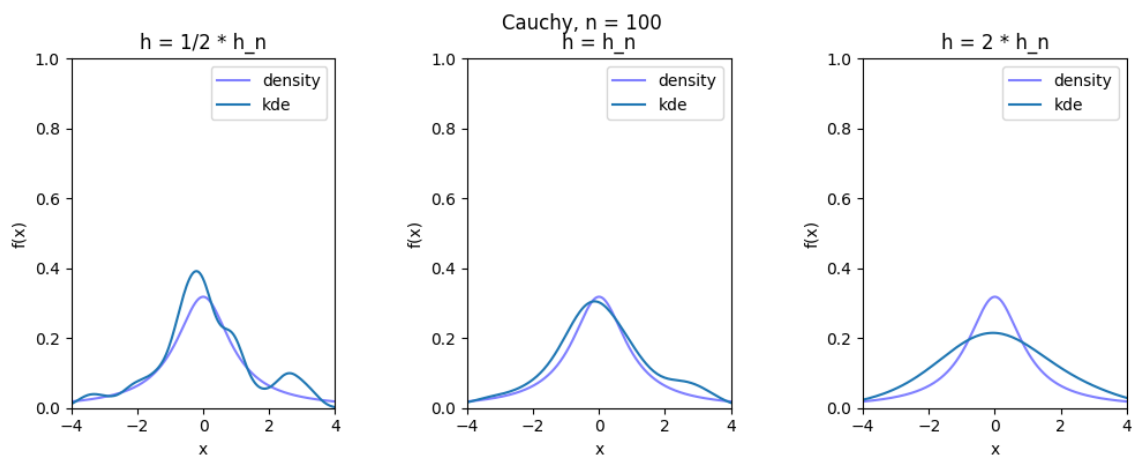


Рис. 21: Распределение Коши, $n = 100$

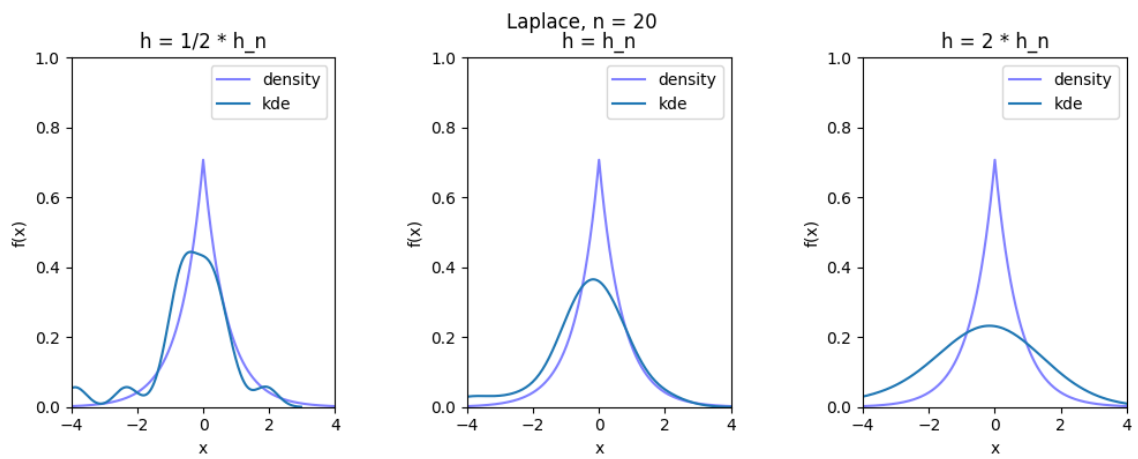


Рис. 22: Распределение Лапласа, $n = 20$

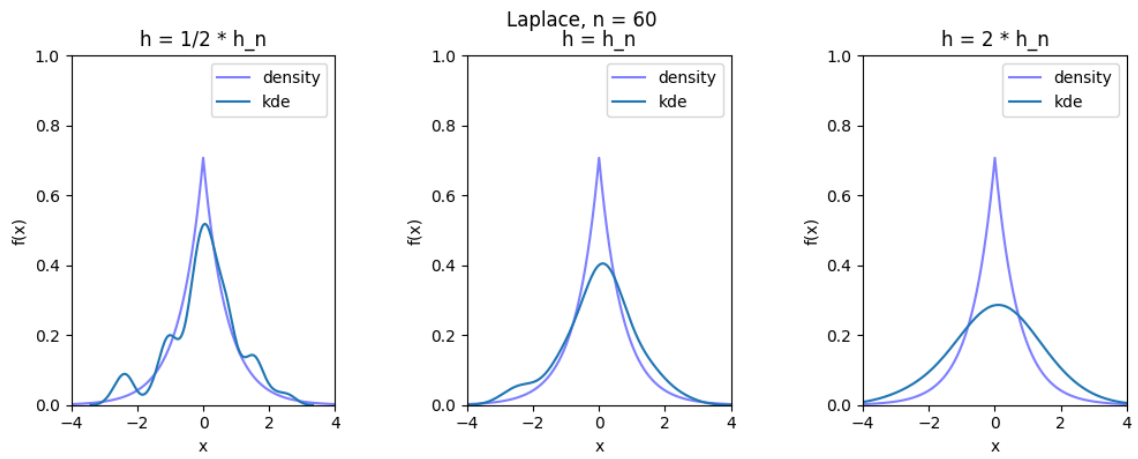


Рис. 23: Распределение Лапласа, $n = 60$

