

Санкт-Петербургский Политехнический университет  
Петра Великого  
Институт Прикладной Математики и Механики  
Кафедра «Прикладная Математика и Информатика»

Отчет  
По лабораторной работе № 1-4  
По Дисциплине «Математическая статистика»

Выполнил:  
Студент Селянкин Федор  
Группа 3630102/70301  
Проверил:  
к.ф. – м.н., доцент  
Баженов Александр Николаевич

Санкт-Петербург  
2020 г.

## Содержание

Постановка задачи .....	5
Теория .....	5
Распределения .....	5
Гистограмма .....	6
Определение .....	6
Графическое описание .....	6
Использование .....	6
Статистические характеристики.....	7
Характеристики рассеяния .....	7
Боксплот Тьюки .....	7
Определение .....	7
Описание .....	7
Построение .....	8
Теоретическая вероятность выбросов.....	8
Эмпирическая функция распределения.....	8
Статистический ряд.....	8
Определение .....	8
Описание .....	9
Оценки плотности вероятности .....	9
Определение .....	9
Ядерные оценки .....	9
Реализация.....	10
Результаты.....	11
Гистограмма и график плотности распределения.....	11
Нормальное распределение.....	11
Распределение Коши .....	12
Распределение Лапласа .....	12
Распределение Пуассона .....	13
Равномерное распределение .....	14
Характеристики положения и рассеяния .....	15
Боксплот Тьюки .....	15
Доля Выбросов .....	17
Теоретическая вероятность выбросов .....	17
Характеристики положения и рассеяния .....	17
Нормальное распределение.....	17
Распределение Коши .....	18

Распределение Лапласа .....	18
Распределение Пуассона .....	18
Равномерное распределение .....	18
Эмпирическая функция распределения.....	19
Ядерные плотности распределения .....	20
Литература .....	22
Обсуждения.....	22
Гистограмма и график плотности распределения.....	22
Характеристики положения и рассеяния .....	22
Доля и теоретическая вероятность выбросов.....	22
Эмпирическая функция и ядерные оценки плотности распределения .....	22

### **Список Иллюстраций:**

Рисунок 1 Нормальное распределение. Мощность выборки 10 .....	11
Рисунок 2 Нормальное распределение. Мощность выборки 50 .....	11
Рисунок 3 Нормальное распределение. Мощность выборки 1000 .....	11
Рисунок 4 Распределение Коши. Мощность выборки 10 .....	12
Рисунок 5 Распределение Коши. Мощность выборки 50 .....	12
Рисунок 6 Распределение Коши. Мощность выборки 1000.....	12
Рисунок 7 Распределение Лапласа. Мощность выборки 10 .....	12
Рисунок 8 Распределение Лапласа. Мощность выборки 50 .....	13
Рисунок 9 Распределение Лапласа. Мощность выборки 1000 .....	13
Рисунок 10 Распределение Пуассона. Мощность выборки 10 .....	13
Рисунок 11 Распределение Пуассона. Мощность выборки 50 .....	13
Рисунок 12 Распределение Пуассона. Мощность выборки 1000 .....	14
Рисунок 13 Нормальное распределение. Мощность выборки 10 .....	14
Рисунок 14 Нормальное распределение. Мощность выборки 50 .....	14
Рисунок 15 Нормальное распределение. Мощность выборки 1000 .....	14
Рисунок 16 Равномерное распределение .....	15
Рисунок 17 Распределение Пуассона .....	15
Рисунок 18 Распределение Лапласа.....	16
Рисунок 19 Нормальное распределение.....	16
Рисунок 20 Распределение Коши .....	17
Рисунок 21 ЭФР: Равномерное распределение .....	19
Рисунок 22 ЭФР: Распределение Пуассона.....	19
Рисунок 23 Распределение Лапласа.....	19
Рисунок 24 ЭФР: Нормальное распределение .....	20
Рисунок 25 ЭФР: Распределение Коши.....	20
Рисунок 26 ЯПР: Равномерное распределение.....	20
Рисунок 27 ЯПР: Распределение Пуассона.....	21
Рисунок 28 ЯПР: Распределение Лапласа .....	21
Рисунок 29 ЯПР: Нормальное распределение .....	21
Рисунок 30 ЯПР: Распределение Коши .....	22

### **Список Таблиц:**

Таблица 1 Статистический ряд.....	8
Таблица 2 Таблица распределений.....	9

Таблица 3 Доля Выбросов.....	17
Таблица 4 Теоретическая вероятность выбросов .....	17
Таблица 5 Нормальное распределение .....	18
Таблица 6 Распределение Коши.....	18
Таблица 7 Распределение Лапласа .....	18
Таблица 8 Распределение Пуассона.....	18
Таблица 9 Равномерное распределение.....	18

## 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 и их квадратов:

(1)

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

Вычислить оценку дисперсии по формуле:

(2)

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

Представить полученные данные в виде таблиц.

- 3) Сгенерировать выборки размером 20, 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.

## Теория

### Распределения

- Нормальное распределение

(3)

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

- Распределение Коши

(4)

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

- Распределение Лапласа

(5)

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

- Распределение Пуассона

(6)

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

- Равномерное распределение

(7)

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

## Гистограмма

### Определение

Гистограмма в математической статистике – это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него.

### Графическое описание

Графически гистограмма строится следующим образом. Сначала 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.

### 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.

## Статистические характеристики

- Выборочное среднее

(8)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Выборочная медиана

(9)

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

- Полусумма экстремальных выборочных элементов

(10)

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

- Полусумма квартилей

Выборочная квартиль  $z_p$  порядка  $p$  определяется формулой.

(11)

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

- Полусумма квартилей

(12)

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2}$$

- Усеченное среднее

(13)

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

## Характеристики рассеяния

- Выборочная дисперсия

(14)

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

## Боксплот Тьюки

### Определение

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

### 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.

### 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 и полутора межквартильных расстояний. Формула имеет вид

(15)

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

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.

### Теоретическая вероятность выбросов

Зная, первый и третий квартили по формуле (15) можно вычислить теоретически нижнюю и верхнюю границы уса ( $X_1^T$  и  $X_2^T$  соответственно). Выбросами считаются такие величины  $x$ , что

(16)

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

Теоретическая вероятность выбросов для непрерывных распределений

(17)

$$P_b^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)),$$

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

Теоретическая вероятность выбросов для дискретных распределений

(18)

$$P_b^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)),$$

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

### Эмпирическая функция распределения

#### 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

#### Определение

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



(19)

$$F_n^*(x) = P^*(X < x)$$

### 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

(20)

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

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

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

Таблица 2 Таблица распределений

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

(21)

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

### Оценки плотности вероятности

#### Определение

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

(22)

$$\hat{f}(x) \approx f(x)$$

#### Ядерные оценки

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

(23)

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

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

(24)

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

Такие оценки называются непрерывными ядерными.

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

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

Правило Сильвермана

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

Где  $\hat{\sigma}$  – выборочное стандартное отклонение.

## Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm, с использованием дополнительных библиотек для отображения и расчетов. Исходный код лабораторной выложен на веб-сервисе GitHub [6].

