

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Прикладная математика»

Отчет
по лабораторным работам №1-2
по дисциплине
«Математическая статистика»

Выполнил студент: Колосков Александр

Группа: 3630102/80301

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

Санкт-Петербург

2021 г.

Содержание

	Страница
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 рассеивания	6
2.5 Ящик с «усами»	6
2.5.1 Построение	6
2.6 Теоретическая вероятность выбросов	6
2.7 Эмпирическая функция распределения	6
2.7.1 Статистический ряд	6
2.7.2 Эмпирическая функция распределения	6
2.7.3 Нахождение э. ф. р.	6
2.8 Оценки плотности вероятности	7
2.8.1 Определение	7
2.8.2 Ядерные оценки	7
3 Реализация	7
4 Результаты	8
4.1 Гистограммы и графики плотности распределения	8
4.2 Характеристики положения и рассеивания	13
4.3 Боксплот Тьюки	16
4.4 Доля выбросов	18
4.5 Теоретическая вероятность выбросов	18
4.6 Эмпирическая функция распределения	19
4.7 Ядерные оценки плотности распределения	21
4.7.1 Нормальное распределение	21
4.7.2 Распределение Коши	23
4.7.3 Распределение Лапласа	24
4.7.4 Распределение Пуассона	26
4.7.5 Равномерное распределение	27
5 Обсуждение	29
5.1 Гистограмма и график плотности распределения	29
5.2 Характеристики положения и рассеяния	29
5.3 Доля и теоретическая вероятность выбросов	29
5.4 Эмпирическая функция и ядерные оценки плотности распределения	29

Список иллюстраций

	Страница
1 Гистограмма. Нормальное распределение (3)	8
2 Гистограмма. Распределение Коши (4)	9
3 Гистограмма. Распределение Лапласа (5)	10
4 Гистограмма. Распределение Пуассона (6)	11
5 Гистограмма. Равномерное распределение (7)	12
6 Боксплот. Распределение Коши (4)	16
7 Боксплот. Распределение Лапласа (5)	16
8 Боксплот. Распределение Пуассона (6)	17
9 Боксплот. Равномерное распределение (7)	17
10 ECDF. Нормальное распределение (3)	19
11 ECDF. Распределение Коши (4)	19
12 ECDF. Распределение Лапласа (5)	20
13 ECDF. Распределение Пуассона (6)	20
14 ECDF. Равномерное распределение (7)	21
15 Ядерные оценки. Нормальное распределение (3), $h = 0.5h_n^S$	21
16 Ядерные оценки. Нормальное распределение (3), $h = h_n^S$	22
17 Ядерные оценки. Нормальное распределение (3), $h = 2h_n^S$	22
18 Ядерные оценки. Распределение Коши (4), $h = 0.5h_n^S$	23
19 Ядерные оценки. Распределение Коши (4), $h = h_n^S$	23
20 Ядерные оценки. Распределение Коши (4), $h = 2h_n^S$	24
21 Ядерные оценки. Распределение Лапласа (5), $h = 0.5h_n^S$	24
22 Ядерные оценки. Распределение Лапласа (5), $h = h_n^S$	25
23 Ядерные оценки. Распределение Лапласа (5), $h = 2h_n^S$	25
24 Ядерные оценки. Распределение Пуассона (6), $h = 0.5h_n^S$	26
25 Ядерные оценки. Распределение Пуассона (6), $h = h_n^S$	26
26 Ядерные оценки. Распределение Пуассона (6), $h = 2h_n^S$	27
27 Ядерные оценки. Равномерное распределение (7), $h = 0.5h_n^S$	27
28 Ядерные оценки. Равномерное распределение (7), $h = h_n^S$	28
29 Ядерные оценки. Равномерное распределение (7), $h = 2h_n^S$	28

Список таблиц

	Страница
1 Таблица распределения	6
2 Выборочные характеристики. Нормальное распределение (3)	13
3 Выборочные характеристики. Распределение Коши (4)	14
4 Выборочные характеристики. Распределение Лапласа (5)	14
5 Выборочные характеристики. Распределение Пуассона (6)	15
6 Выборочные характеристики. Равномерное распределение (7)	15
7 Доля выбросов	18
8 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} , $medx$, 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:

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

Вычислить порядковую оценку для средних $\hat{E}(z)$ на интервале $(E(z) - \sqrt{D(z)}, E(z) + \sqrt{D(z)})$. 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

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

- 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

2.3 Вариационный ряд

Вариационным ряд - последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются.

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

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

- Полусумма экстремальных выборочных элементов

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

- Полусумма квартилей

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

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

- Усечённое среднее

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

2.4.2 Характеристики рассеивания

Выборочная дисперсия

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

2.5 Ящик с «усами»

2.5.1 Построение

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.

Выбросы отражены на графике в виде маленьких кружков за границами «усов».

2.6 Теоретическая вероятность выбросов

- для непрерывных распределений:

$$P_{\text{в}}^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)) \quad (16)$$

- для дискретных распределений:

$$P_{\text{в}}^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)) \quad (17)$$

2.7 Эмпирическая функция распределения

2.7.1 Статистический ряд

Статистическим рядом назовем совокупность, состоящую из последовательности $\{z_i\}_{i=1}^k$ попарно различных элементов выборки, расположенных по возрастанию, и последовательности $\{n_i\}_{i=1}^k$ частот, с которыми эти элементы содержатся в выборке.

2.7.2 Эмпирическая функция распределения

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

$$F_n^*(x) = P^*(X < x). \quad (18)$$

2.7.3 Нахождение э. ф. р.

$$F^*(x) = \frac{1}{n} \sum_{z_i < x} n_i. \quad (19)$$

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

X^*	z_1	z_2	\dots	z_k
P	n_1/n	n_2/n	\dots	n_k/n

Таблица 1: Таблица распределения

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

$$F_n^*(x) \approx F_X(x). \quad (20)$$

2.8 Оценки плотности вероятности

2.8.1 Определение

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

$$\hat{f}(x) \approx f(x). \quad (21)$$

2.8.2 Ядерные оценки

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

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right). \quad (22)$$

$K(u)$ - ядро, т. е. непрерывная функция, являющаяся плотностью вероятности, x_1, \dots, x_n - элементы выборки, а $\{h_n\}_{n \in \mathbb{N}}$ - последовательность элементов из \mathbb{R}_+ такая, что

$$h_n \xrightarrow{n \rightarrow \infty} 0; \quad nh_n \xrightarrow{n \rightarrow \infty} \infty. \quad (23)$$

Такие оценки называются непрерывными ядерными.

Гауссово ядро:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}. \quad (24)$$

Правило Сильвермана:

$$h_n^S = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5} \approx 1.06\hat{\sigma}n^{-1/5}, \quad (25)$$

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

3 Реализация

Лабораторная работа выполнена на языке Python в среде PyCharm с использованием библиотек numpy, scipy.stats, matplotlib.pyplot, statsmodels

4 Результаты

4.1 Гистограммы и графики плотности распределения

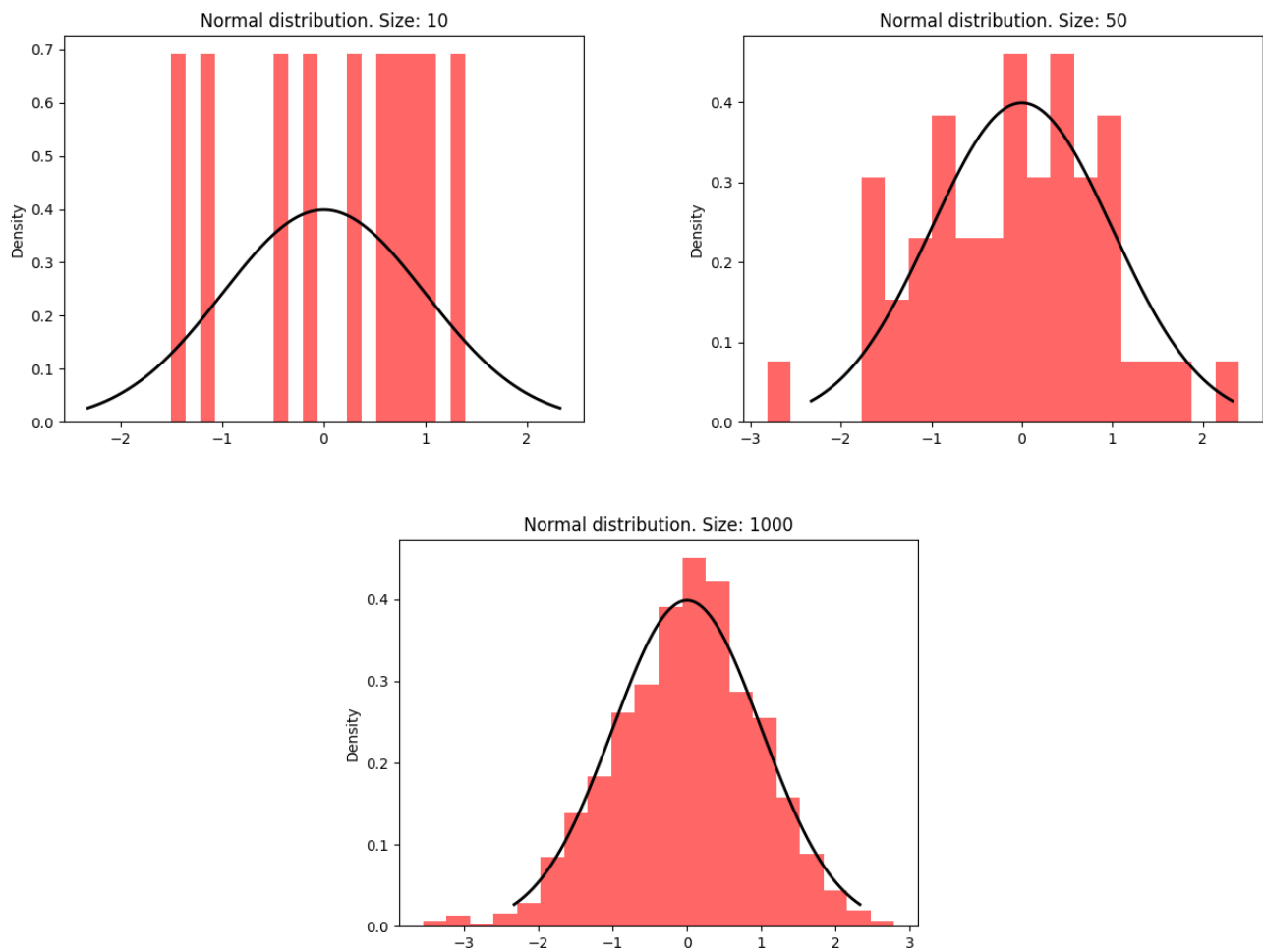


Рис. 1: Гистограмма. Нормальное распределение (3)

