

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

Отчёт по лабораторной работе №1-4
по дисциплине Математическая статистика

Выполнил студент:
Курова Анна Николаевна
группа: 3630102/70401
Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Содержание

1	Постановка задачи	5
2	Теория	5
2.1	Плотности распределения	5
2.2	Гистограмма	6
2.2.1	Определение	6
2.2.2	Графическое описание	6
2.2.3	Использование	6
2.3	Вариационный ряд	6
2.4	Выборочные числовые характеристики	7
2.4.1	Характеристики положения	7
2.4.2	Характеристики рассеяния	7
2.5	Боксплот Тьюки	8
2.5.1	Определение	8
2.5.2	Описание	8
2.5.3	Построение	8
2.6	Теоретическая вероятность выбросов	8
2.7	Эмпирическая функция распределения	9
2.7.1	Статистический ряд	9
2.7.2	Определение	9
2.7.3	Описание	9
2.8	Оценки плотности вероятности	9
2.8.1	Определение	9
2.8.2	Ядерные оценки	10
3	Реализация	10
4	Результаты	11
4.1	График плотности и гистограмма распределений	11
4.2	Характеристики положения и рассеяния	13
4.3	Боксплот Тьюки	15
4.4	Доля выбросов	17
4.5	Теоретическая вероятность выбросов	17
4.6	Эмпирическая функция распределения	18
4.7	Ядерные оценки плотности распределения	20
5	Обсуждения	29
5.1	Графики плотностей и гистограммы	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
6	Приложения	31

Список иллюстраций

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

Список таблиц

1	Статистический ряд	9
2	Таблица распределения	9
3	Нормальное распределение	13
4	Распределение Коши	13
5	Распределение Лапласа	14
6	Распределение Пуассона	14
7	Равномерное распределение	15
8	Доля выбросов	17
9	Теоретическая вероятность выбросов	18

1 Постановка задачи

1. Для 5 видов различных распределений сгенерировать выборки, состоящие из 10, 50 и 1000 элементов.

На одном графике построить плотность и гистограмму распределения для каждой из выборок

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:

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

Представить полученные данные в виде таблиц

3. Для каждого распределения изобразить boxplot и вычислить долю выбросов. Сравнить практическую и теоретическую долю выбросов.
4. Вычислить эмпирические функции распределения и ядерные оценки плотности распределения

2 Теория

2.1 Плотности распределения

Формулы вычисления плотности для каждого распределения:

1. Нормальное распределение:

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

2. Распределение Коши:

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

3. Распределение Лапласа:

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

4. Распределение Пуассона:

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

5. Равномерное распределение:

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

2.2 Гистограмма

2.2.1 Определение

Гистограмма - способ графического представления табличных данных. Количественные соотношения некоторого показателя представлены в виде прямоугольников, площади которых пропорциональны. Чаще всего для удобства восприятия ширину прямоугольников берут одинаковую, при этом их высота определяет соотношения отображаемого параметра.

2.2.2 Графическое описание

Гистограмма строится следующим образом. Сначала множество значений, которое может принимать элемент выборки, разбивается на несколько интервалов (bins). 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.

2.2.3 Использование

Гистограммы применяются в основном для визуализации данных на начальном этапе статистической обработки. Построение гистограмм используется для получения эмпирической оценки плотности распределения случайной величины. Для построения гистограммы наблюдаемый диапазон изменения случайной величины разбивается на несколько интервалов и подсчитывается доля от всех измерений, попавшая в каждый из интервалов. Величина каждой доли, отнесенная к величине интервала, принимается в качестве оценки значения плотности распределения на соответствующем интервале

2.3 Вариационный ряд

Вариационным рядом называется последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются [2,

с. 409].

Запись вариационного ряда: $x_{(1)}, x_{(2)}, \dots, x_{(n)}$

Элементы вариационного ряда $x_{(i)}$ ($i = 1, 2, \dots, n$) называются порядковыми статистиками.

2.4 Выборочные числовые характеристики

С помощью выборки образуются её числовые характеристики. Это числовые характеристики дискретной случайной величины X^* , принимающей выборочные значения x_1, x_2, \dots, x_n [5]

2.4.1 Характеристики положения

1. Выборочное среднее

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

2. Выборочная медиана

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

3. Полусумма экстремальных выборочных элементов

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

4. Полусумма квартилей

Выборочный квартиль z_p порядка p определяется формулой

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

Тогда полусумма квартилей

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

5. Усеченное среднее

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

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 Определение

Боксплот (англ. box plot) — график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей.

2.5.2 Описание

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Несколько таких ящиков можно нарисовать бок о бок, чтобы визуально сравнивать одно распределение с другим; их можно располагать как горизонтально, так и вертикально. Расстояния между различными частями ящика позволяют определить степень разброса (дисперсии) и асимметрии данных и выявить выбросы

2.5.3 Построение

Границами ящика служат первый и третий квартили, линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выбросов). Длину «усов» определяют разность первого квартиля и полутора межквартильных расстояний и сумма третьего квартиля и полутора межквартильных расстояний. Формула имеет вид:

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

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 ромбов

2.6 Теоретическая вероятность выбросов

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

$$\begin{cases} x < X_1 \\ x > X_2 \end{cases}$$

Теоретическая вероятность выбросов для непрерывных распределений

$$P_B = P(x < X_1) + P(x > X_2) = F(X_1) - (1 - F(X_2)) \quad (16)$$

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

Теоретическая вероятность выбросов для дискретных распределений

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

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

Таблица 1: Статистический ряд

z	z_1	z_2	...	z_k
n	n_1	n_2	...	n_k

2.7.2 Определение

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

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

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

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

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

Таблица 2: Таблица распределения

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

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:

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

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

$$h_n \xrightarrow{n \rightarrow \infty} \infty; \quad \frac{h_n}{n^{-1}} \xrightarrow{n \rightarrow \infty} \infty \quad (23)$$

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 [6, с. 38]

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

Правило Сильвермана [6, с. 44]

$$h_n = 1.06 \hat{\sigma} n^{-\frac{1}{5}} \quad (25)$$

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

3 Реализация

Для реализации использовался высокоуровневый язык python.

Для наглядного представления графиков использовались библиотеки matplotlib и seaborn

Так же использовалась библиотека math

Для реализации нормального распределения использовался метод Бокса-Мюллера.

Для реализации распределения Коши модифицированный метод Бокса-Мюллера, только в данном случае мы получали тангенс и возвращали результат деления двух нормальных случайных величин

Для распределения Пуассона использовался метод на основе инверсии путем последовательного поиска (основан на алгоритме Кнута)

Для равномерного распределения одна из простейших схем: $x_m = ax_{m-1} + b \pmod{m}$

4 Результаты

4.1 График плотности и гистограмма распределений

Красным обозначен график плотности распределения. Синяя линия является вспомогательной

