

Санкт-Петербургский Политехнический Университет  
им. Петра Великого

Институт прикладной математики и механики  
Кафедра прикладной математики

**Отчёт по лабораторным работам №1-4 по дисциплине  
“Математическая статистика”**

Выполнил студент:

Мишутин Д. В.

Группа:

3630102/70301

Проверил:

К.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2020 г.

## Оглавление

1 Постановка задачи.....	4
2 Теория.....	5
2.1 Распределения.....	5
2.2 Функции плотности вероятности.....	5
2.2.1 Определение.....	5
2.3 Характеристики положения и рассеяния.....	5
2.3.1 Характеристики положения.....	5
2.3.2 Характеристики рассеяния.....	6
2.4 Боксплот Тьюки.....	6
2.4.1 Определение.....	6
2.4.2 Характеристики положения.....	6
2.5 Эмпирические функции и ядерные оценки.....	7
2.5.1 Эмпирическая функция.....	7
2.5.2 Ядерные функции плотности.....	7
3 Реализация.....	8
4 Результаты.....	8
5 Выводы.....	32
5.1 Функции плотности вероятности и гистограммы.....	32
5.2 Характеристики положения выборки.....	32
5.3 Боксплот Тьюки.....	33
5.4 Эмпирические функции и ядерные оценки.....	33
6 Литература.....	33
7 Приложения.....	33

## Список иллюстраций и таблиц

- 4.1 Функции плотности распределения и гистограммы
- 4.2 Характеристики положения и рассеивания
- 4.3 Боксплот Тьюки
- 4.4 Эмпирические функции распределения (ЭФР)
- 4.5 Ядерные функции плотности (ЯФП)
- 4.6 Таблицы. Средние модулей ошибок

## 1 Постановка задачи

Для 5 распределений:

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

1. Любыми средствами сгенерировать выборки с мощностями 10, 50, 100 и 1000 элементов для 5 распределений. Построить на одном рисунке гистограмму и график плотности для каждого распределения с отдельной мощностью выборки.
2. Любыми средствами сгенерировать выборки с мощностями 10, 100 и 1000 элементов для 5 распределений. Для каждой выборки вычислить следующие характеристики положения:

$avg$  (выборочное среднее),  $med$ ,  $x$ ,  $Z_R$ ,  $Z_Q$ ,  $Z_{tr}$ ,  $n$ ,  $p$ ,  $r \approx \frac{n}{4}$ . Построить по ним таблицы. После вычисления характеристик положения 1000 раз, для каждой характеристики находятся их средние значения и дисперсии:

$$E(z) = \frac{1}{n} \sum_{i=1}^n z_i \quad (1.1)$$

$$D(z) = E(z^2) - E^2(z) \quad (1.2)$$

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 Распределения

- Стандартное нормальное распределение:

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (2.1.1)$$

- Стандартное распределение Коши:

$$C(x, 0, 1) = \frac{1}{\pi(1+x^2)} \quad (2.1.2)$$

- Распределение Лапласа:

$$L\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \quad (2.1.3)$$

- Распределение Пуассона:

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (2.1.4)$$

- Равномерное распределение:

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & \text{при } |x| \leq \sqrt{3} \\ 0, & \text{при } |x| > \sqrt{3} \end{cases} \quad (2.1.5)$$

### 2.2 Функции плотности вероятности

#### 2.2.1 Определение

*Плотность вероятности* есть способ задания вероятностной меры в  $R^n$

### 2.3 Характеристики положения и рассеяния

#### 2.3.1 Характеристики положения

- Выборочное среднее:

$$avg = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.3.1)$$

- Выборочная медиана:

$$med x = \begin{cases} x_{k+1}, & n = 2k + 1 \\ \frac{1}{2}(x_k + x_{k+1}), & n = 2k \end{cases} \quad (2.3.2)$$

- Полусумма экстремальных значений:

$$Z_R = \frac{1}{2}(x_1 + x_n) \quad (2.3.3)$$

- Полусумма кватилей:

$$Z_Q = \frac{1}{2}\left(Z_{\frac{1}{4}} + Z_{\frac{3}{4}}\right) \quad (2.3.4)$$

- Усечённое среднее:

$$Z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_i \quad (2.3.5)$$

### 2.3.2 Характеристики рассеяния

- Выборочная дисперсия

$$D = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.3.6)$$

## 2.4 Боксплот Тьюки

### 2.4.1 Определение

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

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

*Выброс* – результат, сильно выделяющийся из общей выборки.

### 2.4.2 Характеристики положения

- Выборочное среднее:

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

- Выборочная медиана:

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

- Квартиль:

$$Z_{[p]} = \begin{cases} x_{[np]}, np \in Z \\ x_{[np]+1}, np \notin Z \end{cases} \quad (2.4)$$

## 2.5 Эмпирические функции и ядерные оценки

### 2.5.1 Эмпирическая функция

Эмпирической функцией распределения (ЭФР)  $\hat{F}_n$  называется относительная частота события  $X < x$ , полученная по данной выборке суммированием частот  $n_i$ , для которых элементы  $z_i$  статистического ряда меньше  $x$ :

$$\hat{F}_n(x) = \hat{F}(X < x) = \frac{1}{n} \sum_{z_i < x} n_i \quad (2.5.1)$$

ЭФР является оценкой, то есть приближённым значением, генеральной функции распределения.

$$\hat{F}_n(x) \approx F_X(x)$$

### 2.5.2 Ядерные функции плотности

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

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

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

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) \quad (2.5.2)$$

Здесь  $K(u)$  называется ядерной функцией плотности (ЯФП), непрерывна и является плотностью вероятности,  $\{h_n\}$  – любая последовательность положительных чисел, обладающая свойствами:

1.  $\lim_{n \rightarrow \infty} h_n = 0$
2.  $\lim_{n \rightarrow \infty} \frac{h_n}{n^{-1}} = \infty$

Гауссово ядро:

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

Правило Сильвермана:

$$h_n = 1.06 \hat{\sigma} n^{-\frac{1}{5}},$$

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

### 3 Реализация

Был использован язык *Python 3.8.2* и его средства: модуль *numpy* для генерации выборок с различными распределениями и вычисления описательных статистик, модуль *scipy* для функций и плотностей распределений, модуль *matplotlib* для построения и отображения графиков и гистограмм, модуль *math* для математических расчётов, модуль *pandas* для оптимального хранения статистических данных и функция *display* из модуля *IPython.display* для их корректного отображения в таблицах. Исходный код лабораторных приведён в приложении.

