

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ПЕТРА ВЕЛИКОГО

ФИЗИКО-МЕХАНИЧЕСКИЙ ИНСТИТУТ

ВЫСШАЯ ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ВЫЧИСЛИТЕЛЬНОЙ ФИЗИКИ

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

Выполнил студент:
Иванова А.С.
группа: 5030102/00101

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

Санкт-Петербург
2023 г.

Содержание

1	Постановка задачи	4
2	Теория	5
2.1	Рассматриваемые распределения	5
2.2	Гистограмма	5
2.3	Вариационный ряд	5
2.4	Выборочные числовые характеристики	5
2.4.1	Характеристики положения	6
2.4.2	Характеристики рассеяния	6
2.5	Боксплот Тьюки	7
2.5.1	Построение	7
2.6	Теоретическая вероятность выбросов	7
2.7	Эмпирическая функция распределения	7
2.7.1	Статистический ряд	7
2.7.2	Определение	8
2.7.3	Описание	8
2.8	Оценки плотности вероятности	8
2.8.1	Определение	8
2.8.2	Ядерные оценки	8
3	Реализация	10
4	Результаты	11
4.1	Гистограмма и график плотности распределения	11
4.2	Характеристики положения и рассеяния	12
4.3	Боксплот Тьюки	15
4.4	Доля выбросов	18
4.5	Теоретическая вероятность выбросов	18
4.6	Эмпирическая функция распределения	18
4.7	Ядерные оценки плотности распределения	20
5	Обсуждение	26
5.1	Гистограмма и график плотности распределения	26
5.2	Характеристики положения и рассеяния	26
5.3	Боксплот Тьюки	26
5.4	Доля выбросов	26
5.5	Эмпирическая функция распределения	26
5.6	Ядерные оценки плотности распределения	26
	Литература	27

Список иллюстраций

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

Список таблиц

1	Статистический ряд	7
2	Таблица распределения	8
3	Характеристики положения и рассеяния нормального распределения	13
4	Характеристики положения и рассеяния распределения Коши	13
5	Характеристики положения и рассеяния распределения Лапласа	14
6	Характеристики положения и рассеяния распределения Пуассона	14
7	Характеристики положения и рассеяния равномерного распределения	15
8	Доля выбросов	18
9	Теоретическая вероятность выбросов	18

1 Постановка задачи

Для 5 распределений:

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

1. Сгенерировать выборки размером 10, 50 и 1000 элементов. Построить на одном рисунке гистограмму и график плотности распределения.
2. Сгенерировать выборки размером 10, 100 и 1000 элементов. Для каждой выборки вычислить следующие статистические характеристики положения данных: \bar{x} , $med\ x$, z_R , z_Q , z_{tr} . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения их квадратов:

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

Вычислить оценку дисперсии по формуле:

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

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 вероятности рассматриваемых распределений:

- Нормальное распределение

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

- Распределение Коши

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

- 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

Множество значений, которое может принимать элемент выборки разбивается на несколько интервалов. Чаще всего эти интервалы берутся одинаковыми (но не обязательно). Данные интервалы откладываются на горизонтальной оси, затем над каждым рисуется прямоугольник. Если все интервалы одинаковые, то высота каждого прямоугольника пропорциональна числу элементов выборки, попадающих в соответствующий интервал. Если интервалы разные, то высота прямоугольника выбирается так, чтобы его площадь была пропорциональна числу элементов выборки, попавших в данный интервал. [1].

2.3 Вариационный ряд

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

2.4 Выборочные числовые характеристики

[2, с. 409]

2.4.1 Характеристики положения

- Выборочное среднее

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

- Выборочная медиана

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

- Полусумма экстремальных выборочных элементов

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

- Полусумма квартилей

Выборочная квартиль z_p порядка p определяется формулой

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

Полусумма квартилей

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

- Усечённое среднее

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

2.4.2 Характеристики рассеяния

Выборочная дисперсия

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

2.5 Боксплот Тьюки

2.5.1 Построение

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

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

где X_1 — нижняя граница уса, X_2 — верхняя граница уса, Q_1 — первый квартиль, Q_3 — третий квартиль. Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков.[3]

2.6 Теоретическая вероятность выбросов

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

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

Теоретическая вероятность выбросов для непрерывных распределений

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

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

Теоретическая вероятность выбросов для дискретных распределений

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

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

2.7 Эмпирическая функция распределения

2.7.1 Статистический ряд

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

Статистический ряд обычно записывается в виде таблицы

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

Таблица 1: Статистический ряд

2.7.2 Определение

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

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

2.7.3 Описание

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

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

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

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

Таблица 2: Таблица распределения

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

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

2.8 Оценки плотности вероятности

2.8.1 Определение

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

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

2.8.2 Ядерные оценки

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

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

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

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

Такие оценки называются непрерывными ядерными [2, с. 421-423].

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

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

Гауссово (нормальное) ядро [4, с. 38]

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

Правило Сильвермана [4, с. 44]

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

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

3 Реализация

Лабораторная работа была выполнена с помощью встроенных средств языка программирования Python (библиотеки: NumPy, SciPy, Matplotlib, Seaborn) в среде разработки Visual Studio Code. Исходный код работы приведен в приложении.

Ссылка на репозиторий с исходным кодом: <https://github.com/anivse/MathematicalStatistics>