## Результаты

### Гистограмма и график плотности распределения

#### Нормальное распределение

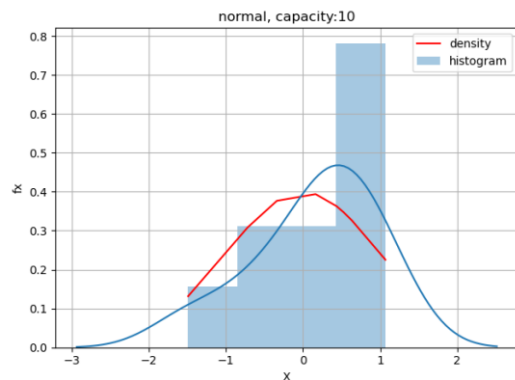


Рисунок 1 Нормальное распределение. Мощность выборки 10

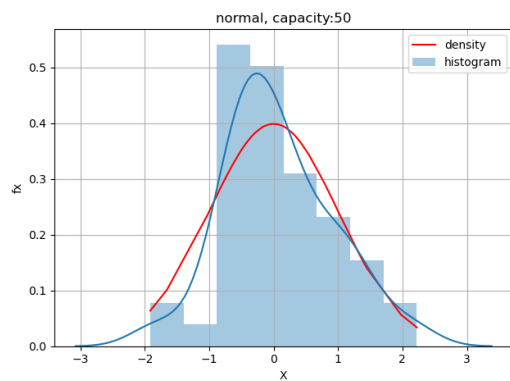


Рисунок 2 Нормальное распределение. Мощность выборки 50

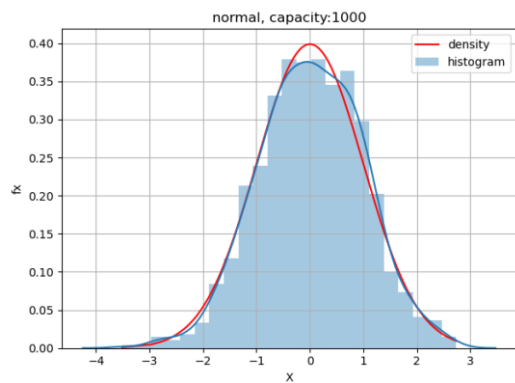


Рисунок 3 Нормальное распределение. Мощность выборки 1000

## Распределение Коши

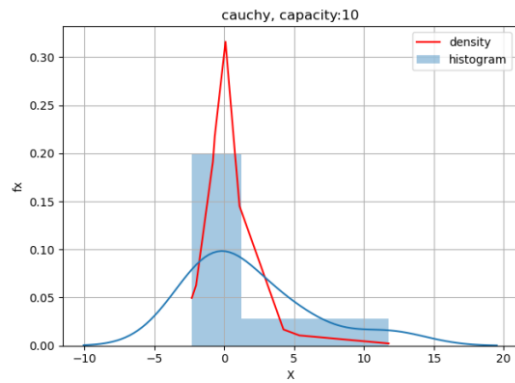


Рисунок 4 Распределение Коши. Мощность выборки 10

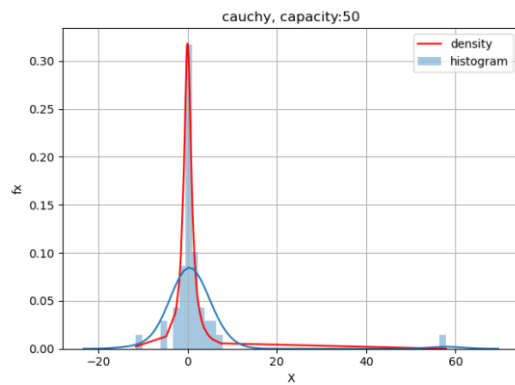


Рисунок 5 Распределение Коши. Мощность выборки 50

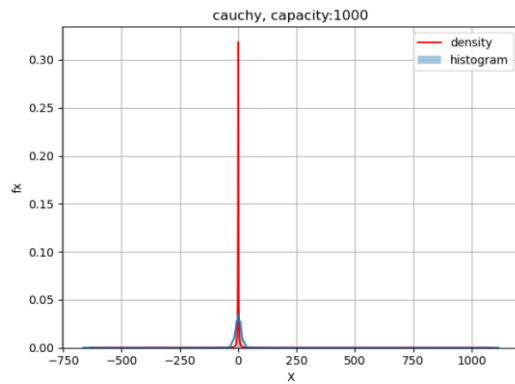


Рисунок 6 Распределение Коши. Мощность выборки 1000

## Распределение Лапласа

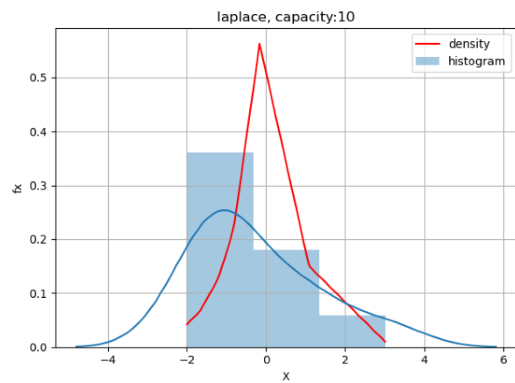


Рисунок 7 Распределение Лапласа. Мощность выборки 10

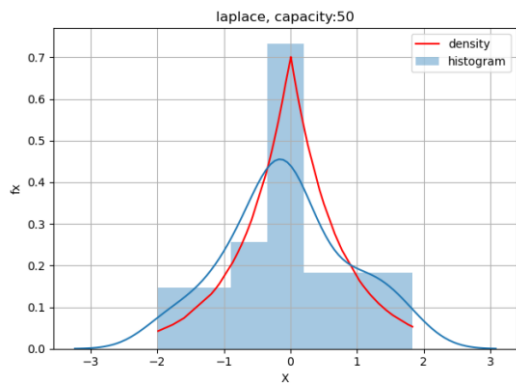


Рисунок 8 Распределение Лапласа. Мощность выборки 50

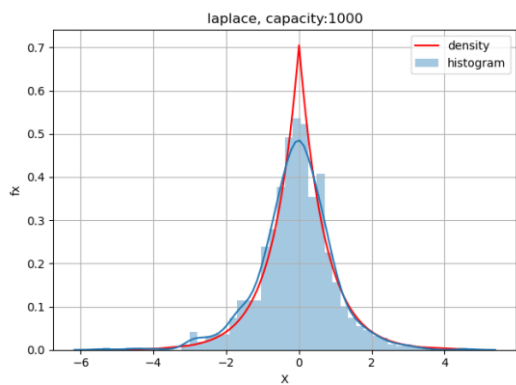


Рисунок 9 Распределение Лапласа. Мощность выборки 1000

### Распределение Пуассона

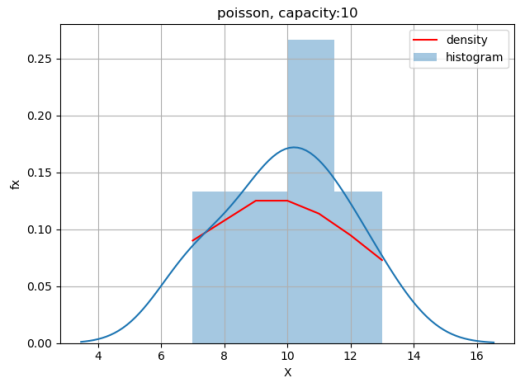


Рисунок 10 Распределение Пуассона. Мощность выборки 10