### 4 Результаты

#### 4.1 Функции плотности распределения и гистограммы

Рис. 1 Стандартное нормальное распределение при  $n=10$

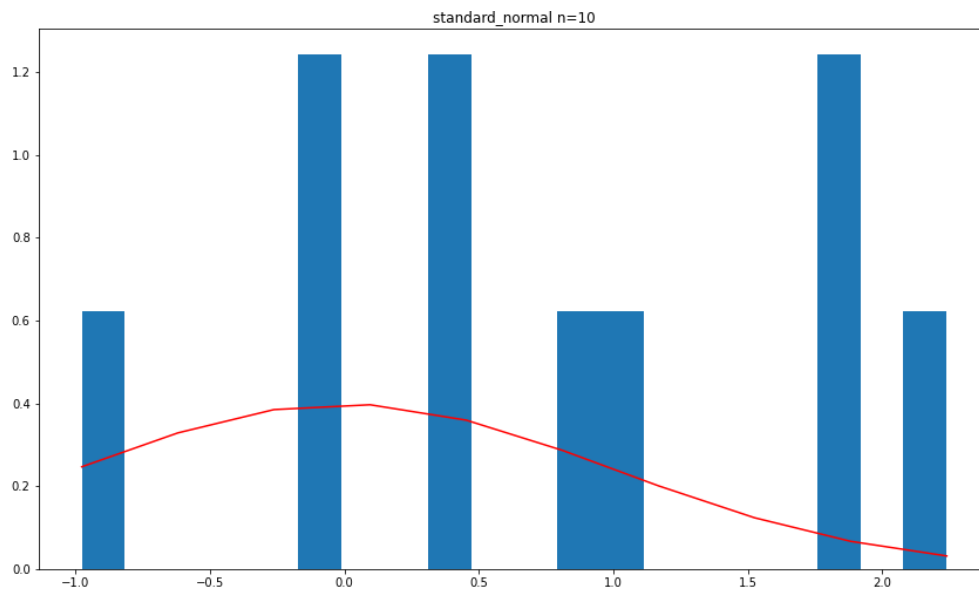


Рис. 2 Стандартное нормальное распределение при  $n=50$

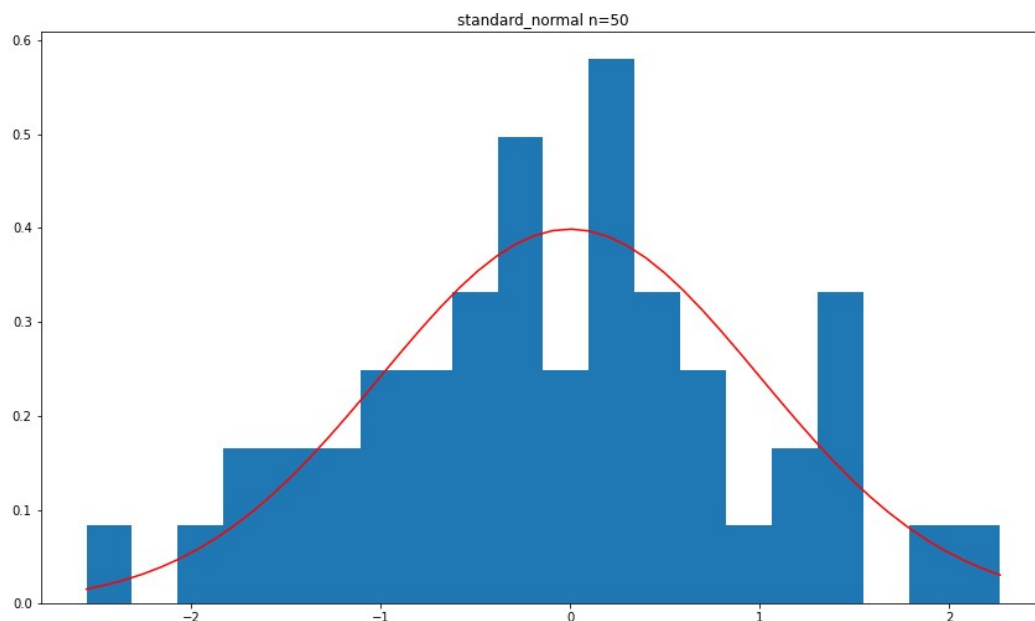


Рис. 3 Стандартное нормальное распределение при  $n=100$

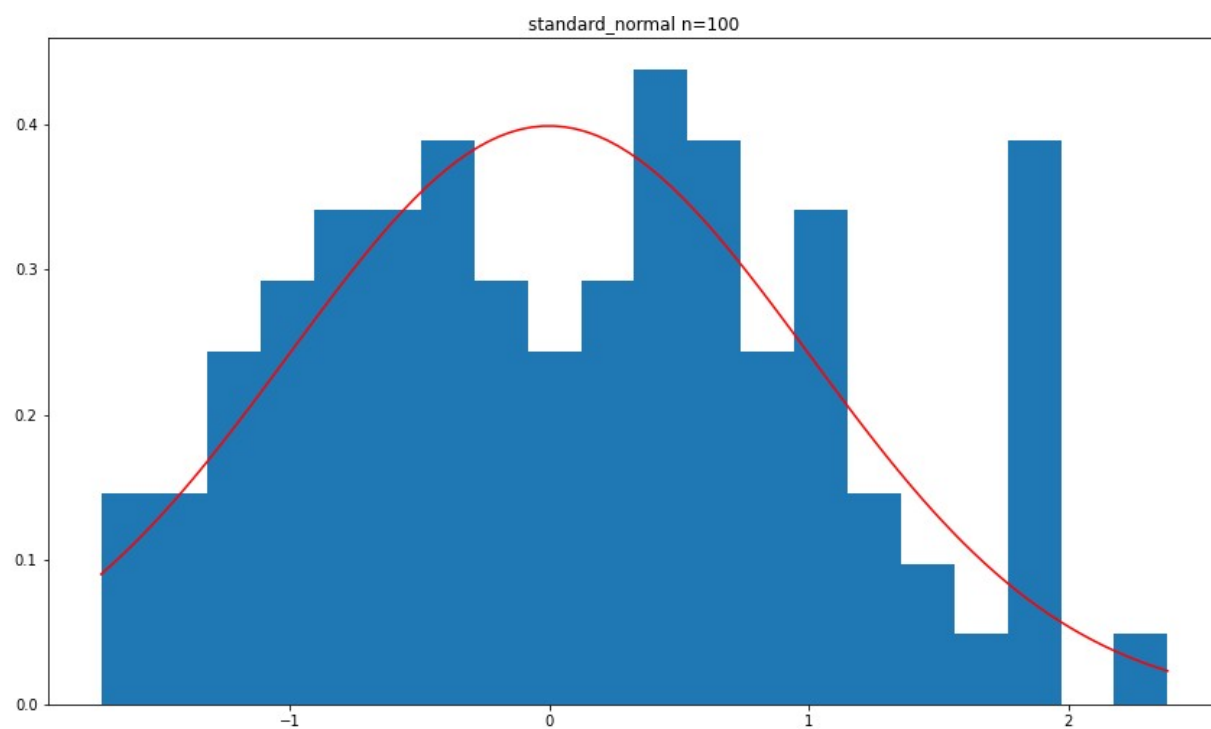


Рис. 4 Стандартное нормальное распределение при  $n=1000$

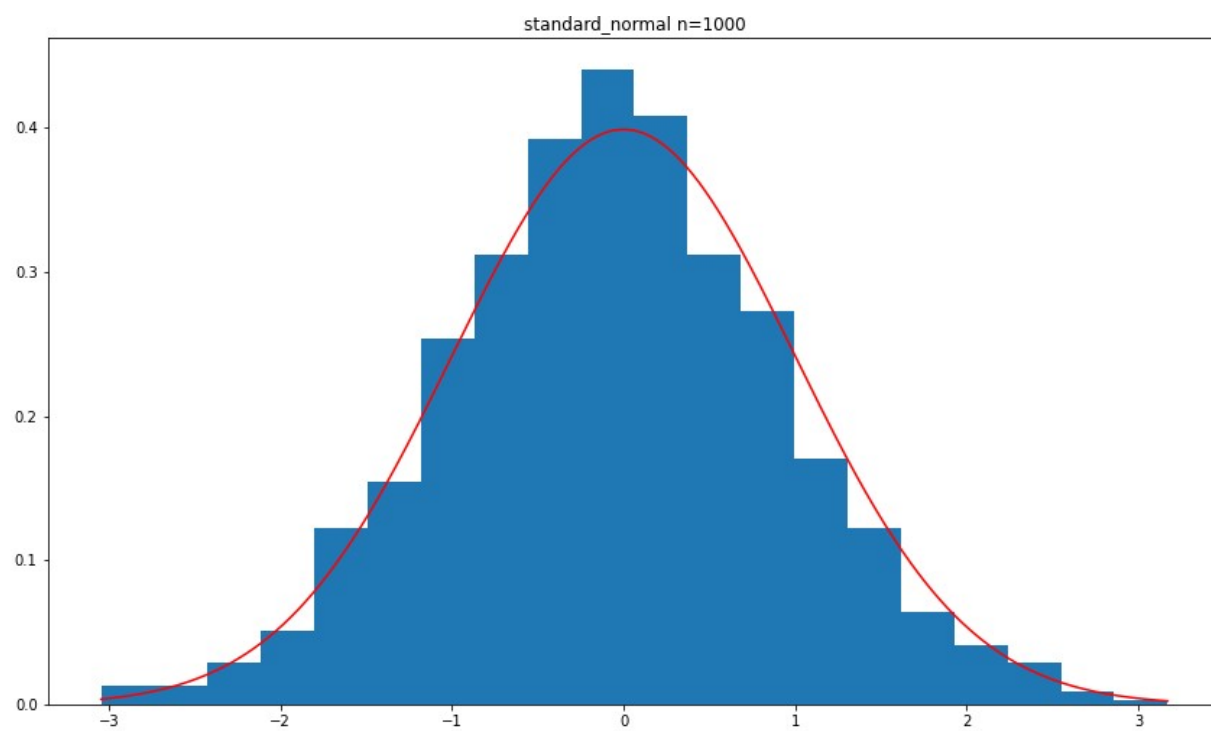




Рис. 5 Стандартное распределение Коши при  $n=10$

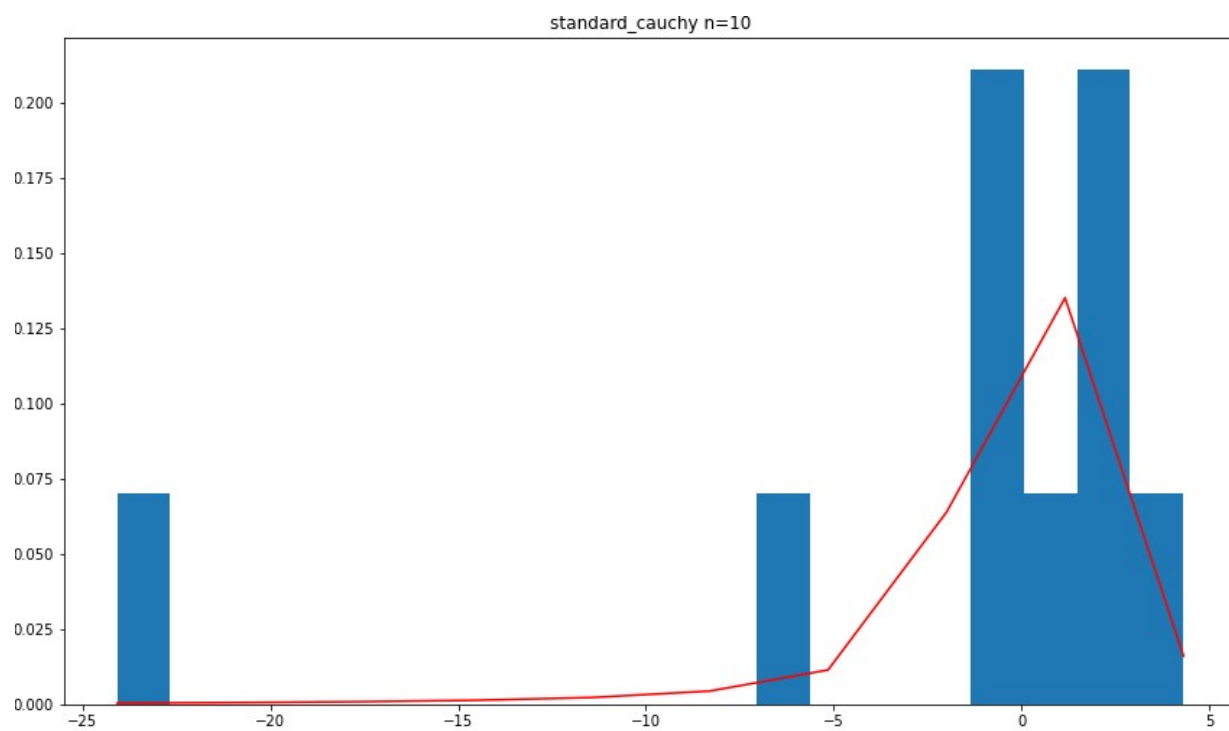


Рис. 6 Стандартное распределение Коши при  $n=50$

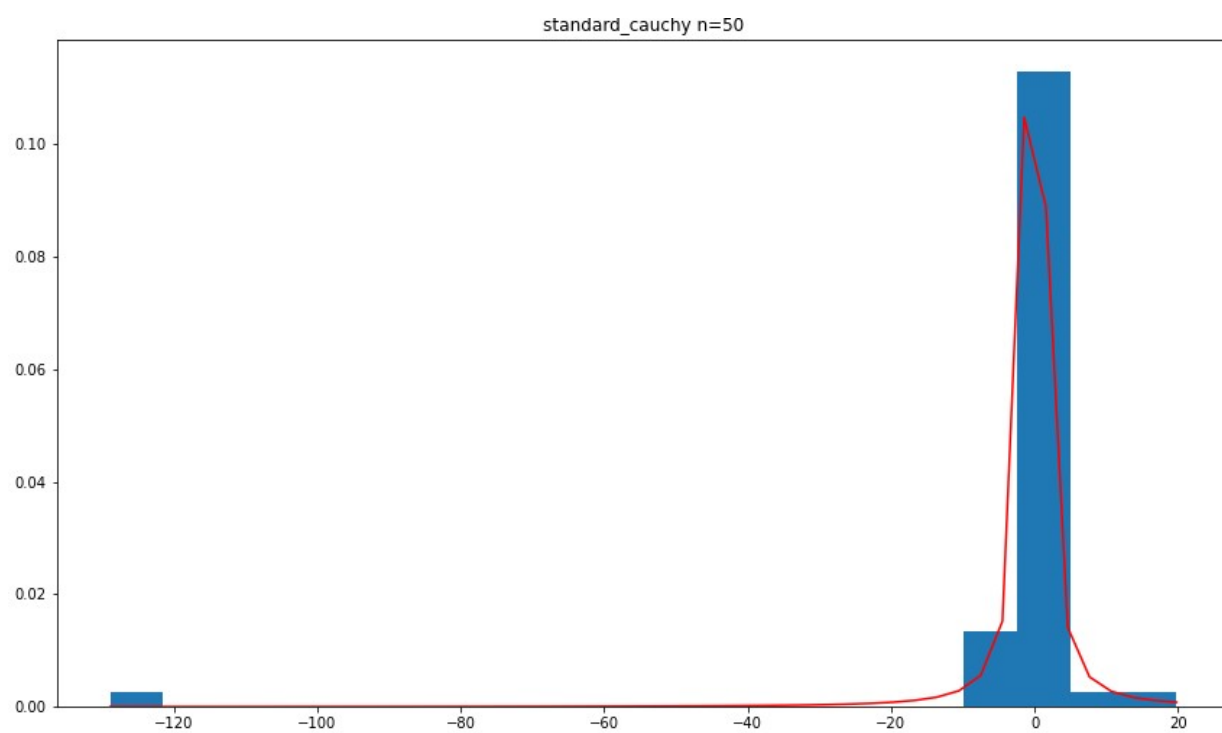


Рис. 7 Стандартное распределение Коши при  $n=100$

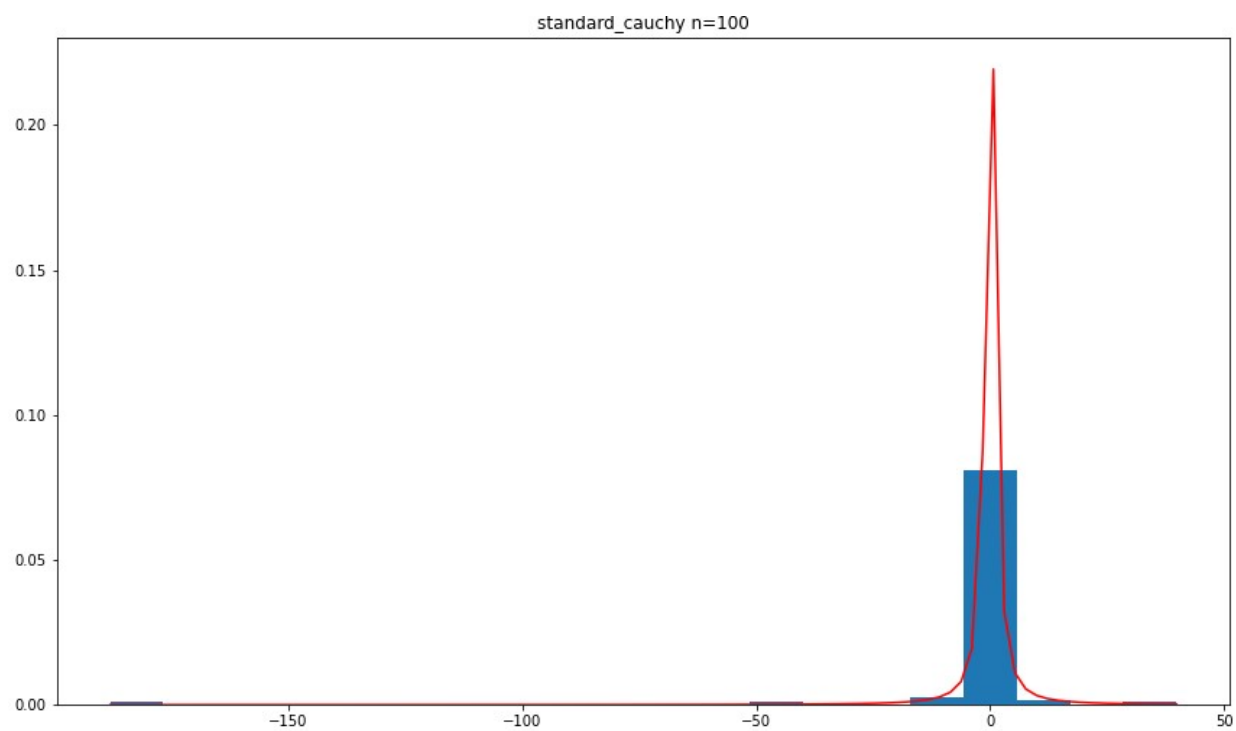


Рис. 8 Стандартное распределение Коши при  $n=1000$