4 Результаты

4.1 Гистограмма и график плотности распределения

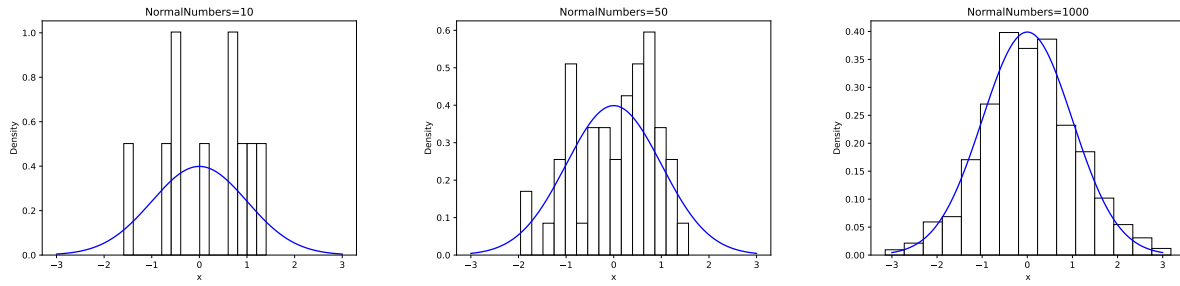


Рис. 1: Нормальное распределение

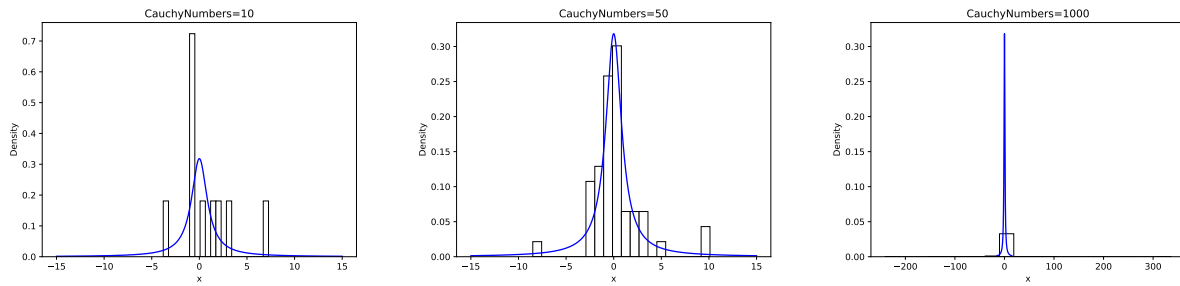


Рис. 2: Распределение Коши

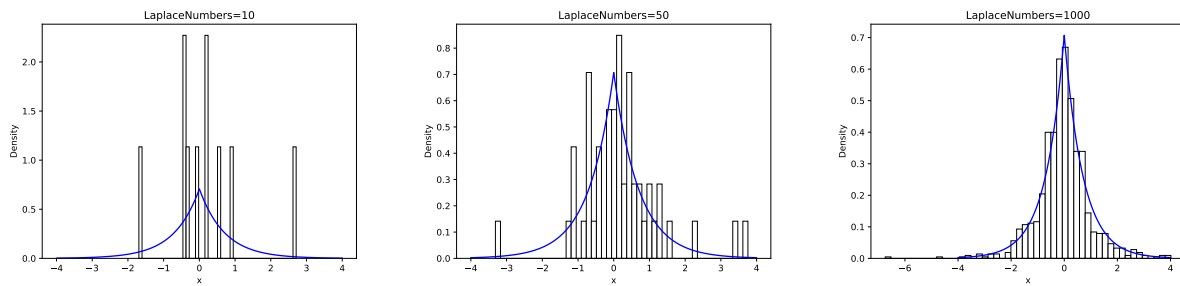


Рис. 3: Распределение Лапласа

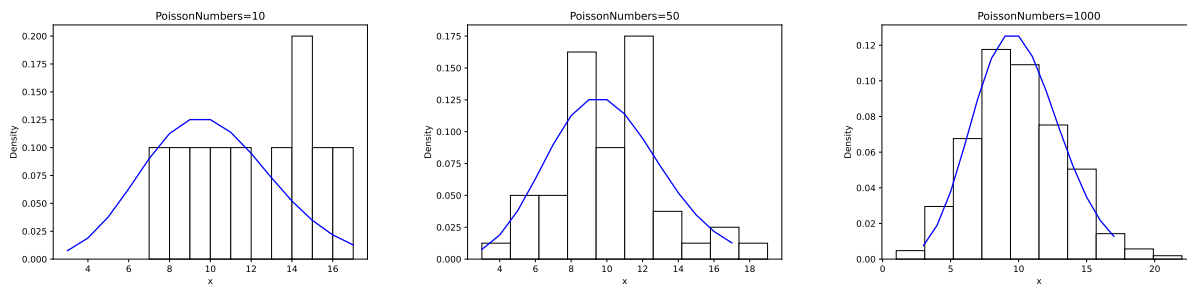


Рис. 4: Распределение Пуассона

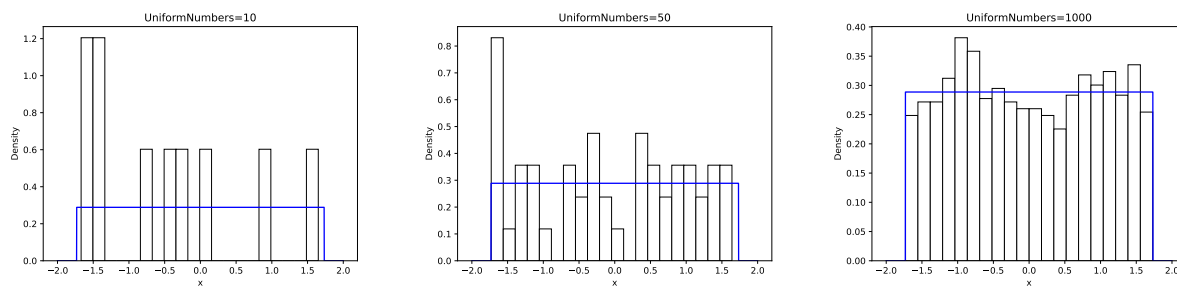


Рис. 5: Равномерное распределение

4.2 Характеристики положения и рассеяния

Округление проводилось следующим образом:

В оценке $x = E \pm D$ вариации подлежит первая цифра после точки. В данном случае $x = 0.0 \pm 0.1k$, k зависит от доверительной вероятности и вида распределения (рассматривается в дальнейшем цикле лабораторных работ).

Округление сделано для $k = 1$

Normal n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.00423	0.00676	-0.01363	-0.00839	0.00216
$D(z)$	0.09888	0.13638	0.17744	0.11849	0.11997
$E(z) \pm \sqrt{D(z)}$	[-0.31868;0.31023]	[-0.36254;0.37605]	[-0.43486;0.40760]	[-0.35262;0.33584]	[-0.34421;0.34852]
$\hat{E}(z)$	0.0 ± 0.3	0.0 ± 0.4	0.0 ± 0.4	0.0 ± 0.3	0.0 ± 0.3
Normal n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.00267	-0.00356	-0.01646	-0.01707	-0.00476
$D(z)$	0.01051	0.01671	0.08515	0.01307	0.01265
$E(z) \pm \sqrt{D(z)}$	[-0.10519;0.09985]	[-0.13281;0.12570]	[-0.30826;0.27534]	[-0.13138;0.09724]	[-0.11723;0.10770]
$\hat{E}(z)$	0.0 ± 0.1	0.0 ± 0.1	0.0 ± 0.3	0.0 ± 0.1	0.0 ± 0.1
Normal n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00003	0.00090	0.01124	-0.00193	0.00065
$D(z)$	0.00099	0.00161	0.06110	0.00121	0.00122
$E(z) \pm \sqrt{D(z)}$	[-0.03145;0.03152]	[-0.03923;0.04102]	[-0.23595;0.25843]	[-0.03674;0.03287]	[-0.03428;0.03557]
$\hat{E}(z)$	0.0	0.0	0.0 ± 0.2	0.0	0.0

