

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ им. ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ

КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ

СВОДНЫЙ ОТЧЕТ ПО ЛАБОРАТОРНЫМ 1-4.

3 КУРС, ГРУППА 3630102/70301

Студент группы 3630102/70301

Лебедев К.С.

Преподаватель

Баженов А. Н.

САНКТ-ПЕТЕРБУРГ
2020 г.

Содержание

1. Список иллюстраций	3
2. Постановка задачи	4
3. Теория	4
3.1. Плотности распределений	4
3.2. Характеристики положения	4
3.3. Боксплот Тьюки	5
3.4. Эмпирическая функция и ядерная оценка плотности	5
4. Реализация	5
5. Результаты	6
5.1. Стандартное нормальное распределение	6
5.2. Распределение Коши	11
5.3. Распределение Лапласа	16
5.4. Равномерное распределение	21
5.5. Распределение Пуассона	26
6. Выводы	31
7. Список литературы	32

1 Список иллюстраций

1	Гистограмма нормального распределения	6
2	Характеристики положения нормального распределения	7
3	Boxplot нормальное распределение	7
4	Эмпирическая функция для нормального стандартного распределения ..	8
5	Ядерные функции плотности для нормального распределения	9
6	Гистограмма распределения Коши	11
7	Характеристики положения распределения Коши	11
8	Boxplot распределения Коши	12
9	Эмпирическая функция для распределения Коши	13
10	Ядерные функции плотности для распределения Коши	14
11	Гистограмма распределения Лапласа	16
12	Характеристики положения распределения Лапласа	16
13	Boxplot распределения Лапласа	17
14	Эмпирическая функция для распределения Лапласа.....	18
15	Ядерные функции плотности для распределения Лапласа	19
16	Гистограмма равномерного распределения	21
17	Характеристики положения равномерного распределения	21
18	Boxplot равномерного распределения	22
19	Эмпирическая функция для равномерного распределения	23
20	Ядерные функции плотности для равномерного распределения	24
21	Гистограмма Распределения Пуассона	26
22	Характеристики положения распределения Пуассона	26
23	Boxplot распределения Пуассона	27
24	Эмпирическая функция для распределения Пуассона	28
25	Ядерные функции плотности для распределения Пуассона.....	29

2 Постановка задачи

Сгенерировать выборки различных размеров для 5-ти распределений. На основе этих выборок:

1. Построить гистограммы распределения
2. Вычислить характеристики положения (\bar{x} , $med\ x$, Z_R , Z_Q , Z_{tr} , при $r = \frac{n}{4}$.)
3. Построить боксплот и исследовать распределение на выбросы
4. Построить эмпирические функции распределения и ядерные оценки

3 Теория

3.1 Плотности распределений

1. Стандартное нормальное распределения [4]:

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

2. Распределение Коши[4]:

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

3. Распределение Лапласа[4]:

$$L\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \quad (3)$$

4. Распределение Пуассона[4]:

$$P(\lambda, k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (4)$$

5. Равномерное распределение[4]:

$$M(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & |x| \leq \sqrt{3} \\ 0 & |x| > \sqrt{3} \end{cases} \quad (5)$$

3.2 Характеристики положения

1. Выборочное среднее [8]:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

2. Выборочная медиана [9]:

$$med\ x = \begin{cases} x_{k+1}, & n = 2k + 1 \\ \frac{1}{2}(x_k + x_{k+1}), & n = 2k \end{cases} \quad (7)$$

3. Полусумма экстремальных значений [10]:

$$Z_R = \frac{1}{2}(x_1 + x_n) \quad (8)$$

4. Полусумма квартилей [11]:

$$Z_Q = \frac{1}{2} \left(Z_{\frac{1}{4}} + Z_{\frac{3}{4}} \right) \quad (9)$$

5. Усечённое среднее [12]:

$$Z_{tr} = \frac{1}{n - 2r} \sum_{i=r+1}^{n-r} x_i \quad (10)$$

3.3 Боксплот Тьюки

Боксплот Тьюки - график, использующийся в описательной статистике, изображающий одномерное распределение вероятностей.

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

1. Выборочная медиана [9]:

$$\text{med } x = \begin{cases} x_{k+1}, & n = 2k + 1 \\ \frac{1}{2} (x_k + x_{k+1}), & n = 2k \end{cases} \quad (11)$$

2. Квартиль [?]:

$$z_{[p]} = \begin{cases} x_{np}, & np \in \mathbb{Z} \\ x_{[np]+1}, & np \notin \mathbb{Z} \end{cases} \quad (12)$$

3.4 Эмпирическая функция и ядерная оценка плотности

Эмпирическая функция распределения [5], построенная по выборке $X = (X_1, \dots, X_n)$ есть случайная функция $F_n(y)$, определённая на \mathbb{R} :

$$F_n(y) = \sum_{i=1}^n I(X_i < y) \quad \text{где } I(X_i < y) = \begin{cases} 1, & X_i < y \\ 0, & \text{иначе} \end{cases} \quad (13)$$

$X = (X_1, \dots, X_n)$ есть одномерная выборка одинаково распределённых элементов, с плотностью распределения f .

Ядерная оценка плотности [6]:

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (14)$$

где K является ядром, а $h > 0$ является сглаживающим параметром, и называется шириной полосы.

В данной работе в качестве ядра была выбрана плотность вероятности стандартного нормального распределения [7]:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (15)$$

4 Реализация

Для генерации выборки был использован *Python 3.6*: модуль *random* библиотеки *numpy* для генерации случайных чисел с различными распределениями и библиотека *matplotlib* для построения графиков и гистограмм. Функции распределения обрабатывались при помощи библиотеки *scipy.stats*.

5 Результаты

5.1 Стандартное нормальное распределение

Рис. 1: Гистограмма нормального распределения

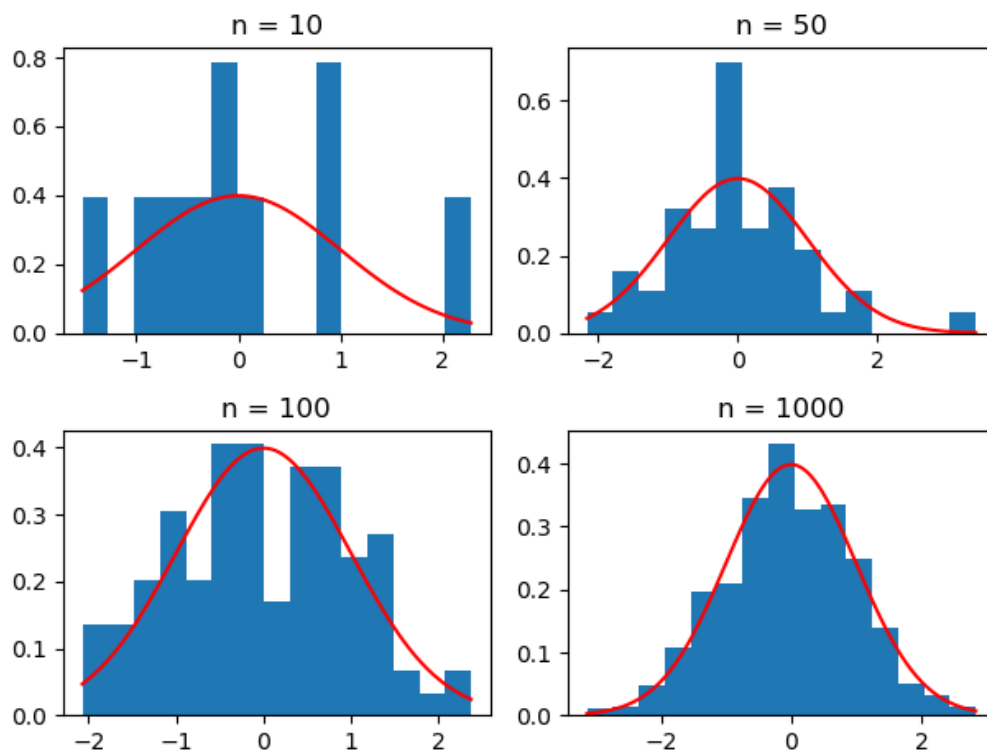
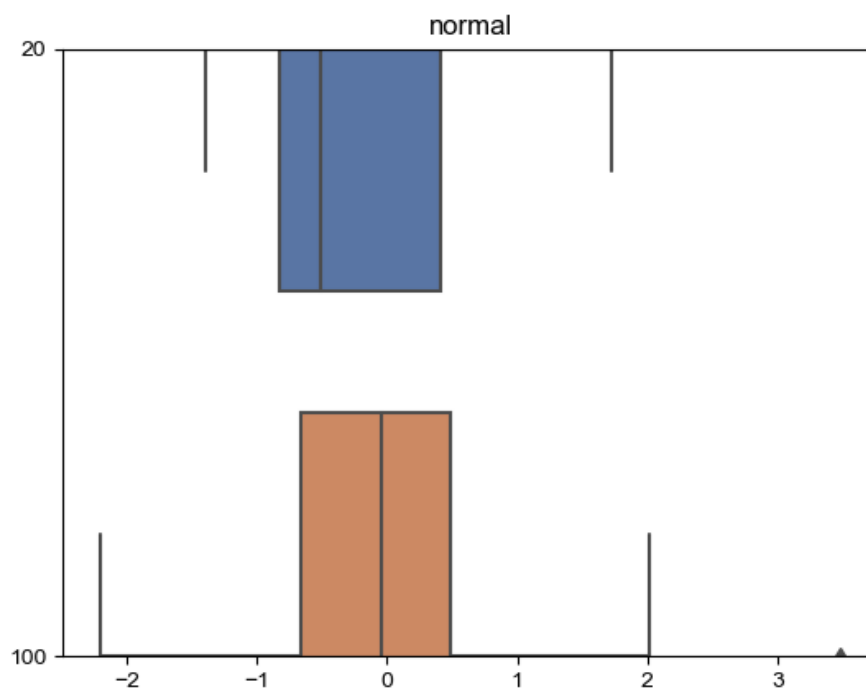