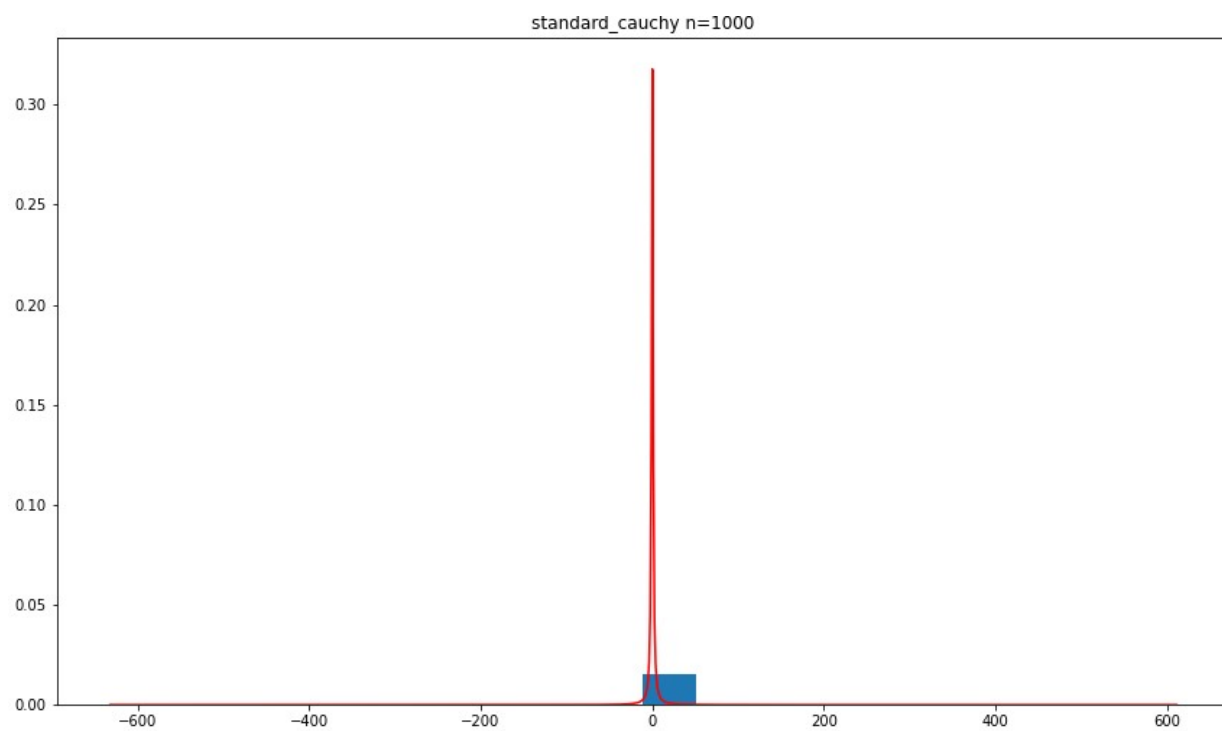


Рис. 9 Распределение Лапласа при  $n=10$

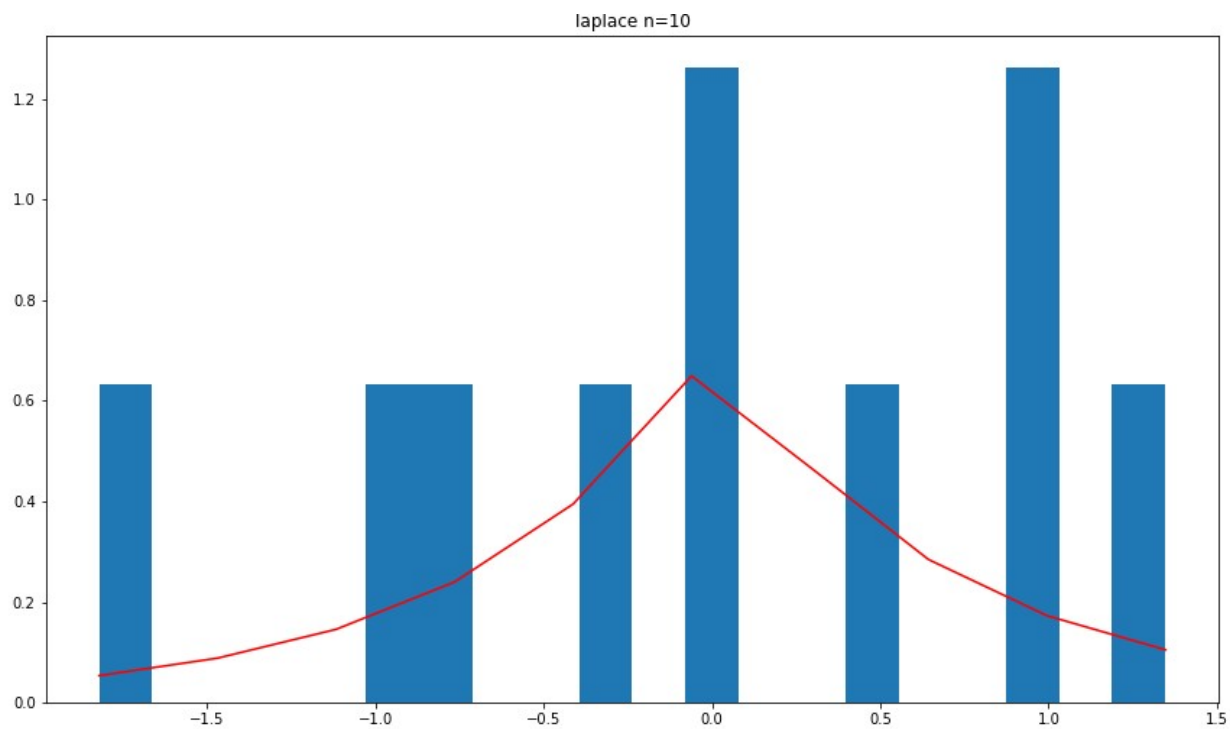


Рис. 10 Распределение Лапласа при  $n=50$

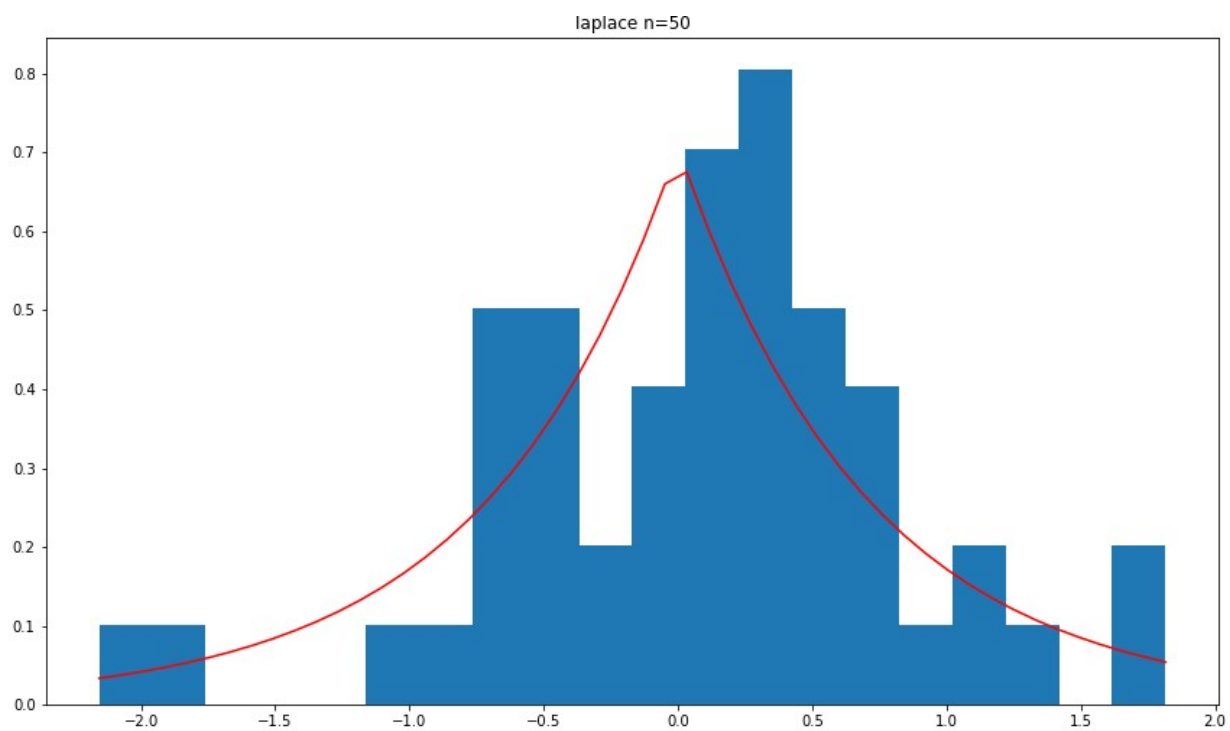


Рис. 11 Распределение Лапласа при  $n=100$

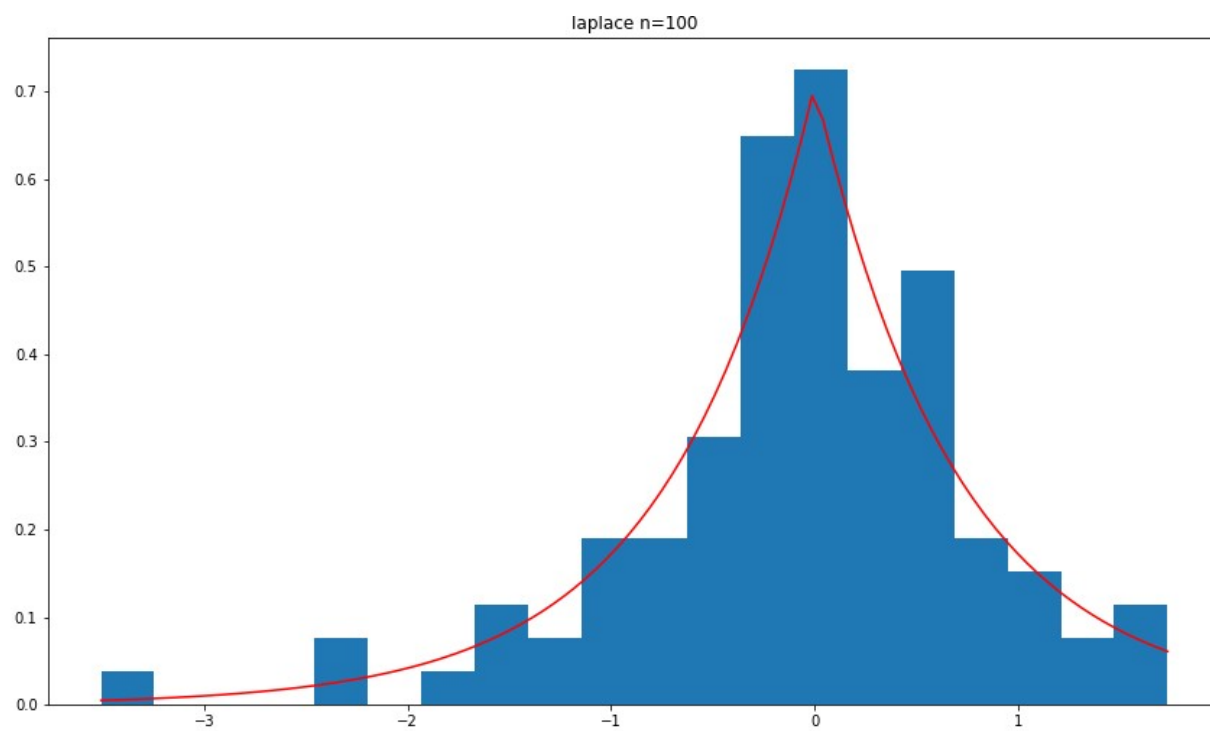


Рис. 12 Распределение Лапласа при  $n=1000$

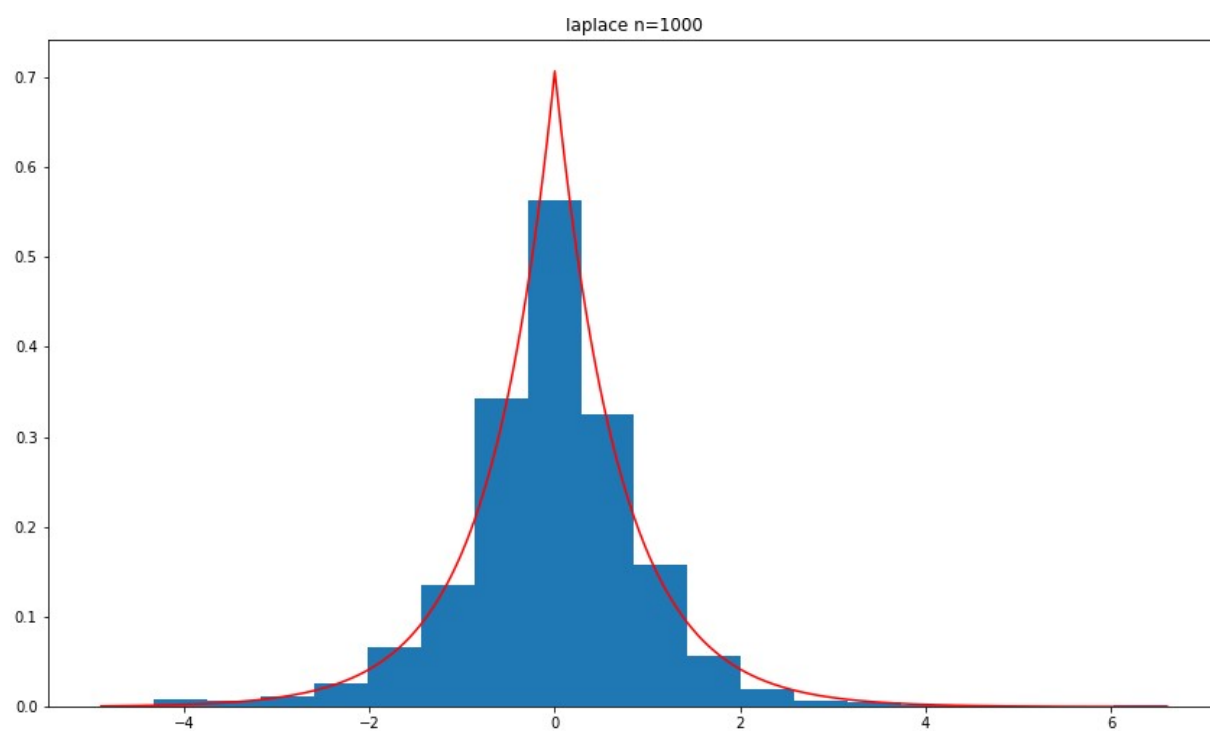


Рис. 13 Распределение Пуассона при  $n=10$

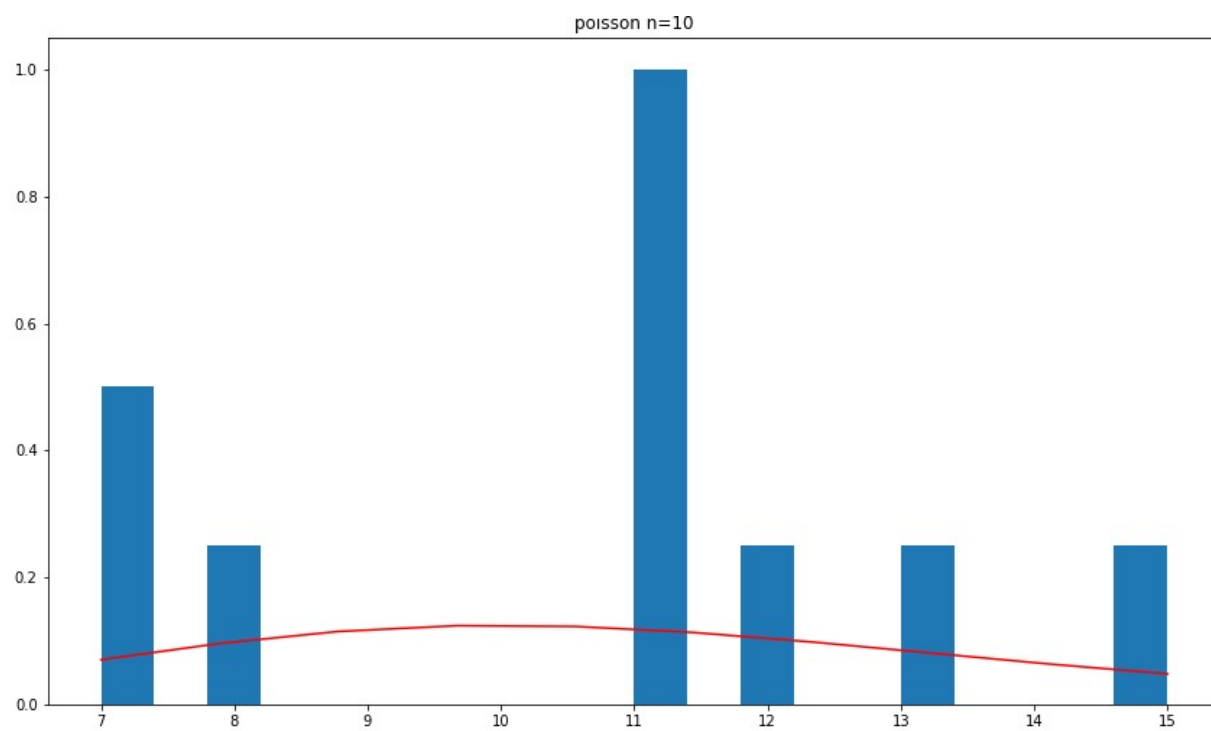


Рис. 14 Распределение Пуассона при  $n=50$

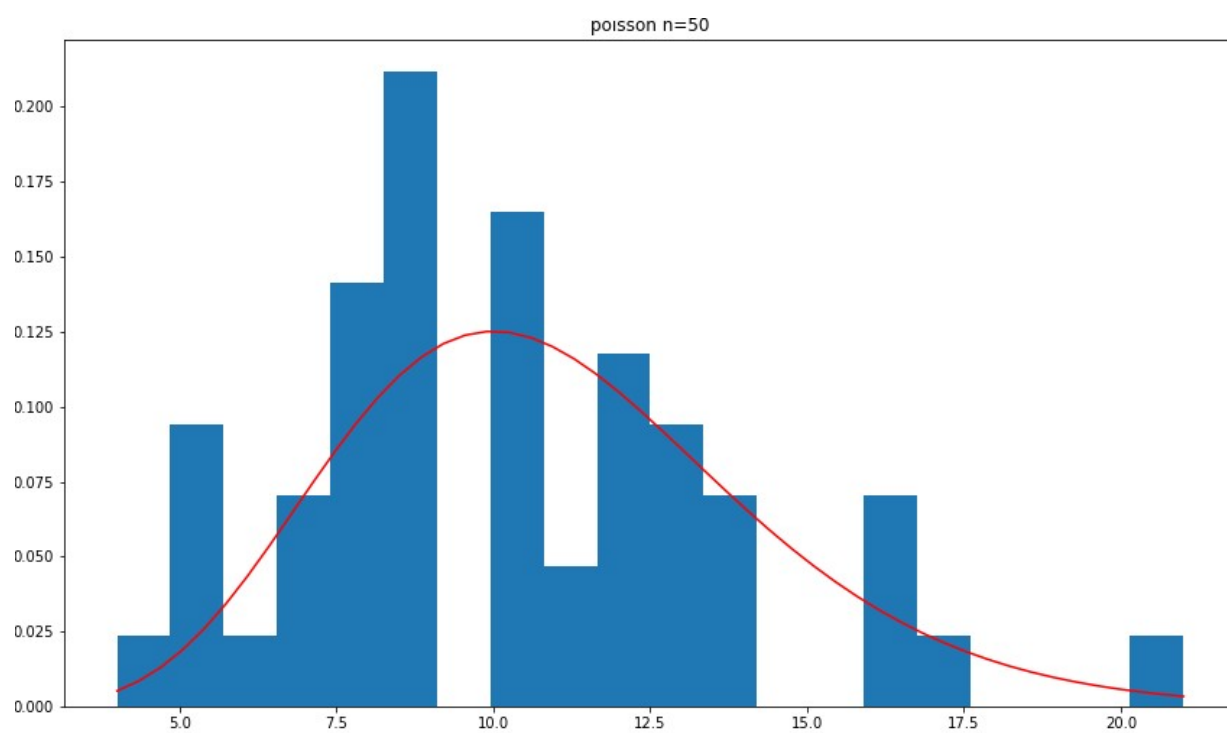


Рис. 15 Распределение Пуассона при  $n=100$

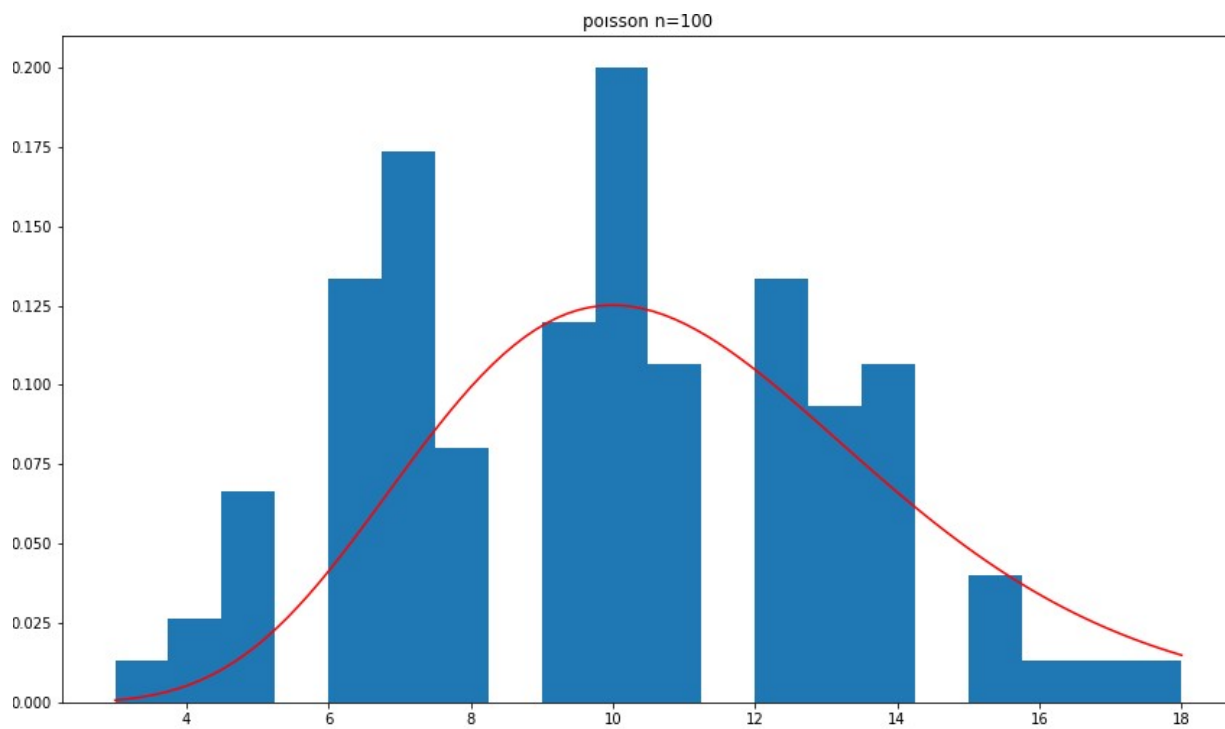


Рис. 16 Распределение Пуассона при  $n=1000$

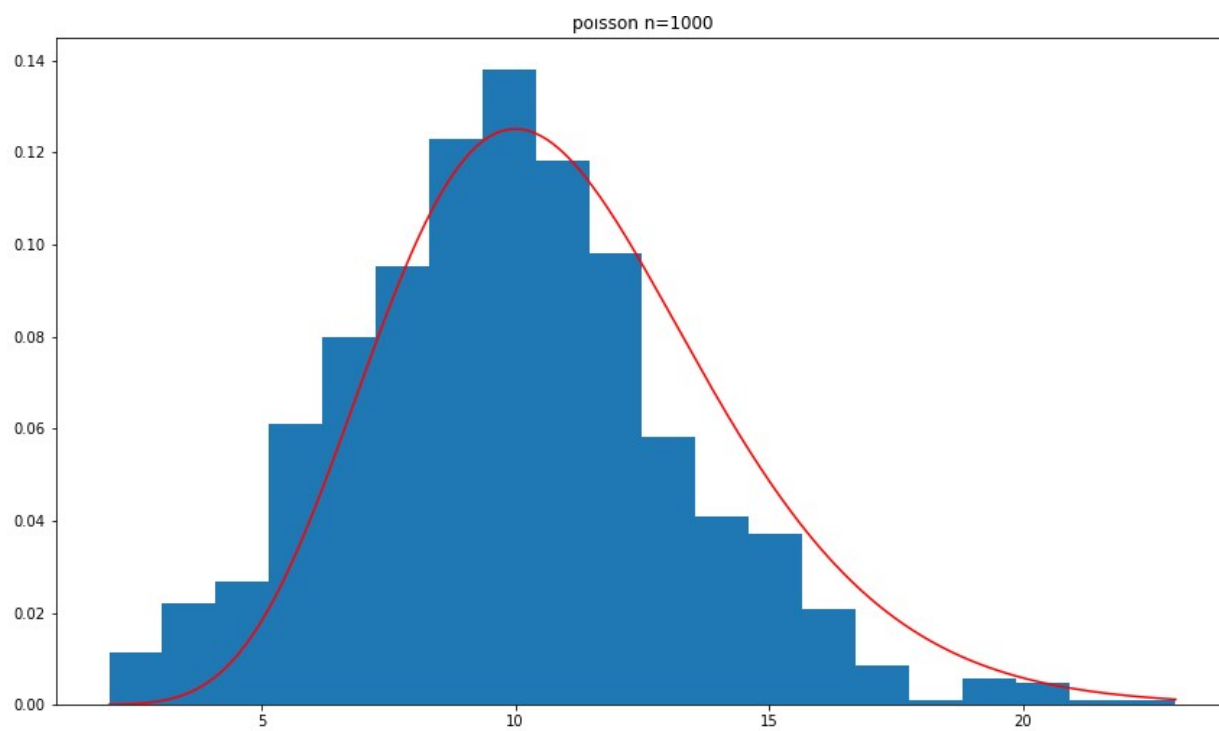


Рис. 17 Равномерное распределение при  $n=10$

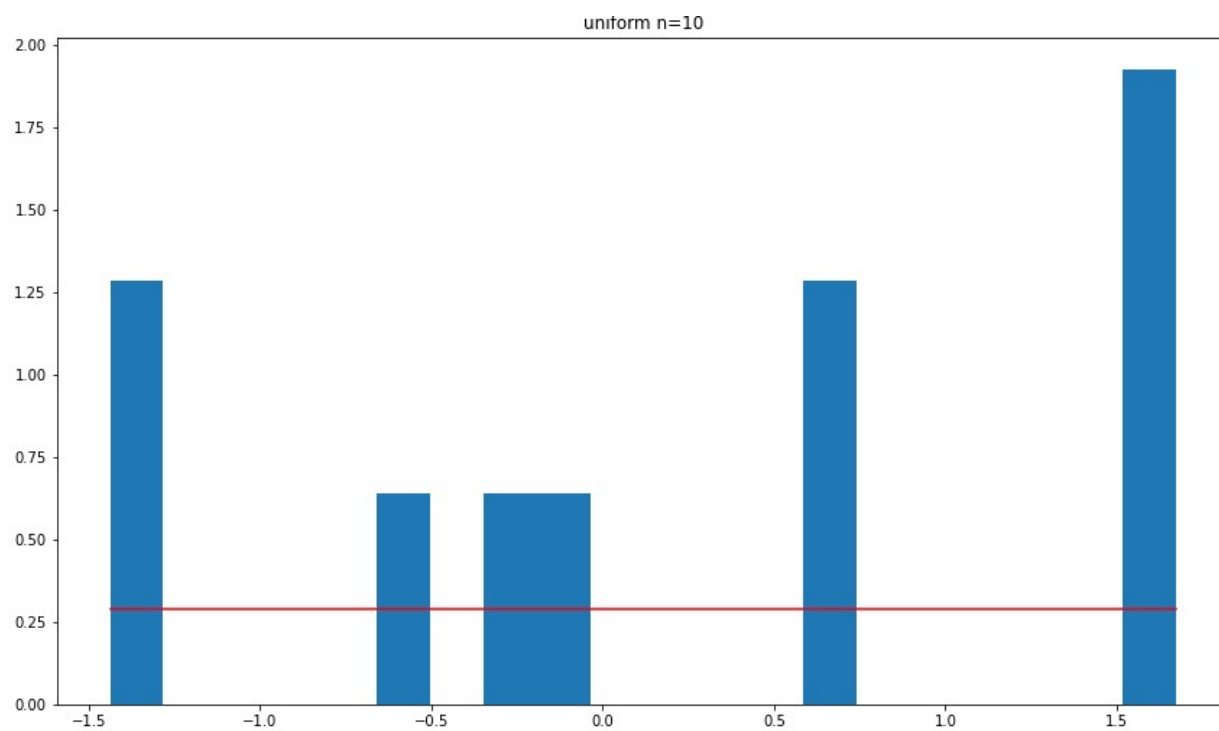