Таблица 3: Характеристики положения и рассеяния нормального распределения

Cauchy n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-2.05988	-0.00508	-10.24303	-0.02834	-0.00392
$D(z)$	4882.32839	0.34732	121894.61704	1.24367	0.37423
$E(z) \pm \sqrt{D(z)}$	[-71.93354;67.81378]	[-0.59441;0.58426]	[-359.37713;338.89107]	[-1.14354;1.08686]	[-0.61566;0.60782]
$\hat{E}(z)$	-	0.0 ± 0.5	-	-	0.0 ± 0.6
Cauchy n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.74021	-0.00696	-35.05494	-0.03106	-0.00405
$D(z)$	191.82278	0.02725	458909.38994	0.05454	0.02816
$E(z) \pm \sqrt{D(z)}$	[-14.59022;13.10980]	[-0.17204;0.15812]	[-712.48345;642.37357]	[-0.26461;0.20248]	[-0.17185;0.16375]
$\hat{E}(z)$	-	0.0 ± 0.2	-	-	0.0 ± 0.2
Cauchy n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-1.49646	0.00078	-749.72204	-0.00382	0.00007
$D(z)$	5237.69953	0.00240	1303730937.09335	0.00490	0.00248
$E(z) \pm \sqrt{D(z)}$	[-73.86841;70.87550]	[-0.04825;0.04981]	[-36856.93652;35357.49243]	[-0.07381;0.06618]	[-0.04974;0.04988]
$\hat{E}(z)$	-	0.0	-	-	0.0

Таблица 4: Характеристики положения и рассеяния распределения Коши

Laplace n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.00209	0.00125	-0.01499	0.00119	0.00307
$D(z)$	0.10094	0.07944	0.38017	0.10796	0.07744
$E(z) \pm \sqrt{D(z)}$	[-0.31980;0.31563]	[-0.28061;0.28310]	[-0.63157;0.60160]	[-0.32738;0.32976]	[-0.27522;0.28135]
$\hat{E}(z)$	0.0 ± 0.3	0.0 ± 0.3	0.0 ± 0.6	0.0 ± 0.3	0.0 ± 0.3
Laplace n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00179	-0.00190	-0.01973	-0.00883	0.00074
$D(z)$	0.01082	0.00611	0.40428	0.01029	0.00666
$E(z) \pm \sqrt{D(z)}$	[-0.10222;0.10580]	[-0.08005;0.07626]	[-0.65556;0.61610]	[-0.11025;0.09258]	[-0.08084;0.08232]
$\hat{E}(z)$	0.0 ± 0.1	0.0	0.0 ± 0.6	0.0 ± 0.1	0.0
Laplace n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00092	0.00041	-0.00812	-0.00043	0.00060
$D(z)$	0.00105	0.00052	0.40209	0.00104	0.00062
$E(z) \pm \sqrt{D(z)}$	[-0.03150;0.03335]	[-0.02232;0.02313]	[-0.64223;0.62598]	[-0.03261;0.03176]	[-0.02429;0.02549]
$\hat{E}(z)$	0.0	0.0	0.0 ± 0.6	0.0	0.0

Таблица 5: Характеристики положения и рассеяния распределения Лапласа

Poisson n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	10.00820	9.86250	10.29550	9.93450	9.87850
$D(z)$	1.04423	1.49934	1.86593	1.23946	1.33149
$E(z) \pm \sqrt{D(z)}$	[8.98632;11.03008]	[8.63802;11.08698]	[8.92951;11.66149]	[8.82119;11.04781]	[8.72460;11.03240]
$\hat{E}(z)$	10 ± 1	10 ± 1	10 ± 1	10 ± 1	10 ± 1
Poisson n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	9.99391	9.84300	10.88600	9.87050	9.84868
$D(z)$	0.09509	0.19885	0.85600	0.14998	0.11589
$E(z) \pm \sqrt{D(z)}$	[9.68554;10.30228]	[9.39707;10.28893]	[9.96080;11.81120]	[9.48323;10.25777]	[9.50826;10.18910]
$\hat{E}(z)$	10 ± 0.3	10 ± 0.4	10 ± 0.5	10 ± 0.3	10 ± 0.3
Poisson n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	10.00147	9.99300	11.66950	9.99500	9.86030
$D(z)$	0.01015	0.00695	0.67602	0.00298	0.01094
$E(z) \pm \sqrt{D(z)}$	[9.90074;10.10219]	[9.90963;10.07637]	[10.84730;12.49170]	[9.94046;10.04954]	[9.75572;9.96488]
$\hat{E}(z)$	10 ± 0.1	10 ± 0.1	11.5 ± 1	10 ± 0.0	9.8 ± 0.2

Таблица 6: Характеристики положения и рассеяния распределения Пуассона

Uniform n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.01019	0.01713	0.00280	0.01554	0.01411
$D(z)$	0.10163	0.23375	0.04321	0.13959	0.19896
$E(z) \pm \sqrt{D(z)}$	[-0.30861;0.32899]	[-0.46635;0.50060]	[-0.20507;0.21068]	[-0.35808;0.38917]	[-0.43193;0.46015]
$\hat{E}(z)$	0.0 ± 0.3	$0.0 \pm 0.$	0.0 ± 0.2	0.0 ± 0.4	0.0 ± 0.4
Uniform n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.00082	0.00055	0.00004	-0.01693	0.00053
$D(z)$	0.00933	0.02793	0.00051	0.01384	0.01814
$E(z) \pm \sqrt{D(z)}$	[-0.09740;0.09576]	[-0.16658;0.16768]	[-0.02264;0.02272]	[-0.13459;0.10072]	[-0.13414;0.13520]
$\hat{E}(z)$	0.0	0.0 ± 0.1	0.0	0.0 ± 0.1	0.0 ± 0.1
Uniform n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.00012	0.00082	-0.00005	-0.00204	-0.00020
$D(z)$	0.00099	0.00283	0.00001	0.00151	0.00198
$E(z) \pm \sqrt{D(z)}$	[-0.03152;0.03128]	[-0.05235;0.05399]	[-0.00236;0.00225]	[-0.04087;0.03678]	[-0.04473;0.04433]
$\hat{E}(z)$	0.0	0.0	0.0	0.0	0.0