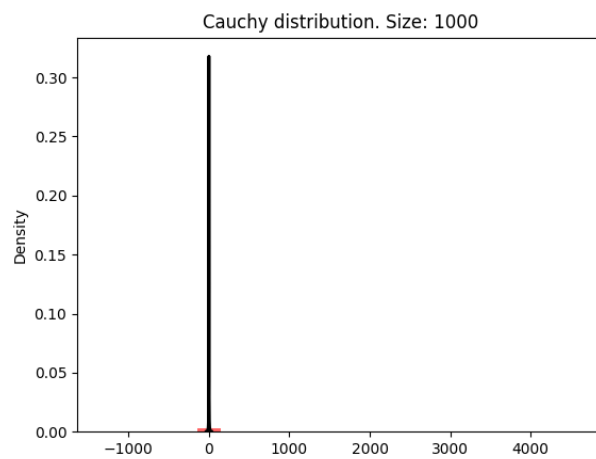
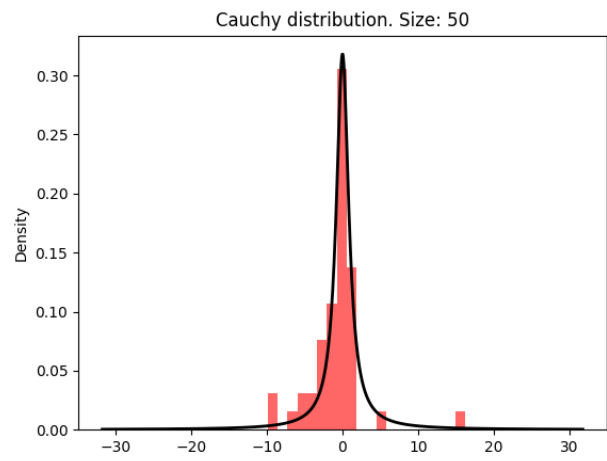
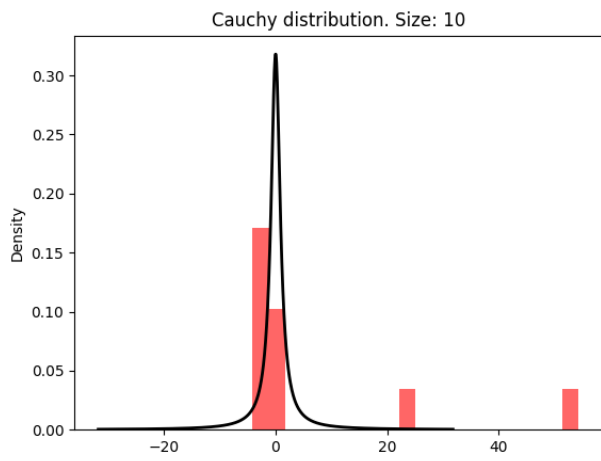


Рис. 2: Гистограмма. Распределение Коши (4)

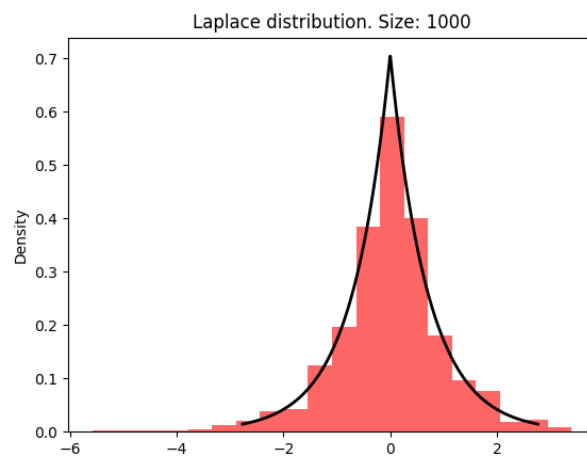
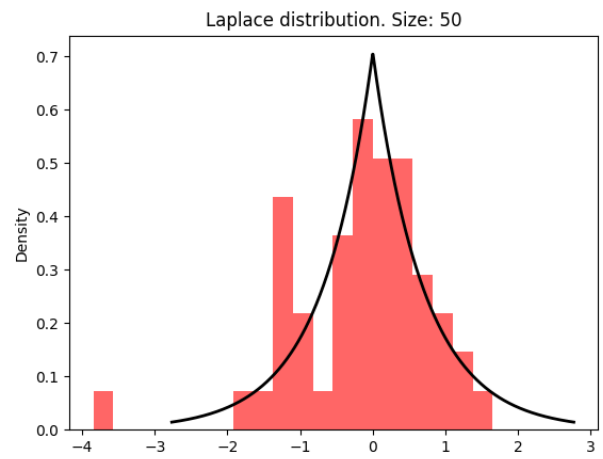
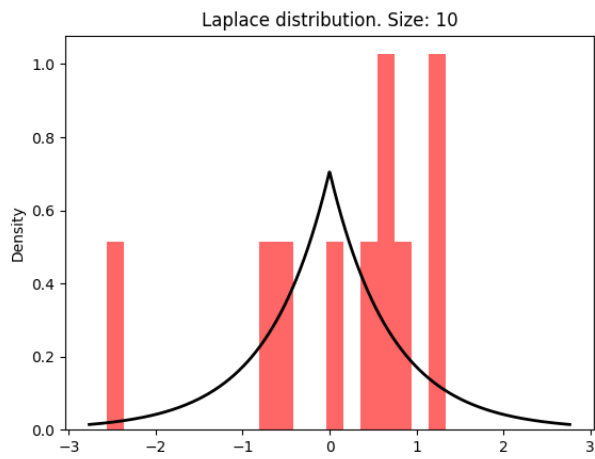


Рис. 3: Гистограмма. Распределение Лапласа (5)

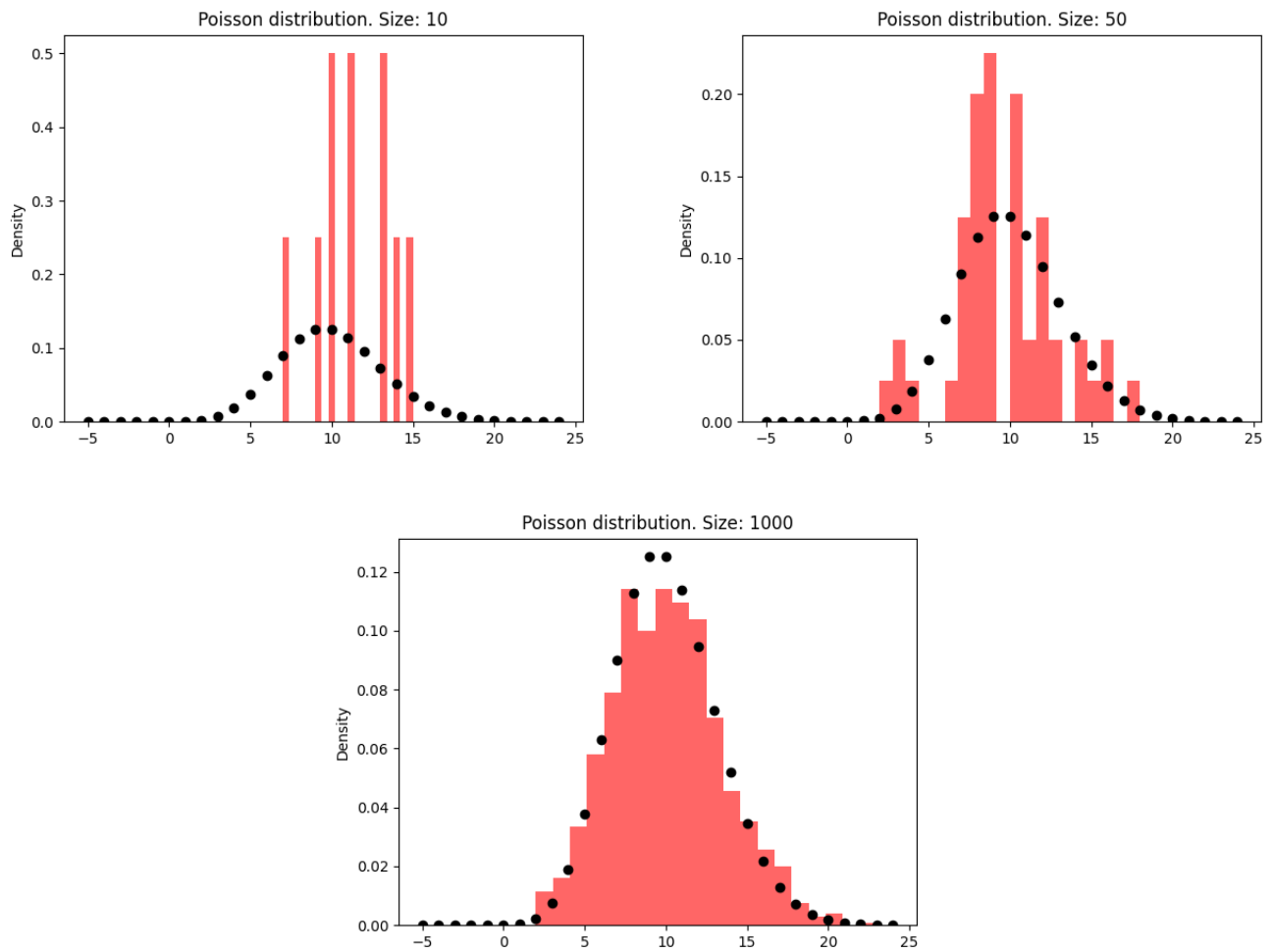


Рис. 4: Гистограмма. Распределение Пуассона (6)

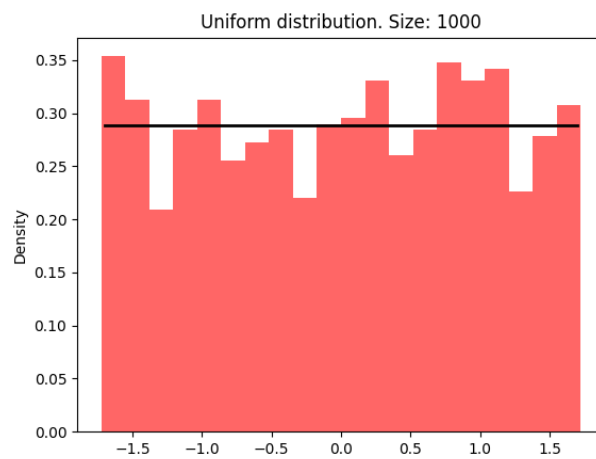
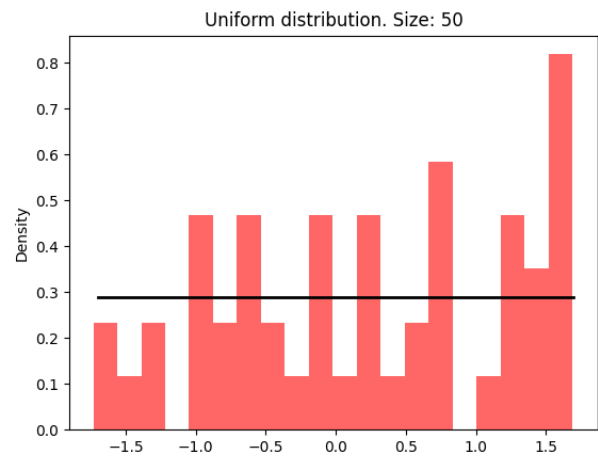
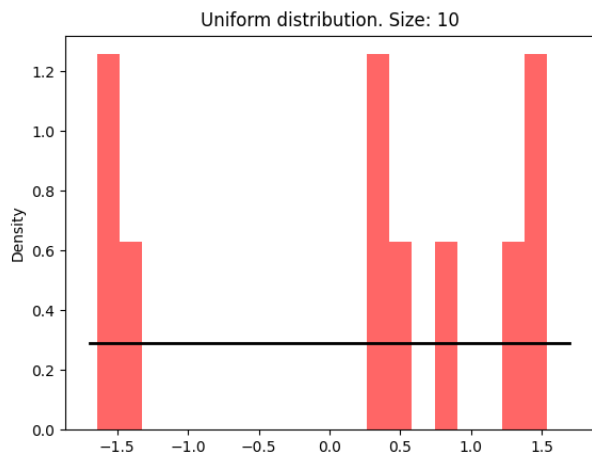


Рис. 5: Гистограмма. Равномерное распределение (7)