Рис. 18 Равномерное распределение при  $n=50$

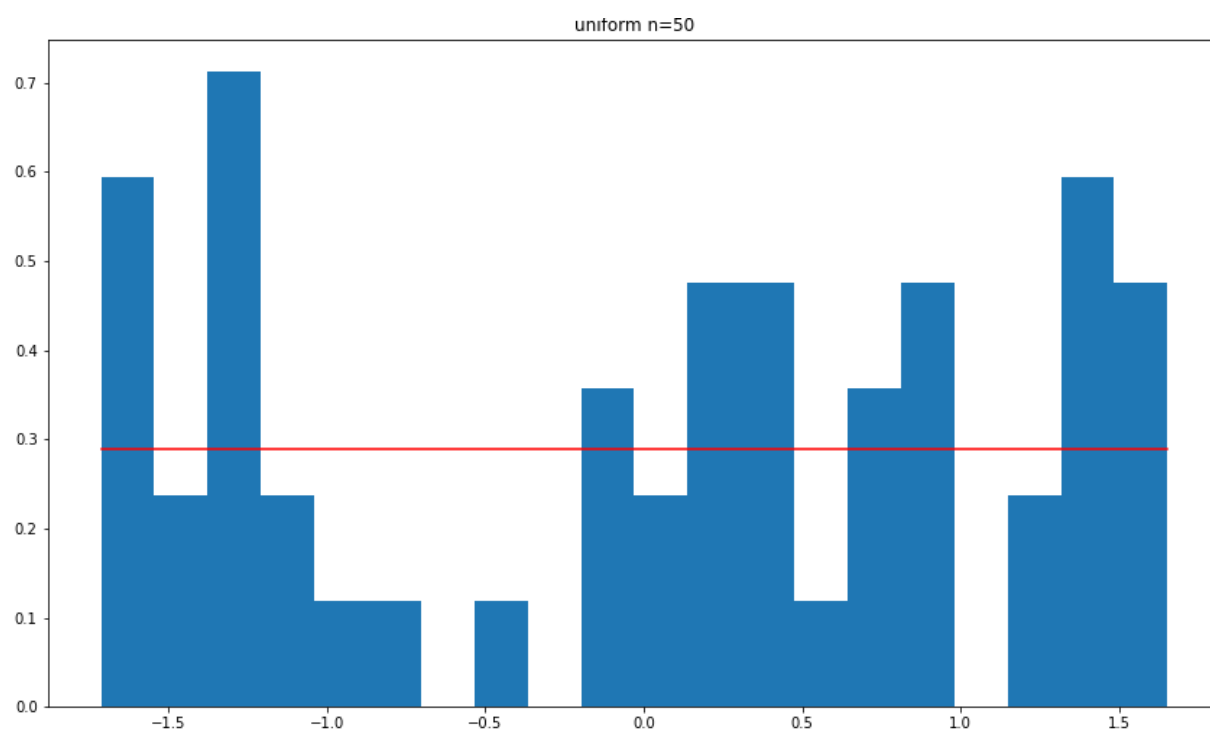


Рис. 19 Равномерное распределение при  $n=100$

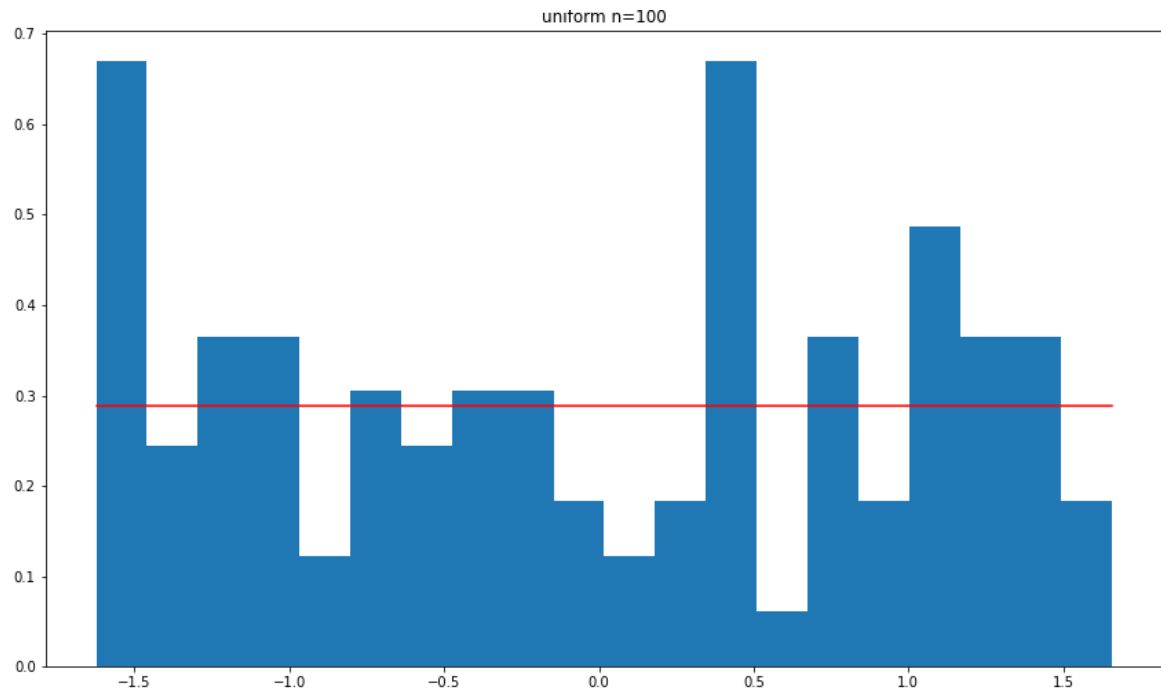
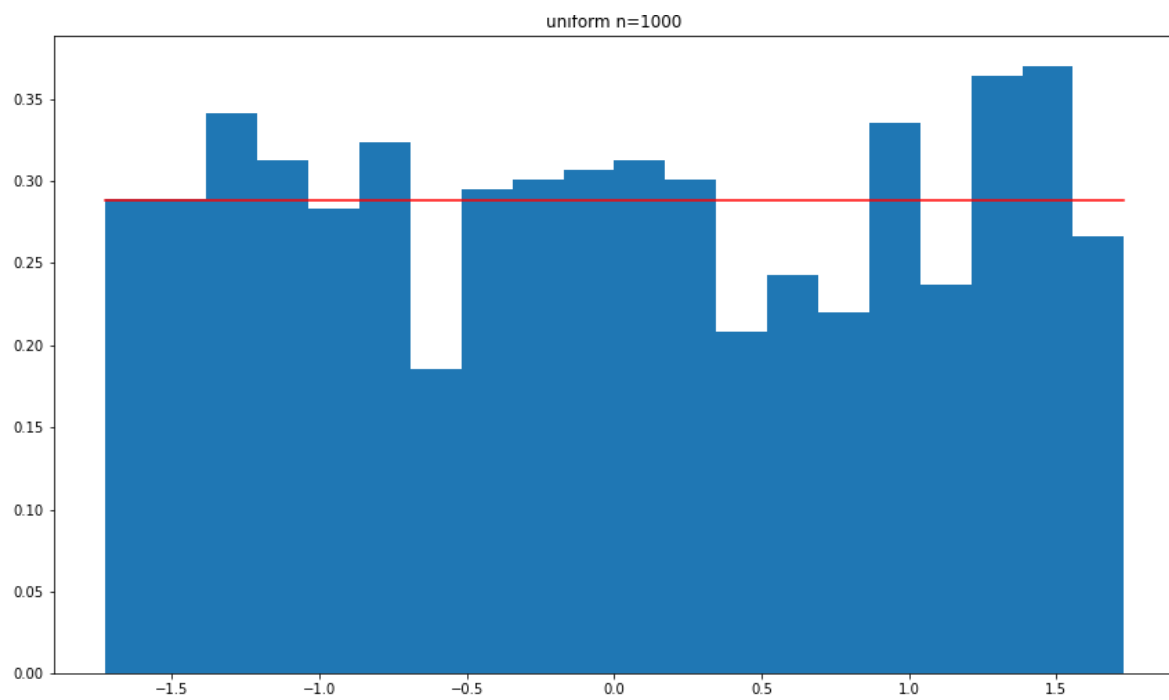


Рис. 20 Равномерное распределение при  $n=1000$



## 4.2 Характеристики положения и рассеивания

Таблица 1 Стандартное нормальное распределение

$n=10$	$avrg$	$med\ x$	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0.01	0.01	0.02	0	0.28
$D(z)$	0.097848	0.12874	0.190887	0.11015	0.109265
$n=100$	$avrg$	$med\ x$	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0	0	0	-0	0.03
$D(z)$	0.009964	0.015673	0.093959	0.012457	0.011819



n=1000	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0	0	-0	0	0
$D(z)$	0.00094	0.001552	0.062108	0.00116	0.001157

Таблица 2 Стандартное распределение Коши