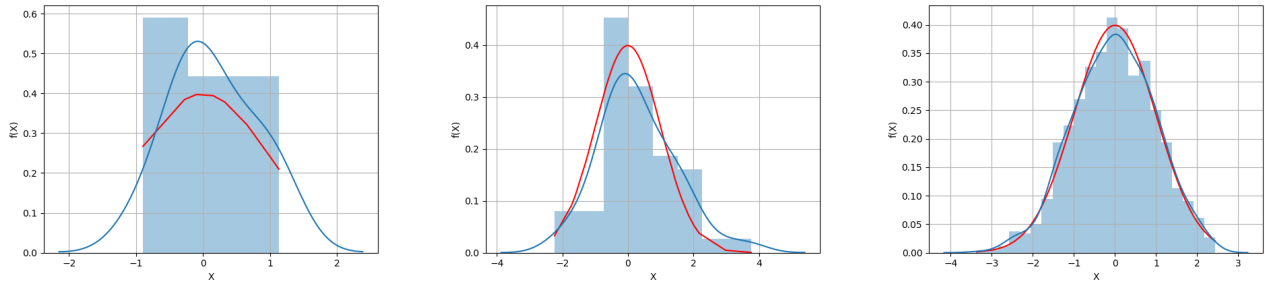


Рис. 1: Нормальное распределение

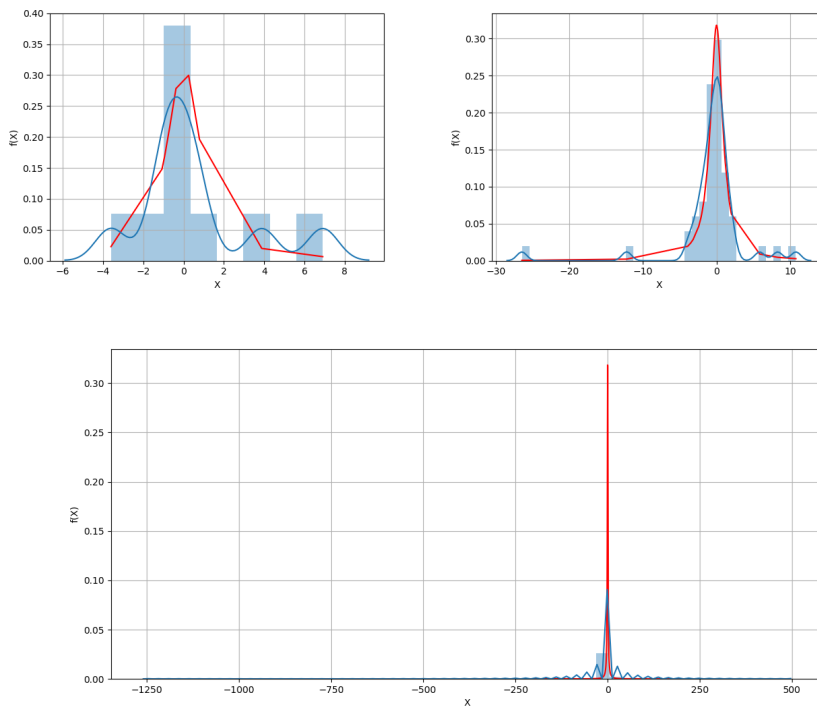


Рис. 2: Распределение Коши

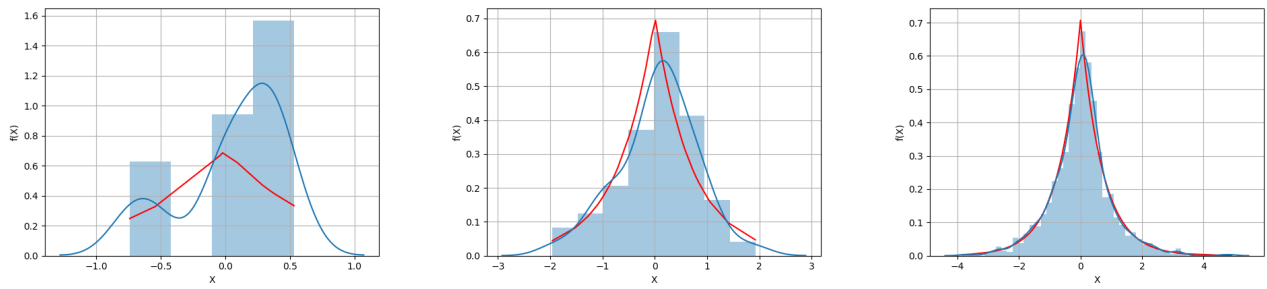


Рис. 3: Распределение Лапласа

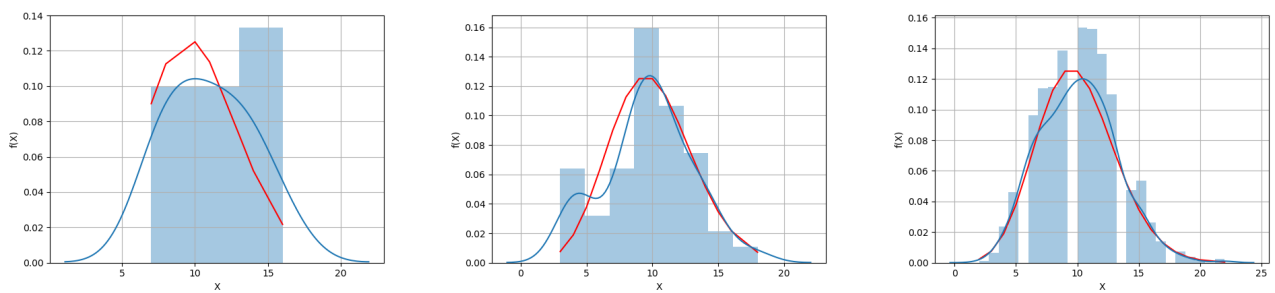


Рис. 4: Распределение Пуассона

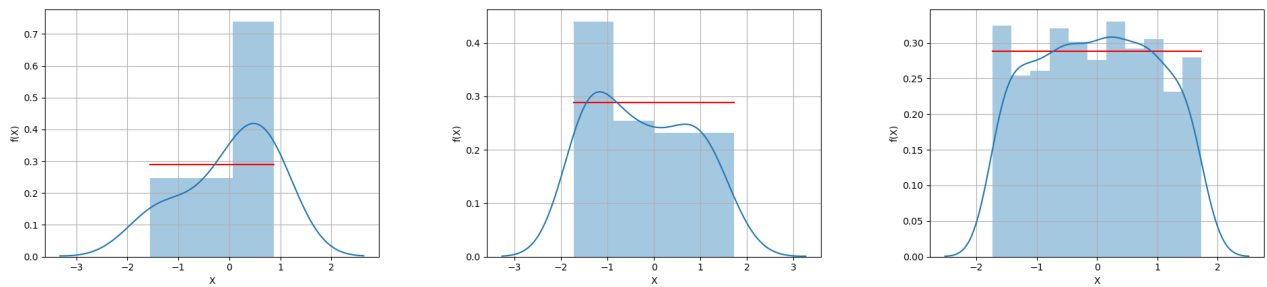


Рис. 5: Равномерное распределение

4.2 Характеристики положения и рассеяния

В скобках указаны ссылки на формулы.

Таблица 3: Нормальное распределение

normal n = 10					
	\bar{x} (8)	med x (9)	z_R (10)	z_Q (12)	z_{tr} (13)
$E(z)$ (1)	-0.0	-0.15	-0.0	0.2	-0.0
$D(z)$ (2)	0.1	0.15	0.178	0.129	0.08
normal n = 100					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.0	-0.0	-0.0	0.01	0.0
$D(z)$	0.01	0.01	0.09	0.012	0.011
normal n = 1000					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.0	-0.001	-0.0	0.001	0.0
$D(z)$	0.001	0.001	0.061	0.001	0.001

Таблица 4: Распределение Коши

cauchy n = 10					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	1.526	-0.1	7.456	1.247	0.0
$D(z)$	3384	0.393	84407	2.941	0.240
cauchy n = 100					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.21	-0.01	-8.964	0.02	0.01
$D(z)$	4221	0.021	10517251	0.048	0.021
cauchy n = 1000					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	0.372	0.0	182.17	0.0	0.0
$D(z)$	571042	0.002	1.42e+11	0.005	0.002

Таблица 5: Распределение Лапласа

laplace n = 10					
	\bar{x}	<i>med</i> x	z_R	z_Q	z_{tr}
$E(z)$	-0.0	-0.08	0.0	0.2	-0.0
$D(z)$	0.1	0.082	0.426	0.121	0.045
laplace n = 100					
	\bar{x}	<i>med</i> x	z_R	z_Q	z_{tr}
$E(z)$	-0.0	-0.009	-0.0	0.01	0.009
$D(z)$	0.01	0.005	0.385	0.01	0.005
laplace n = 1000					
	\bar{x}	<i>med</i> x	z_R	z_Q	z_{tr}
$E(z)$	0.0	0.0004	0.03	0.001	0.001
$D(z)$	0.001	0.0005	0.397	0.001	0.0006

Таблица 6: Распределение Пуассона

poisson n = 10					
	\bar{x}	<i>med</i> x	z_R	z_Q	z_{tr}
$E(z)$	9.96	9.424	10.284	10.921	7.86
$D(z)$	1.03	1.57	1.84	1.42	0.81
poisson n = 100					
	\bar{x}	<i>med</i> x	z_R	z_Q	z_{tr}
$E(z)$	9.992	9.791	10.908	9.961	9.691
$D(z)$	0.097	0.213	0.947	0.144	0.111
poisson n = 1000					
	\bar{x}	<i>med</i> x	z_R	z_Q	z_{tr}
$E(z)$	9.999	9.997	11.7035	9.9925	9.842
$D(z)$	0.009	0.006	0.62	0.002	0.01