Таблица 7: Характеристики положения и рассеяния равномерного распределения

4.3 Боксплот Тьюки

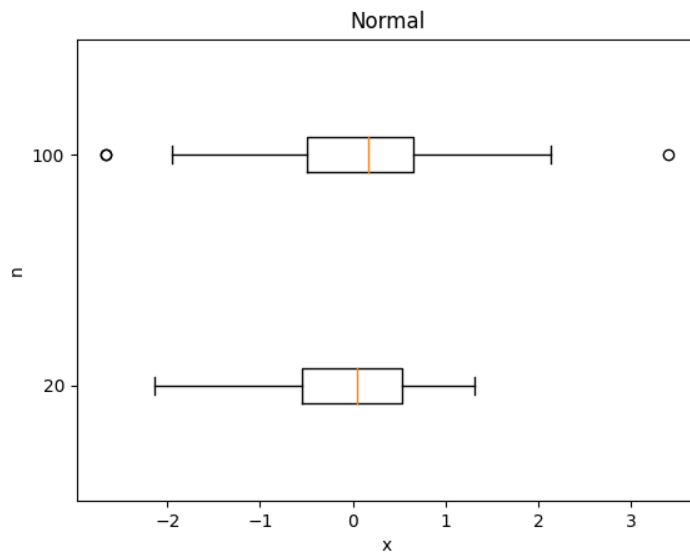


Рис. 6: Боксплот нормального распределения

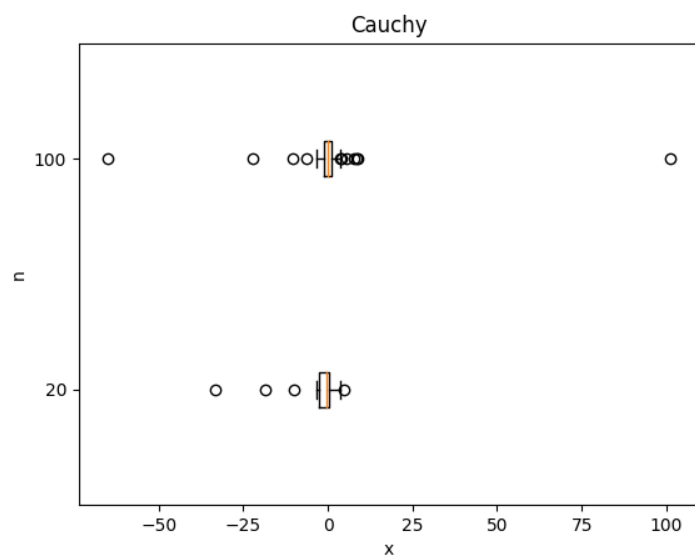


Рис. 7: Боксплот распределения Коши

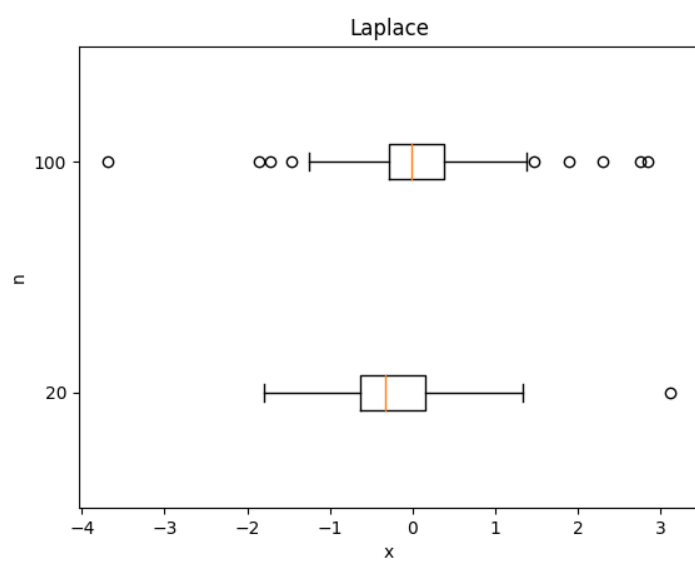


Рис. 8: Боксплот распределения Лапласа

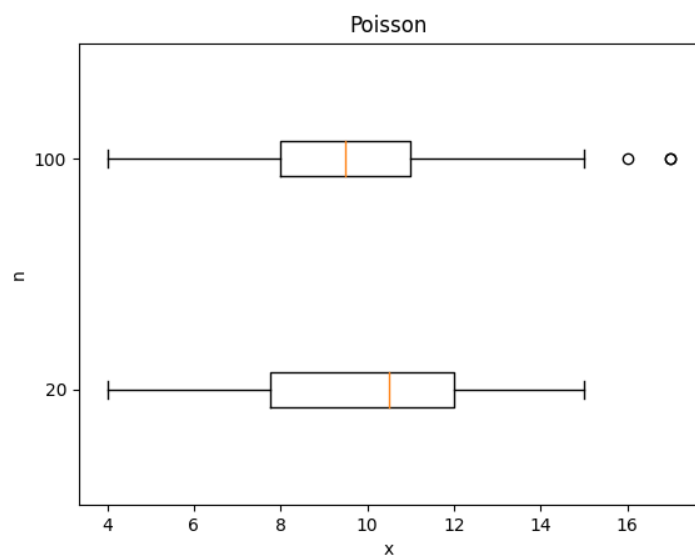


Рис. 9: Боксплот распределения Пуассона

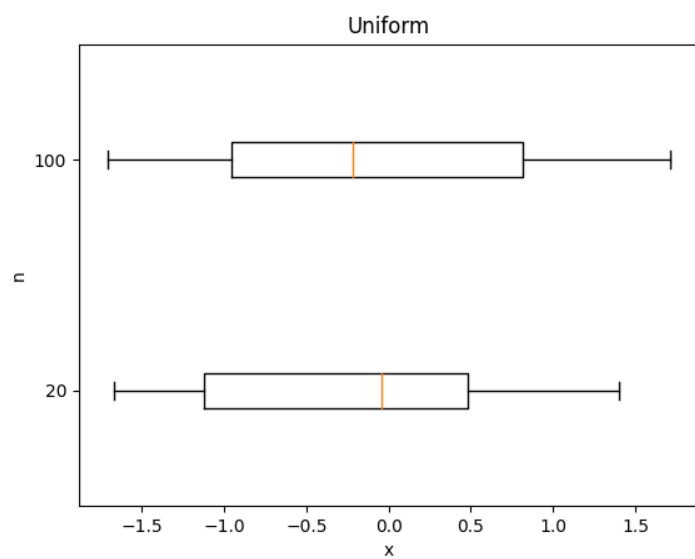


Рис. 10: Боксплот равномерного распределения

4.4 Доля выбросов

Округление осуществлялось следующим образом:

Выборка случайна, поэтому в качестве оценки рассеяния можно взять дисперсию пуассоновского потока:

$$D_n \approx \sqrt{n}$$

$$\text{Доля } p_n = \frac{D_n}{n} = \frac{1}{\sqrt{n}}$$

Для n=20: $p_n = \frac{1}{\sqrt{20}}$ – примерно 0.2 или 20 %

Для n=100 $p_n = 0.1$ или 10 %

Исходя из этого можно решить, сколько знаков оставлять в доле выбросов

Выборка	Доля выбросов
Normal n=20	0.02
Normal n=100	0.01
Cauchy n=20	0.15
Cauchy n=100	0.15
Laplace n=20	0.07
Laplace n=100	0.06
Poisson n=20	0.02
Poisson n=100	0.01
Uniform n=20	0.00
Uniform n=100	0.00

Таблица 8: Доля выбросов

4.5 Теоретическая вероятность выбросов

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

Таблица 9: Теоретическая вероятность выбросов

4.6 Эмпирическая функция распределения

Графики теоретической функции распределения синего цвета, графики эмперической функции распределения – черного.