n=10	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	-12.31	0	-61.48	-0.02	0.65
$D(z)$	161834.047069	0.276286	4045976.36516	0.750808	0.918042
n=100	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	-0.52	-0	-24.32	-0.01	0.03
$D(z)$	875.221257	0.02564	2090211.365057	0.05555	0.028177
n=1000	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	-0.4	0	-198.59	-0	0
$D(z)$	520.951391	0.002476	128259873.764481	0.004587	0.002515

Таблица 3 Распределение Лапласа

n=10	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0	0.01	0	0	0.23
$D(z)$	0.098106	0.066493	0.405711	0.085741	0.078464
n=100	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0	0	-0	0	0.02
$D(z)$	0.010295	0.005968	0.419241	0.010031	0.006175
n=1000	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0	0	0.02	0	0
$D(z)$	0.00102	0.000534	0.404961	0.00103	0.000639

Таблица 4 Распределение Пуассона

n=10	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	9.98	9.82	10.29	9.89	10.74
$D(z)$	0.939362	1.378308	1.653056	1.114103	1.202277
n=100	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	10.01	9.85	11	9.91	9.95
$D(z)$	0.098454	0.202596	0.987738	0.156002	0.121359
n=1000	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	10	10	11.66	9.99	9.87
$D(z)$	0.010283	0.002991	0.689351	0.004562	0.010918

Таблица 5 Равномерное распределение

n=10	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	-0.03	-0.03	-0.01	-0.03	0.29
$D(z)$	0.09994	0.228635	0.048268	0.138834	0.156282
n=100	avg	med x	$Z_R$	$Z_Q$	$Z_{tr}$

$E(z)$	0	0	0	0	0.04
$D(z)$	0.009809	0.028060	0.000538	0.014772	0.019582
n=1000	<i>avrg</i>	<i>med x</i>	$Z_R$	$Z_Q$	$Z_{tr}$
$E(z)$	0	-0	-0	0	0
$D(z)$	0.00096	0.00287	0.000006	0.001504	0.001887