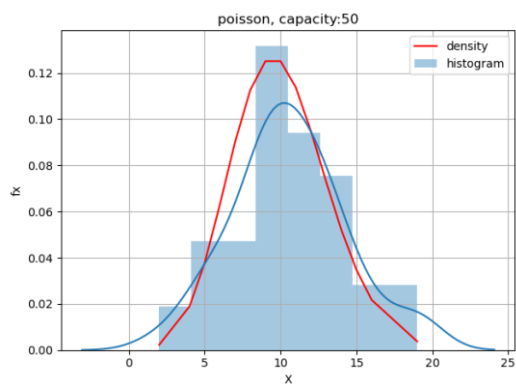


Рисунок 11 Распределение Пуассона. Мощность выборки 50

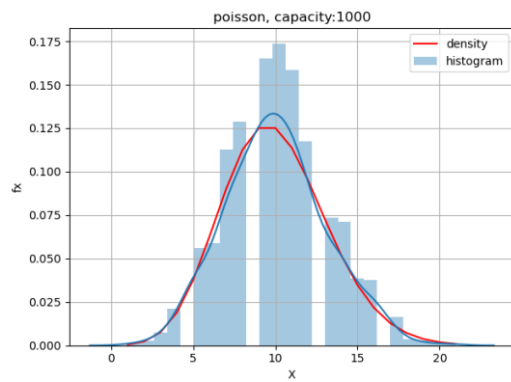


Рисунок 12 Распределение Пуассона. Мощность выборки 1000

Равномерное распределение

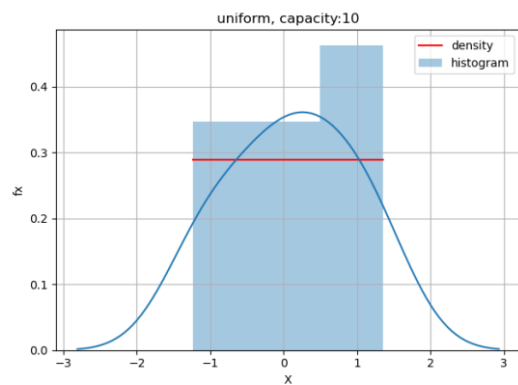


Рисунок 13 Нормальное распределение. Мощность выборки 10

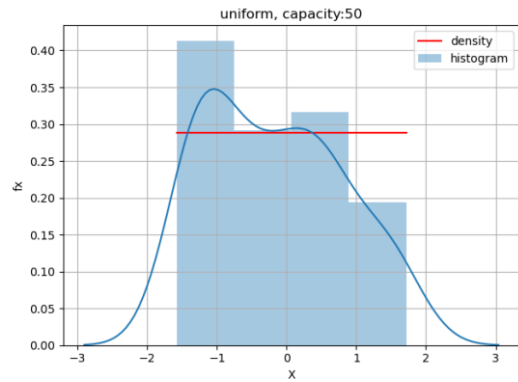


Рисунок 14 Нормальное распределение. Мощность выборки 50

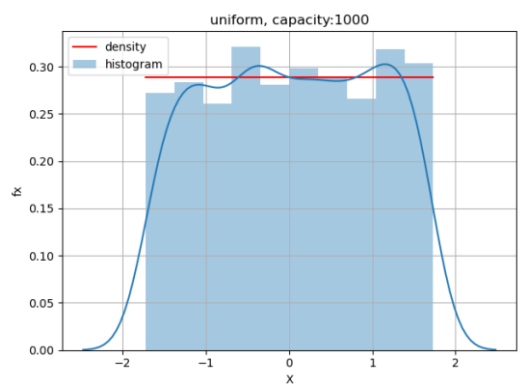


Рисунок 15 Нормальное распределение. Мощность выборки 1000

## Характеристики положения и рассеяния

### Боксплот Тьюки

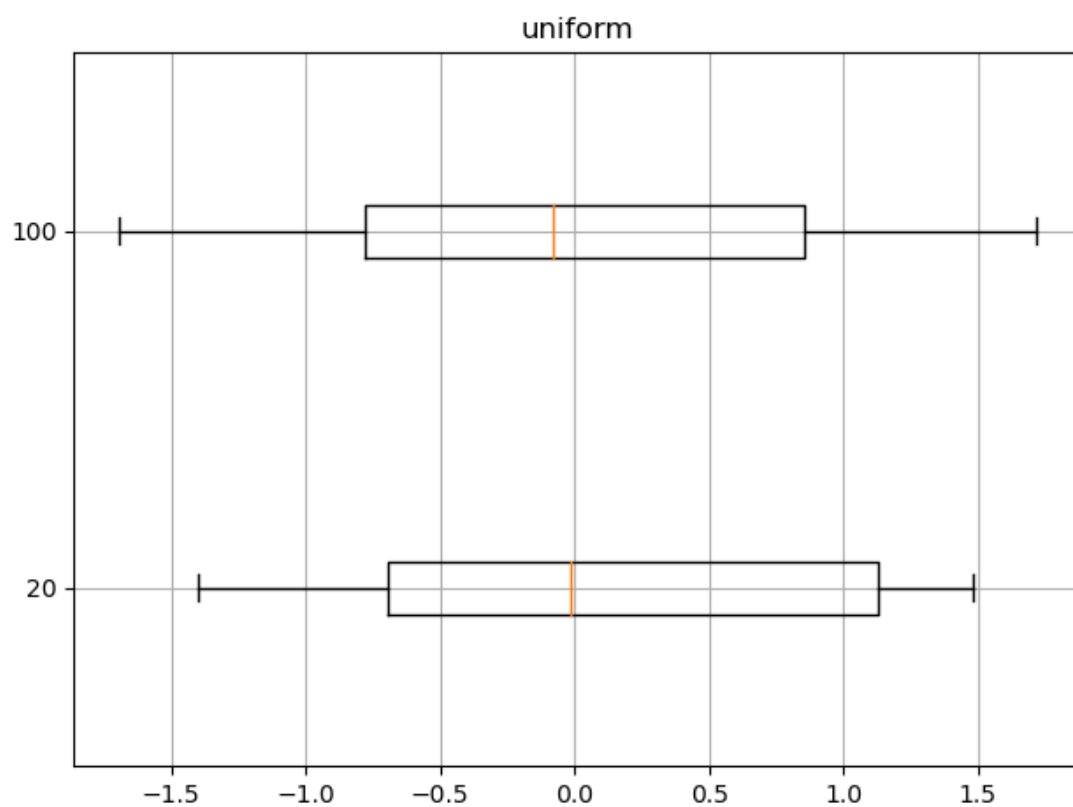


Рисунок 16 Равномерное распределение

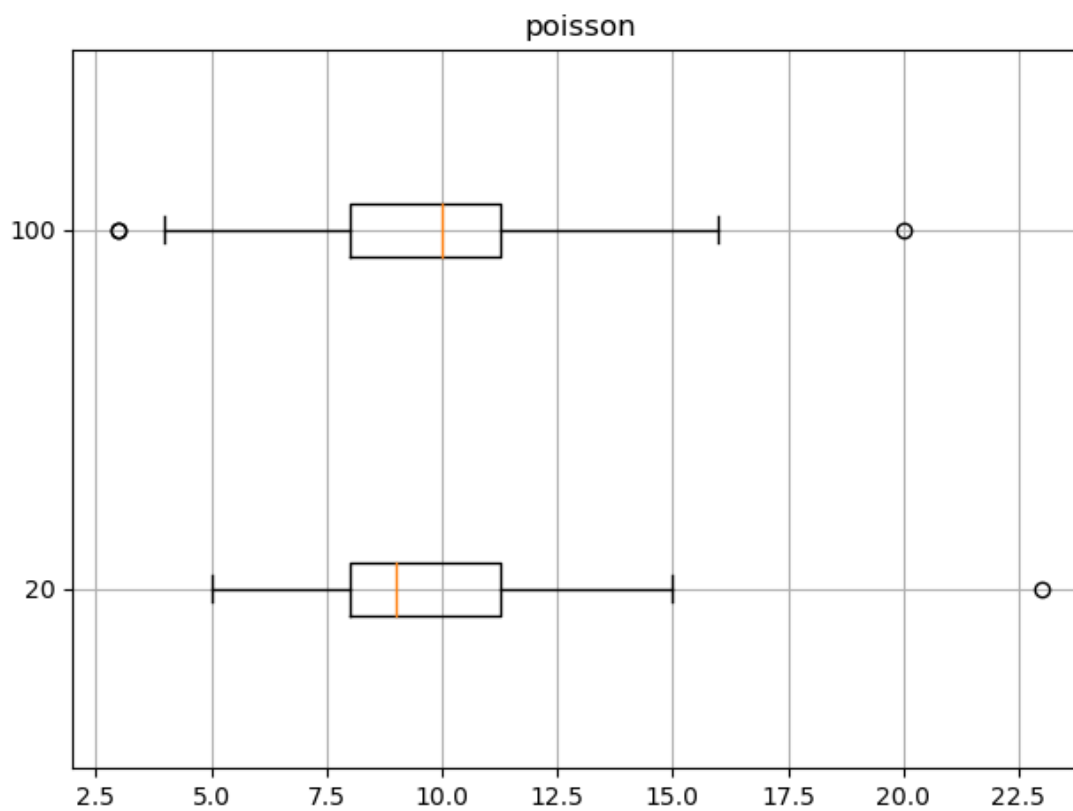


Рисунок 17 Распределение Пуассона

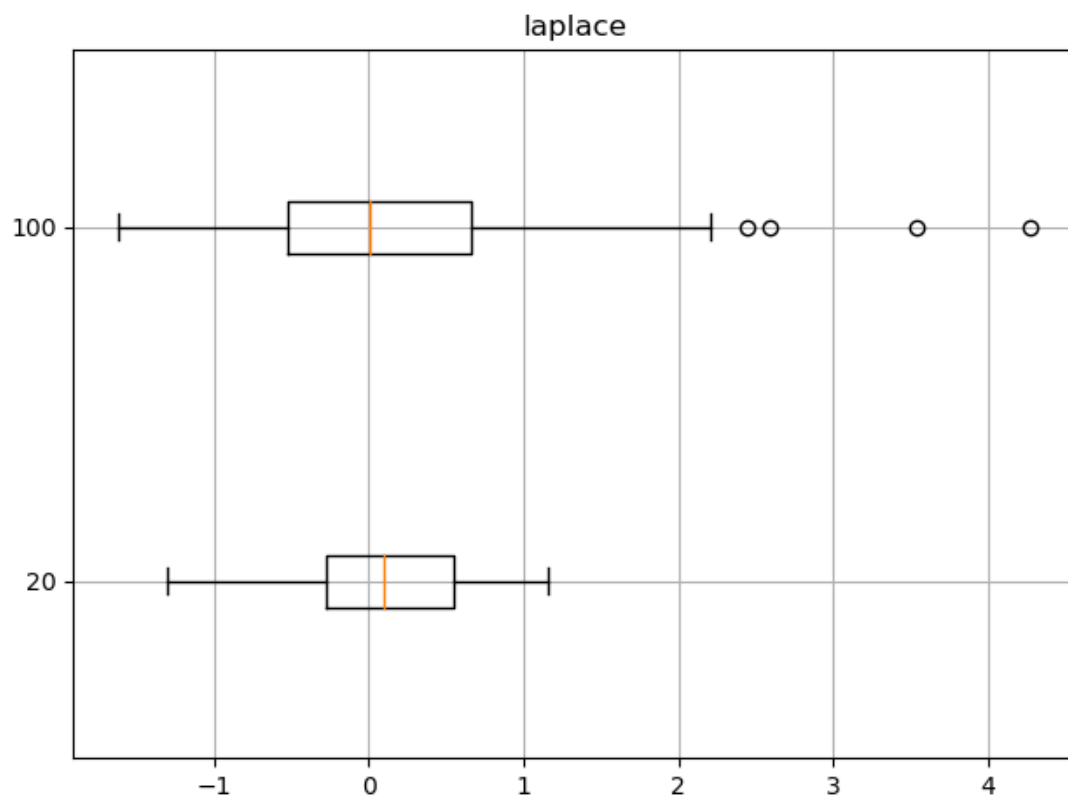


Рисунок 18 Распределение Лапласа

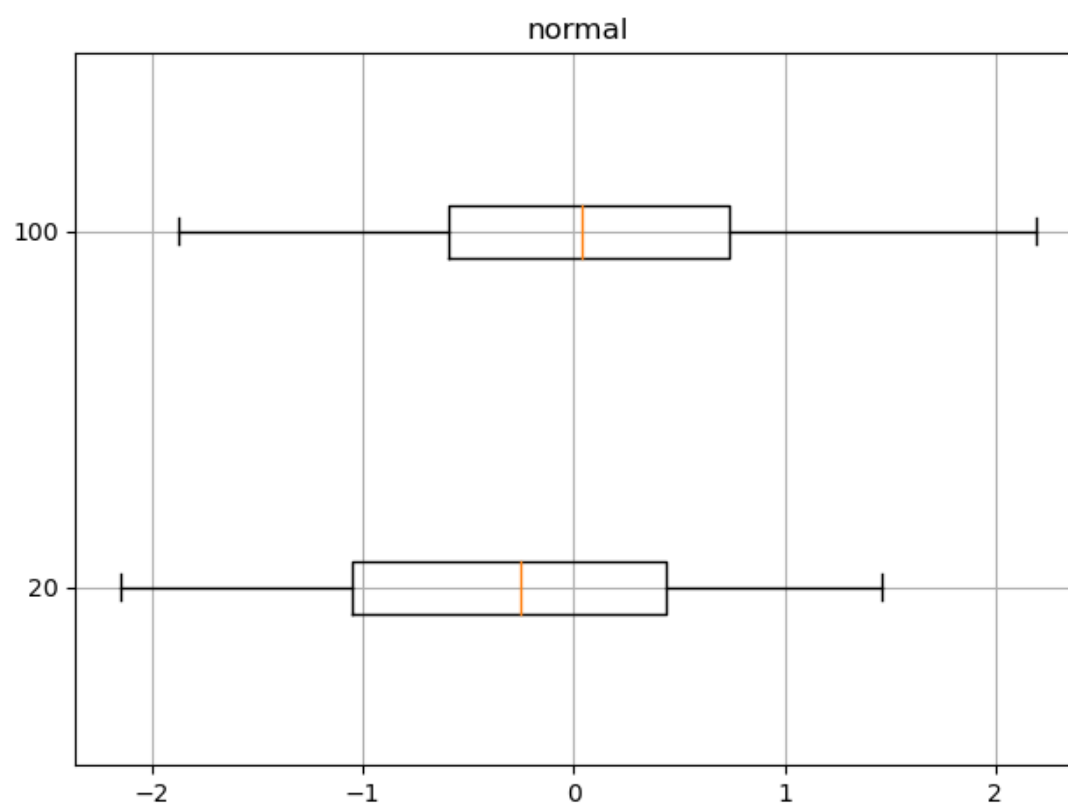


Рисунок 19 Нормальное распределение



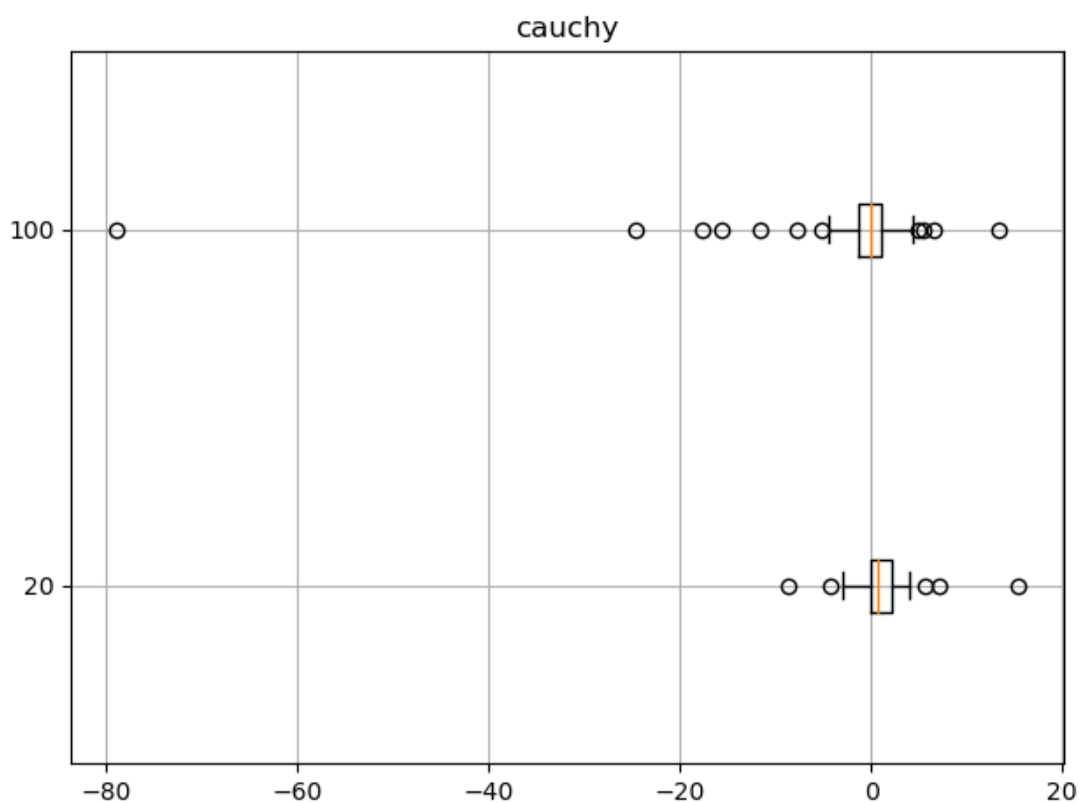


Рисунок 20 Распределение Коши

#### Доля Выбросов

Выборка	Доля выбросов
Нормальное распределение n = 20	0.02
Нормальное распределение n = 100	0.01
Распределение Коши n = 20	0.15
Распределение Коши n = 100	0.15
Распределение Лапласа n = 20	0.08