4.2 Характеристики положения и рассеивания

Normal $n=10$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00959	-0.01475	-0.01115	0.30273	0.26314
$D(z)$	0.09256	0.13242	0.17131	0.11861	0.10969
$E(z) - \sqrt{D(z)}$	-0.31382	-0.37864	-0.42504	-0.04167	-0.06805
$E(z) + \sqrt{D(z)}$	0.29464	0.34914	0.40274	0.64712	0.59433
$\hat{E}(z)$	0.	0.	0.	0.	0.
Normal $n=100$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00016	0.00253	-0.00293	0.01528	0.0274
$D(z)$	0.00994	0.01476	0.09291	0.01253	0.01146
$E(z) - \sqrt{D(z)}$	-0.09987	-0.11898	-0.30774	-0.09664	-0.07964
$E(z) + \sqrt{D(z)}$	0.09955	0.12403	0.30188	0.12721	0.13444
$\hat{E}(z)$	0.0	0.	0.	0.	0.
Normal $n=1000$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	0.00089	0.00022	0.00681	0.00256	0.00338
$D(z)$	0.00094	0.00157	0.06392	0.00117	0.00111
$E(z) - \sqrt{D(z)}$	-0.02985	-0.03939	-0.24601	-0.03161	-0.02998
$E(z) + \sqrt{D(z)}$	0.03162	0.03983	0.25962	0.03673	0.03675
$\hat{E}(z)$	0.0	0.0	0.	0.0	0.0

Таблица 2: Выборочные характеристики. Нормальное распределение (3)

Cauchy $n=10$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	3.1961	-0.02349	16.10151	1.13243	0.66878
$D(z)$	3331.29003	0.31953	83018.6462	6.69709	1.48047
$E(z) - \sqrt{D(z)}$	-54.52123	-0.58875	-272.02805	-1.45544	-0.54797
$E(z) + \sqrt{D(z)}$	60.91343	0.54178	304.23108	3.7203	1.88553
$\hat{E}(z)$	—	0.	—	0.	0.
Cauchy $n=100$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	10.40686	0.01087	515.42425	0.05097	0.05566
$D(z)$	62846.66993	0.02341	157101747.21806	0.05404	0.0256
$E(z) - \sqrt{D(z)}$	-240.28552	-0.14213	-12018.59933	-0.1815	-0.10434
$E(z) + \sqrt{D(z)}$	261.09924	0.16388	13049.44783	0.28343	0.21566
$\hat{E}(z)$	—	0.	—	0.	0.
Cauchy $n=1000$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	4.22379	0.00052	2115.88208	0.00372	0.00493
$D(z)$	19584.49654	0.00268	4894410646.01298	0.00465	0.00256
$E(z) - \sqrt{D(z)}$	-135.72083	-0.05125	-67844.18257	-0.06448	-0.04565
$E(z) + \sqrt{D(z)}$	144.16841	0.0523	72075.94673	0.07191	0.05551
$\hat{E}(z)$	—	0.	—	0.	0.

Таблица 3: Выборочные характеристики. Распределение Коши (4)

Laplace $n=10$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.01398	-0.0048	-0.01905	0.28314	0.22082
$D(z)$	0.10128	0.0727	0.4051	0.11619	0.07966
$E(z) - \sqrt{D(z)}$	-0.33222	-0.27443	-0.65553	-0.05772	-0.06142
$E(z) + \sqrt{D(z)}$	0.30426	0.26483	0.61742	0.62401	0.50307
$\hat{E}(z)$	0.	0.	0.	0.	0.
Laplace $n=100$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	0.00251	0.00138	0.02867	0.0138	0.02029
$D(z)$	0.01002	0.00558	0.3784	0.01004	0.00623
$E(z) - \sqrt{D(z)}$	-0.09759	-0.07334	-0.58647	-0.0864	-0.05867
$E(z) + \sqrt{D(z)}$	0.10261	0.07609	0.64382	0.114	0.09925
$\hat{E}(z)$	0.	0.0	0.	0.	0.0
Laplace $n=1000$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00198	-0.00097	-0.05338	0.00139	0.00085
$D(z)$	0.00098	0.00055	0.42766	0.00108	0.00064
$E(z) - \sqrt{D(z)}$	-0.03322	-0.0244	-0.70734	-0.03142	-0.02452
$E(z) + \sqrt{D(z)}$	0.02926	0.02245	0.60058	0.0342	0.02621
$\hat{E}(z)$	0.0	0.0	0.	0.0	0.0

Таблица 4: Выборочные характеристики. Распределение Лапласа (5)

Poisson $n=10$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	9.98641	9.83217	10.2997	10.91608	10.75907
$D(z)$	0.98827	1.33247	1.99809	1.39655	1.19212
$E(z) - \sqrt{D(z)}$	8.9923	8.67784	8.88616	9.73432	9.66723
$E(z) + \sqrt{D(z)}$	10.98053	10.9865	11.71324	12.09784	11.85092
$\hat{E}(z)$	10_{-2}^{+1}	10_{-2}^{+1}	10_{-2}^{+2}	10_{-1}^{+2}	10_{-1}^{+1}
Poisson $n=100$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	10.02141	9.88312	10.86863	9.98901	9.96923
$D(z)$	0.1048	0.20512	0.96301	0.16372	0.12244
$E(z) - \sqrt{D(z)}$	9.69768	9.43022	9.8873	9.58439	9.61931
$E(z) + \sqrt{D(z)}$	10.34514	10.33602	11.84996	10.39363	10.31915
$\hat{E}(z)$	10_{-1}	10_{-1}	10_{-1}^{+1}	10_{-1}	10_{-1}
Poisson $n=1000$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	10.00071	9.995	11.67333	9.99401	9.86692
$D(z)$	0.00989	0.00497	0.71396	0.00296	0.0112
$E(z) - \sqrt{D(z)}$	9.90124	9.92451	10.82836	9.93959	9.76109
$E(z) + \sqrt{D(z)}$	10.10019	10.0655	12.51829	10.04842	9.97275
$\hat{E}(z)$	10_{-1}	10_{-1}	10^{+2}	10_{-1}	9

Таблица 5: Выборочные характеристики. Распределение Пуассона (6)

Uniform $n=10$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	-0.00509	-0.01086	0.00291	0.3088	0.30699
$D(z)$	0.09632	0.22101	0.04288	0.12282	0.14946
$E(z) - \sqrt{D(z)}$	-0.31544	-0.48098	-0.20417	-0.04165	-0.07961
$E(z) + \sqrt{D(z)}$	0.30526	0.45926	0.21	0.65926	0.69359
$\hat{E}(z)$	0.	0.	0.	0.	0.
Uniform $n=100$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	0.001	0.00382	-2e-05	0.01758	0.03607
$D(z)$	0.01011	0.02962	0.00052	0.01475	0.02017
$E(z) - \sqrt{D(z)}$	-0.09952	-0.1683	-0.02287	-0.10389	-0.10595
$E(z) + \sqrt{D(z)}$	0.10153	0.17593	0.02282	0.13905	0.17808
$\hat{E}(z)$	0.	0.	0.0	0.	0.
Uniform $n=1000$					
	\bar{x}	$medx$	z_R	z_Q	z_{tr}
$E(z)$	0.00056	0.00166	-9e-05	0.00168	0.00475
$D(z)$	0.00105	0.00305	1e-05	0.00163	0.00205
$E(z) - \sqrt{D(z)}$	-0.03178	-0.05353	-0.00245	-0.03874	-0.04053
$E(z) + \sqrt{D(z)}$	0.0329	0.05684	0.00228	0.0421	0.05003
$\hat{E}(z)$	0.0	0.0	0.00	0.0	0.0

Таблица 6: Выборочные характеристики. Равномерное распределение (7)

4.3 Боксплот Тьюки

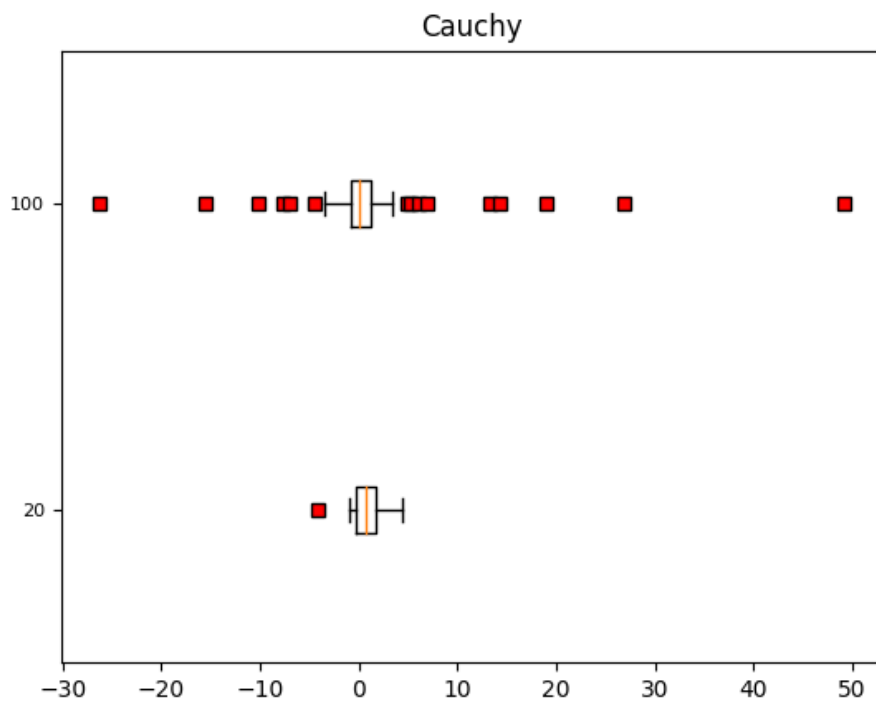


Рис. 6: Боксплот. Распределение Коши (4)

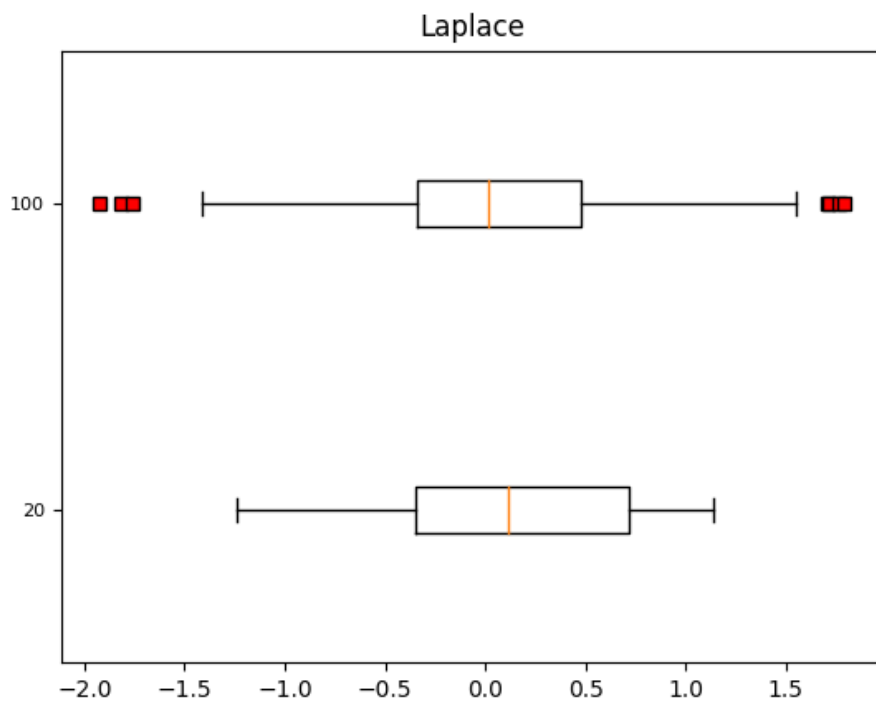


Рис. 7: Боксплот. Распределение Лапласа (5)