4.3 Боксплот Тьюки

Рис. 21 Стандартное нормальное распределение

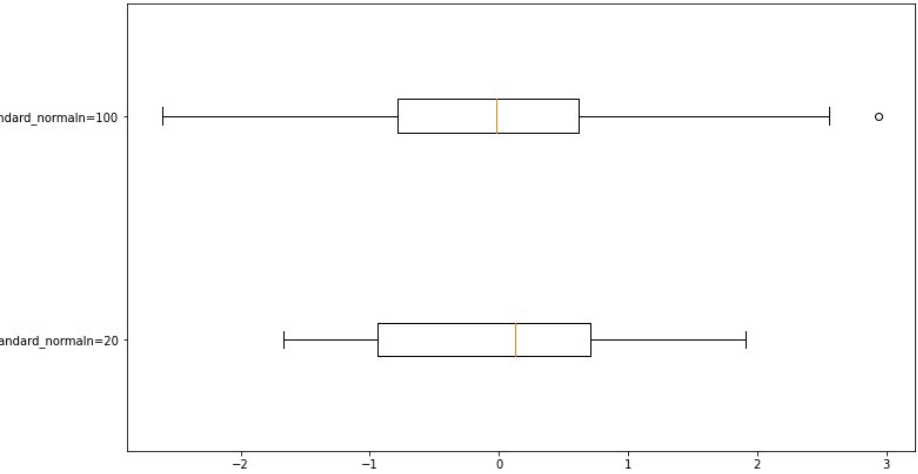


Рис. 22 Стандартное распределение Коши

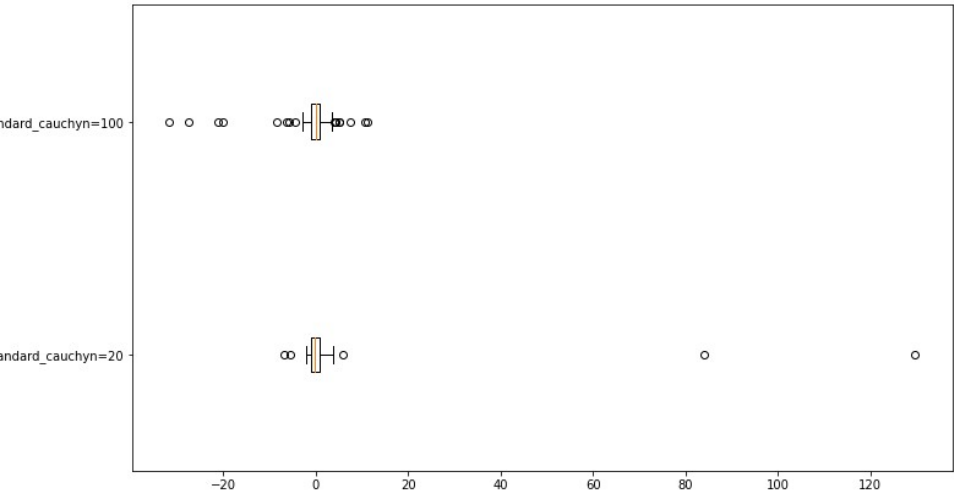


Рис. 23 Распределение Лапласа

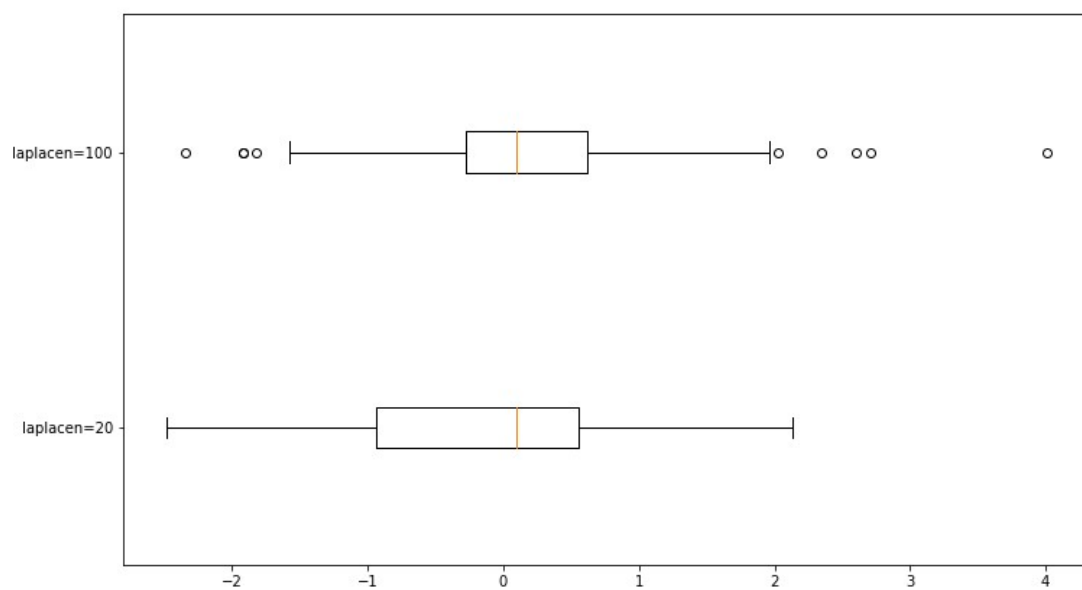


Рис. 24 Распределение Пуассона

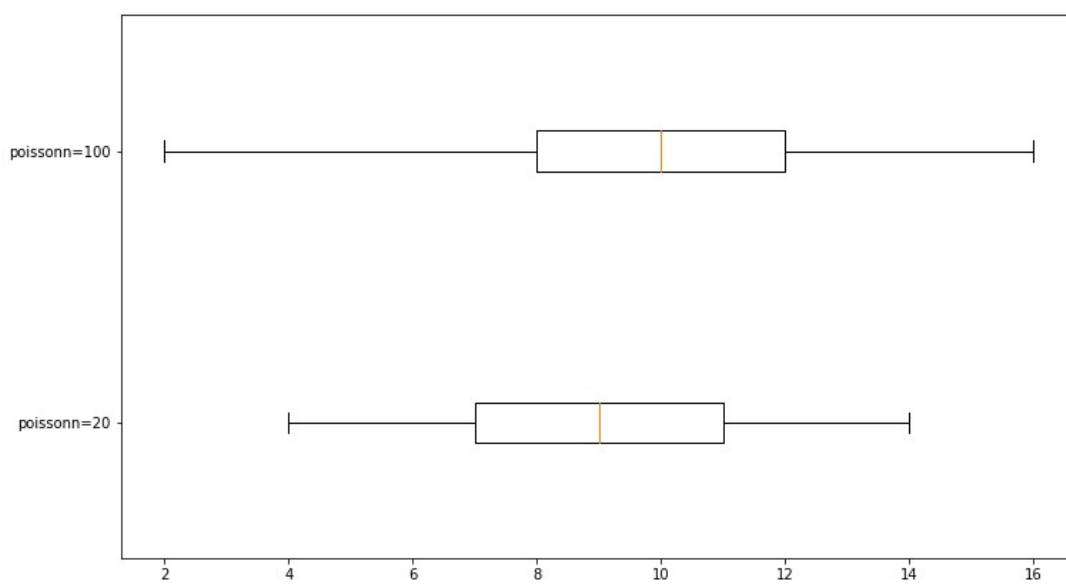


Рис. 25 Равномерное распределение

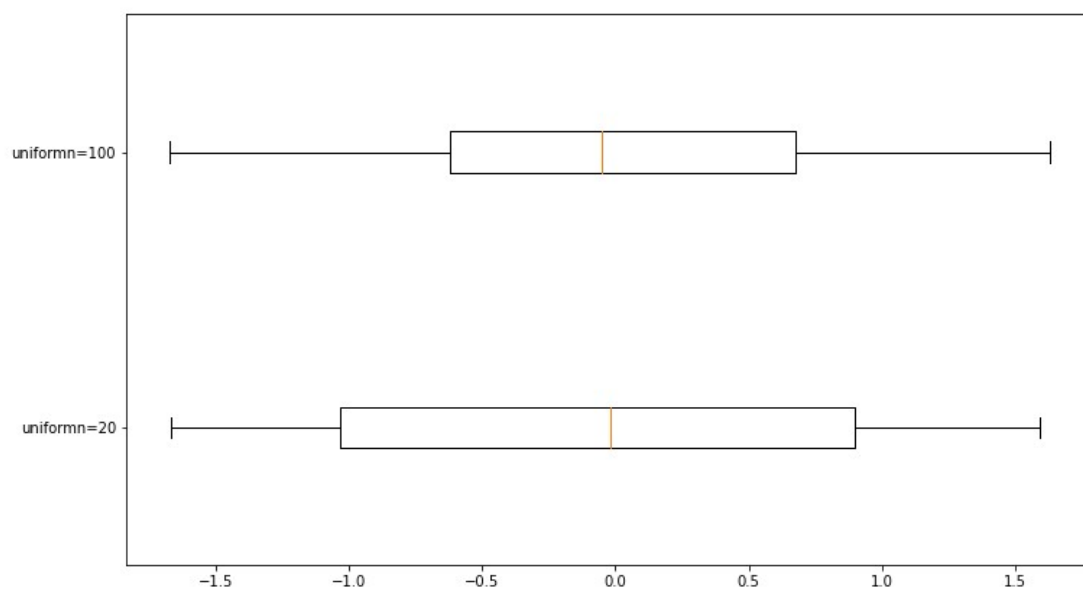


Таблица 6 Средние проценты выбросов

Выборка	Средний процент выбросов
Стандартное нормальное распределение (1.1)	
N = 20	2
N = 100	1
Стандартное распределение Коши (1.2)	
N = 20	15
N = 100	16
Распределение Лапласа (1.3)	
N = 20	8
N = 100	6
Распределение Пуассона (1.4)	
N = 20	1
N = 100	0
Равномерное распределение (1.5)	
N = 20	0
N = 100	0