Распределение Лапласа n = 100	0.07
Распределение Пуассона n = 20	0.02
Распределение Пуассона n = 100	0.01
Равномерное распределение n = 20	0.00
Равномерное распределение n = 100	0.00

Таблица 3 Доля Выбросов

#### Теоретическая вероятность выбросов

Распределение	$Q_1^T$	$Q_3^T$	$X_1^T$	$x_2^T$	$P_v^T$
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 4 Теоретическая вероятность выбросов

#### Характеристики положения и рассеяния

##### Нормальное распределение

n		$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
10	E	0.01	0	0	-0.3	0
	D	0.096971	0.142454	0.193487	0.113430	0.166238
100	E	0	0	0	-0.01	0
	D	0.010519	0.015980	0.103325	0.012640	0.011316
1000	E	0	0	0	0	0
	D	0.000981	0.001644	0.062102	0.001258	0.001219
$\bar{x} < z_{tr} < z_Q < med\ x < z_R$						

Таблица 5 Нормальное распределение

#### Распределение Коши

n		$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
10	E	0	0	0	-1	0
	D	70974.612820	0.306827	31123.008220	5.088886	0.833265
100	E	0	0	0	-0.04	0
	D	340057.722086	0.025110	4790055.844103	0.050678	0.025347
1000	E	0	0	0	0	0
	D	3152.073514	0.002294	36577755.775665	0.005139	0.002566
$med\ x < z_{tr} < z_Q < \bar{x} < z_R$						