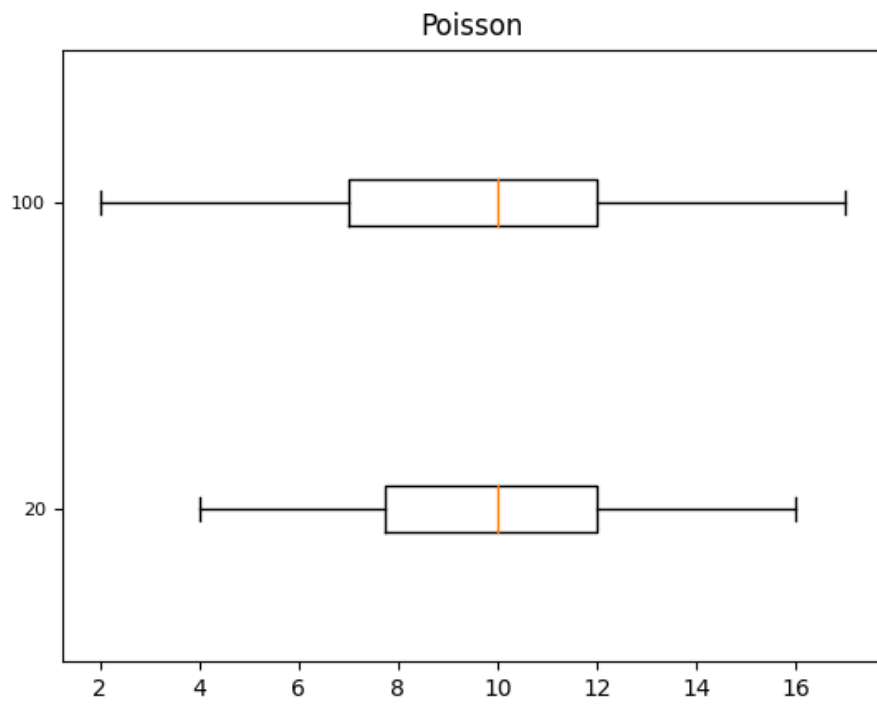


Рис. 8: Боксплот. Распределение Пуассона (6)

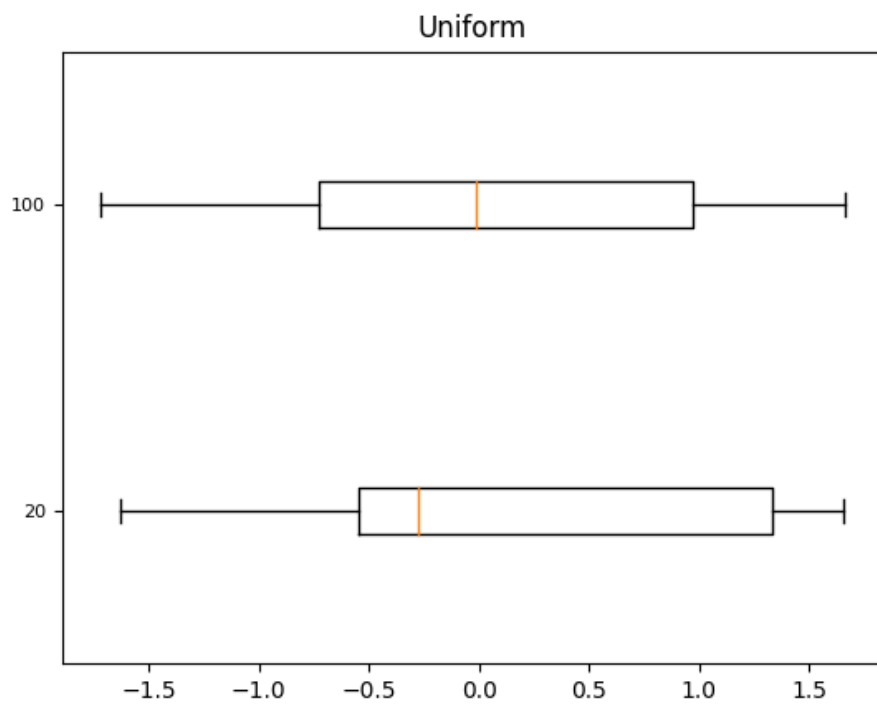


Рис. 9: Боксплот. Равномерное распределение (7)

4.4 Доля выбросов

Sample	Outliers
Normal $n = 20$	0.02
Normal $n = 100$	0.01
Laplace $n = 20$	0.067
Laplace $n = 100$	0.063
Poisson $n = 20$	0.015
Poisson $n = 100$	0.009
Cauchy $n = 20$	0.141
Cauchy $n = 100$	0.154
Uniform $n = 20$	0.002
Uniform $n = 100$	0.0

Таблица 7: Доля выбросов

4.5 Теоретическая вероятность выбросов

Distribution	Q_1^T	Q_3^T	X_1^T	X_2^T	P^T
Normal	-0.674	0.674	-2.698	2.698	0.007
Laplace	-0.49	0.49	-1.961	1.961	0.062
Poisson	8.0	12.0	2.0	18.0	0.01
Cauchy	-1.0	1.0	-4.0	4.0	0.156
Uniform	-0.866	0.866	-3.464	3.464	0.0

Таблица 8: Теоретическая вероятность выбросов

4.6 Эмпирическая функция распределения

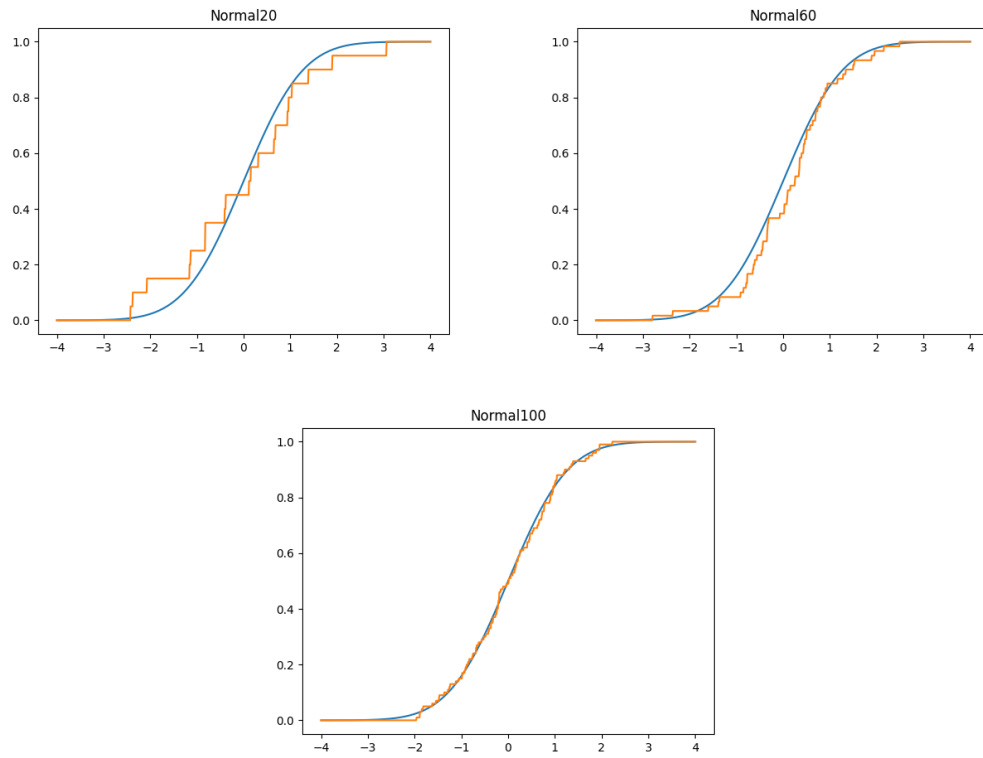


Рис. 10: ECDF. Нормальное распределение (3)

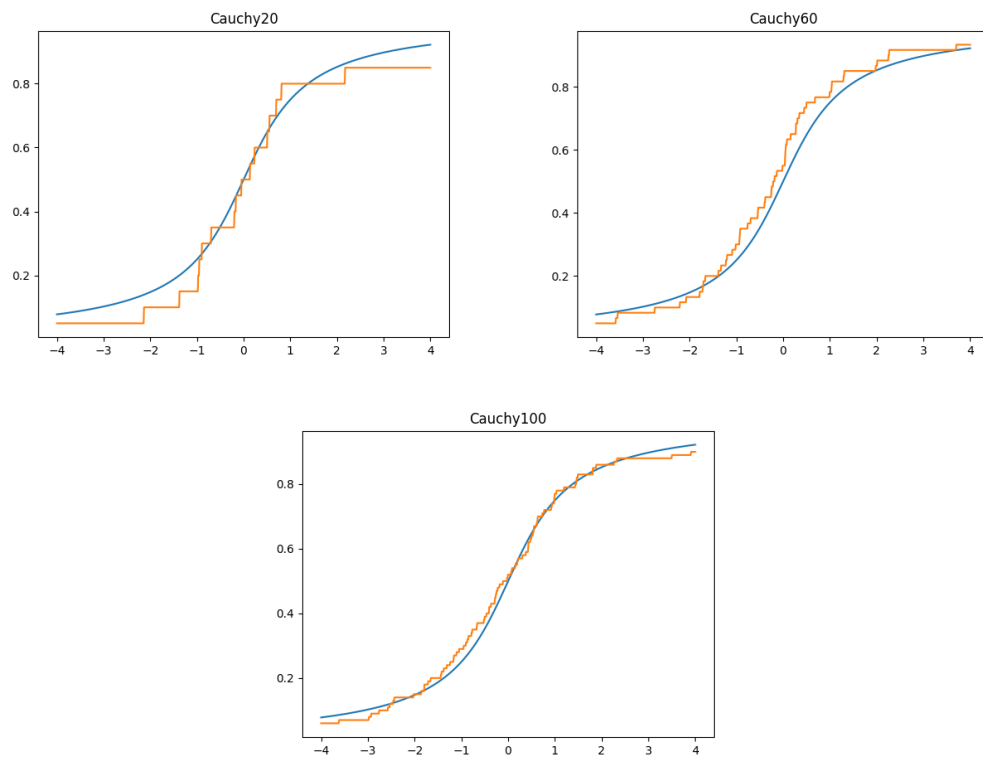


Рис. 11: ECDF. Распределение Коши (4)

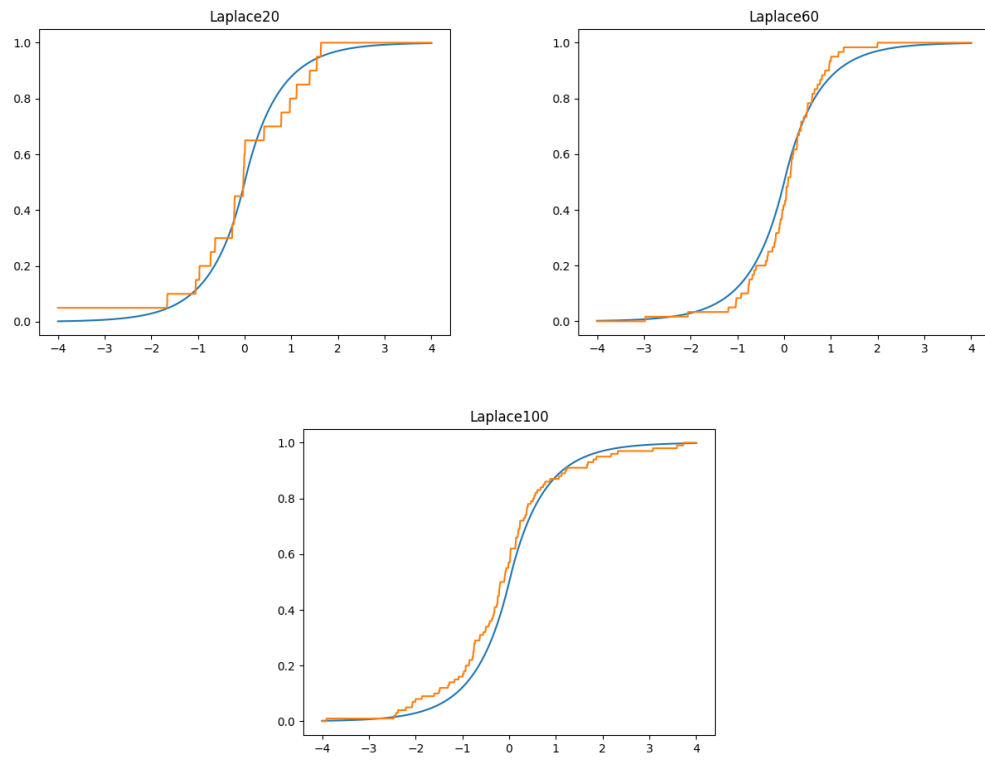


Рис. 12: ECDF. Распределение Лапласа (5)