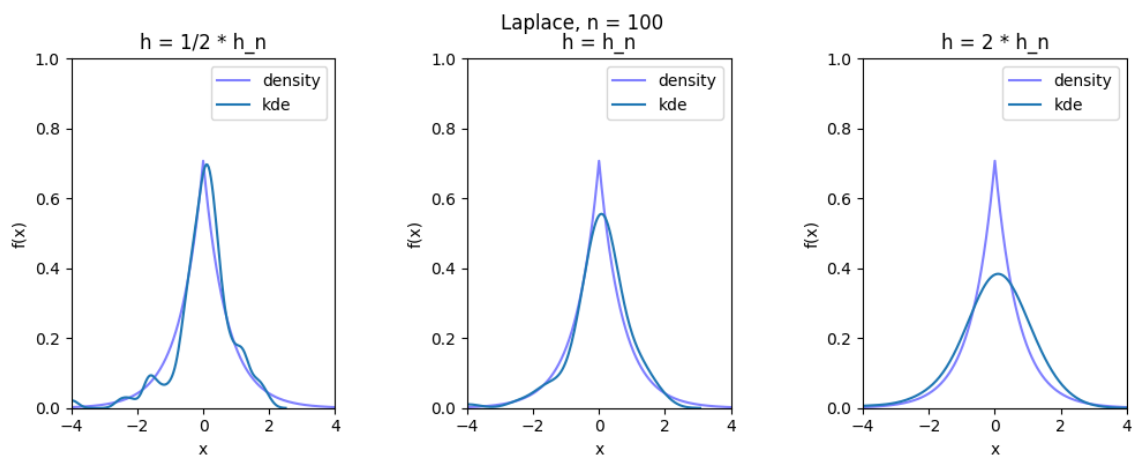


Рис. 24: Распределение Лапласа, $n = 100$

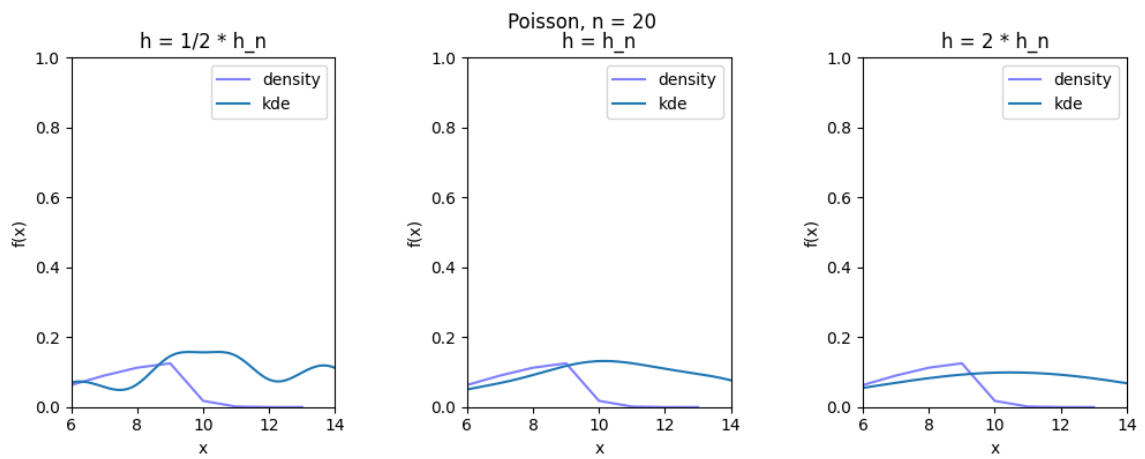


Рис. 25: Распределение Пуассона, $n = 20$

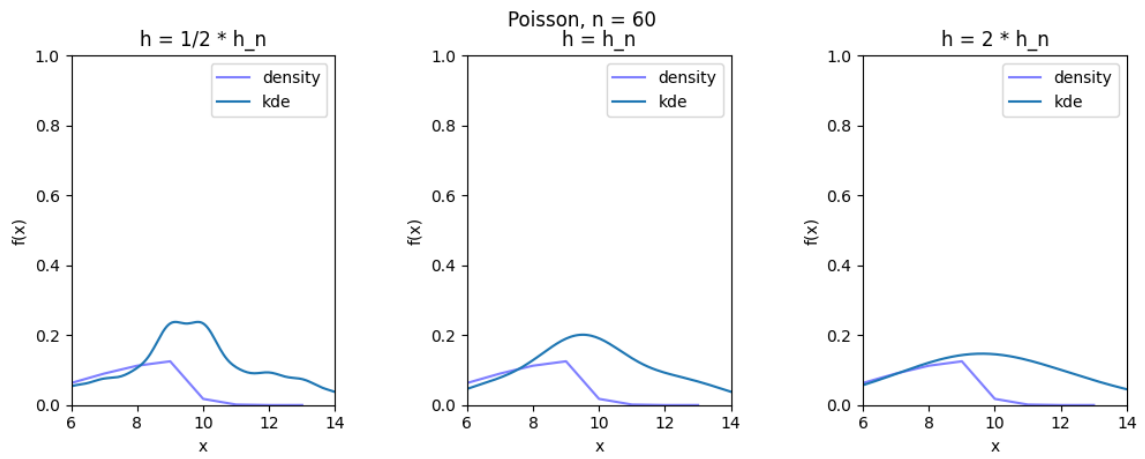


Рис. 26: Распределение Пуассона, $n = 60$

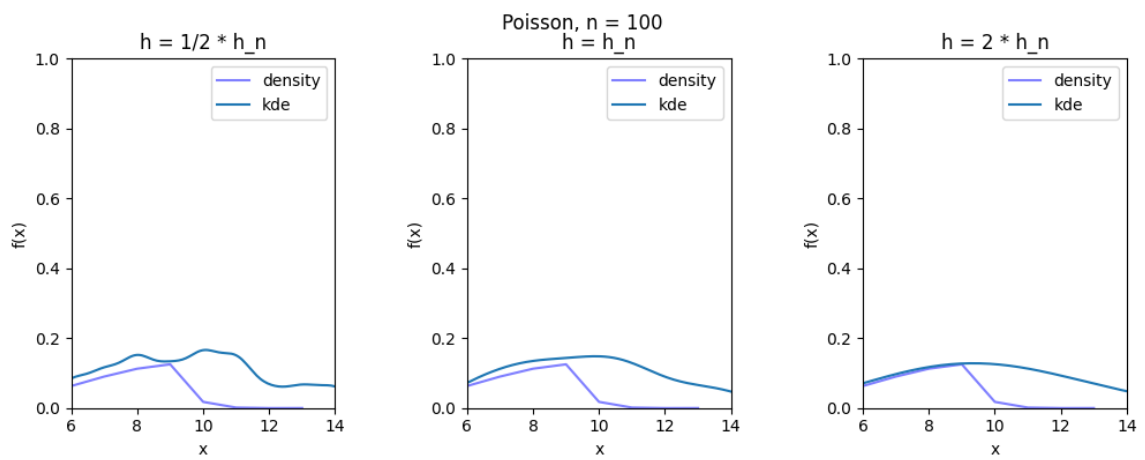