Таблица 6 Распределение Коши

#### Распределение Лапласа

n		$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
10	E	0	0	0	-0.3	0
	D	0.101021	0.076025	0.446803	0.123058	0.098940
100	E	0	0	0	-0.02	0
	D	0.010233	0.005696	0.363117	0.010345	0.005857
1000	E	0	0	0	-0.002	-0.0007
	D	0.001074	0.000531	0.390686	0.001000	0.000597
$med\ x < z_{tr} < z_Q < \bar{x} < z_R$						

Таблица 7 Распределение Лапласа

#### Распределение Пуассона

n		$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
10	E	10.0	10.0	10.0	10.0	11.0
	D	1.008408	1.484790	1.740384	1.231344	1.469999
100	E	10.0	9.9	10.0	10.0	9.9
	D	0.099818	0.224111	0.955558	0.156679	0.118266
1000	E	10.0	10.000000	11.0	9.9	9.9
	D	0.009958	0.000000	0.663388	0.004436	0.011769
$med\ x < z_Q < \bar{x} < z_{tr} < z_R$						

Таблица 8 Распределение Пуассона

#### Равномерное распределение

n		$\bar{x}$	$med\ x$	$z_R$	$z_Q$	$z_{tr}$
10	E	0	0	0	-0.31	0
	D	0.099079	0.227962	0.043533	0.146461	0.228256
100	E	0	0	0	-0.015	0
	D	0.010254	0.028950	0.000637	0.013887	0.019951
1000	E	0.0011	0	-0.000076	-0.002	0
	D	0.000970	0.002742	0.000006	0.001472	0.001991
$z_R < \bar{x} < z_Q < z_{tr} < med\ x$						