Рис. 2: Характеристики положения нормального распределения

n = 20	average	med	Zr	Zq	Ztr r = n/4
E =	-0.006700	0.018546	0.012695	0.000594	0.015402
D =	0.049919	0.072927	0.147937	0.057251	0.057085
n = 60	average	med	Zr	Zq	Ztr r = n/4
E =	0.001934	0.001927	0.001174	0.003355	-0.001168
D =	0.017054	0.024812	0.103528	0.020604	0.018967
n = 100	average	med	Zr	Zq	Ztr r = n/4
E =	0.000708	0.000648	0.010756	0.001811	-0.003775
D =	0.009767	0.015795	0.097672	0.012952	0.012131

Рис. 3: Boxplot нормальное распределение



Доля выбросов при $n = 20$: 0.022250
 Доля выбросов при $n = 100$: 0.009710

Рис. 4: Эмпирическая функция для нормального стандартного распределения

Empirical distribution function for 20, 60, 100 elements. Distribution: normal

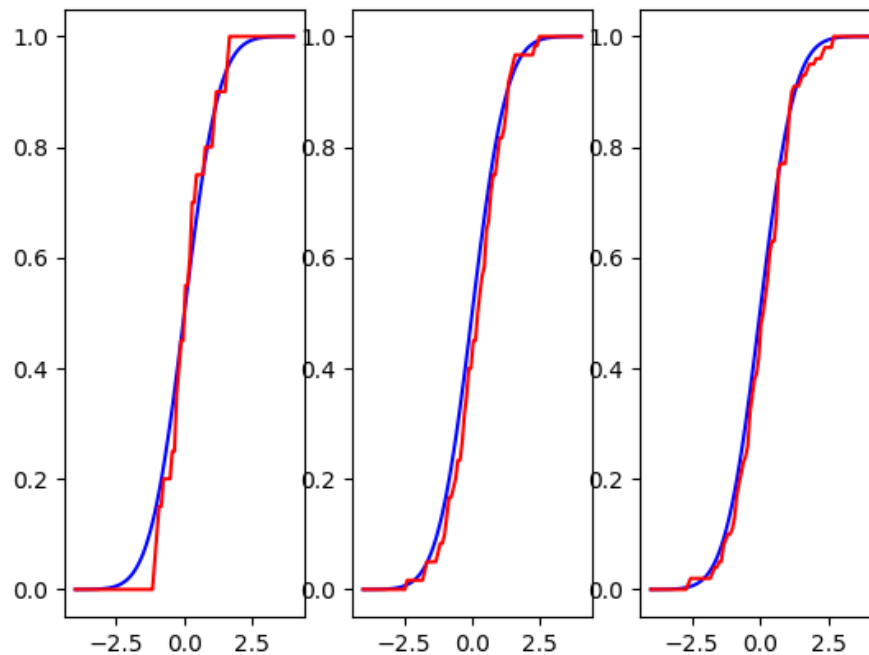
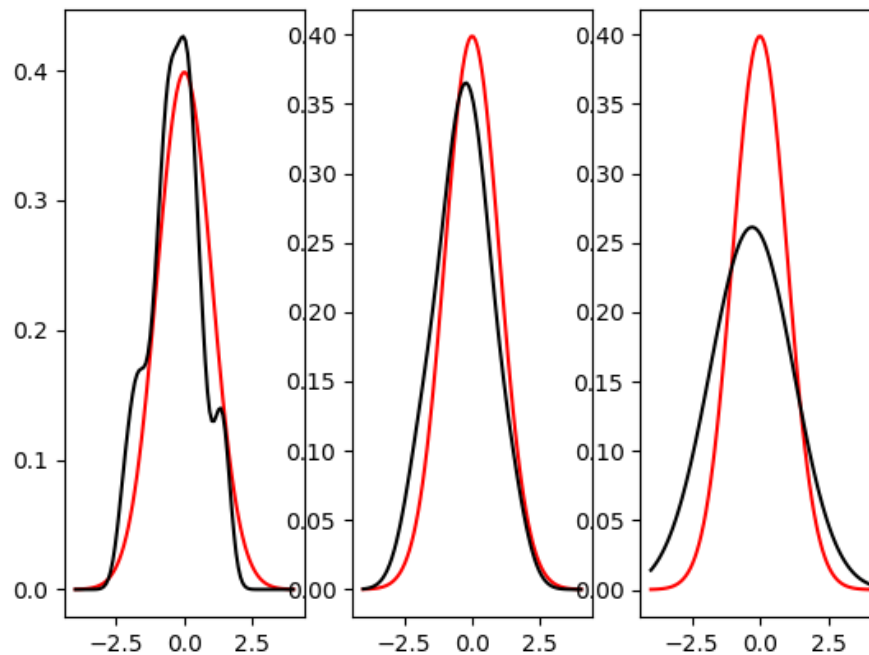
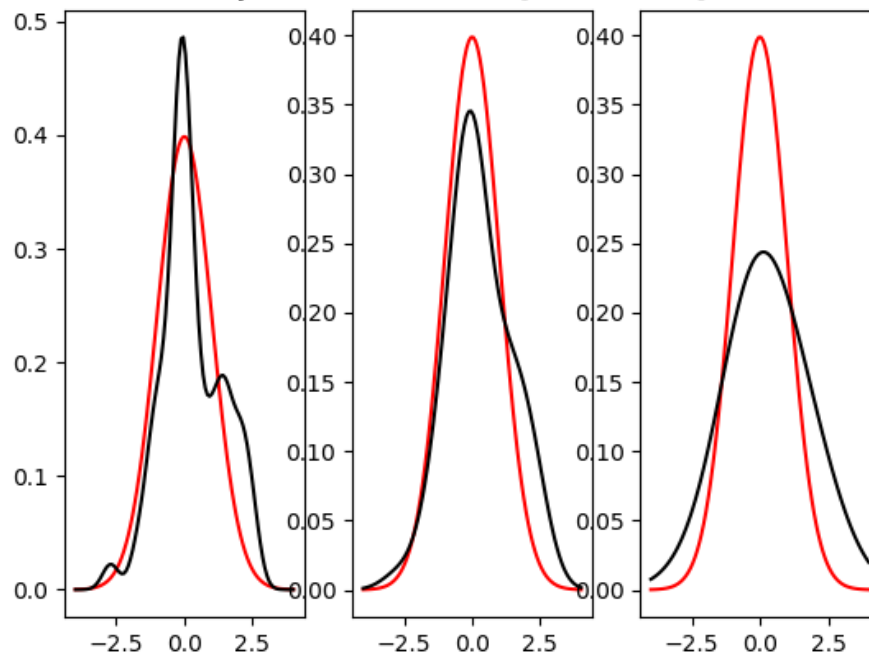


Рис. 5: Ядерные функции плотности для нормального распределения

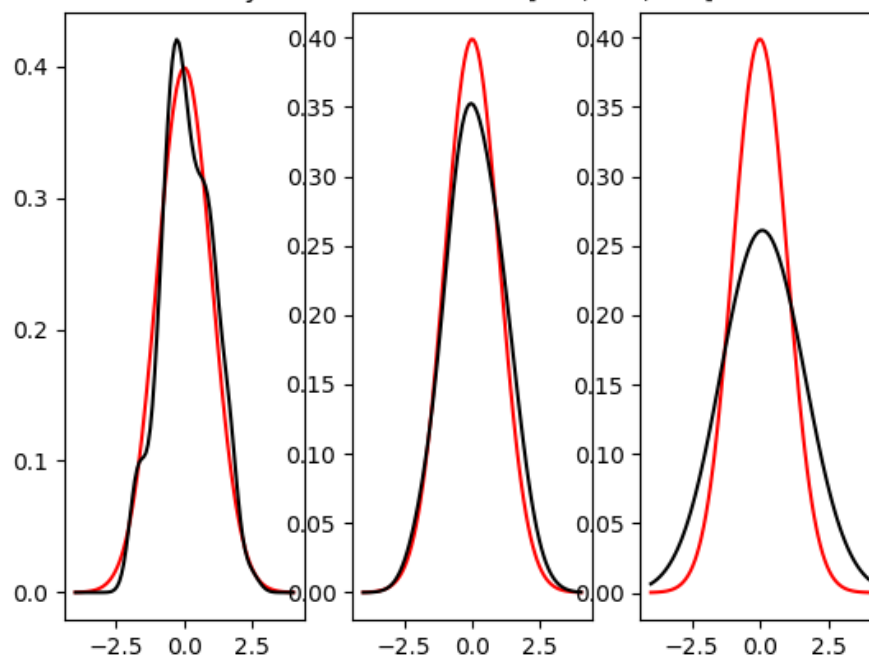
$n = 20$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: normal