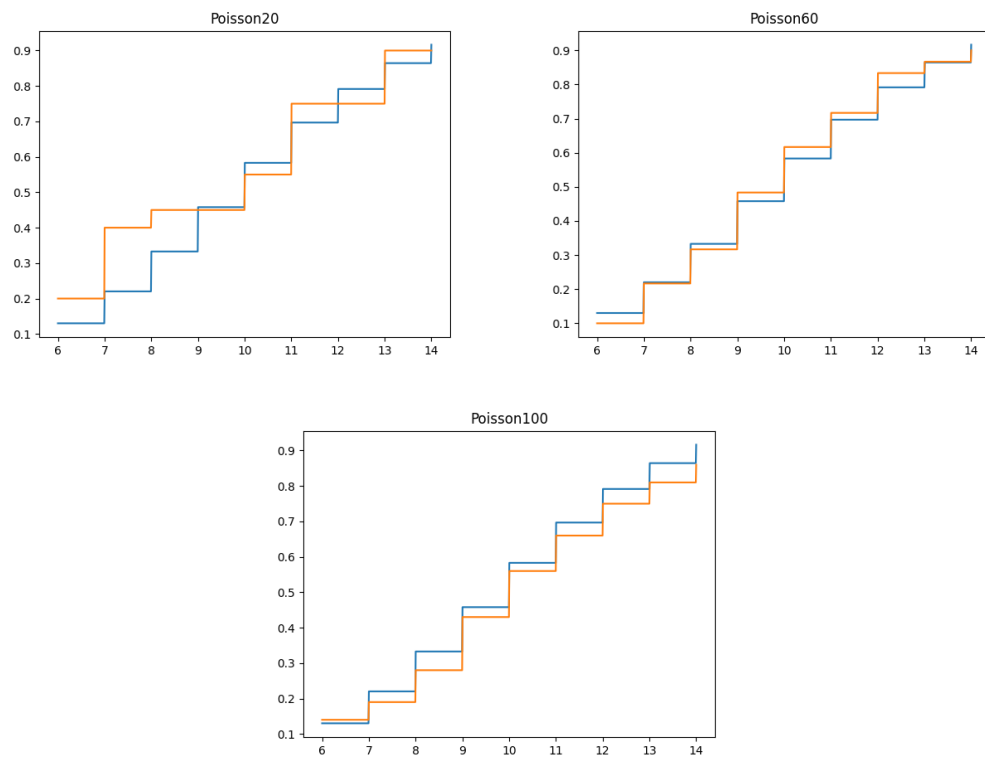


Рис. 13: ECDF. Распределение Пуассона (6)

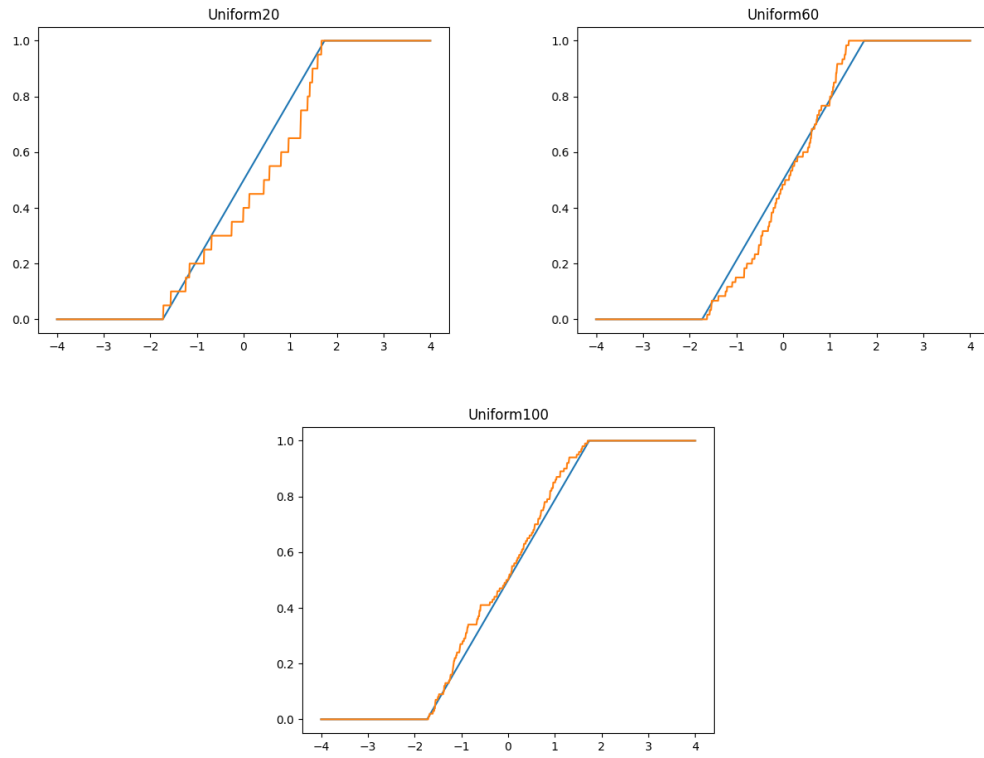


Рис. 14: ECDF. Равномерное распределение (7)