Таблица 9 Равномерное распределение

## Эмпирическая функция распределения

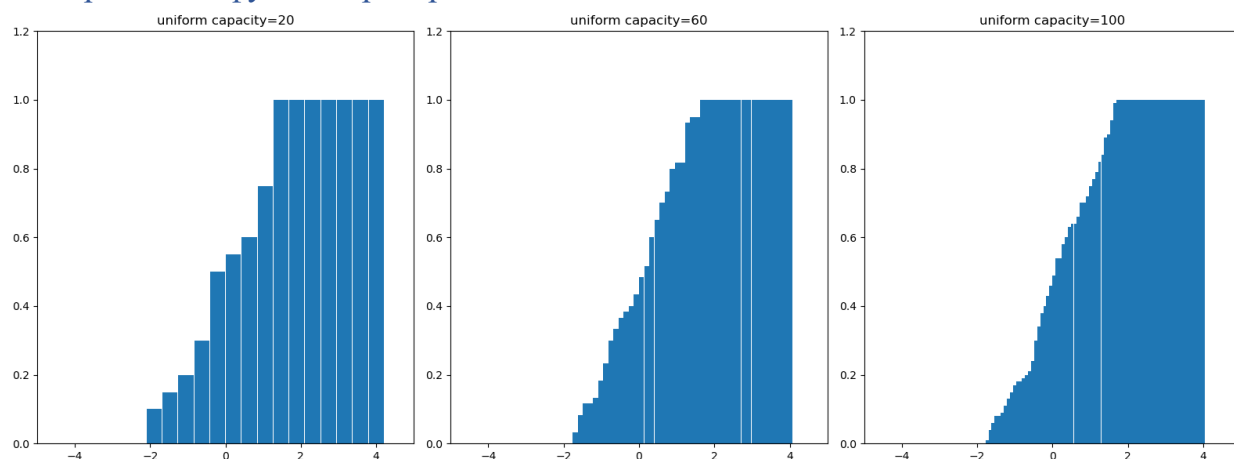


Рисунок 21 ЭФР: Равномерное распределение

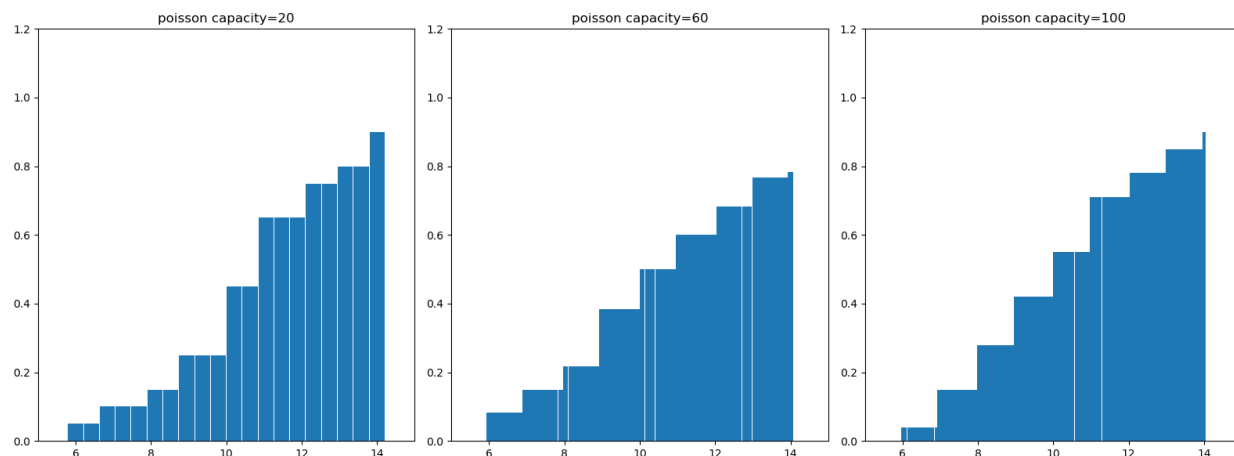


Рисунок 22 ЭФР: Распределение Пуассона

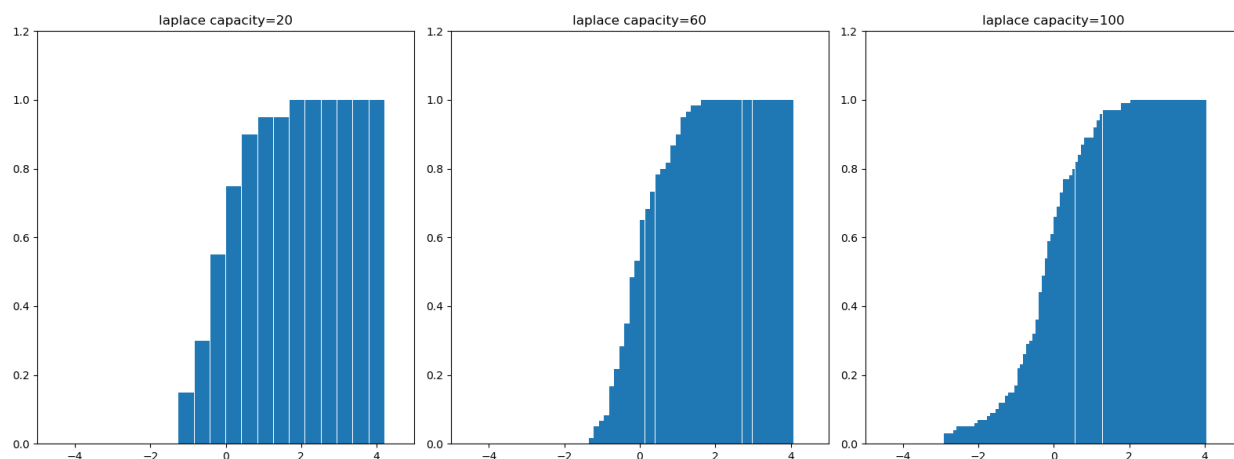


Рисунок 23 Распределение Лапласа

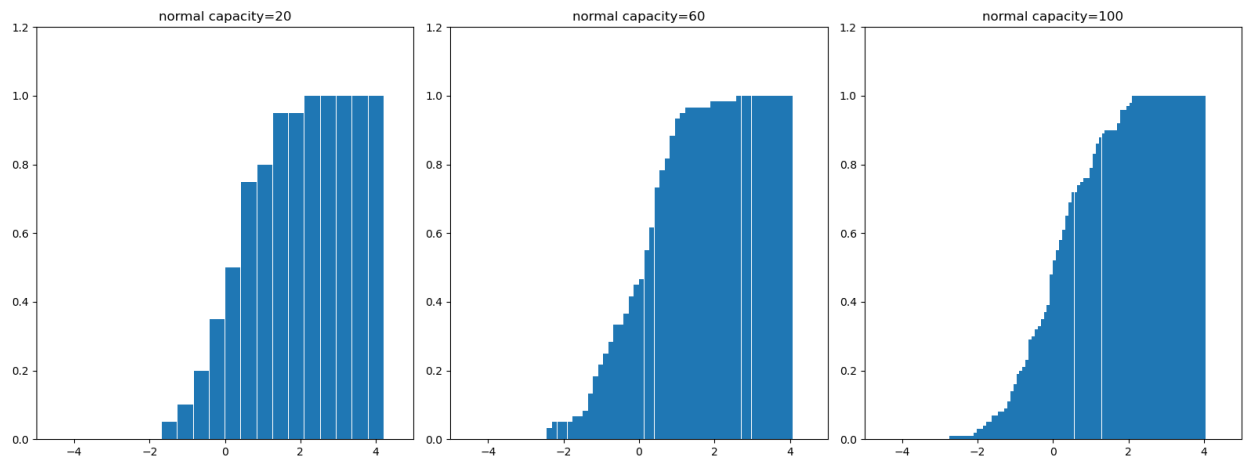


Рисунок 24 ЭФР: Нормальное распределение

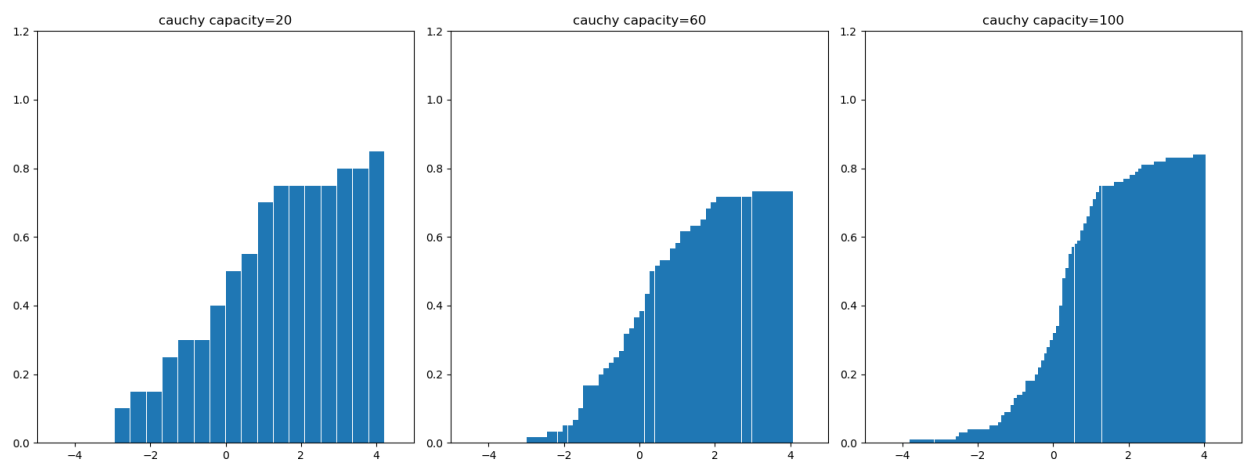


Рисунок 25 ЭФР: Распределение Коши

## Ядерные плотности распределения

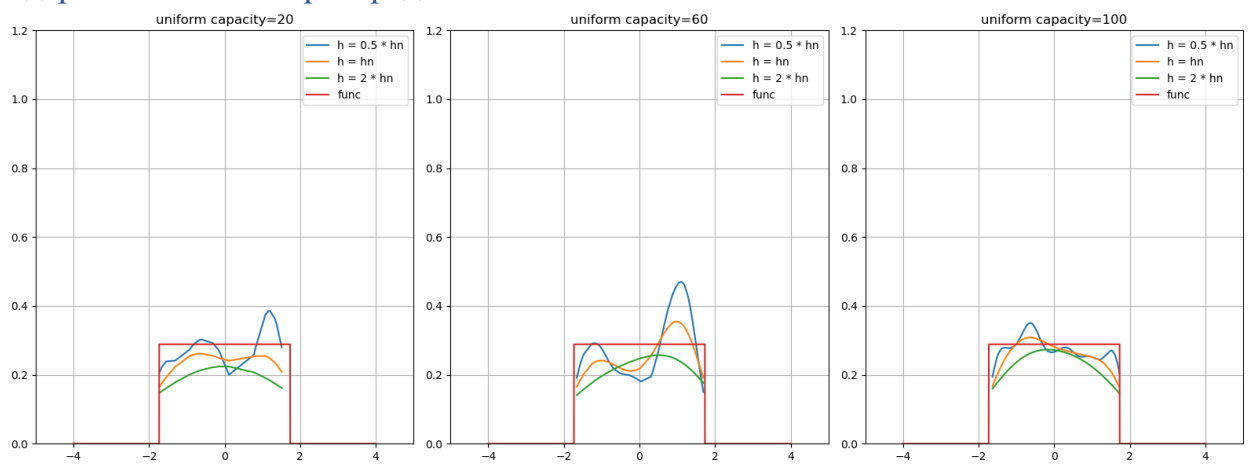


Рисунок 26 ЯПР: Равномерное распределение

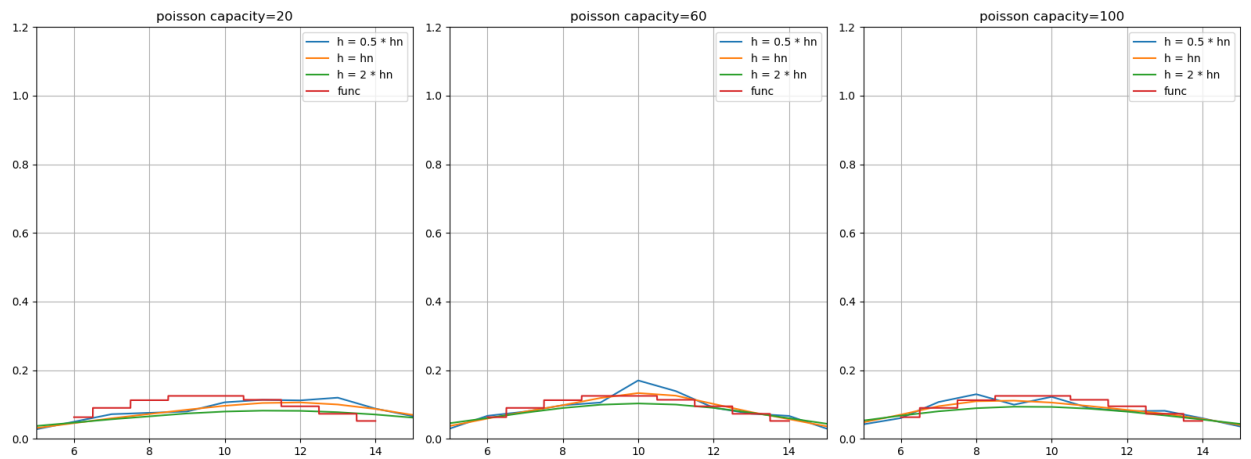


Рисунок 27 ЯПР: Распределение Пуассона

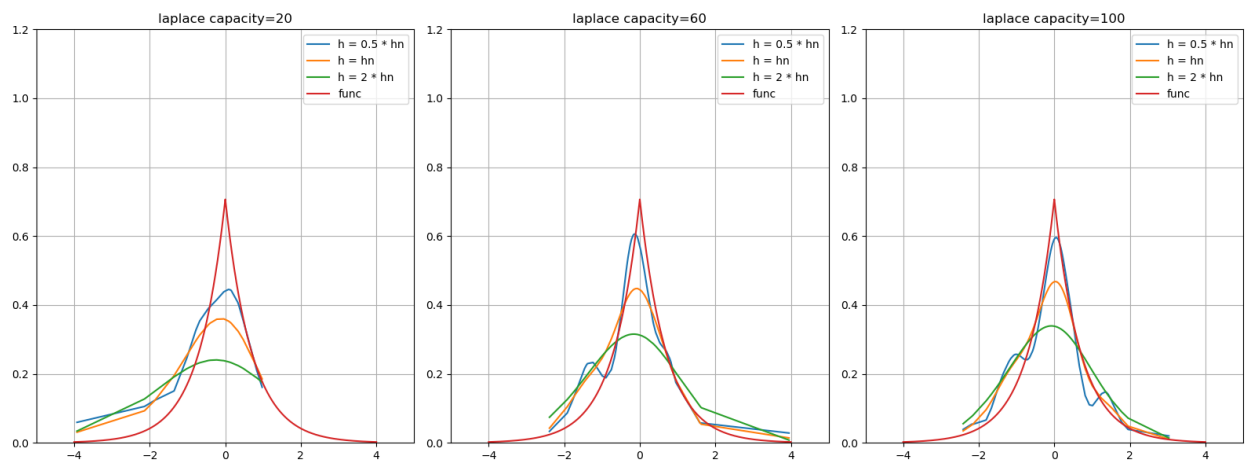


Рисунок 28 ЯПР: Распределение Лапласа

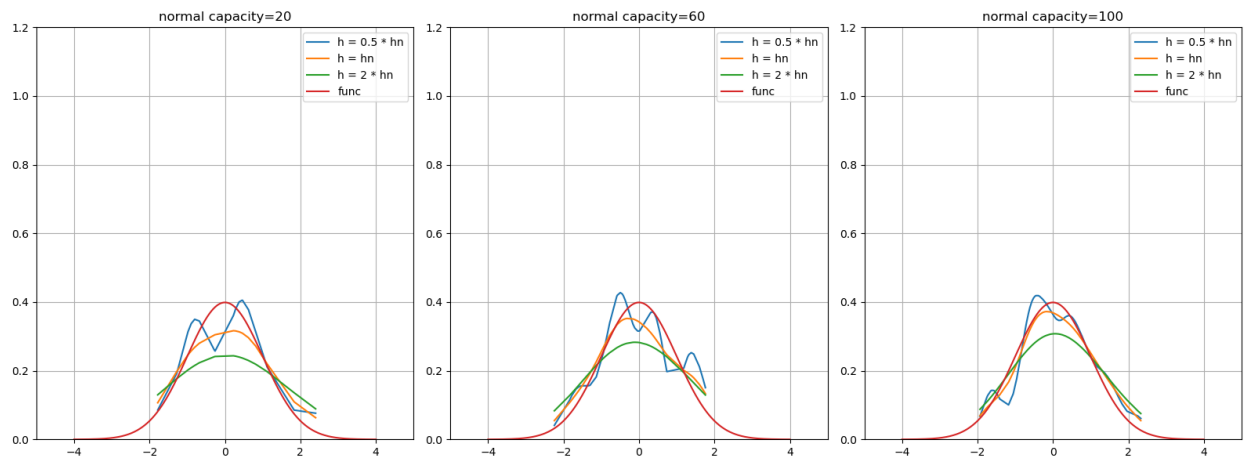


Рисунок 29 ЯПР: Нормальное распределение

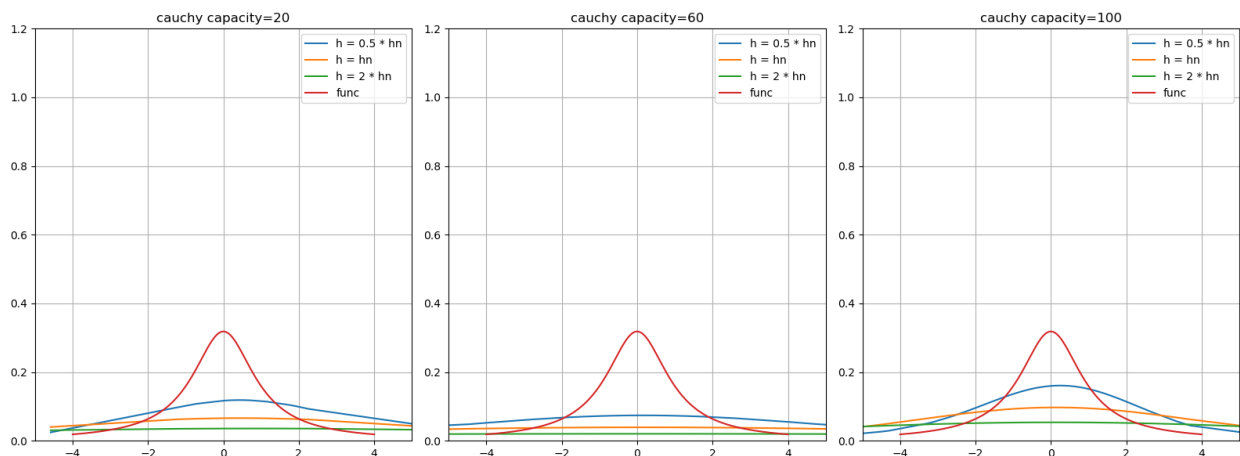


Рисунок 30 ЯПР: Распределение Коши

## Литература

- 1) Гистограмма. <https://en.wikipedia.org/wiki/Histogram>
- 2) Выборочная функция распределения  