$n = 60$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: normal



$n = 100$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: normal



5.2 Распределение Коши

Рис. 6: Гистограмма распределения Коши

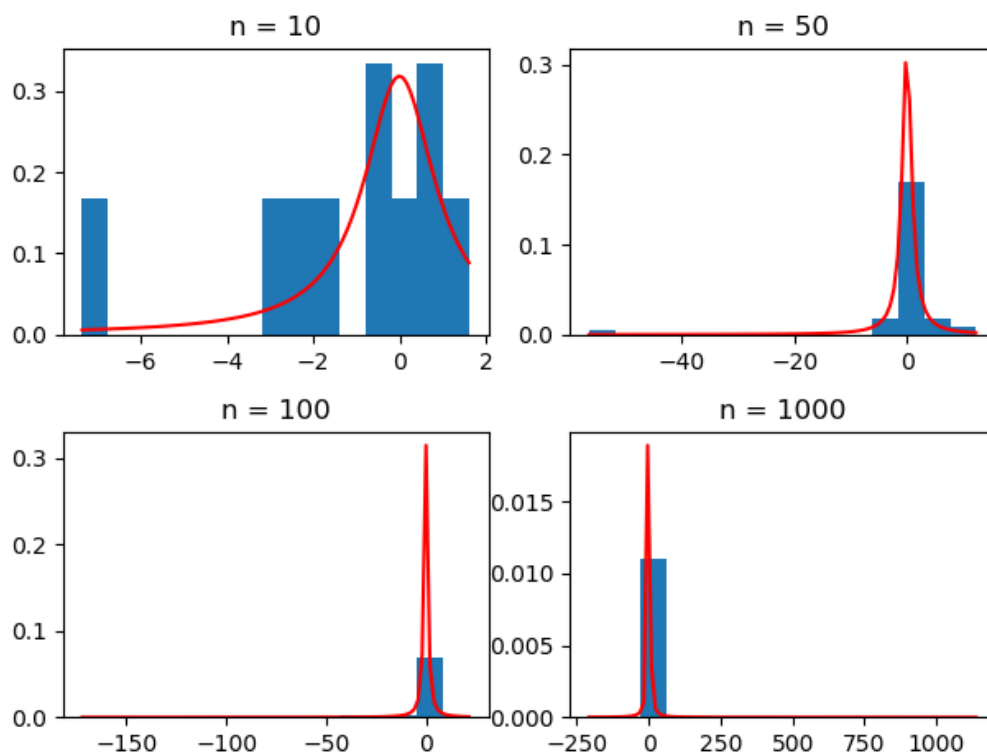
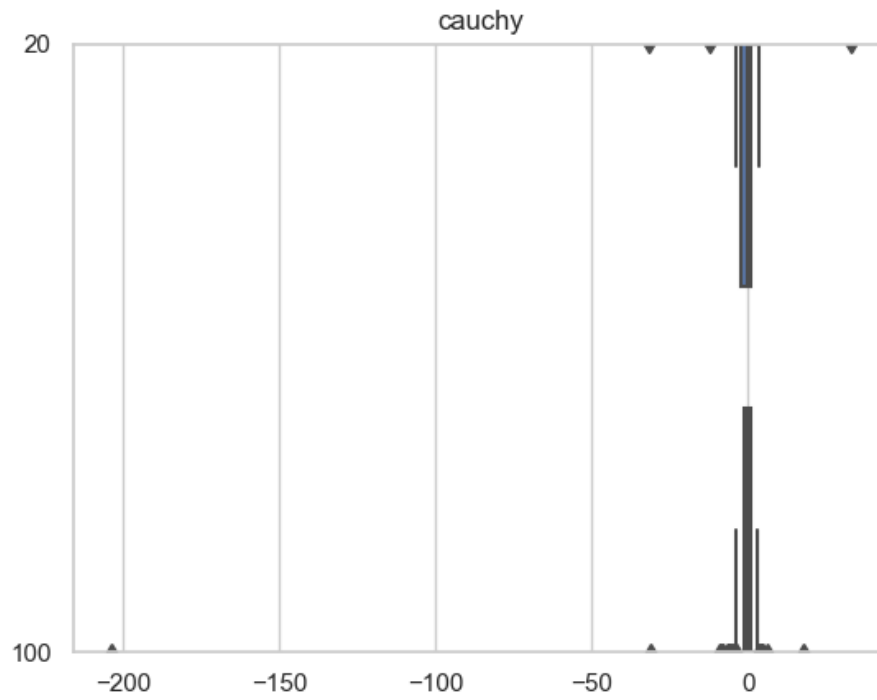


Рис. 7: Характеристики положения распределения Коши

n = 20	average	med	Zr	Zq	Ztr r = n/4
E =	22.758779	-0.006740	78.929010	-0.002305	0.013057
D =	774990.43	0.141492	2423114.9	0.318424	0.178163
n = 60	average	med	Zr	Zq	Ztr r = n/4
E =	0.464150	-0.005508	16.718463	-0.005078	0.007213
D =	599.91366	0.043940	410697.47	0.090639	0.046947
n = 100	average	med	Zr	Zq	Ztr r = n/4
E =	1.667554	-0.003077	128.15939	0.001487	-0.013537
D =	5879.3866	0.024712	9178141.9	0.053342	0.027383

Рис. 8: Boxplot распределения Коши



Доля выбросов при $n = 20$: 0.147700
Доля выбросов при $n = 100$: 0.156480

Рис. 9: Эмпирическая функция для распределения Коши

Empirical distribution function for 20, 60, 100 elements. Distribution:cauchy

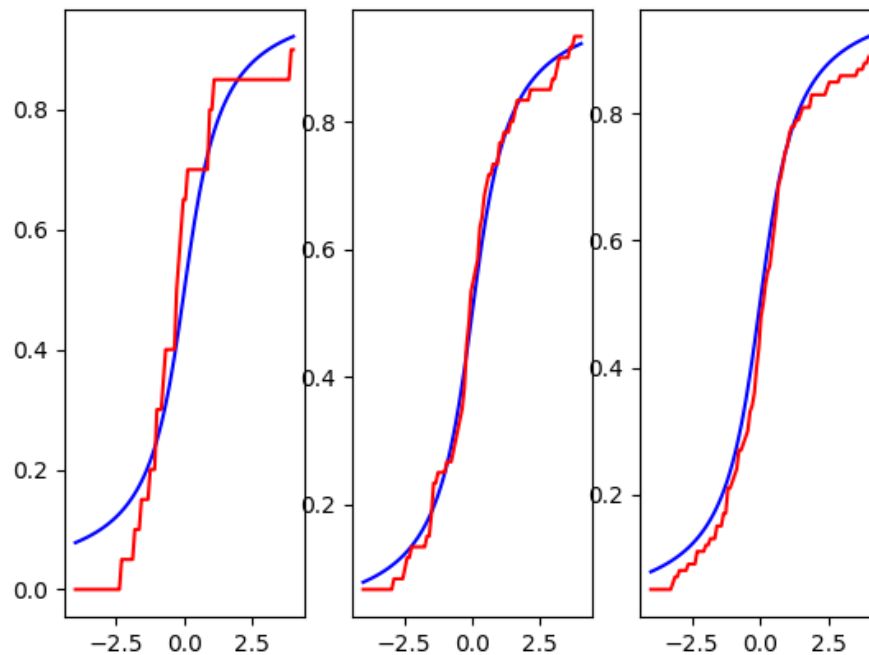
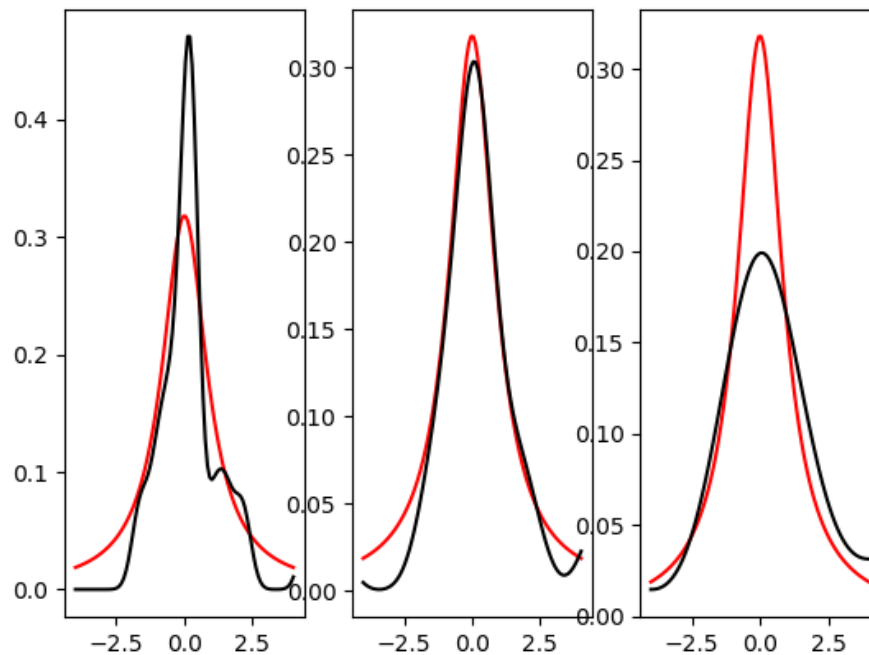
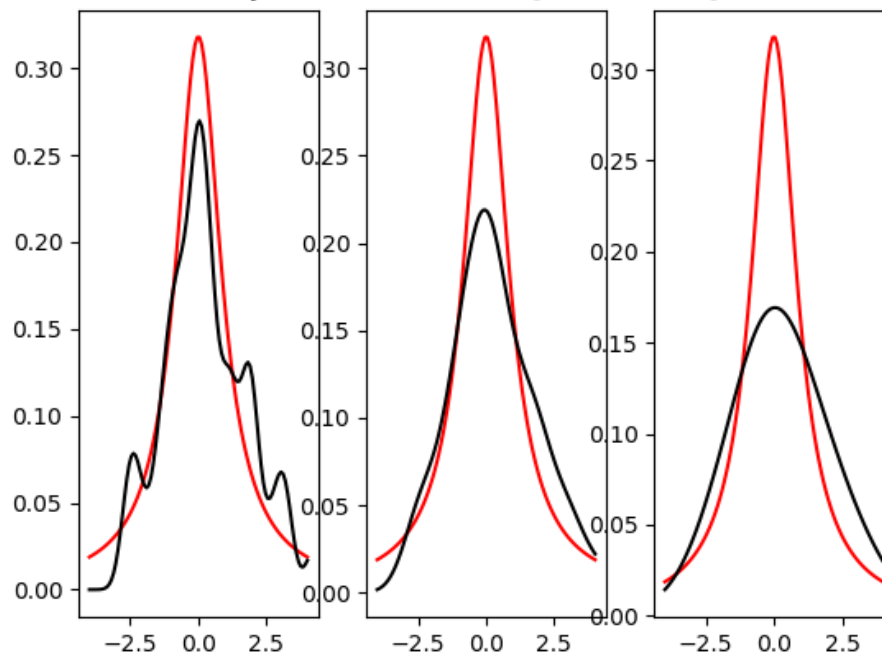