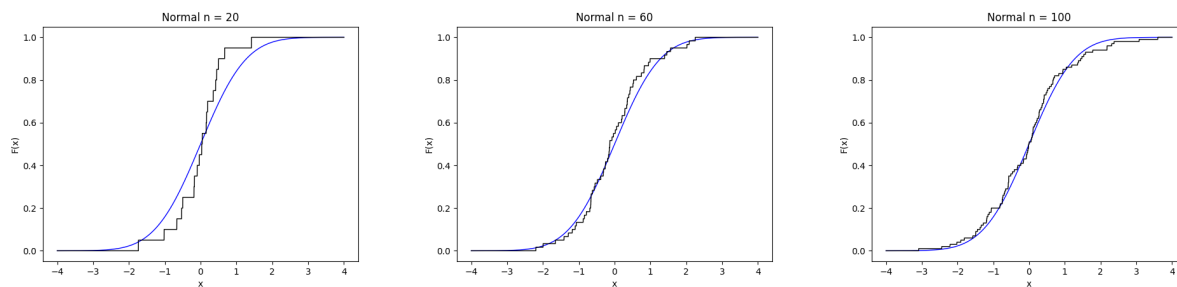


Рис. 11: Функция распределения вероятности нормального p -я

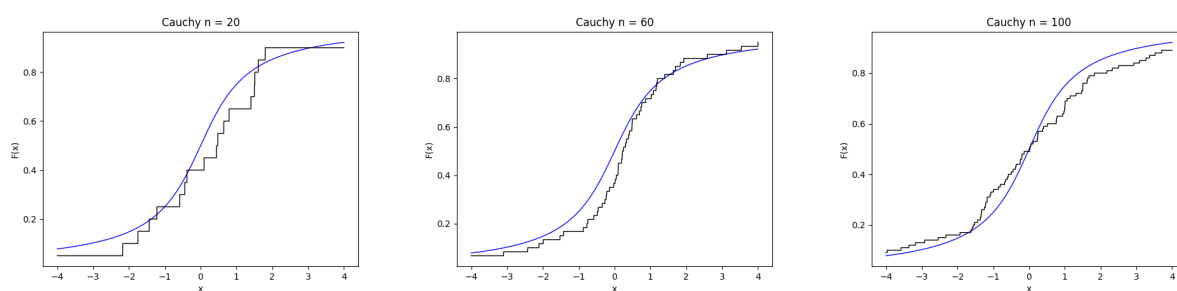


Рис. 12: Функция распределения вероятности p -я Коши

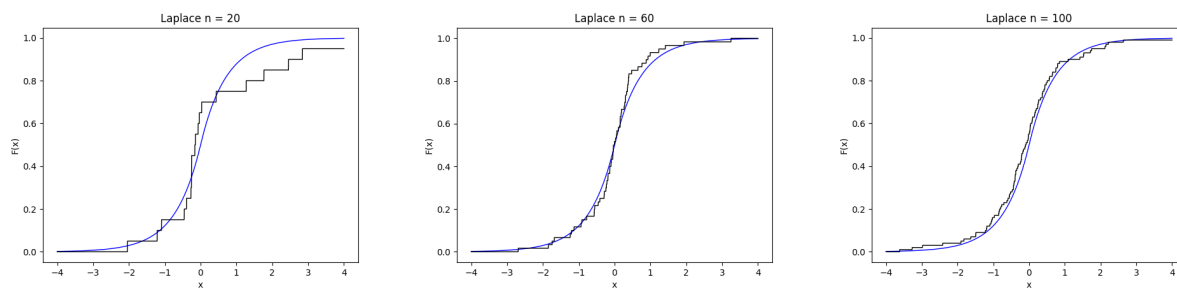


Рис. 13: Функция распределения вероятности p -я Лапласа

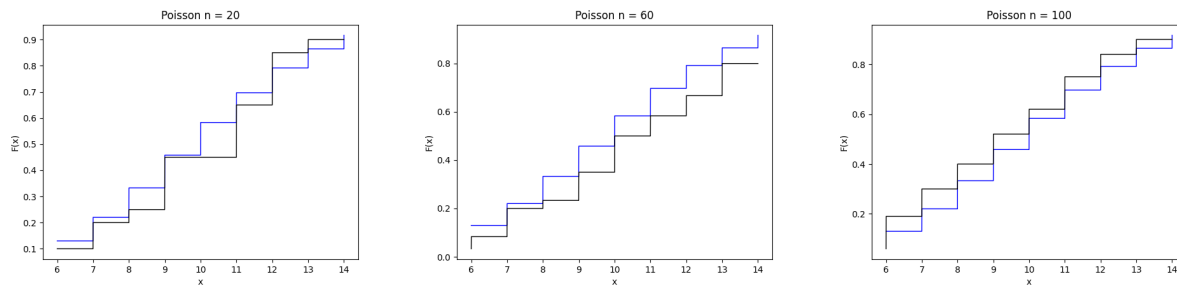


Рис. 14: Функция распределения вероятности р-я Пуассона

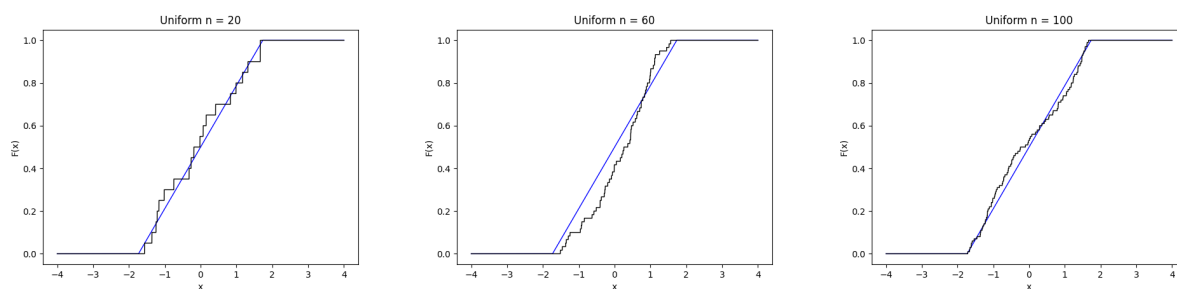


Рис. 15: Функция распределения вероятности равномерного р-я

4.7 Ядерные оценки плотности распределения

Графики плотности распределения красного цвета, графики оценок – черного.