[https://en.wikipedia.org/wiki/Empirical\\_distribution\\_function](https://en.wikipedia.org/wiki/Empirical_distribution_function)
- 3) Распределение Коши  
[https://en.wikipedia.org/wiki/Cauchy\\_distribution](https://en.wikipedia.org/wiki/Cauchy_distribution)
- 4) Вероятностные разделы математики. Учебник для бакалавров технических направлений. //Под ред. Максимова Ю.Д. — Спб.: Иван Федоров., 2001. — 592 с., илл.
- 5) Боксплот  
[https://en.wikipedia.org/wiki/Box\\_plot](https://en.wikipedia.org/wiki/Box_plot)
- 6) Ссылка на репозиторий GitHub  
<https://github.com/SelyankinFyodor/math-statistics>

## Обсуждения

### Гистограмма и график плотности распределения

С увеличением мощности выборки распределения элементов график полученный с помощью распределения все больше совпадает с графиком плотности распределения. Для распределения Коши это «похожесть» выполняется только вблизи нуля.

### Характеристики положения и рассеяния

С увеличением мощности выборки величины среднее и усеченное среднее имеют все меньшее рассеяние (с увеличением мощности в 10 раз, дисперсия уменьшается в 10 раз). Для Распределения Коши в качестве характеристики положения не разумно брать среднее или полусумму экстремалей.

### Доля и теоретическая вероятность выбросов

В распределении Коши доля выбросов меньше, чем в распределении Лапласа или нормальном распределении, но из боксплота видно, что разброс значений намного больше.

### Эмпирическая функция и ядерные оценки плотности распределения

С увеличением мощности выборки график эмпирической функции распределения становится более близким к графику функции распределения.

С увеличением мощности выборки график ядерной плотности распределения все больше становится похож на график плотности распределения. Причем на многих иллюстрациях график с  $h$  выбранном согласно правилу Сильвермана без удваивания или уполовинивания приближен к

графику плотности распределения лучше, чем остальные. Для распределения Коши ядерная плотность имеет не такой сильный изгиб в нуле как плотность распределения.