Рис. 10: Ядерные функции плотности для распределения Коши

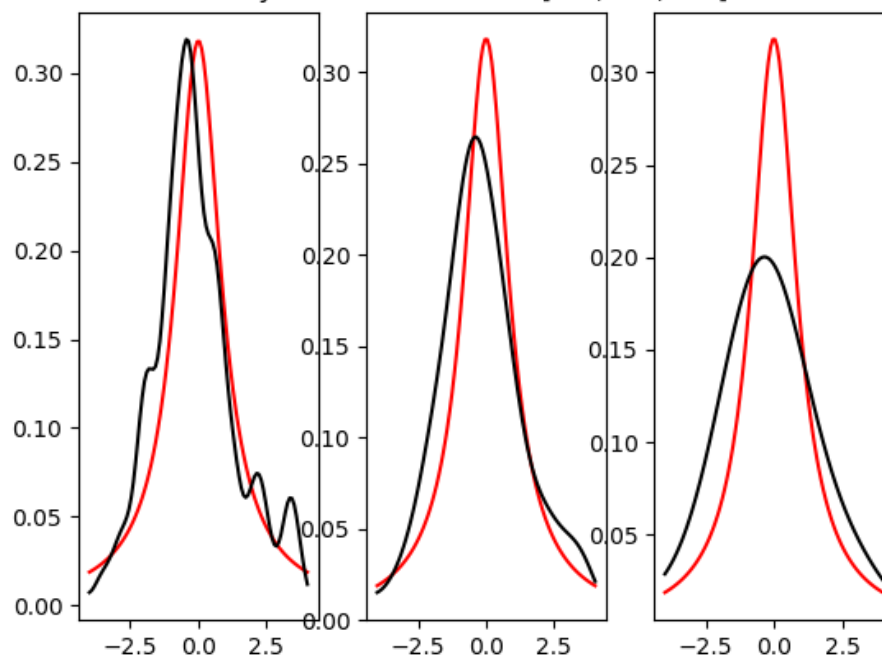
$n = 20$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: cauch



$n = 60$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: cauch



$n = 100$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: cauch



5.3 Распределение Лапласа

Рис. 11: Гистограмма распределения Лапласа

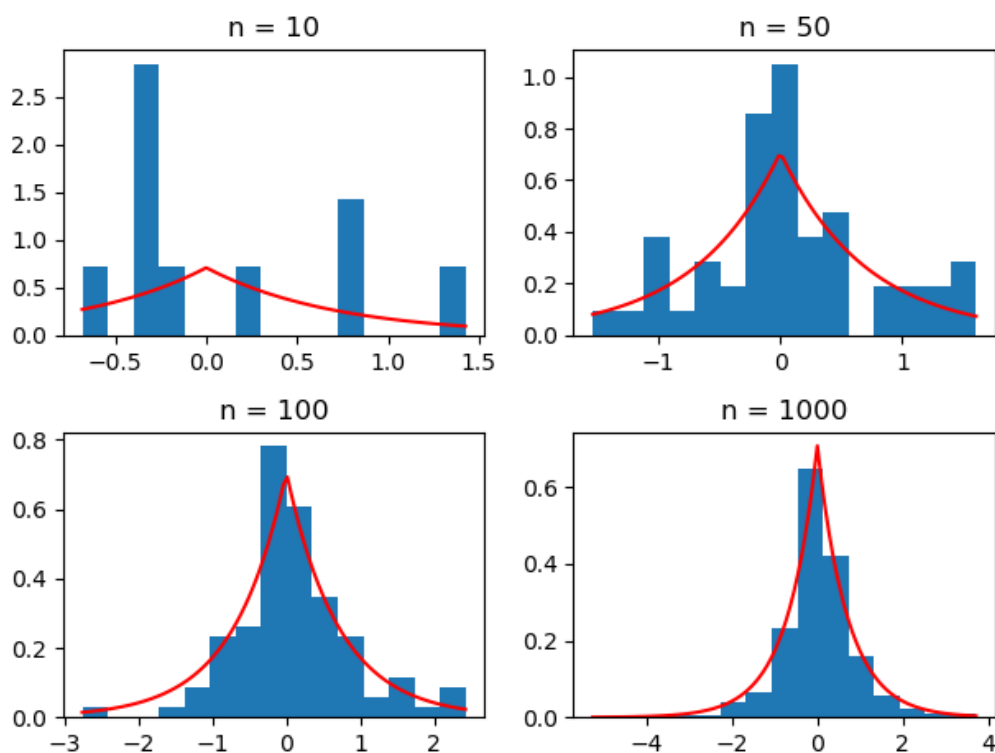
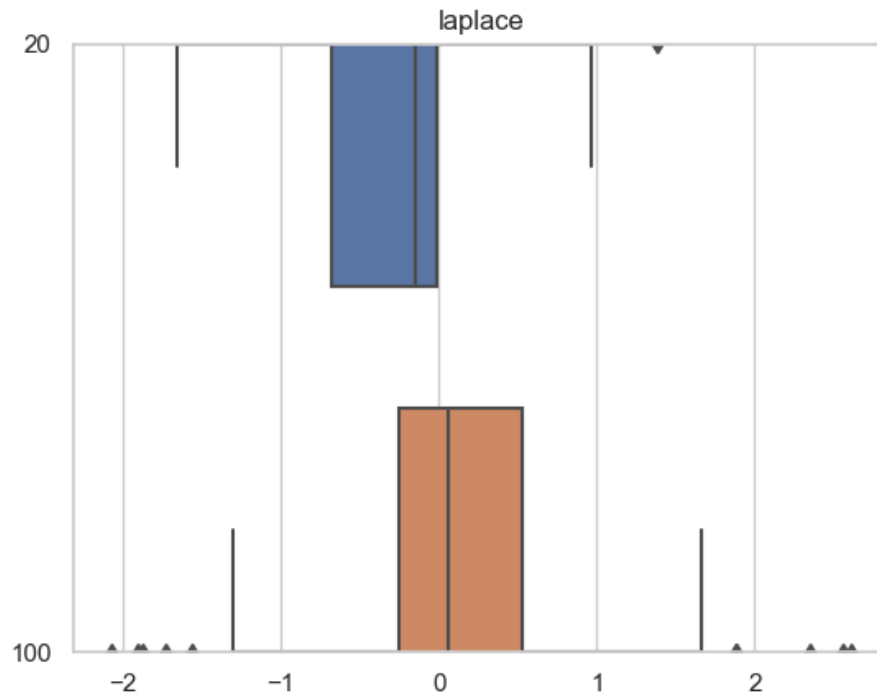


Рис. 12: Характеристики положения распределения Лапласа

$n = 20$	average	med	Zr	Zq	Ztr $r = n/4$
E =	0.006851	0.004852	0.008633	0.000993	0.002342
D =	0.054811	0.034301	0.409060	0.044605	0.034997
$n = 60$	average	med	Zr	Zq	Ztr $r = n/4$
E =	0.003625	-0.002591	-0.003029	-0.003954	-0.003168
D =	0.016641	0.010034	0.417957	0.016484	0.010047
$n = 100$	average	med	Zr	Zq	Ztr $r = n/4$
E =	-0.001436	-0.002020	0.035100	0.006703	0.001789
D =	0.010006	0.006088	0.424029	0.009739	0.005931