4.7 Ядерные оценки плотности распределения

4.7.1 Нормальное распределение

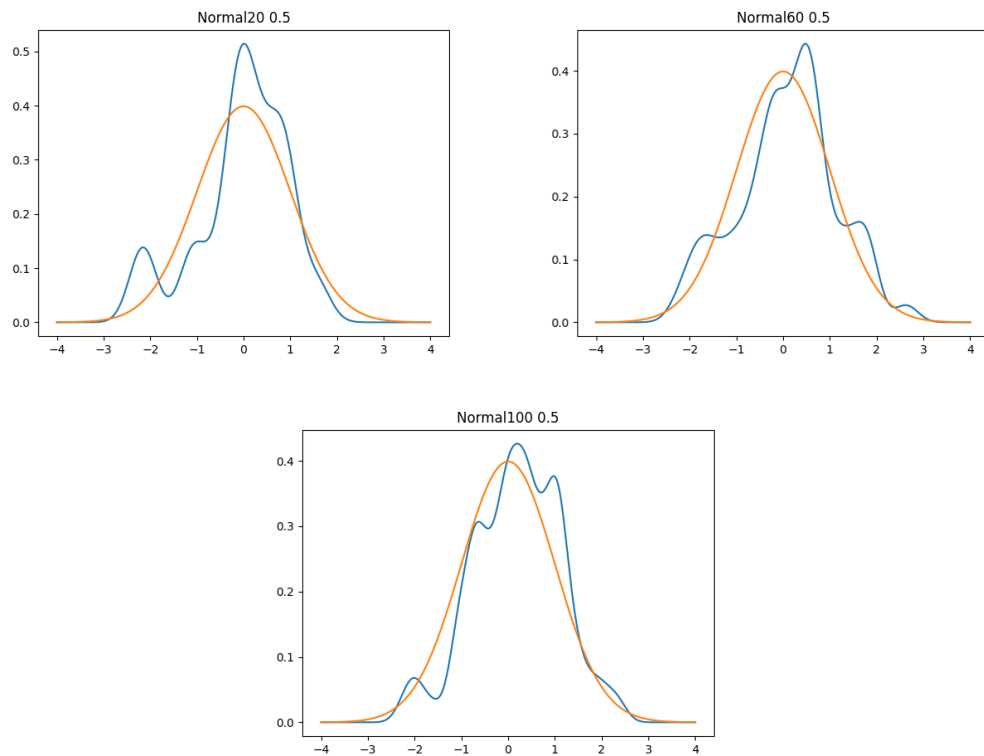


Рис. 15: Ядерные оценки. Нормальное распределение (3), $h = 0.5h_n^S$

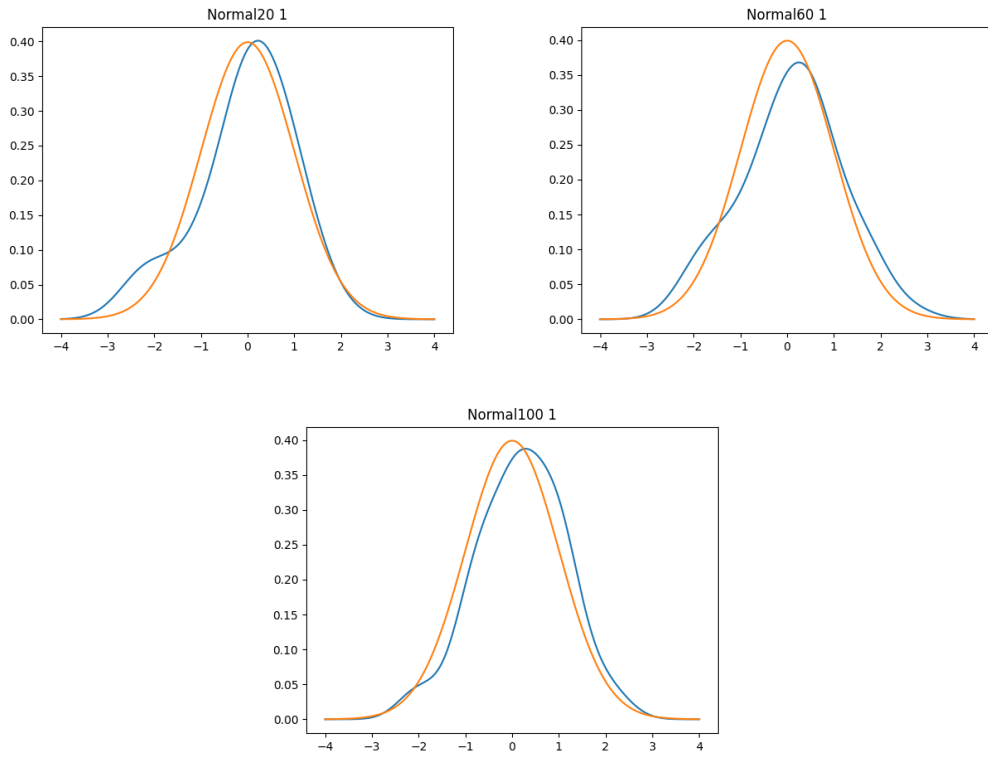


Рис. 16: Ядерные оценки. Нормальное распределение (3), $h = h_n^S$

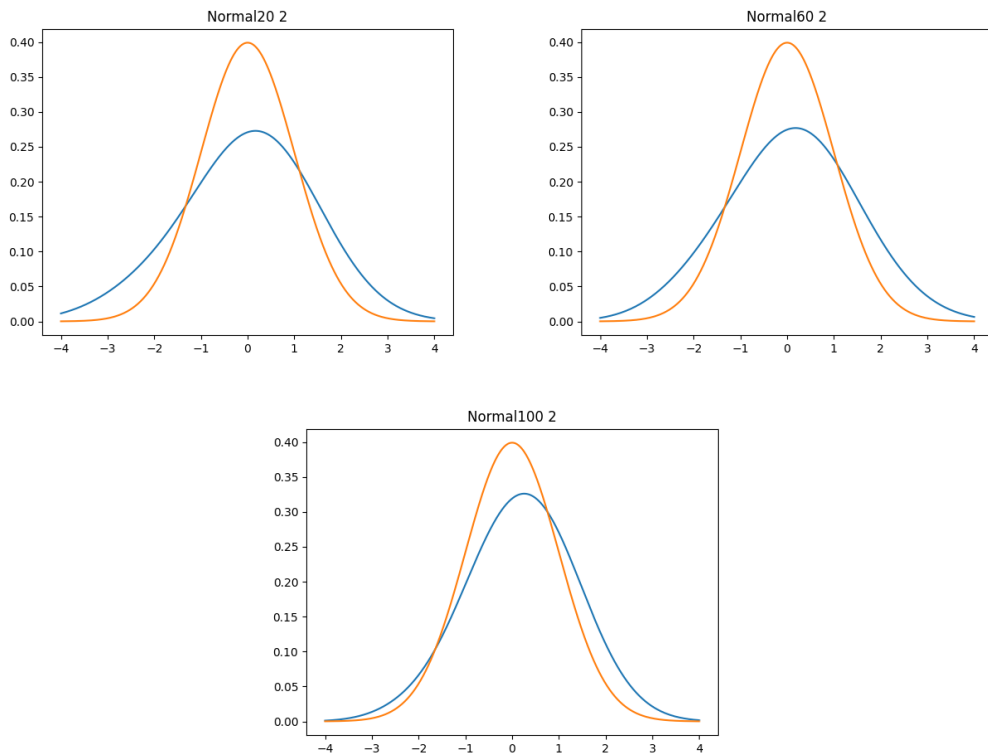


Рис. 17: Ядерные оценки. Нормальное распределение (3), $h = 2h_n^S$

4.7.2 Распределение Коши

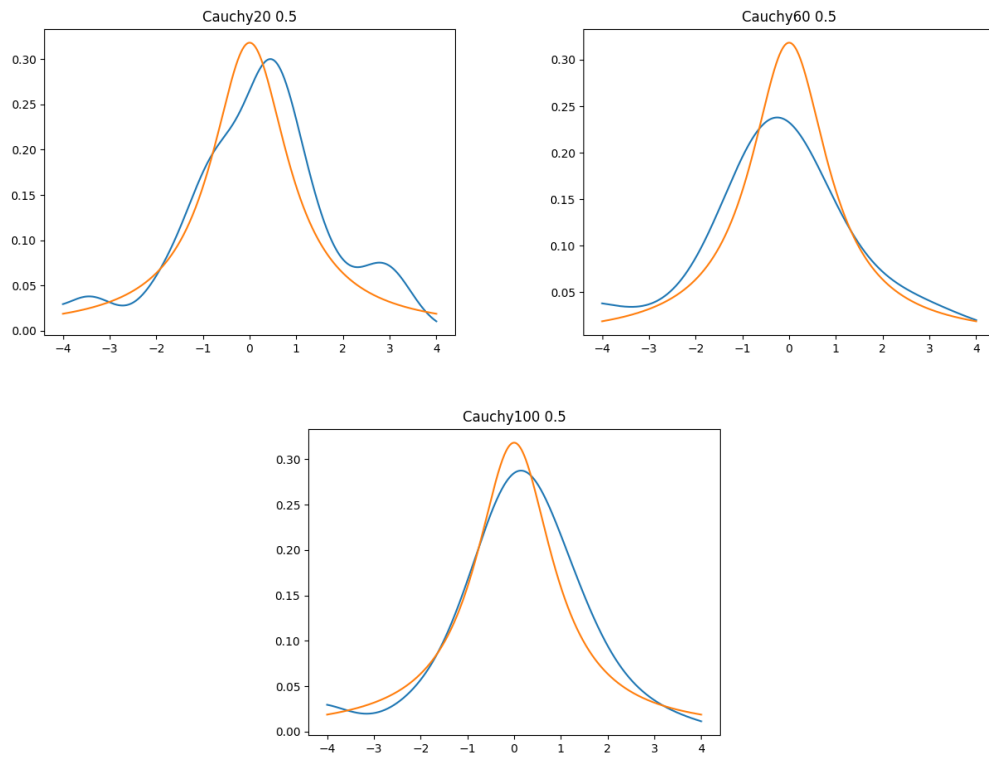


Рис. 18: Ядерные оценки. Распределение Коши (4), $h = 0.5h_n^S$

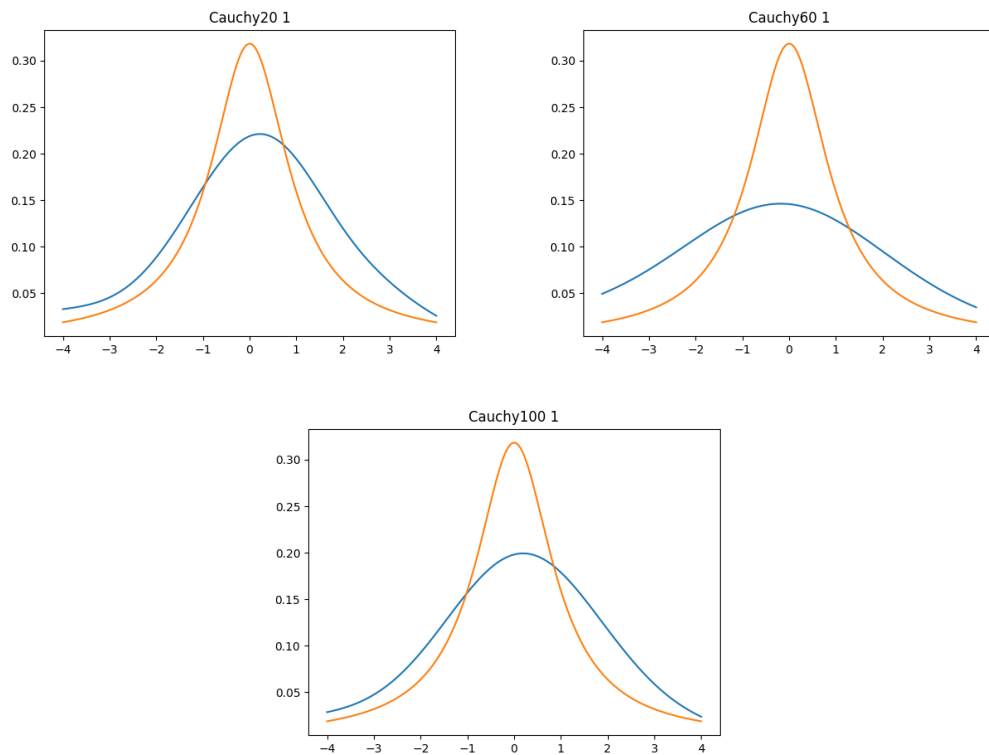


Рис. 19: Ядерные оценки. Распределение Коши (4), $h = h_n^S$

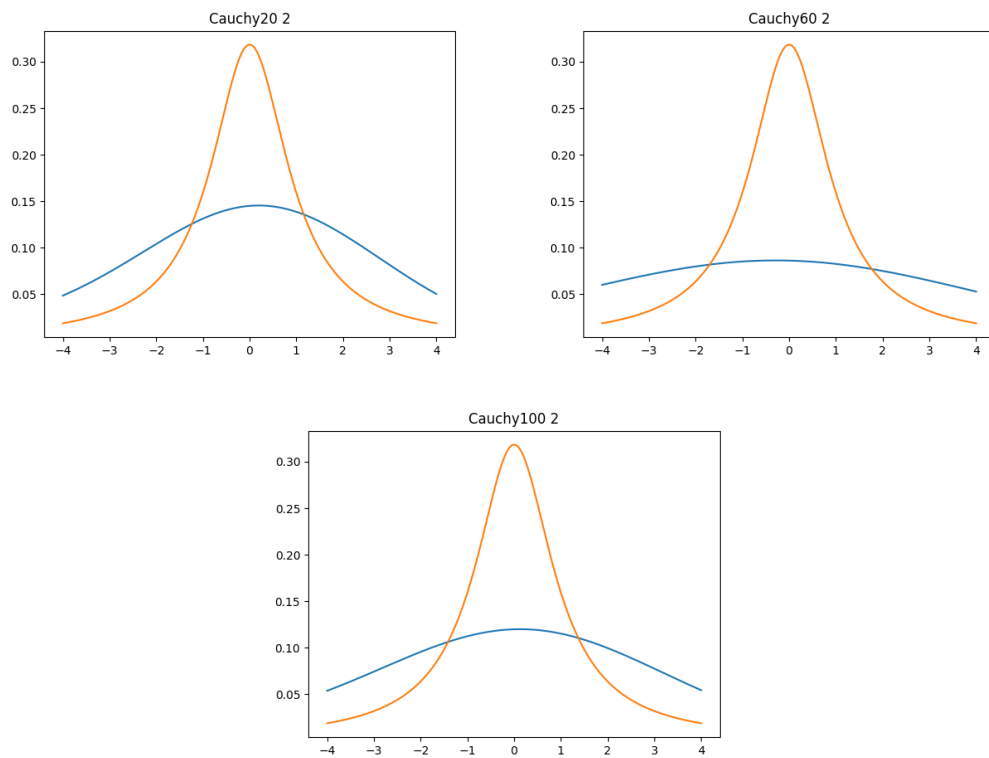


Рис. 20: Ядерные оценки. Распределение Коши (4), $h = 2h_n^S$

4.7.3 Распределение Лапласа

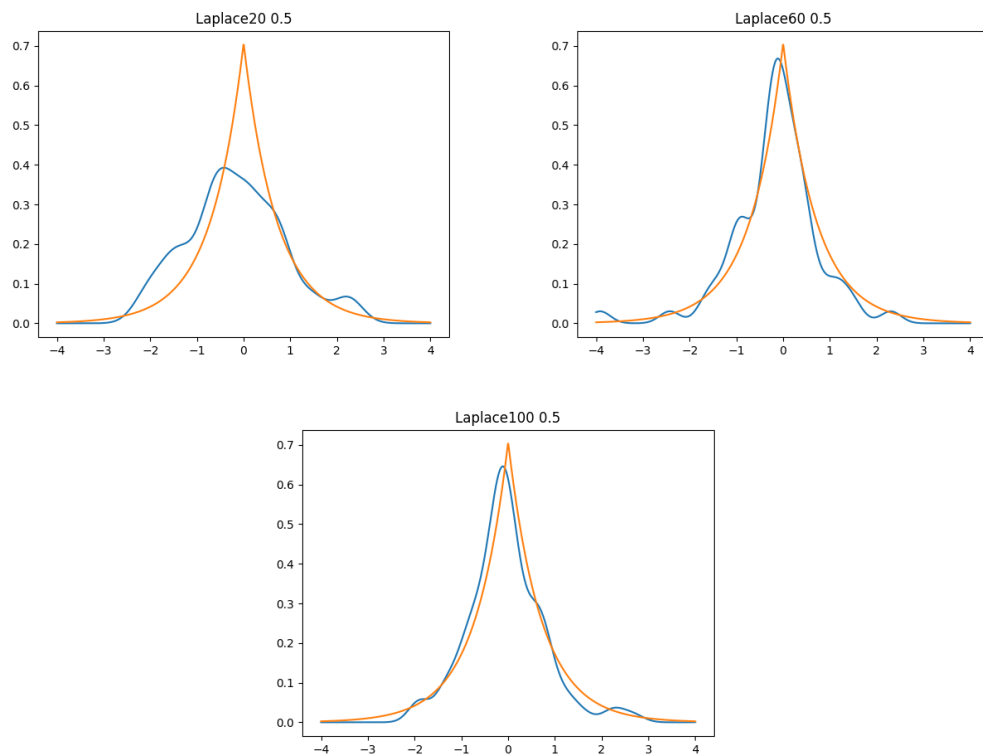


Рис. 21: Ядерные оценки. Распределение Лапласа (5), $h = 0.5h_n^S$

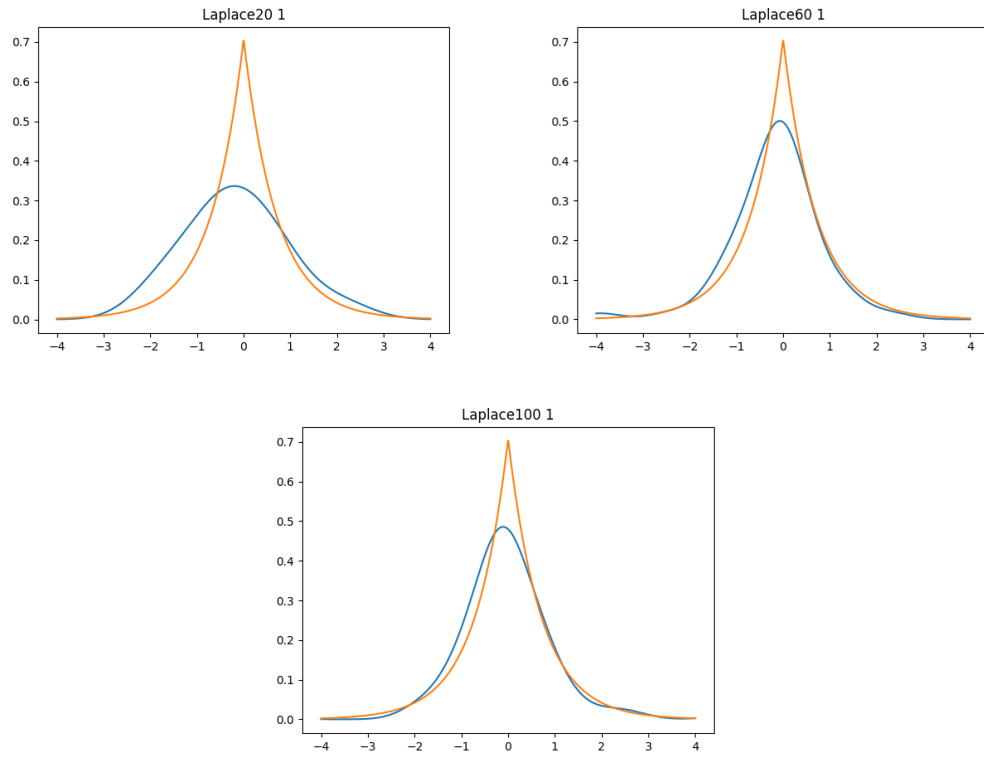


Рис. 22: Ядерные оценки. Распределение Лапласа (5), $h = h_n^S$

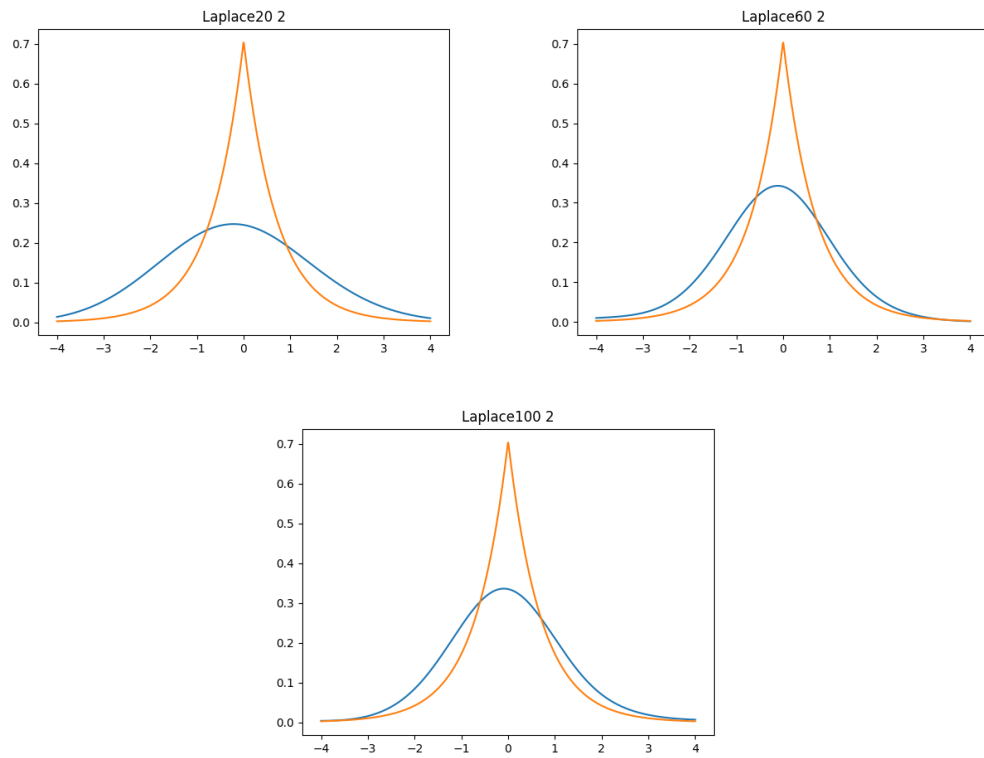


Рис. 23: Ядерные оценки. Распределение Лапласа (5), $h = 2h_n^S$

4.7.4 Распределение Пуассона

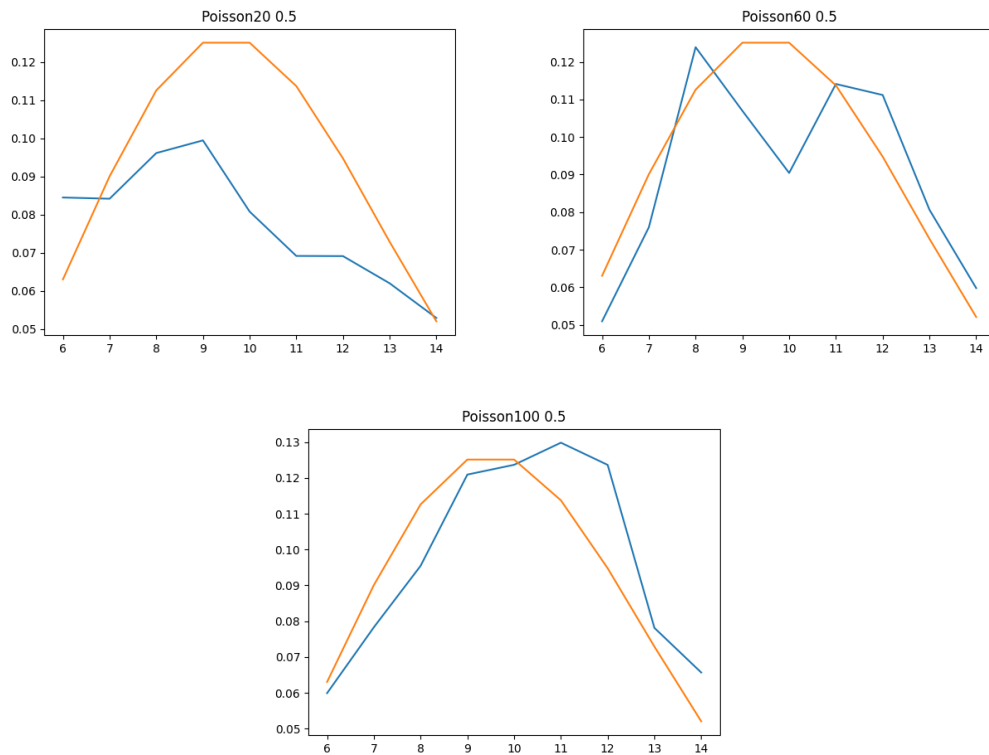


Рис. 24: Ядерные оценки. Распределение Пуассона (6), $h = 0.5h_n^S$

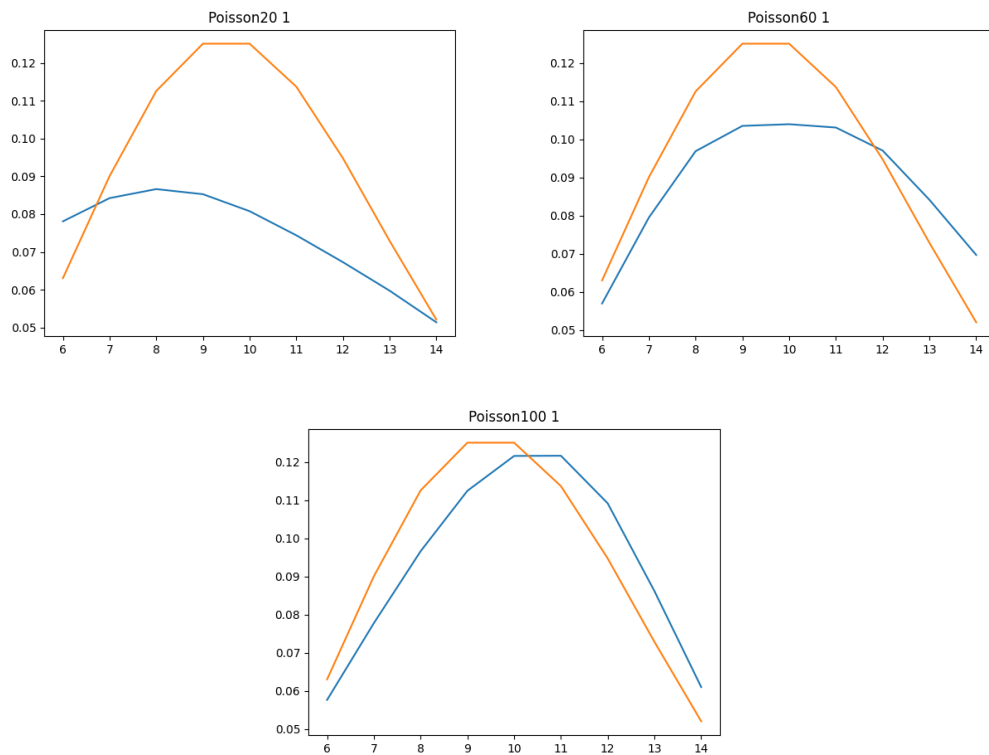


Рис. 25: Ядерные оценки. Распределение Пуассона (6), $h = h_n^S$

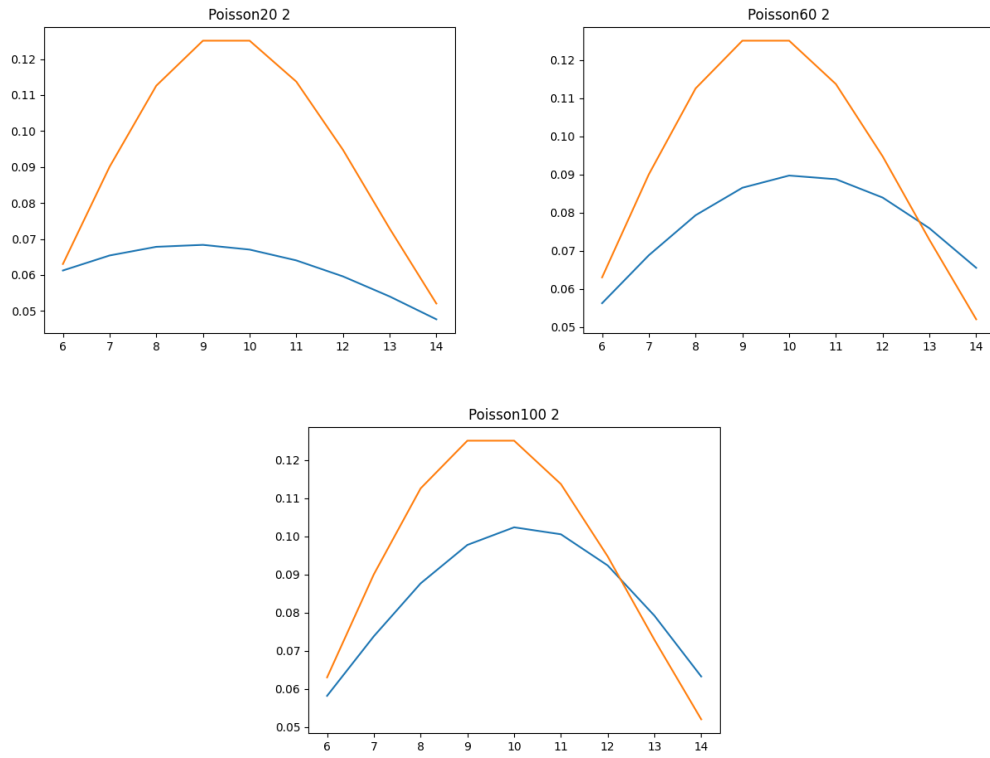


Рис. 26: Ядерные оценки. Распределение Пуассона (6), $h = 2h_n^S$

4.7.5 Равномерное распределение

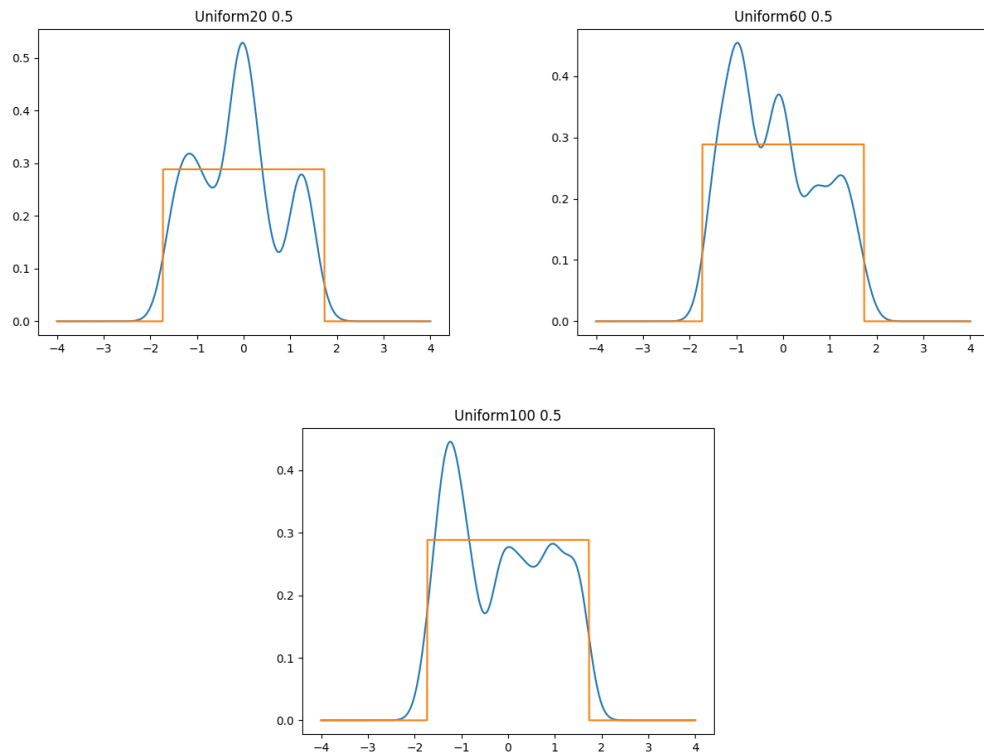


Рис. 27: Ядерные оценки. Равномерное распределение (7), $h = 0.5h_n^S$

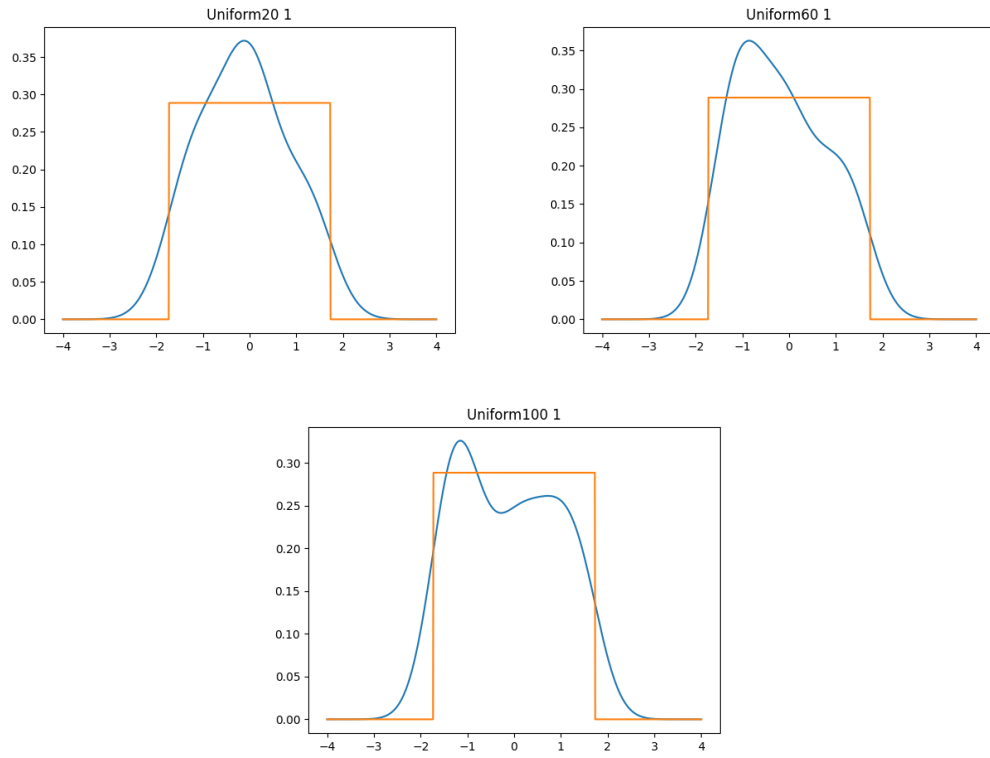


Рис. 28: Ядерные оценки. Равномерное распределение (7), $h = h_n^S$

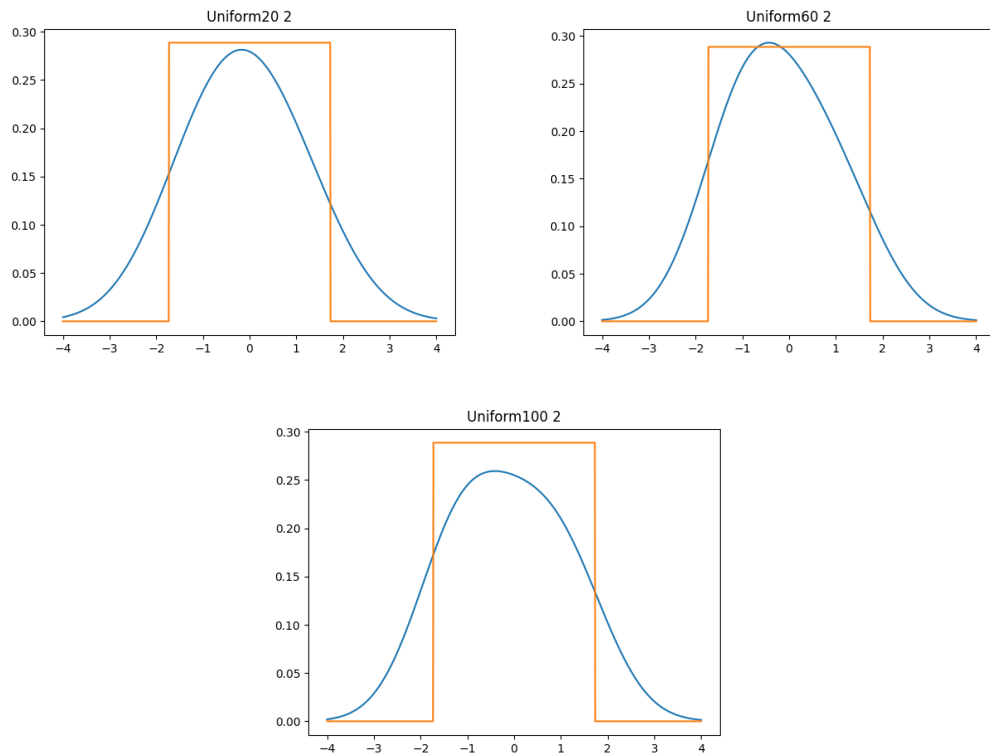