Таблица 7: Равномерное распределение

uniform n = 10					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.004	-0.14	0.0	0.33	0.0
$D(z)$	0.093	0.23	0.046	0.12	0.117
uniform n = 100					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.0	-0.02	-0.0005	0.01	0.01
$D(z)$	0.01	0.028	0.0005	0.014	0.019
uniform n = 1000					
	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.0	-0.001	-0.0001	0.002	0.002
$D(z)$	0.001	0.003	0.0007	0.001	0.002

4.3 Боксплот Тьюки

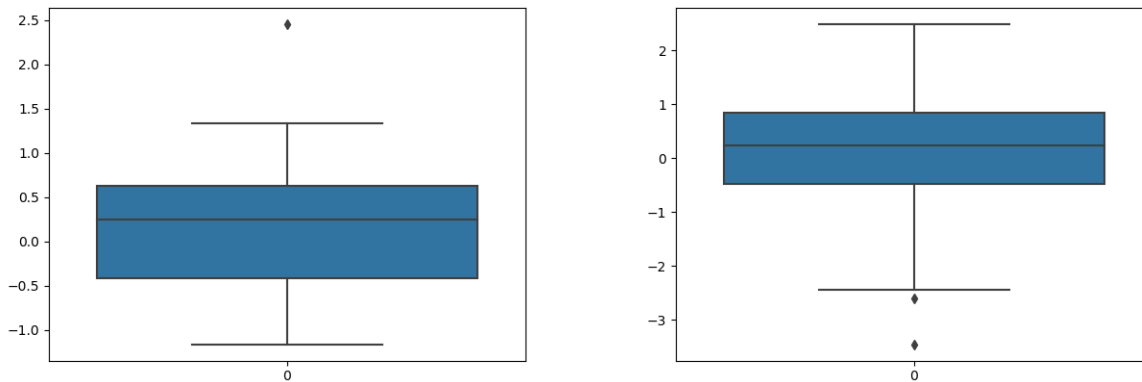


Рис. 6: Нормальное распределение

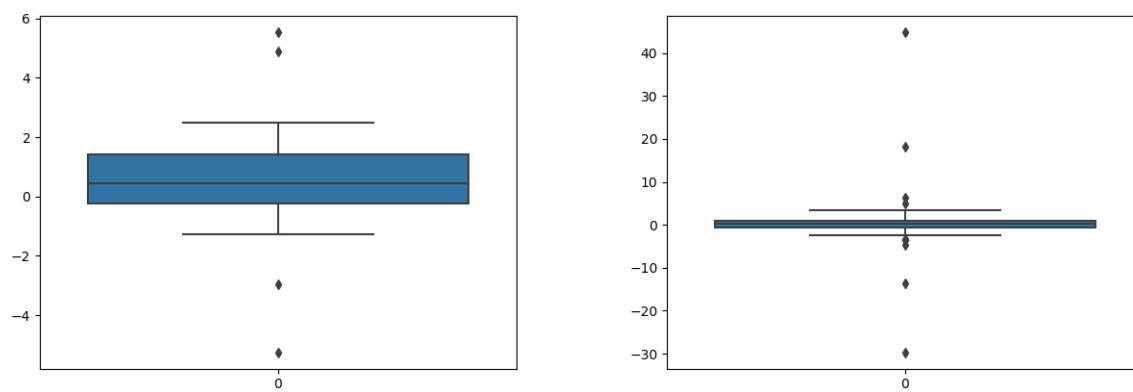


Рис. 7: Распределение Коши

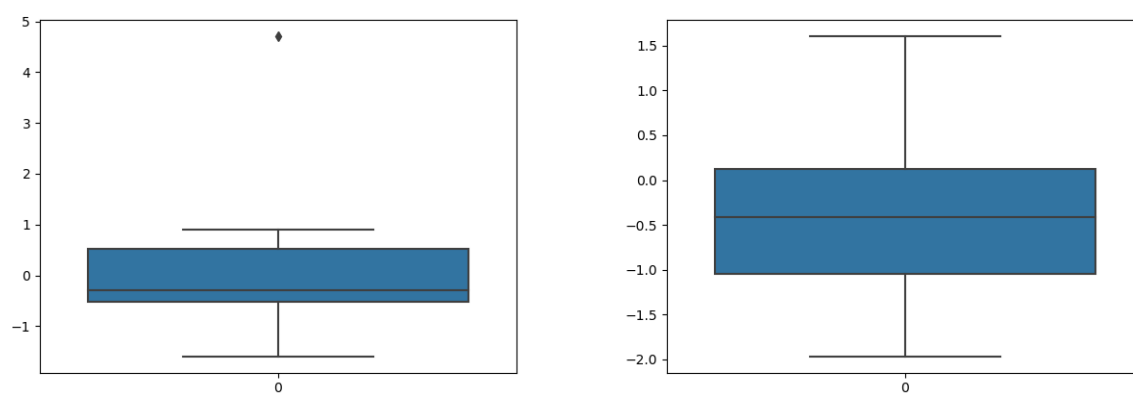


Рис. 8: Распределение Лапласа

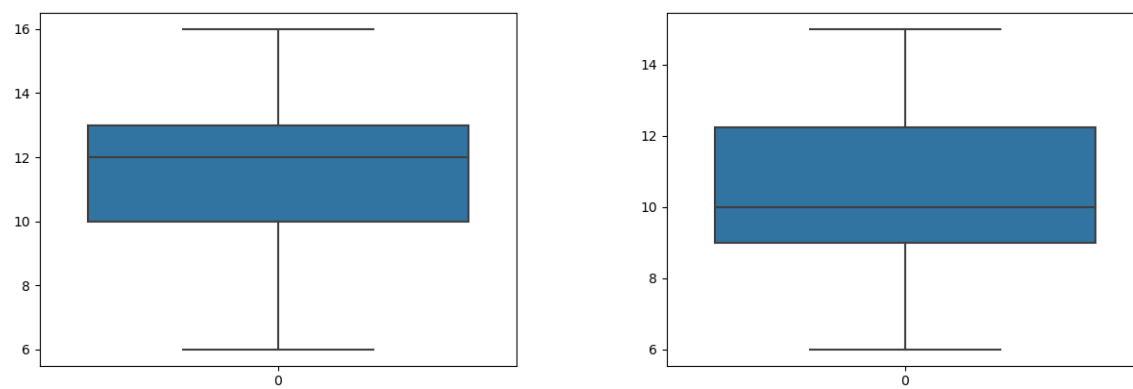


Рис. 9: Распределение Пуассона

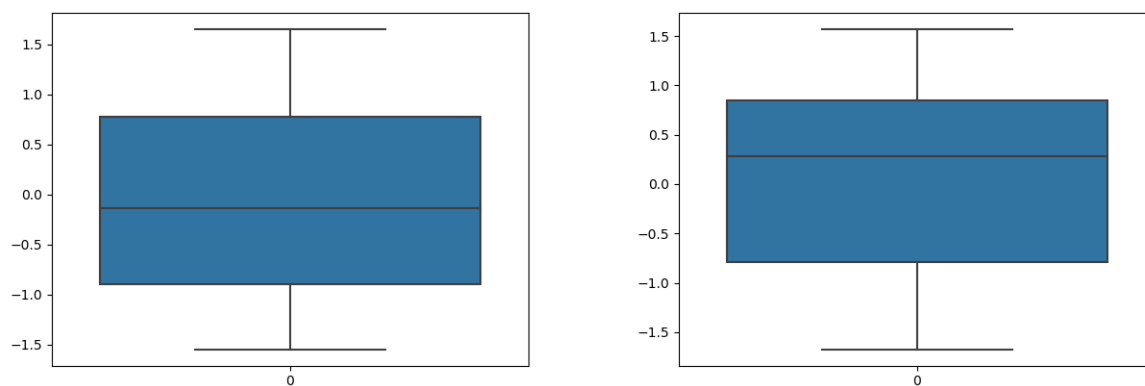


Рис. 10: Равномерное распределение

4.4 Доля выбросов

Выборка	Доля выбросов
normal n = 20	0.01
normal n = 100	0.0
cauchy n = 20	0.2
cauchy n = 100	0.11
laplace n = 20	0.05
laplace n = 100	0.04
poisson n = 20	0.1
poisson n = 100	0.04
uniform n = 20	0.0
uniform n = 20	0.0