$adjust = 0.5$ соответствует $h = \frac{h_n}{2}$

$adjust = 1$ соответствует $h = h_n$

$adjust = 2$ соответствует $h = 2h_n$

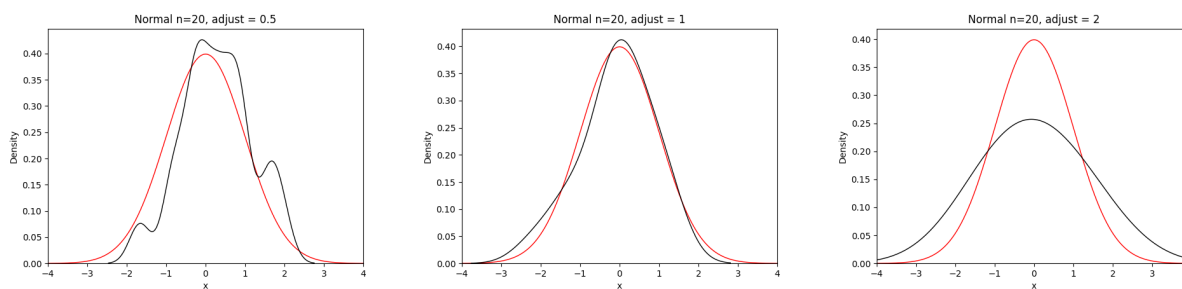


Рис. 16: Нормальное распределение, n=20

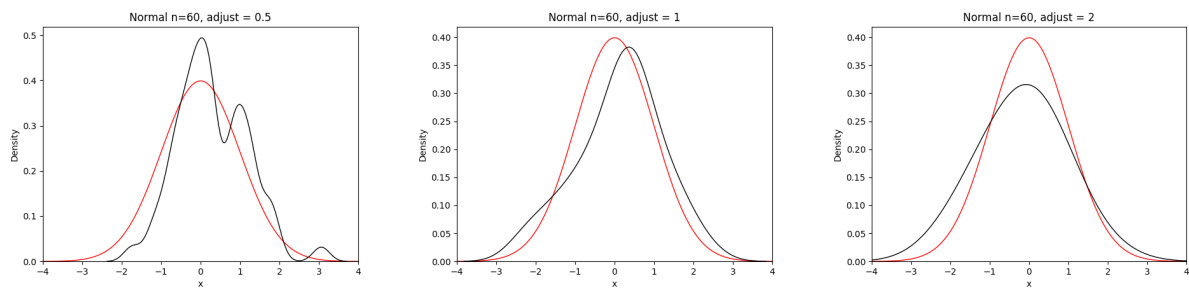


Рис. 17: Нормальное распределение, $n=20$

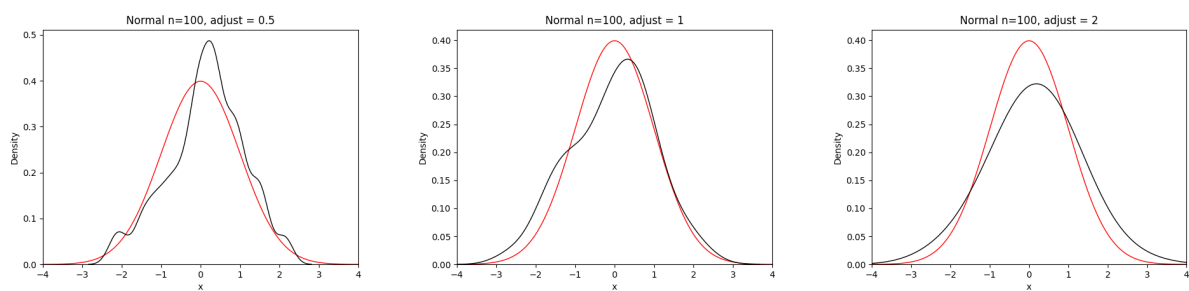


Рис. 18: Нормальное распределение, $n=100$

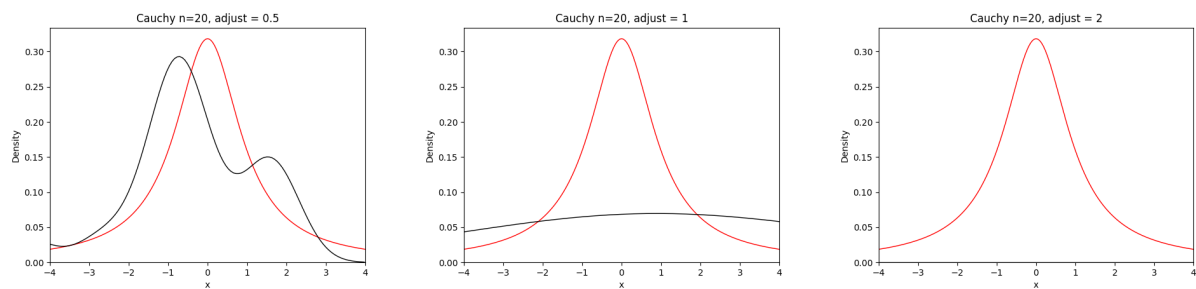


Рис. 19: Распределение Коши, $n=20$

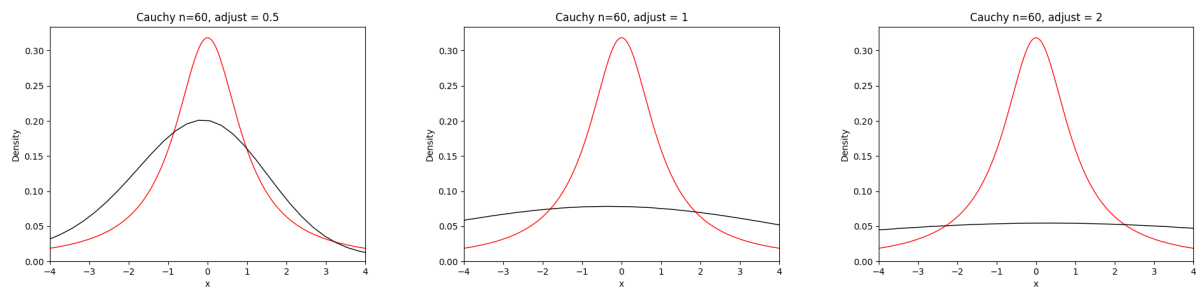


Рис. 20: Распределение Коши, $n=60$

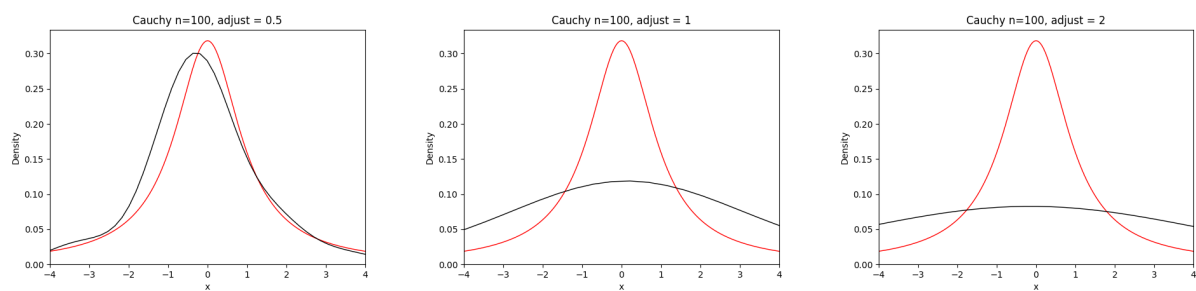


Рис. 21: Распределение Коши, $n=100$

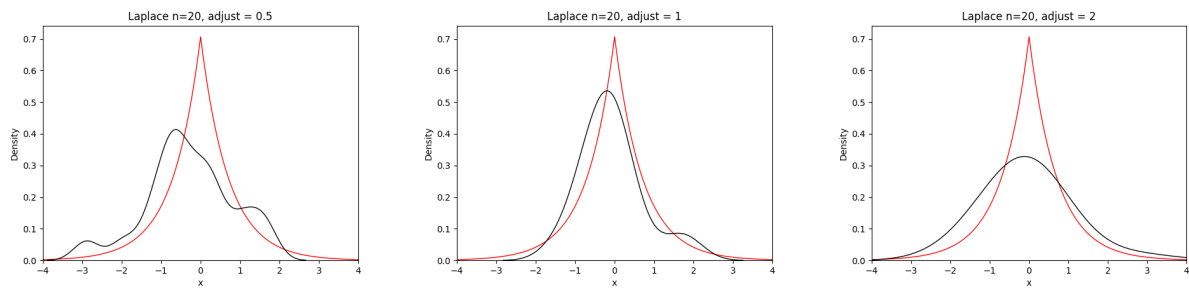


Рис. 22: Распределение Лапласа, $n=20$

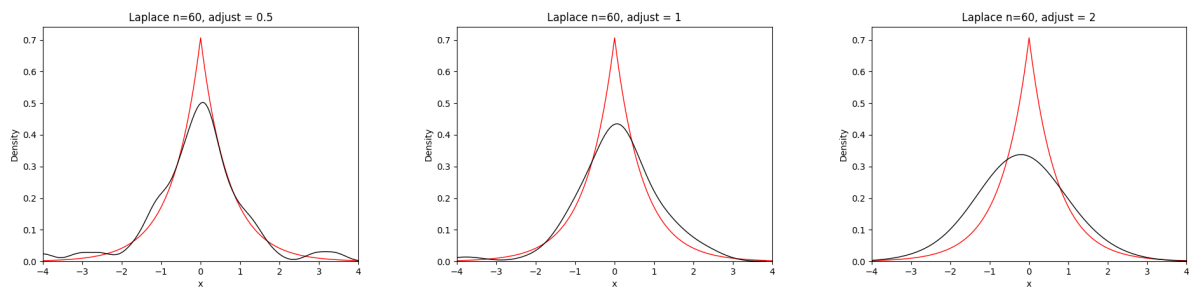


Рис. 23: Распределение Лапласа, $n=60$

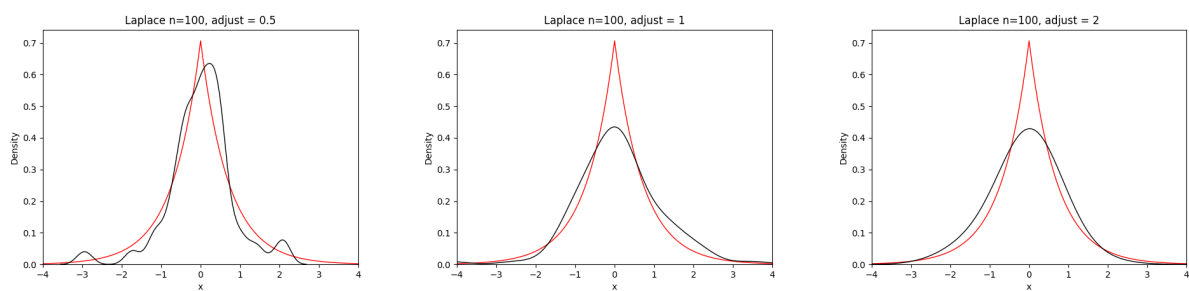


Рис. 24: Распределение Лапласа, $n=100$

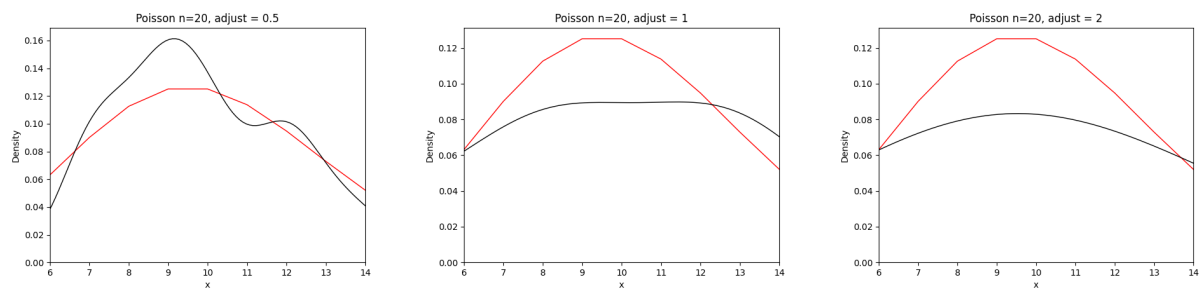


Рис. 25: Распределение Пуассона, $n=20$

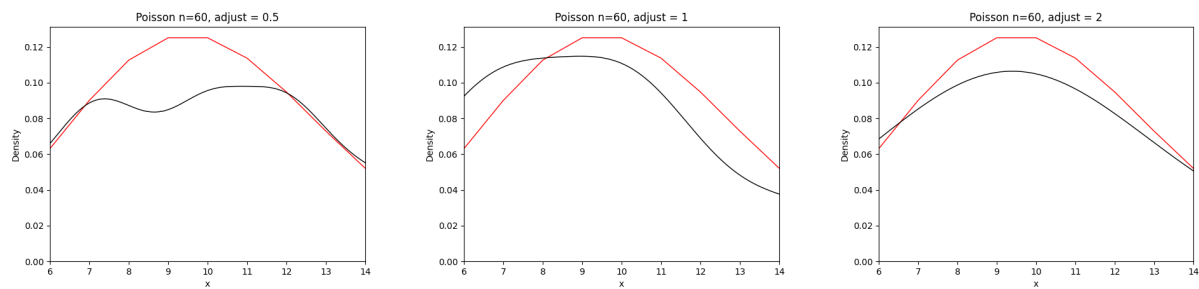


Рис. 26: Распределение Пуассона, $n=60$

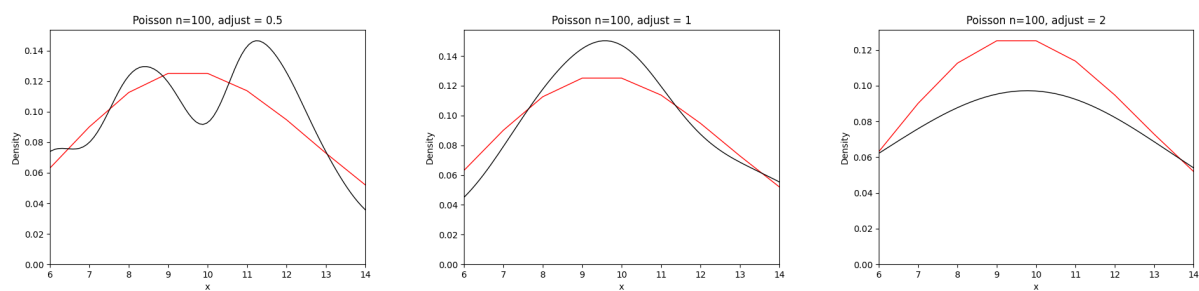


Рис. 27: Распределение Пуассона, $n=100$

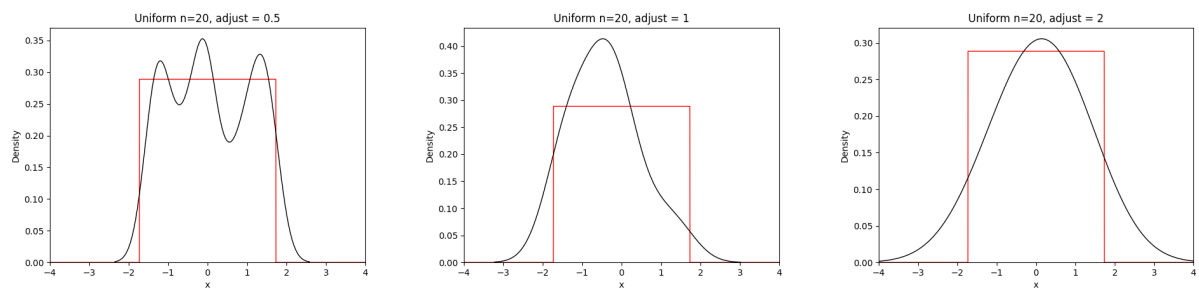


Рис. 28: Равномерное распределение, $n=20$

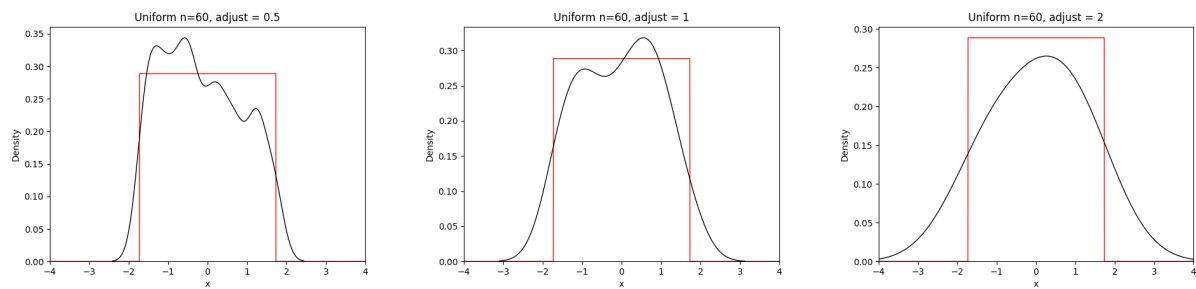


Рис. 29: Равномерное распределение, $n=60$

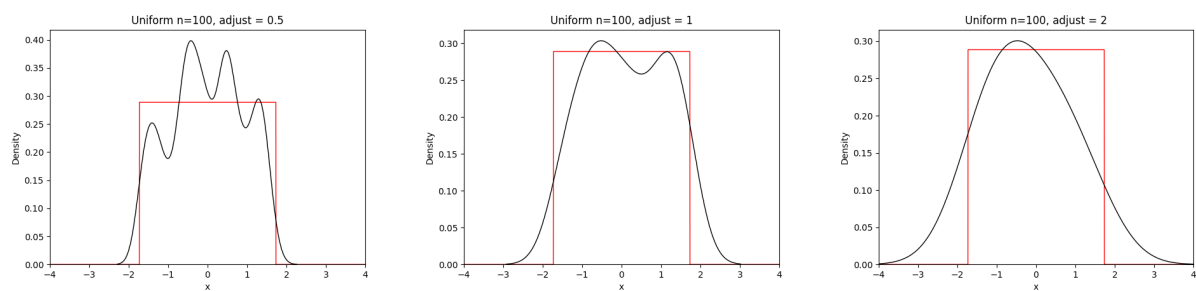