Рис. 29: Ядерные оценки. Равномерное распределение (7), $h = 2h_n^S$

5 Обсуждение

5.1 Гистограмма и график плотности распределения

Результаты проведенных экспериментов подтверждают очевидное предположение о том, что при увеличении мощности выборки случайной величины форма ее гистограммы стремится к графику функции плотности распределения данной случайной величины. Обратно, при малых размерах выборки гистограмма мало совпадает с графиком заданной функцией плотности. В некоторых местах прослеживаются всплески гистограмм, наиболее отчетливо - на распределении Коши.

Стоит также отметить визуальную схожесть для гистограмм нормального, пуассоновского и лапласовского распределений при больших размерах выборки - если распределение Лапласа ещё поддается визуальному различению, то гистограммы для распределений Гаусса и Пуассона практически идентичны.

5.2 Характеристики положения и рассеяния

Сообразно результатам первого эксперимента, в результатах второго также имеется ясная линейная зависимость приближения точных аналитических характеристик заданных случайных величин их статистическими эквивалентами от мощности выборки для всех распределений кроме распределения Коши. На это, в частности, указывает линейное стремление к нулю дисперсии (при увеличении выборки) для заданных характеристик.

Поведение выборочной дисперсии в случае распределения Коши также можно объяснить ссылкой на аналитические свойства данного распределения - отсутствием математического ожидания и бесконечной дисперсией.

5.3 Доля и теоретическая вероятность выбросов

В соответствии с результатами эксперимента (7), (8) для каждого из распределений ясно прослеживается стремление доли выбросов к теоретической оценке при увеличении размеров выборки. Наибольшее количество выбросов за счет «тяжелых хвостов» наблюдается у распределений Коши и Лапласа, наименьшее - околонулевое, - у равномерного распределения.