Таблица 8: Доля выбросов

4.5 Теоретическая вероятность выбросов

В скобках указаны номера формул

Распределение	Q_1	Q_3	X_1 (15)	X_2 (15)	P (16), (17)
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	0.0

Таблица 9: Теоретическая вероятность выбросов

4.6 Эмпирическая функция распределения

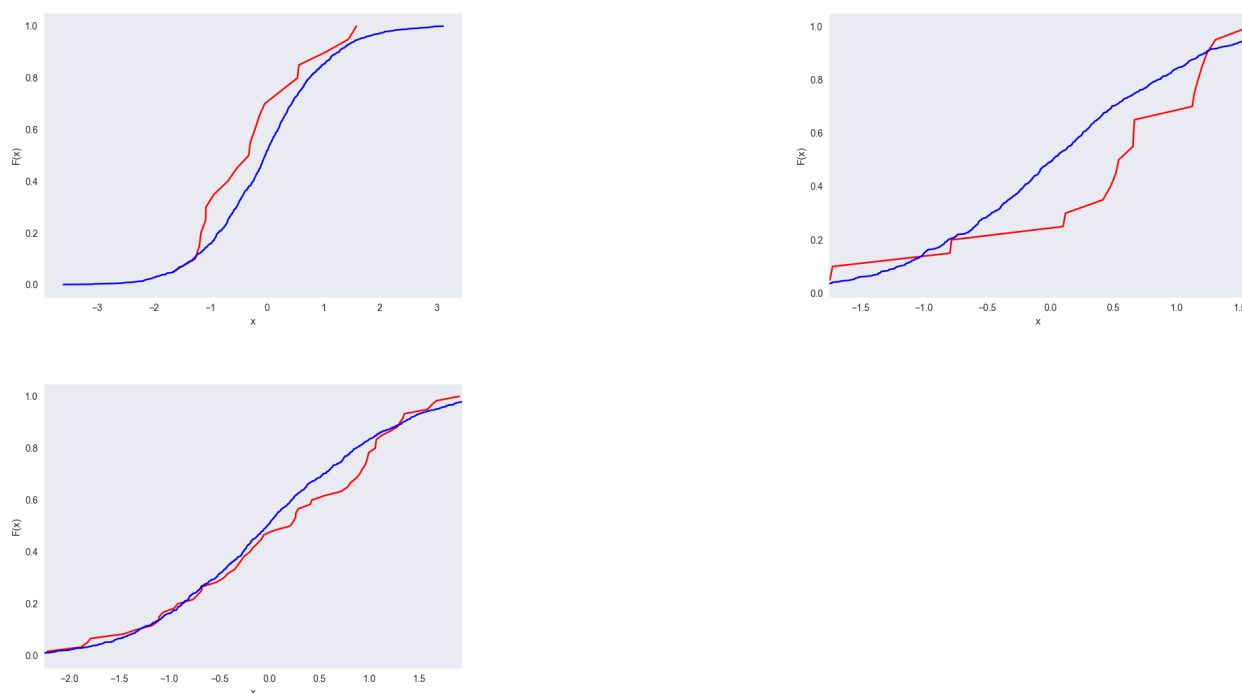


Рис. 11: Нормальное распределение

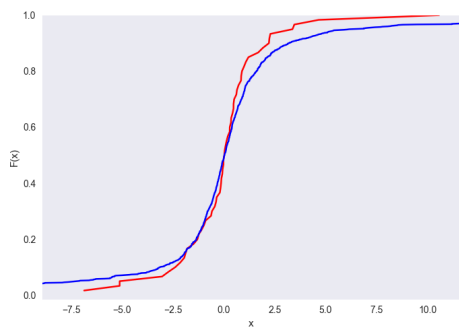
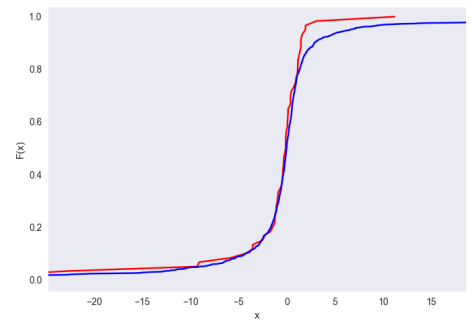
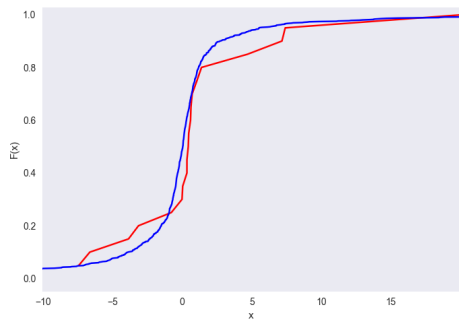


Рис. 12: Распределение Коши

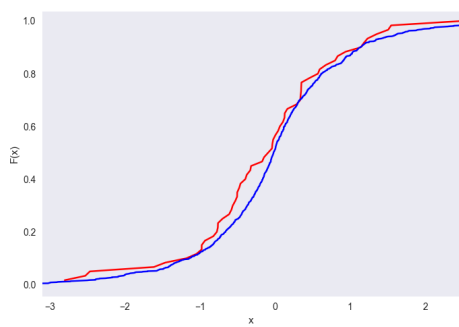
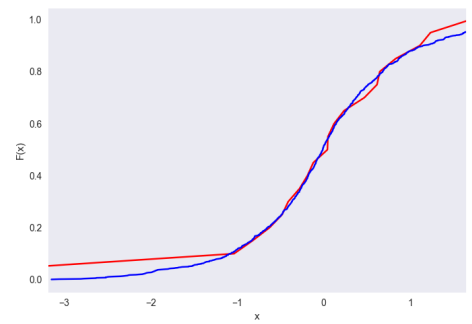
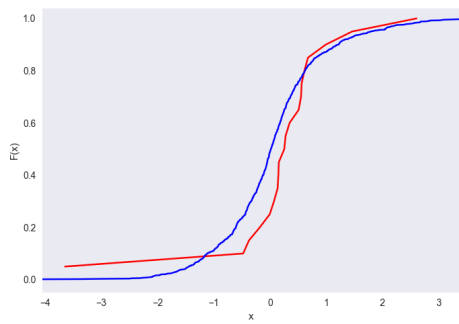


Рис. 13: Распределение Лапласа

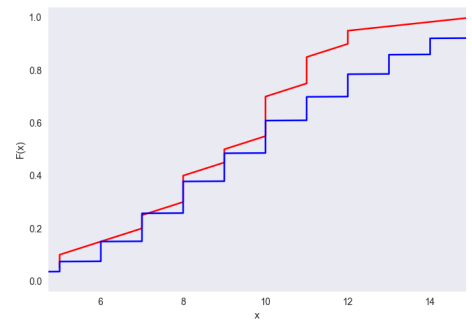
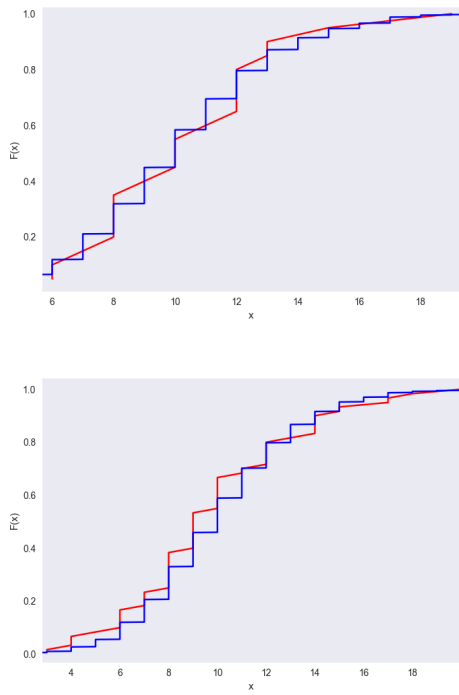