Рис. 13: Boxplot распределения Лапласа



Доля выбросов при $n = 20$: 0.072300
Доля выбросов при $n = 100$: 0.066380

Рис. 14: Эмпирическая функция для распределения Лапласа

Empirical distribution function for 20, 60, 100 elements. Distribution: laplace

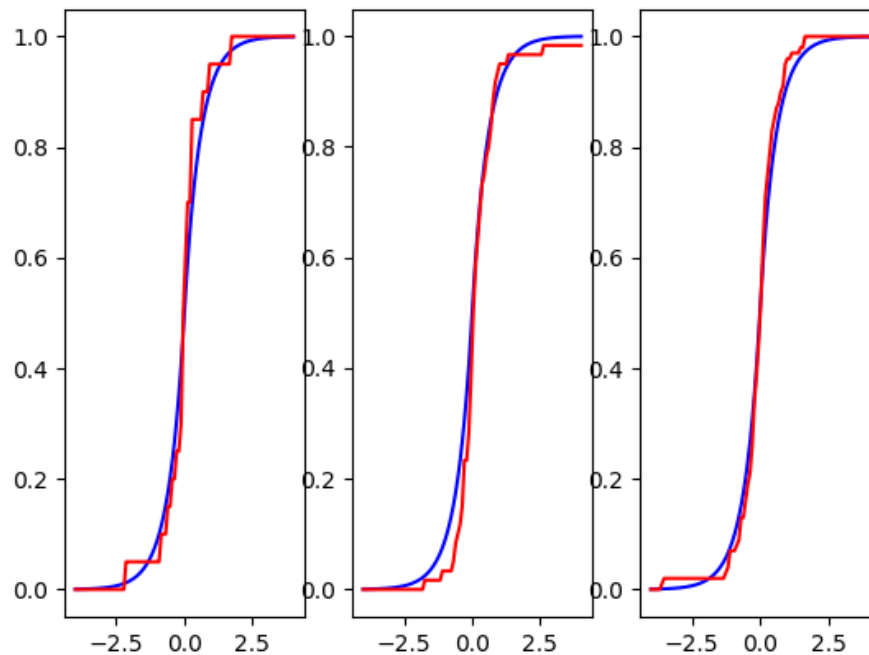
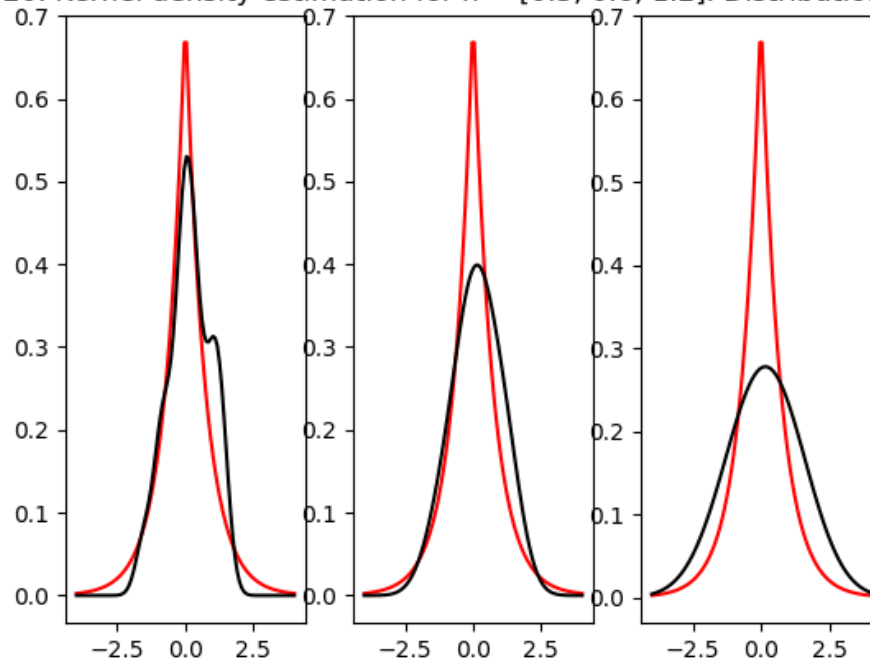
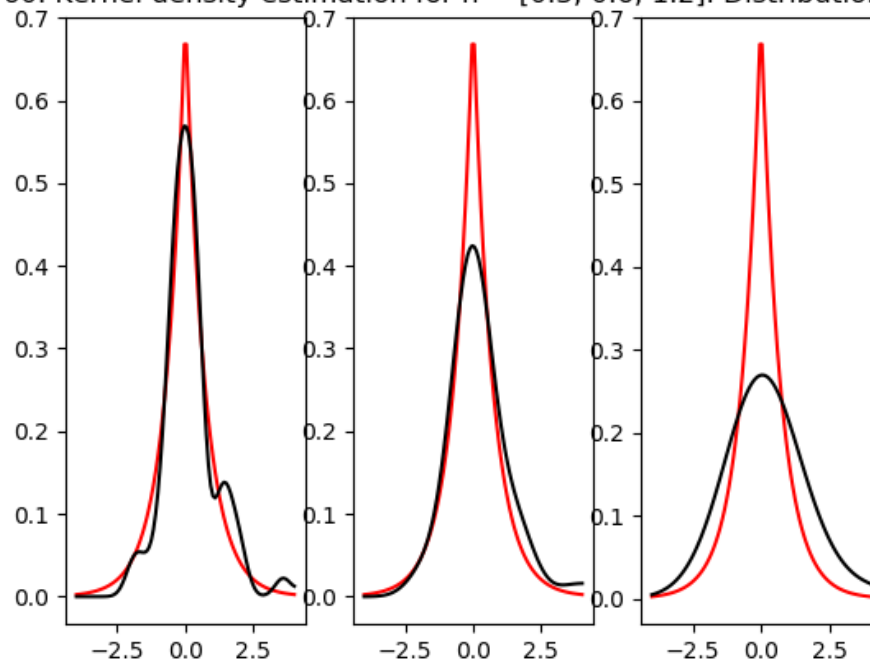


Рис. 15: Ядерные функции плотности для распределения Лапласа

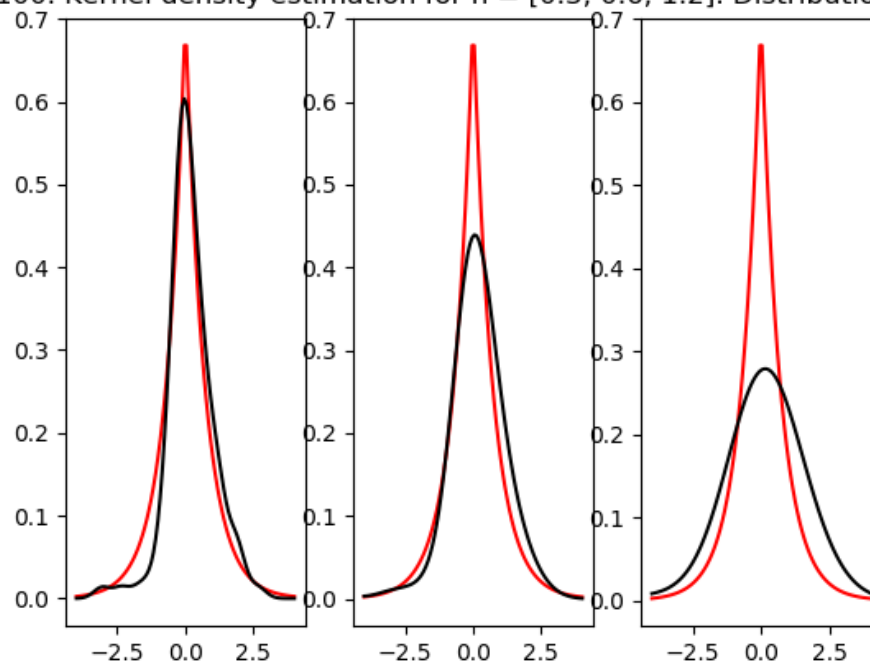
n = 20. Kernel density estimation for h = [0.3, 0.6, 1.2]. Distribution: laplac