Боксплоты Тьюки позволяют более наглядно и с меньшими усилиями оценивать важные характеристики выборочных распределений (медиану, квартили, количество выбросов). Результаты построения боксплотов согласуются с результатами предыдущих исследований.

5.4 Эмпирическая функция и ядерные оценки плотности распределения

Из результатов построения графиков эмпирической функции распределения видно, что для всех исследуемых распределений э.ф.р, имеющая ступенчатую структуру функции дискретного распределения, при увеличении размеров выборки стремится к теоретической функции распределения выборки (дискретной или непрерывной). Наибольшее отклонение э.ф.р от теоретической функции распределения наблюдается в случае распределения Пуассона.

Графики ядерных оценок также показывают стремление ядерных оценок к теоретической плотности вероятности с ростом размера выборки для любого h . В зависимости от распределения

для приближения ядерной оценки к плотности лучше подходят разные параметры h : для равномерного распределения лучше подойдет параметр $h = 2h_n^S$, для распределений Лапласа и Коши $h = 0.5h_n^S$, для нормального распределения и распределения Пуассона $h = h_n^S$. Стоит при этом отметить, что при увеличении параметра сглаживания построенные ядерные оценки для всех распределений, за исключением распределения Коши, становятся визуально неотличимыми друг от друга.

Чем больше коэффициент при параметре сглаживания h , тем меньше изменений знака производной у аппроксимирующей функции, вплоть до того, что при $h = h_n^S$ функция становится унимодальной на рассматриваемом промежутке.

Репозиторий

<https://github.com/KoloskovAleksandr/MathStatLabs2021>