Рис. 14: Распределение Пуассона

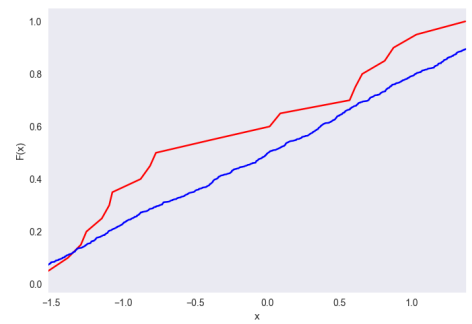
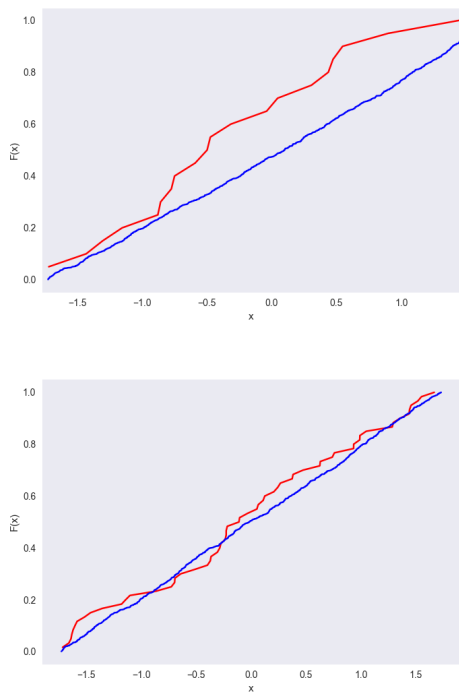


Рис. 15: Равномерное распределение

4.7 Ядерные оценки плотности распределения

На первой изображении в каждом распределении $h = 0.5h_n$, на втором $h = h_n$, на третьем $h = 2h_n$

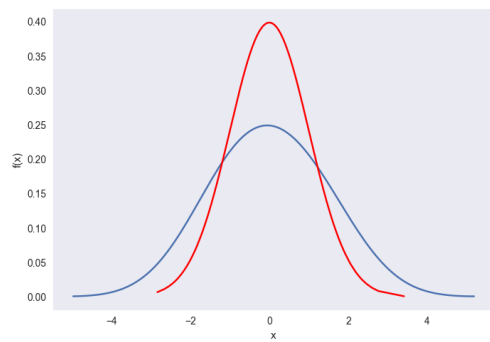
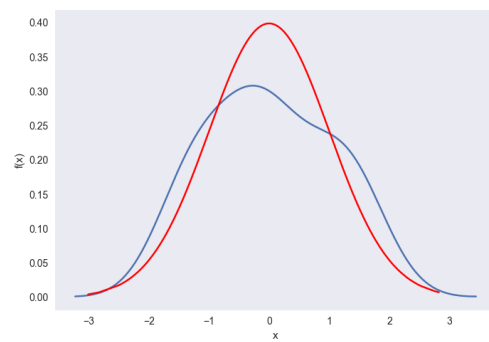
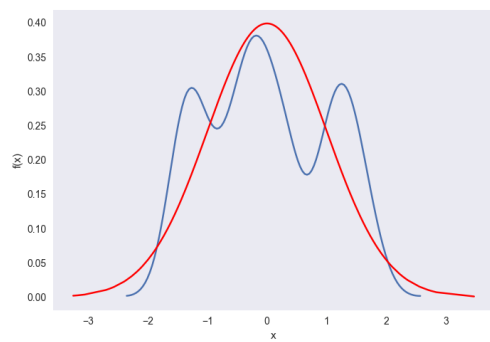


Рис. 16: Нормальное распределение $n = 20$

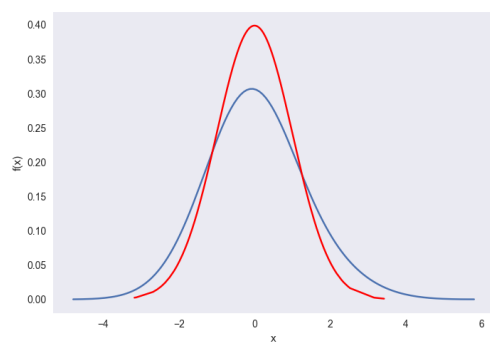
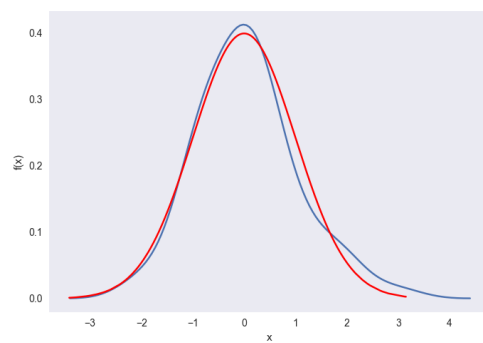
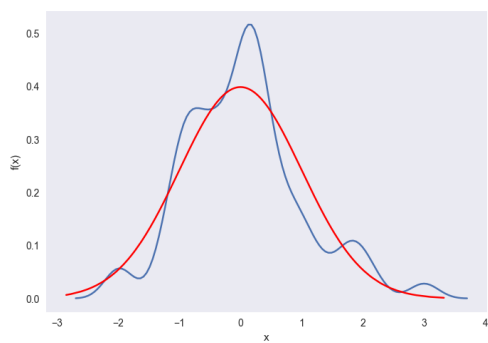


Рис. 17: Нормальное распределение $n = 60$

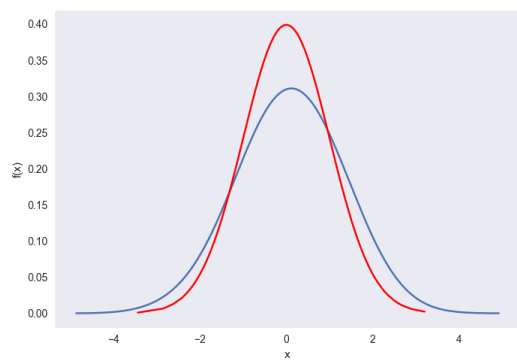
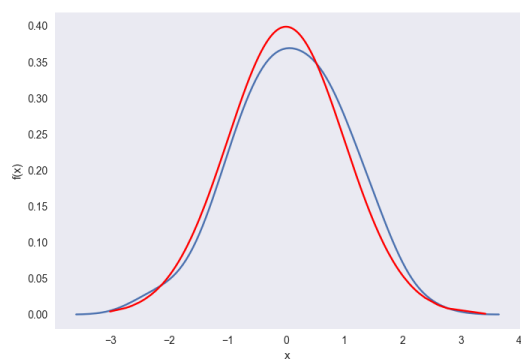
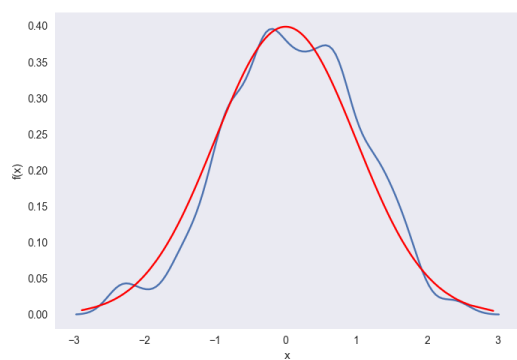


Рис. 18: Нормальное распределение $n = 100$

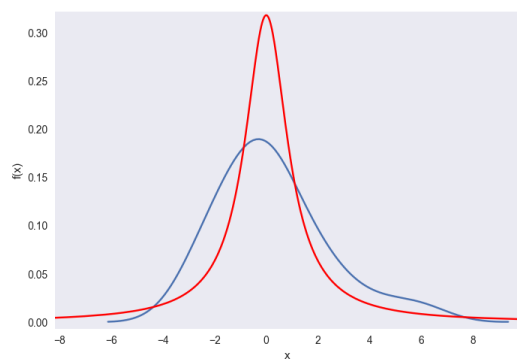
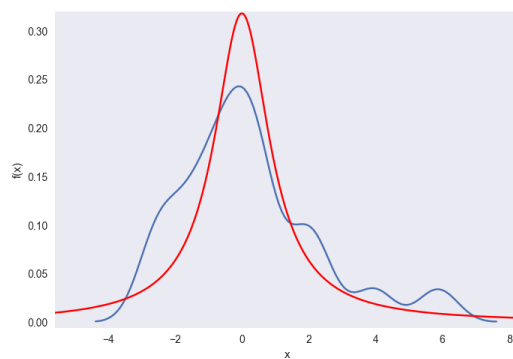
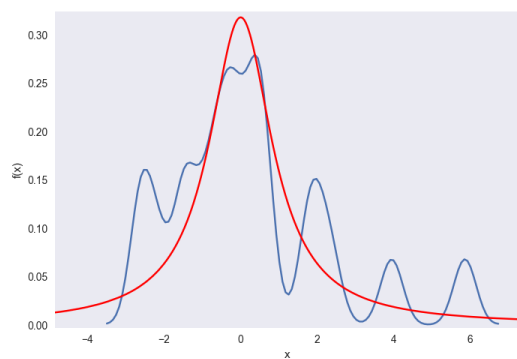