$n = 60$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: laplac



$n = 100$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: laplac



5.4 Равномерное распределение

Рис. 16: Гистограмма равномерного распределения

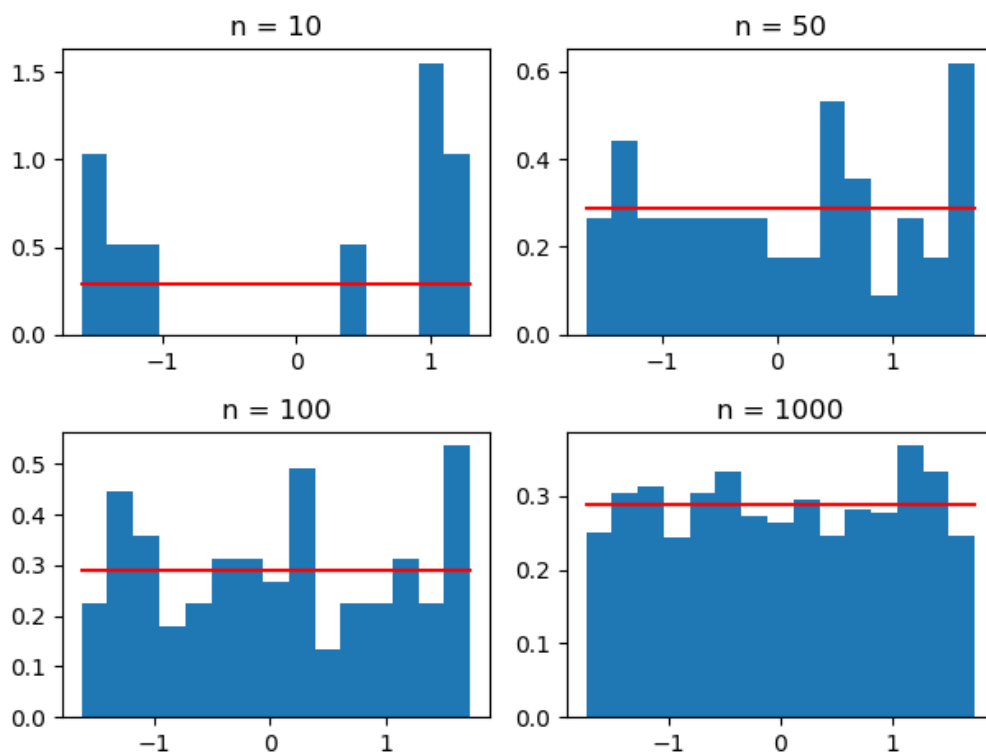
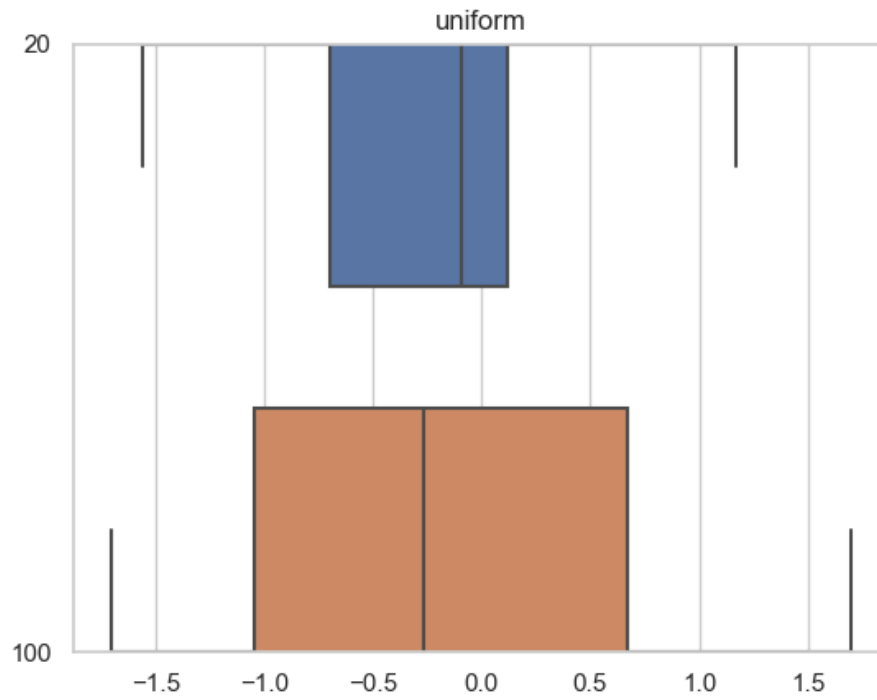


Рис. 17: Характеристики положения равномерного распределения

$n = 20$	average	med	Zr	Zq	Ztr $r = n/4$
E =	0.012552	-0.006810	-0.000923	0.004960	0.009501
D =	0.049610	0.131448	0.013292	0.076374	0.086512
$n = 60$	average	med	Zr	Zq	Ztr $r = n/4$
E =	-0.003837	-0.007174	-0.000018	-0.010948	-0.008353
D =	0.017882	0.047638	0.001874	0.023592	0.031756
$n = 100$	average	med	Zr	Zq	Ztr $r = n/4$
E =	-0.000175	-0.000765	-0.001623	-0.002510	0.002546
D =	0.009693	0.029956	0.000587	0.014596	0.019309

Рис. 18: Вохplot равномерного распределения



Доля выбросов при $n = 20$: 0.002200
Доля выбросов при $n = 100$: 0.000000

Рис. 19: Эмпирическая функция для равномерного распределения

Empirical distribution function for 20, 60, 100 elements. Distribution: uniform

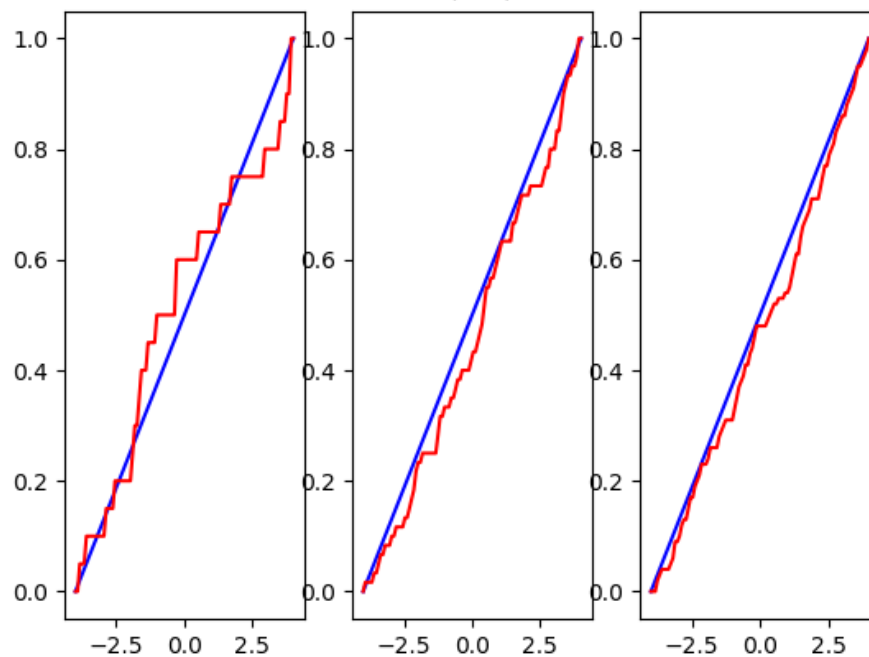
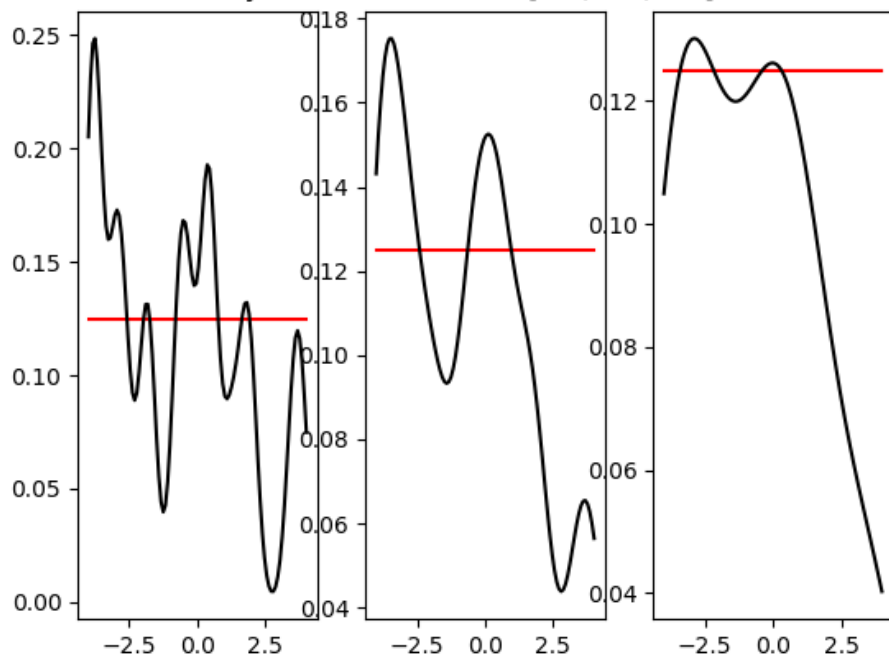
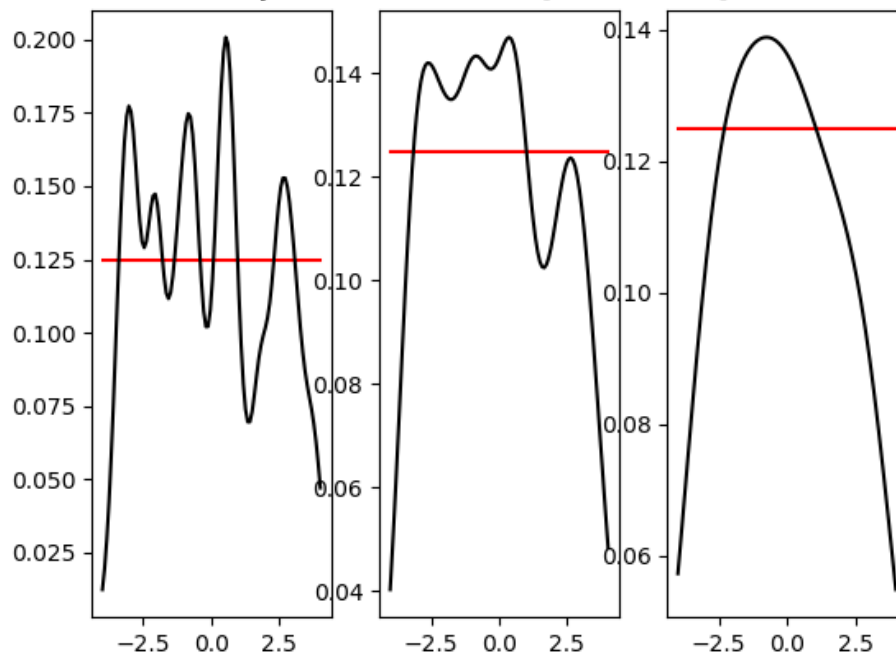