Рис. 27: Распределение Пуассона, $n = 100$

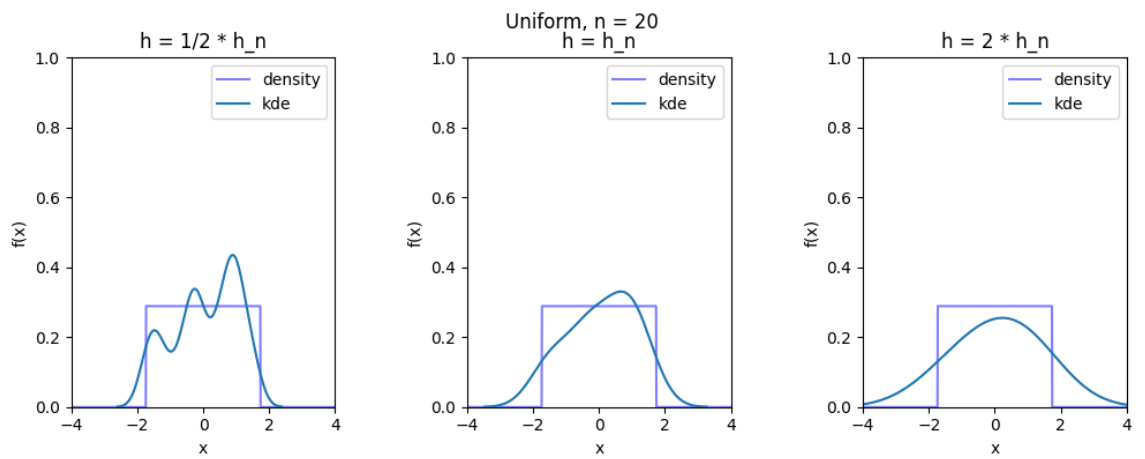


Рис. 28: Равномерное распределение, $n = 20$

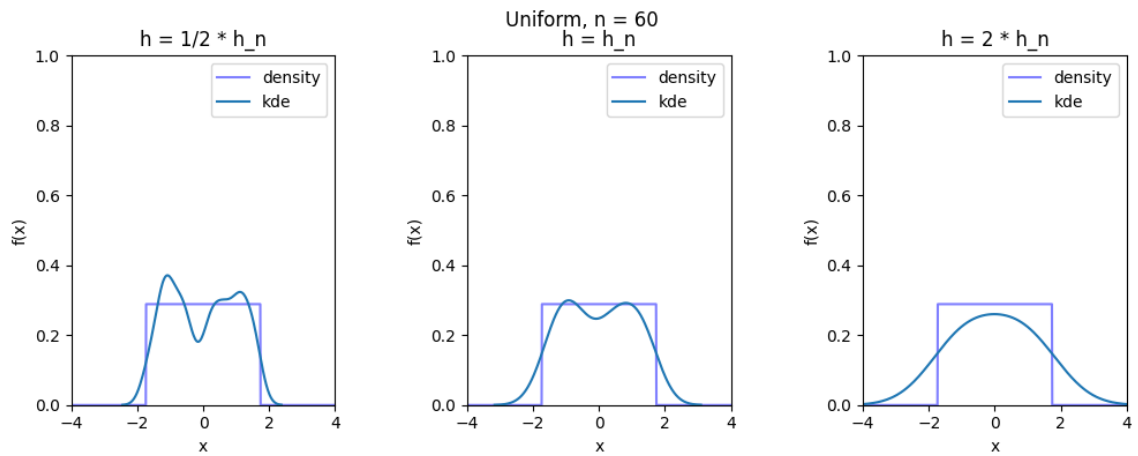


Рис. 29: Равномерное распределение, $n = 60$

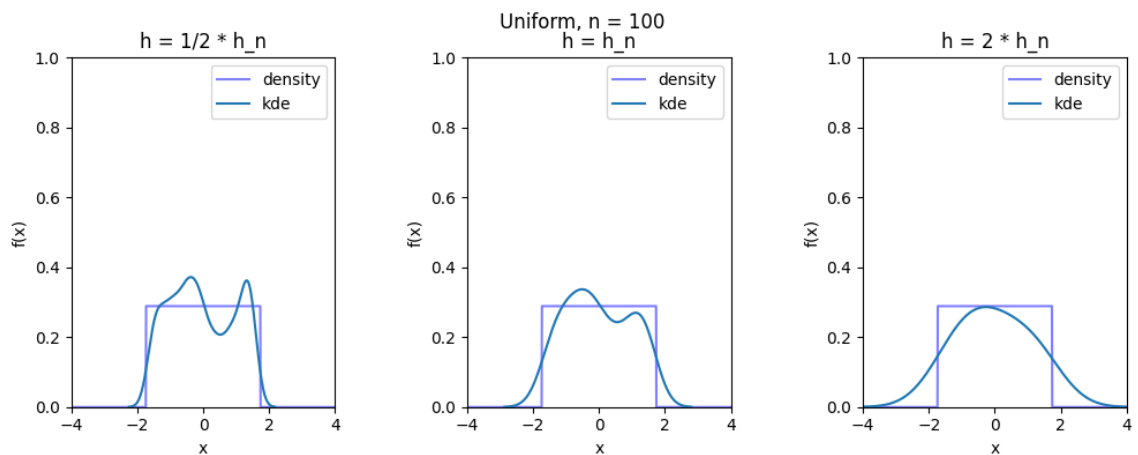


Рис. 30: Равномерное распределение, $n = 100$

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 при-

ближаются 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
- [4] Анатольев, Станислав (2009) «Непараметрическая регрессия», Квантиль, №7, стр. 37-52.