Рис. 19: Распределение Коши $n = 20$

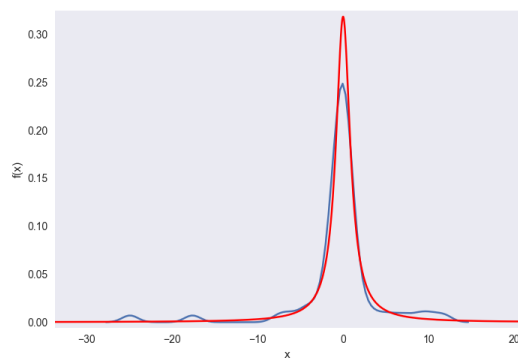
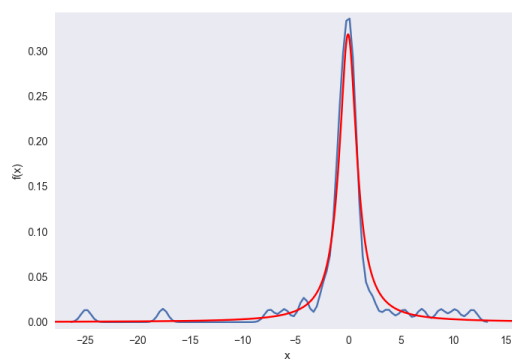
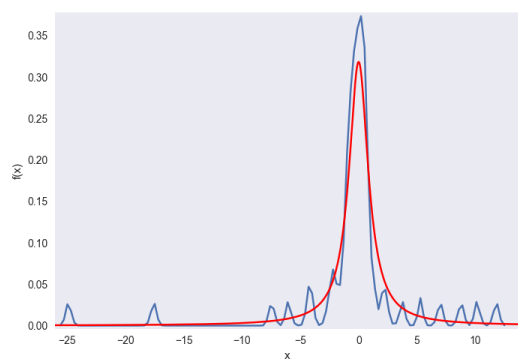


Рис. 20: Распределение Коши $n = 60$

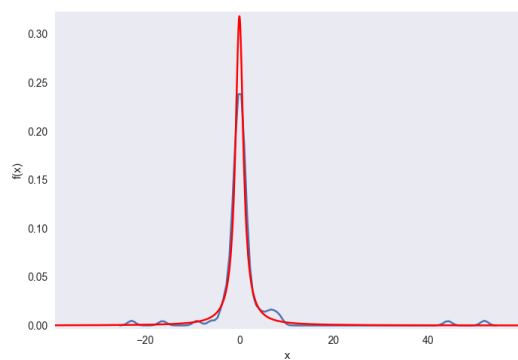
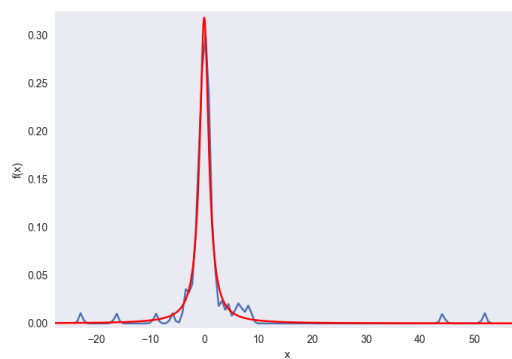
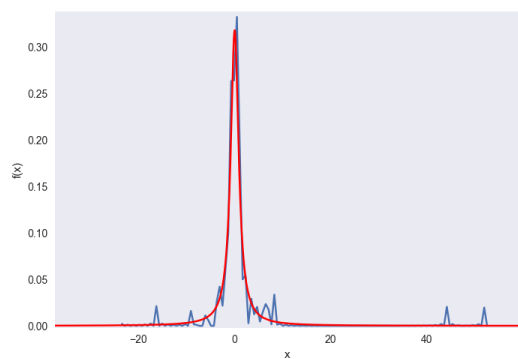


Рис. 21: Распределение Коши $n = 100$

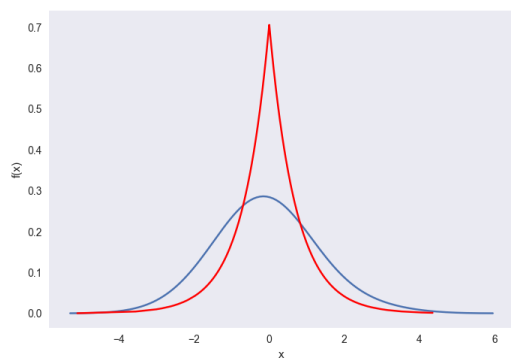
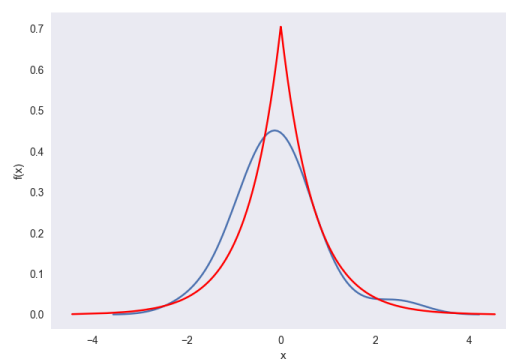
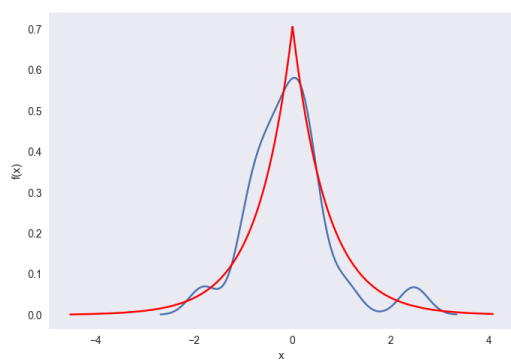


Рис. 22: Распределение Лапласа $n = 20$

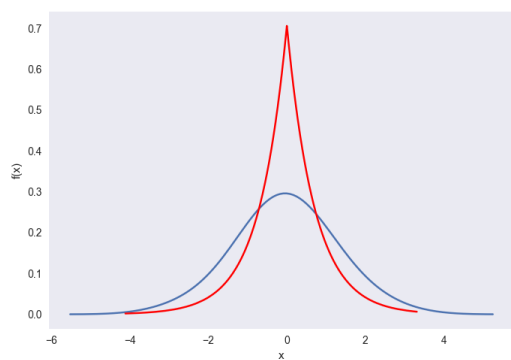
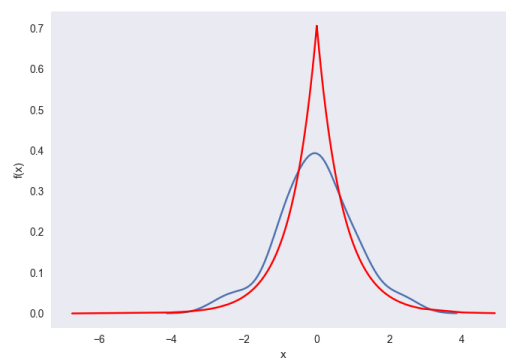
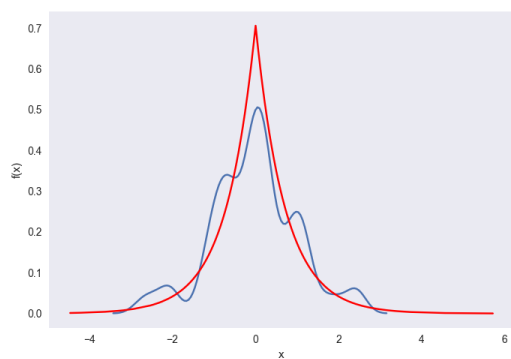