Рис. 20: Ядерные функции плотности для равномерного распределения

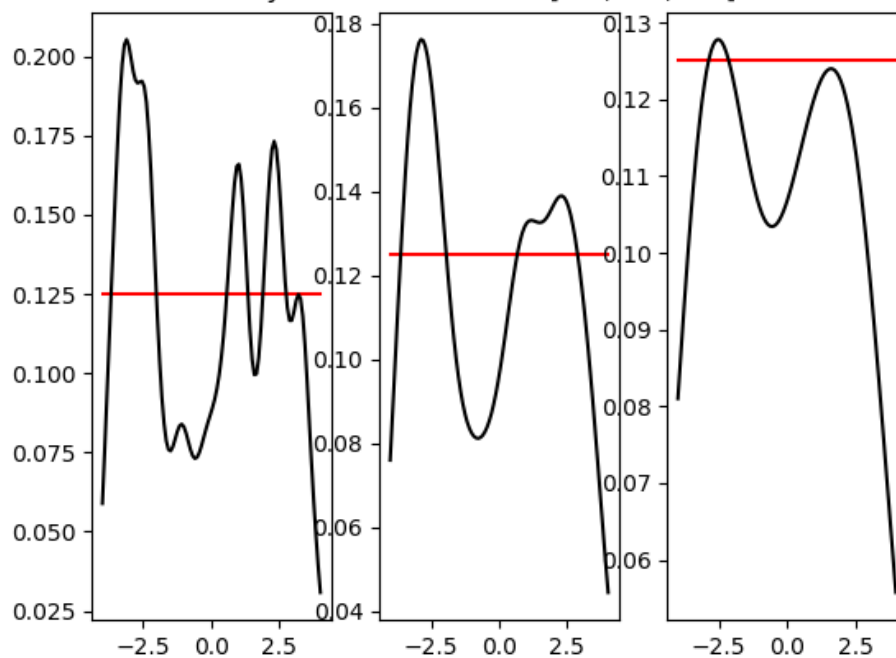
$n = 20$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: uniform



$n = 60$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: uniform



$n = 100$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: uniform



5.5 Распределение Пуассона

Рис. 21: Гистограмма Распределения Пуассона

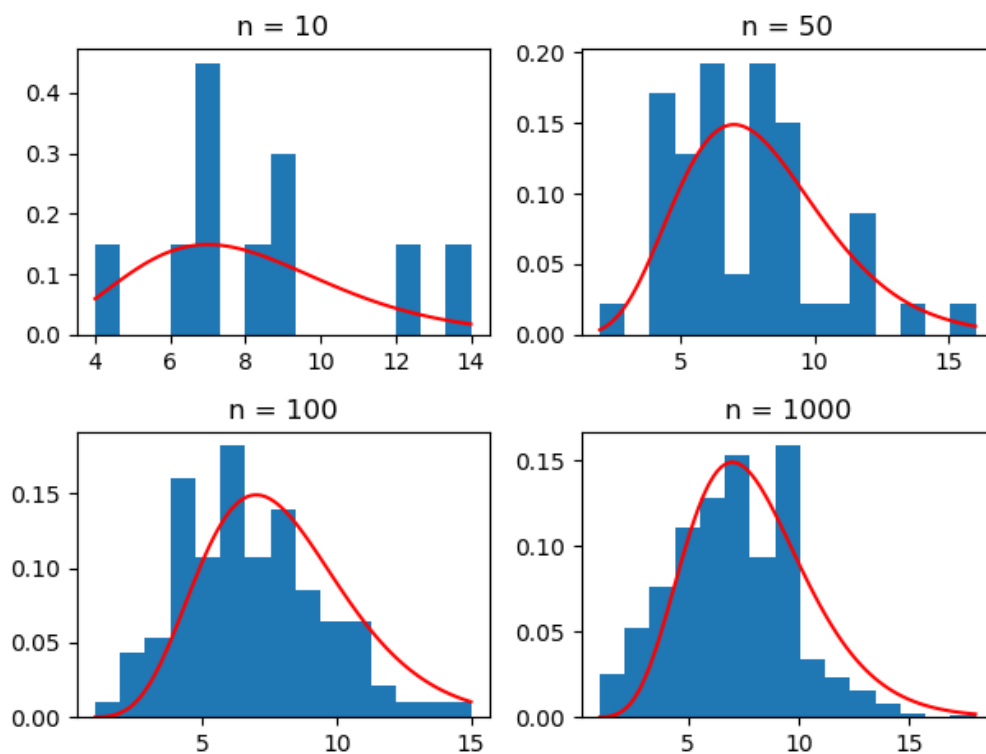
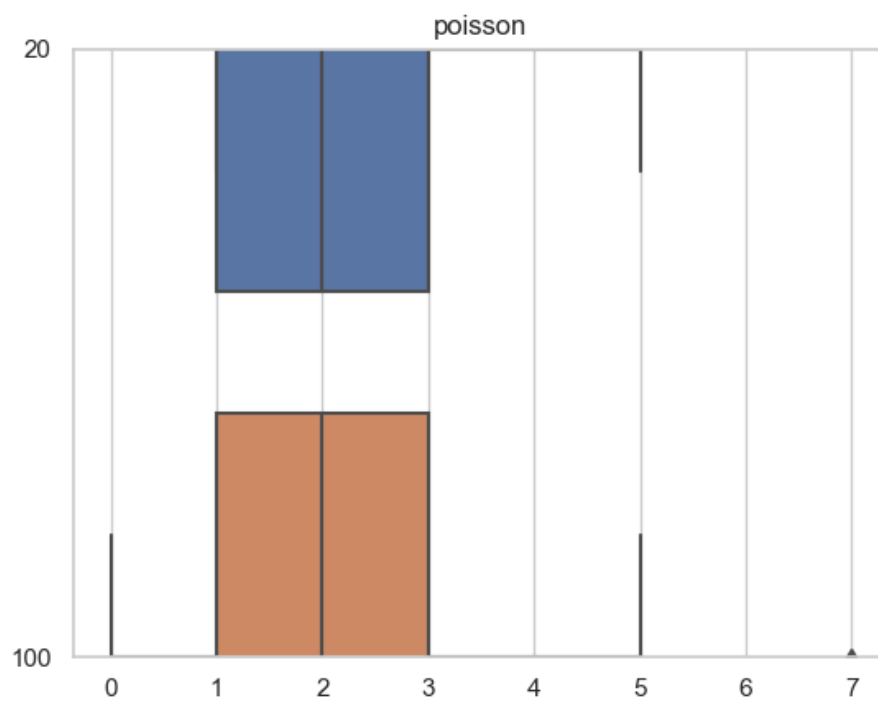


Рис. 22: Характеристики положения распределения Пуассона

n = 20	average	med	Zr	Zq	Ztr r = n/4
E =	1.987600	1.870500	2.515500	1.912250	1.868300
D =	0.099786	0.178480	0.303010	0.121644	0.129565
n = 60	average	med	Zr	Zq	Ztr r = n/4
E =	1.994817	1.936000	2.968500	1.954750	1.828900
D =	0.032801	0.051904	0.241258	0.031765	0.045717
n = 100	average	med	Zr	Zq	Ztr r = n/4
E =	1.996890	1.968500	3.168000	1.971250	1.831700
D =	0.021158	0.027758	0.280776	0.012861	0.026473