Рис. 30: Равномерное распределение, $n=100$

5 Обсуждение

5.1 Гистограмма и график плотности распределения

Исходя из результатов можно сделать вывод, что чем мощнее выборка, тем ближе ее гистограмма к графику плотности вероятности распределения, по которому данная выборка сгенерирована и тем лучше по ней можно определить характер распределения.

На основе выборок с маленьким размером ($n=10$) определить характер распределения почти невозможно. Например при $n=10$ невозможно отличить гистограммы равномерного распределения и распределения Пуассона.

Также на выборках с маленьким размером чаще наблюдаются всплески гистограмм (особенно у распределений Коши и Лапласа).

5.2 Характеристики положения и рассеяния

Полученные результаты дисперсий характеристик рассеяния распределения Коши являются очень большими при всех размерах выборки. Можно сделать вывод о том, что это является следствием большого количества выбросов, которые можно также наблюдать на гистограммах.

5.3 Боксплот Тьюки

Боксплоты Тьюки позволяют наглядно наблюдать характеристики распределений (например большое количество выбросов распределения Коши и отсутствия выбросов у равномерного распределения).

5.4 Доля выбросов

Исходя из полученных результатов мы получили что для всех распределений (за исключением нормального распределения с размером выборки 20) доли выбросов, полученные практически и теоретически приближенно равны.

Экспериментально и теоретически подтверждено большое количество выбросов распределения Коши, замеченное при выполнении предыдущих заданий и отсутствие выбросов у равномерного распределения.

5.5 Эмпирическая функция распределения

На основе полученных результатов можно сделать вывод о том, что чем больше размер выборки, тем лучше эмпирическая функция распределения приближает теоретическую функцию распределения. Наибольшие отклонения при $n=100$ наблюдаются у распределения Пуассона.

5.6 Ядерные оценки плотности распределения

На основе полученных результатов можно сделать вывод, что при увеличении размера выборки при всех h ядерная оценка начинает лучше приближать плотность вероятности.

Также можно сделать вывод, что параметр h в ядерной оценке лучше подбирать в зависимости от характера распределения. Параметр $h = \frac{h_n}{2}$ лучше брать для распределений Коши и Лапласа. $h = h_n$ лучше всего подходит для нормального распределения, а $h = 2h_n$ - для распределений Пуассона и равномерного.

Также можно отметить, что при увеличении параметра h ядерной оценки уменьшается количество изменений знака производной. При $h = 2h_n$ функция становится унимодальной.

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.