Рис. 23: Распределение Лапласа $n = 60$

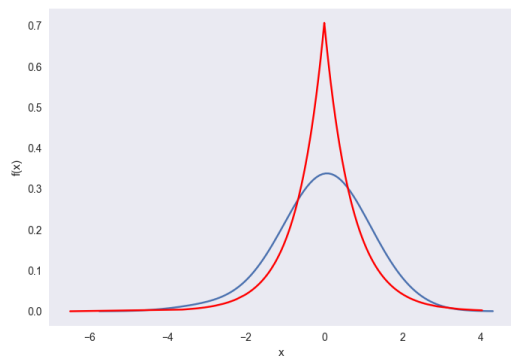
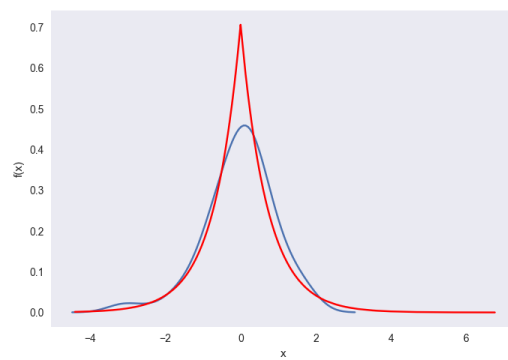
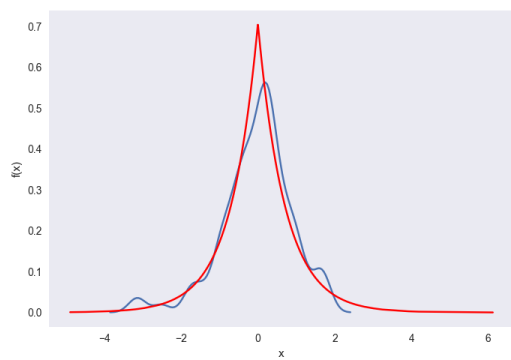


Рис. 24: Распределение Лапласа $n = 100$

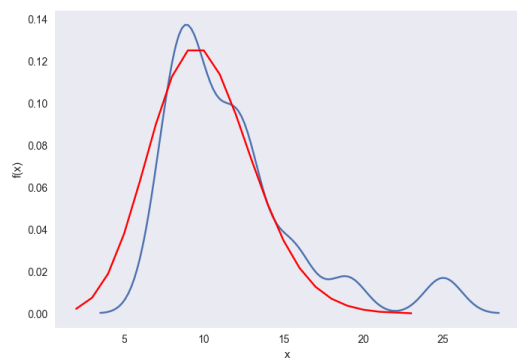
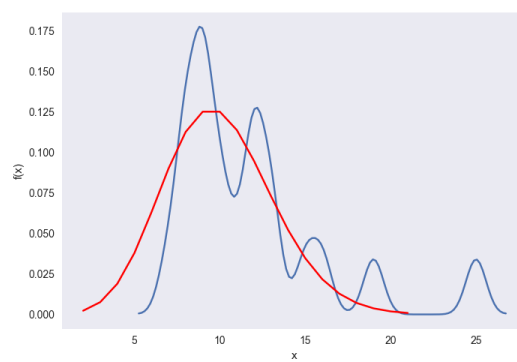
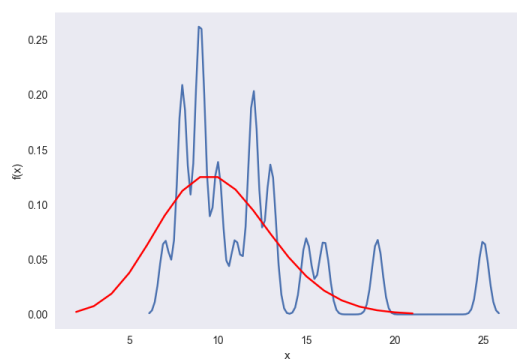


Рис. 25: Распределение Пуассона $n = 20$

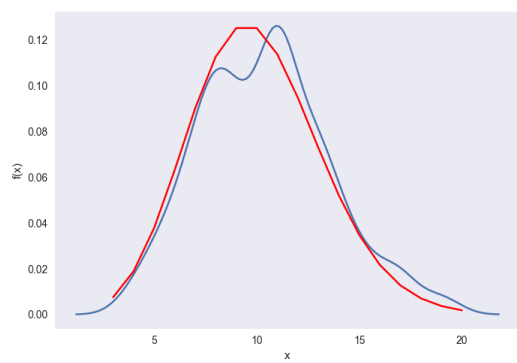
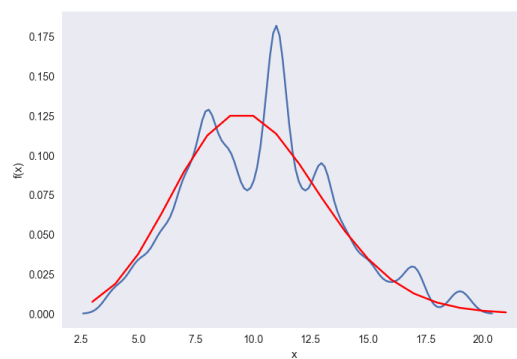
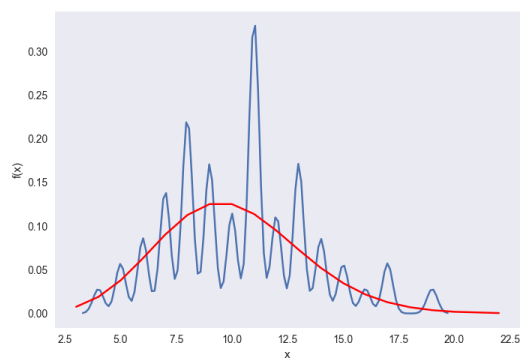


Рис. 26: Распределение Пуассона $n = 60$

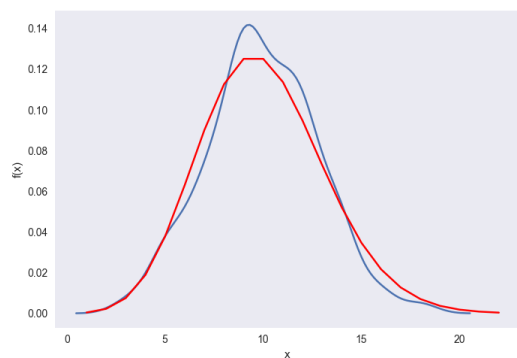
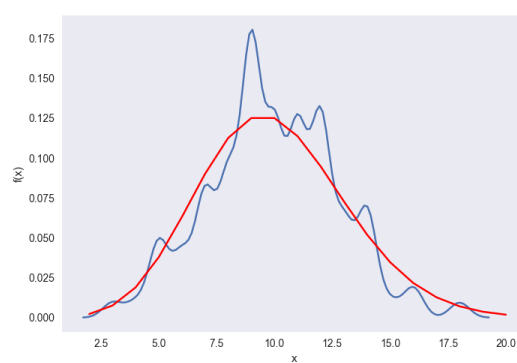
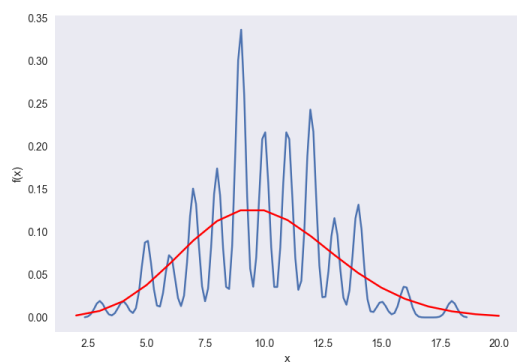