Рис. 23: Вохplot распределения Пуассона



Доля выбросов при $n = 20$: 0.034300
 Доля выбросов при $n = 100$: 0.009650

Рис. 24: Эмпирическая функция для распределения Пуассона

Empirical distribution function for 20, 60, 100 elements.Distribution:poissor

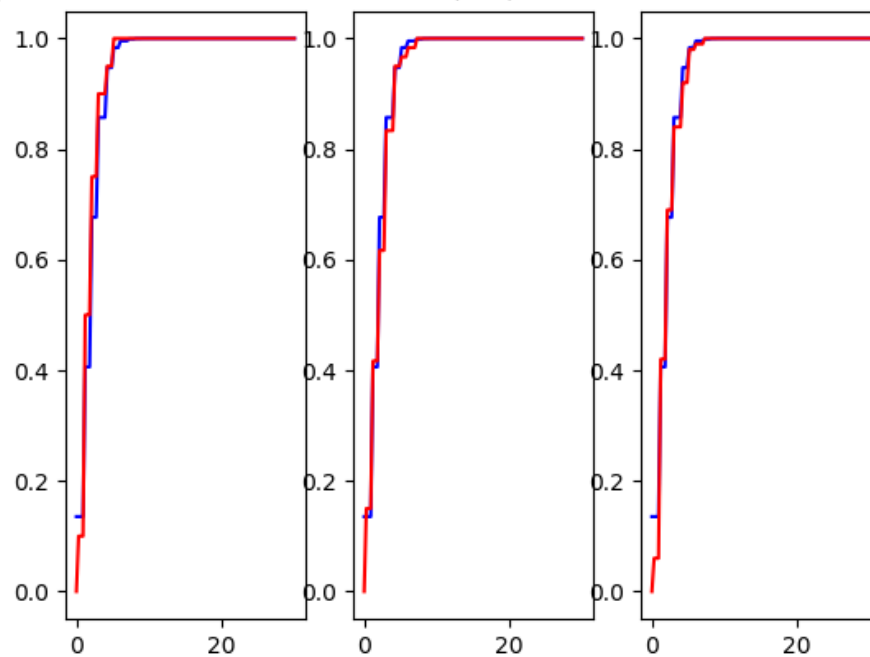
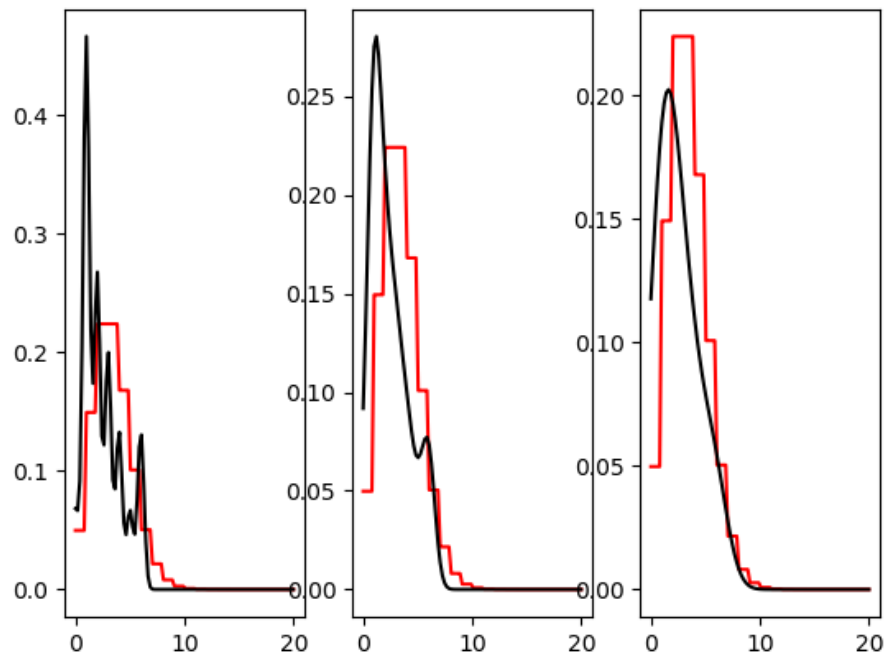
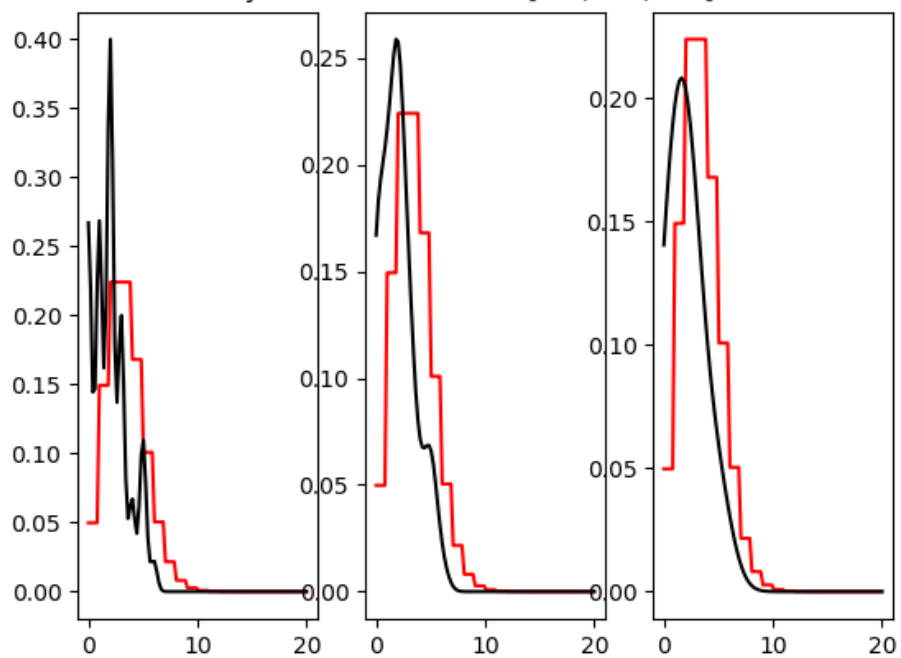


Рис. 25: Ядерные функции плотности для распределения Пуассона

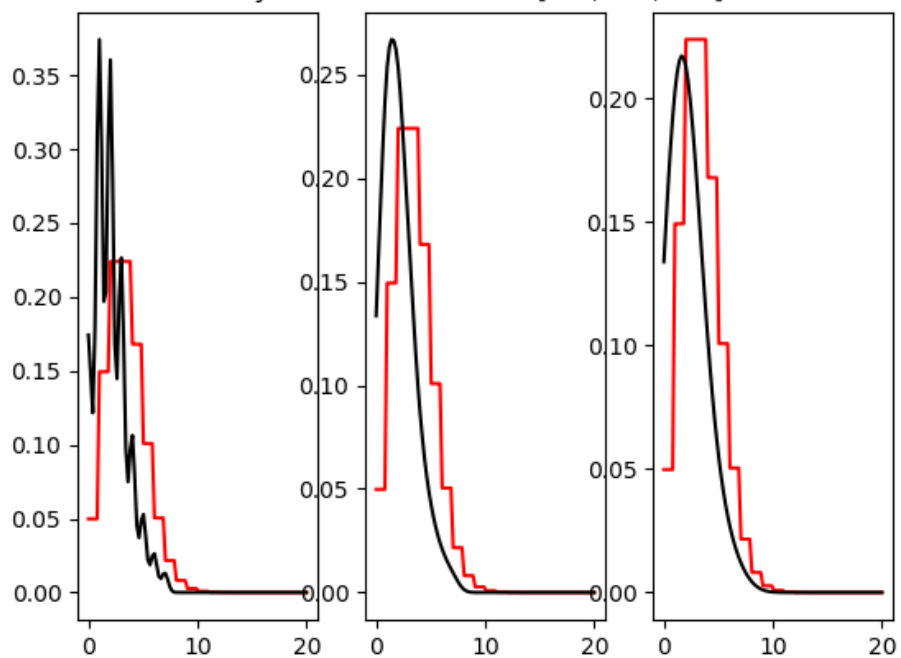
n = 20. Kernel density estimation for h = [0.3, 0.6, 1.2]. Distribution: poisson



$n = 60$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: poisson



$n = 100$. Kernel density estimation for $h = [0.3, 0.6, 1.2]$. Distribution: poisson



6 Выводы

1. При увеличении размера выборки построенная гистограмма приближается к графику плотности.
2. Соотношения для характеристик положения:

- Стандартное нормальное распределение

$$\bar{x} < Z_{tr} < Z_Q < med\ x < Z_R$$

- Стандартное распределение Коши

$$med\ x < Z_Q < Z_{tr} < \bar{x} < Z_R$$

- Распределение Лапласа (коэффициент масштаба $\sqrt{2}$ коэффициент сдвига равен нулю)

$$med\ x < Z_{tr} < \bar{x} < Z_Q < Z_R$$

- Равномерное распределение на отрезке $[-\sqrt{3}, \sqrt{3}]$

$$Z_R < \bar{x} < Z_{tr} < Z_Q < med\ x$$

- Распределение Пуассона (значение мат ожидания равно 3)

$$\bar{x} < Z_{tr} < Z_Q < med\ x < Z_R$$

3. Наименьший процент выбросов у равномерного распределения, а наибольший процент выбросов у распределения Коши
4. Эмпирическая функция лучше приближает эталонную функцию на больших выборках.

Наилучшее приближение функции распределения ядерной функции получено при наибольшей ширине окна. При фиксированной ширине окна точнее приблизить функцию распределения позволяет увеличение выборки.

7 Список литературы

- [1] Модуль numpy - <https://physics.susu.ru/vorontsov/language/numpy.html>
- [2] Модуль matplotlib - <https://matplotlib.org/users/index.html>
- [3] Модуль scipy - <https://docs.scipy.org/doc/scipy/reference/>
- [4] Формулы распределений - https://vk.com/doc184549949_491827451
- [5] <https://nsu.ru/mmftvims/chernova/ms/lec/node4.html>
- [6] <https://www.mql5.com/ru/articles/396>
- [7] <http://users.stat.umn.edu/helwig/notes/den-Notes.pdf>
- [8] Выборочное среднее - https://en.wikipedia.org/wiki/Sample_mean_and_covariance
- [9] Выборочная медиана - http://femto.com.ua/articles/part_1/2194.html
- [10] Полусумма экстремальных значений - <https://studopedia.info/8-56888.html>
- [11] Квартили - <https://studfiles.net/preview/2438125/page:13/>
- [12] Усечённое среднее - <https://ole-olesko.livejournal.com/15773.html>