Рис. 27: Распределение Пуассона $n = 100$

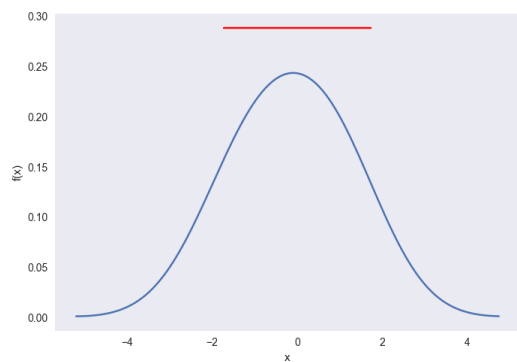
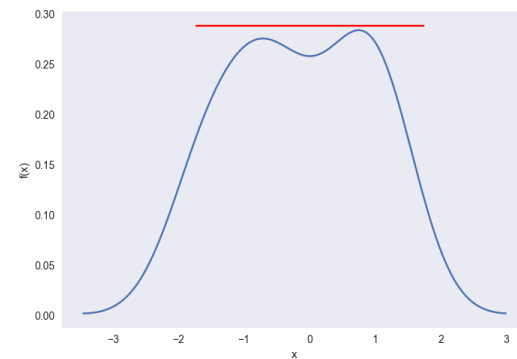
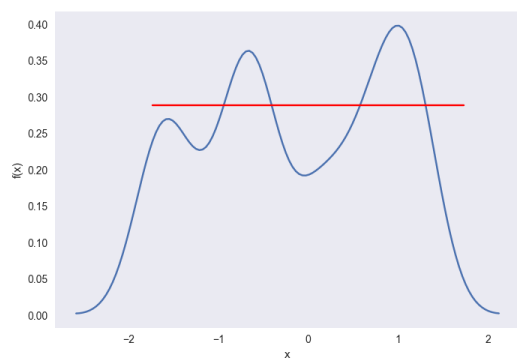


Рис. 28: Равномерное распределение $n = 20$

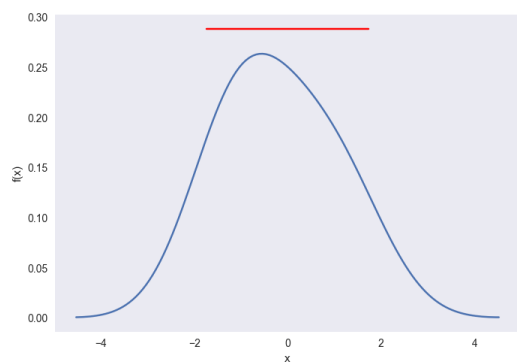
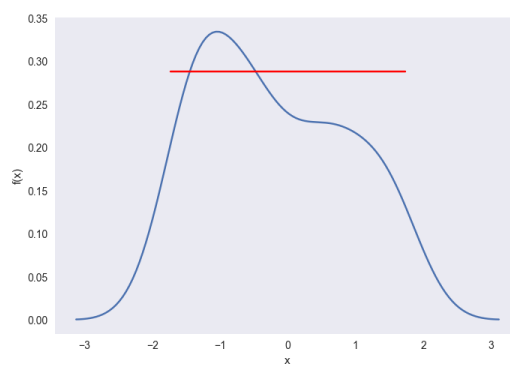
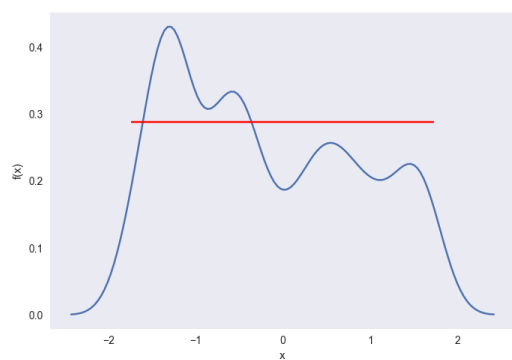


Рис. 29: Равномерное распределение $n = 60$

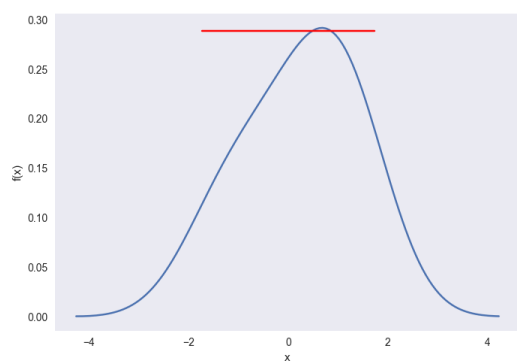
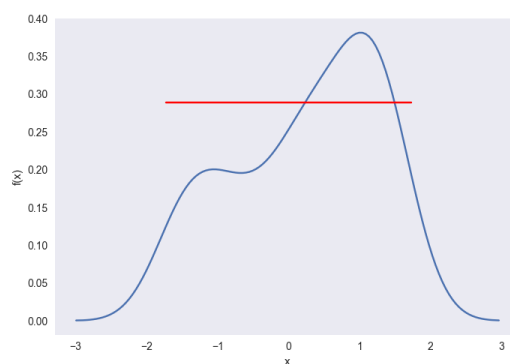
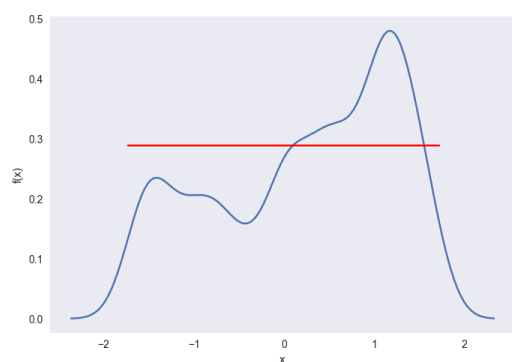


Рис. 30: Равномерное распределение $n = 100$

5 Обсуждения

5.1 Графики плотностей и гистограммы

Исходя из полученных графиков можем сделать некоторые выводы. Чем больше выборка, тем ближе гистограмма и график плотности. Можно заметить, что иногда наблюдаются некоторые "скачки" в гистограмме, особенно при малых выборках. Первые четыре распределения имеют один ярковыраженный пик - они унимодальны.

На рисунке номер пять большое количество пиков - распределение однородно. Все гистограммы имеют симметричную форму относительно центральной линии.

5.2 Характеристики положения и рассеяния

Заметим, что для нормального распределения, распределения Лапласа и равномерного распределения $E(z)$ и $D(z)$ для всех характеристик уменьшаются с ростом выборки.

В распределении Коши значения $D(z)$ для \bar{x} и z_R и $E(z)$ для z_R достигают больших порядков, тогда как остальные величины сопоставимы с распределениями, описанными выше.

Больше всех выделяется таблица значений для распределения Пуассона. Значения $E(z)$ для всех характеристик колеблется около 10, но в $D(z)$ мы видим знакомое нам уменьшение значений при росте выборки.

5.3 Доля и теоретическая вероятность выбросов

Сравним теоретическую и практическую долю выбросов

Так как для генерации нормального распределения использовался метод Бокса-Мюллера, то доля выбросов на достаточно большой выборке равна 0.

Порядок доли выбросов для распределений Коши, Лапласа и Пуассона совпадает.

Равномерное же распределение имеет нулевую долю выбросов.

5.4 Эмпирическая функция и ядерные оценки плотности распределения

Смотря на график эмпирической функции, можем заметить, что чем больше наша выборка, тем ближе график к "идеальному" теоретическому

Наилучшее приближение к теоретическому графику происходит при выборе параметра по правилу Сильвермана. Когда мы увеличиваем или уменьшаем данный параметр точность ухудшается. Исключениями являются распределение Пуассона, при $h = 2h_n$ график получается более точный, и распределение

Лапласа, при $h = 0.5h_n$ график максимально приближен к теоретическому графику.

Для ядерной оценки точность растет с увеличением выборки.

6 Приложения

Код программы: <https://github.com/katonapng/MathStat> Для написания программы использовались источники [3],[4],[5]

Список литературы

- [1] Histogram: <https://ru.wikipedia.org/wiki/Гистограмма>
- [2] Вероятностные разделы математики. Учебник для бакалавров технических направлений. //Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.
- [3] Генераторы дискретно распределенных случайных величин: <https://m.habr.com/ru/post/265321/>
- [4] Генераторы непрерывно распределенных случайных величин: <https://m.habr.com/ru/post/263993/>
- [5] Generating Poisson random values: <https://www.johndcook.com/blog/2010/06/14/generating-poisson-random-values/>
- [6] Анатольев, Станислав (2009) «Непараметрическая регрессия», Квантиль, №7, стр. 37-52.