## 4.4 Эмпирические функции распределения (ЭФР)

Рис. 26 ЭФР для Стандартного нормального распределения

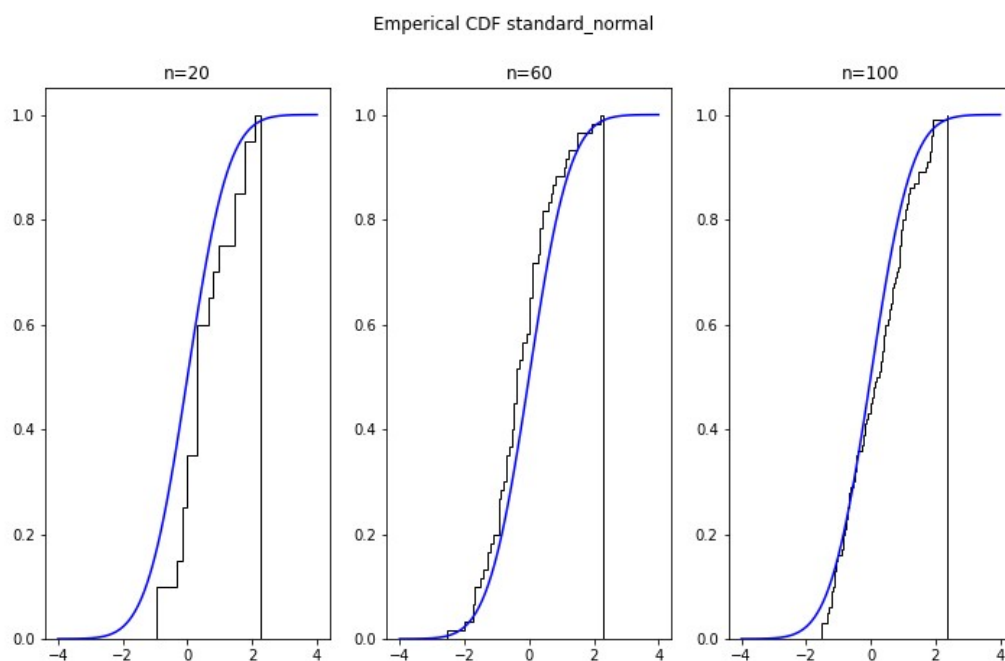


Рис. 27 ЭФР для Стандартного распределения Коши

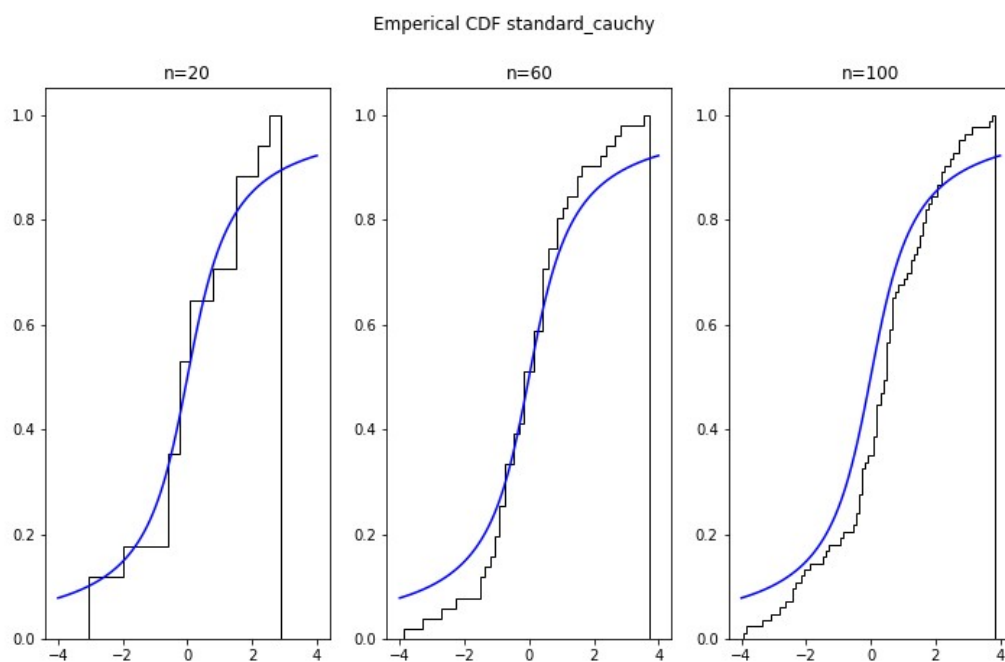


Рис. 28 ЭФР для распределения Лапласа

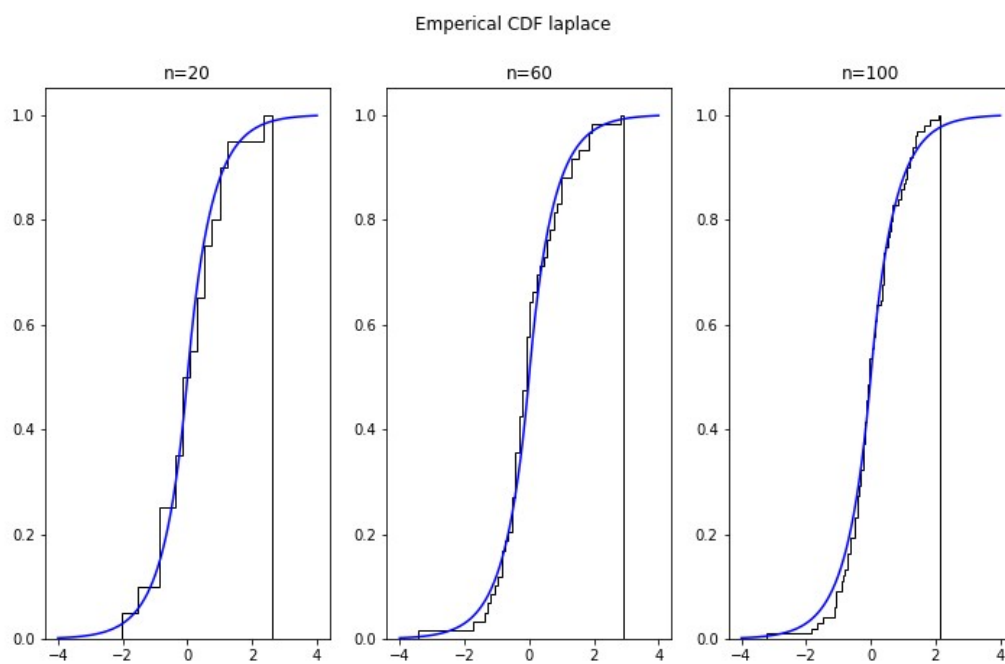


Рис. 29 ЭФР для Равномерного распределения

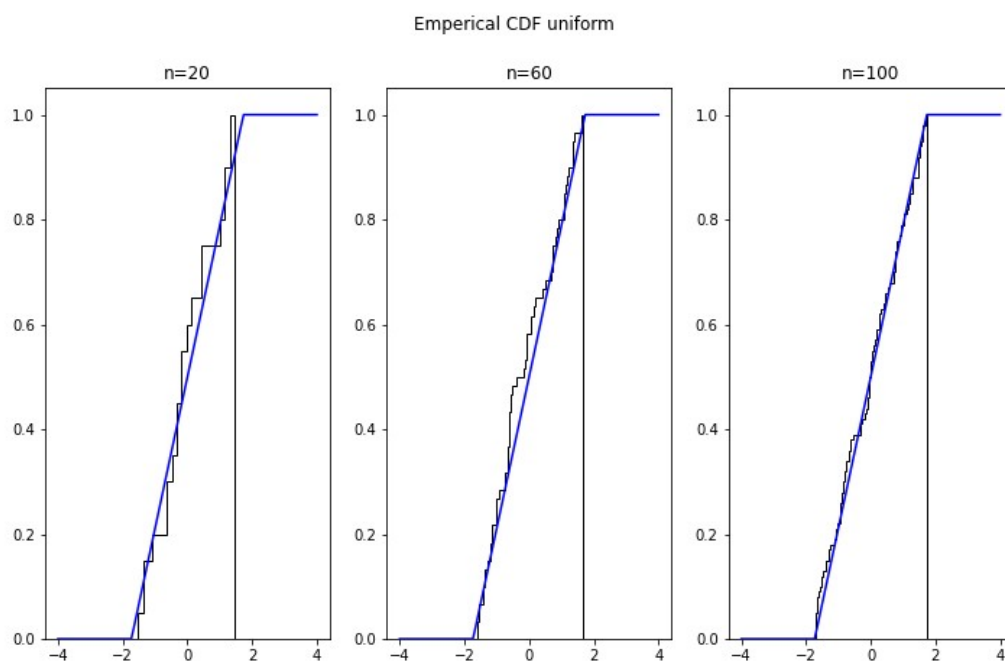
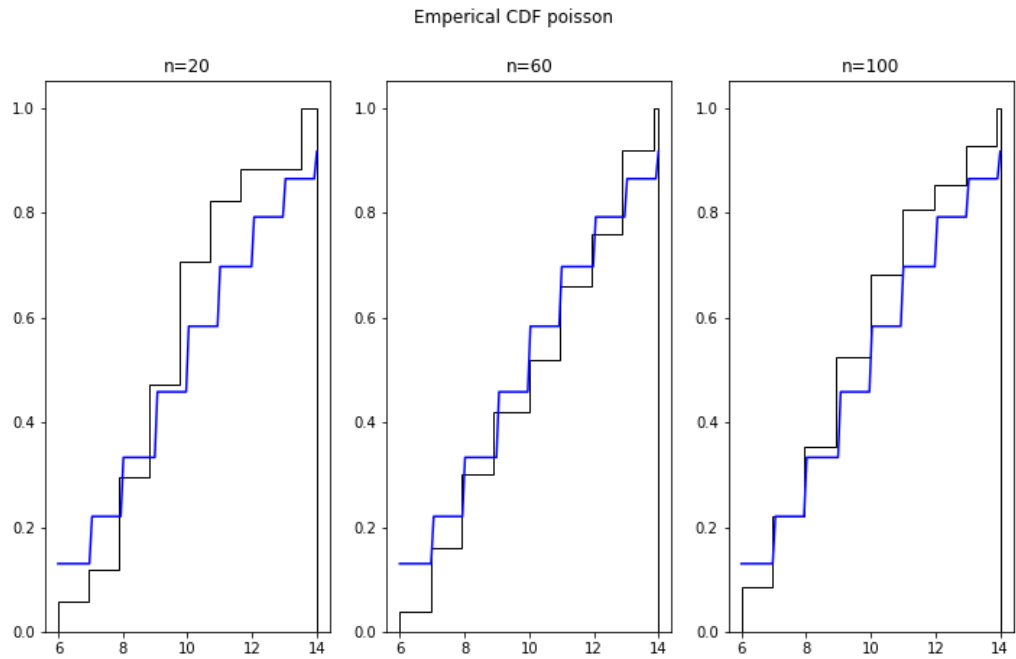


Рис. 30 ЭФР для распределения Пуассона



#### 4.5 Ядерные функции плотности (ЯФП)

Рис. 31 ЯФП для Стандартного нормального распределения при  $n=20$

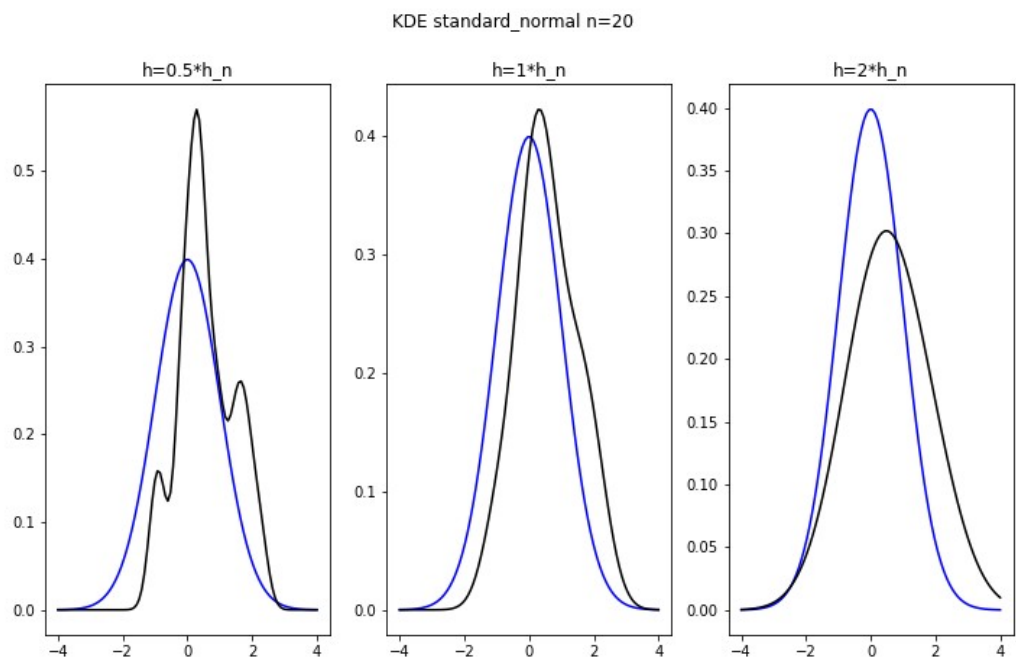


Рис. 32 ЯФП для Стандартного нормального распределения при  $n=60$

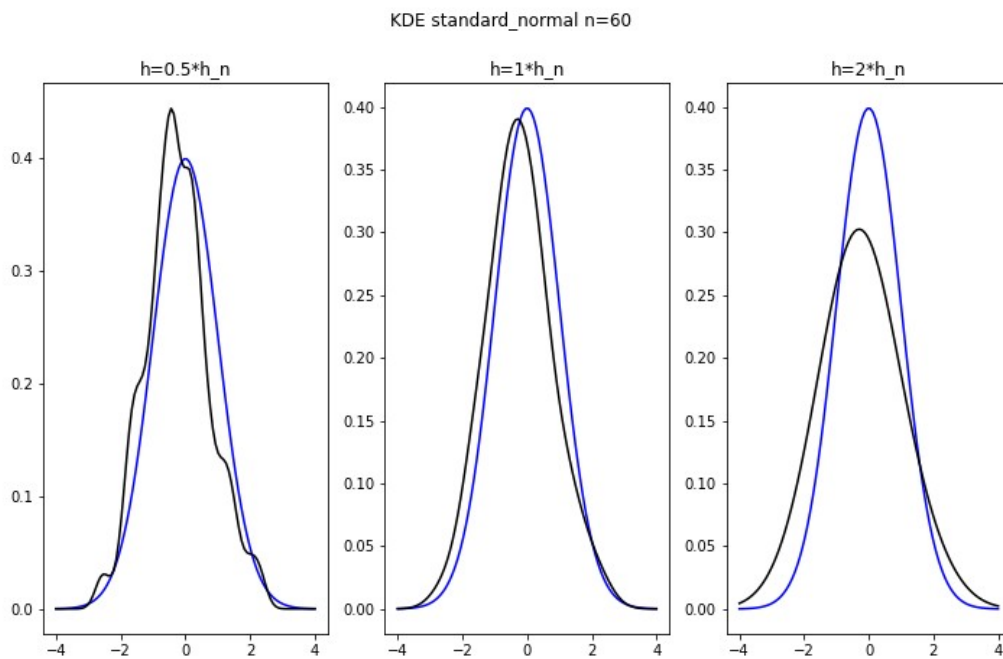


Рис. 33 ЯФП для Стандартного нормального распределения при  $n=100$

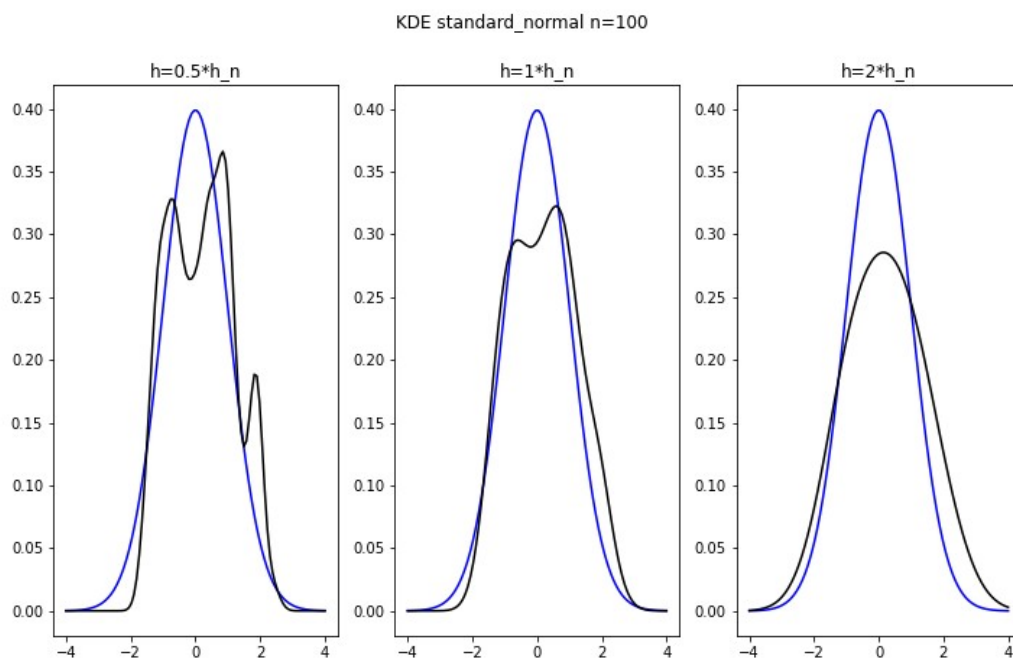




Рис. 34 ЯФП для Стандартного распределения Коши при  $n=20$

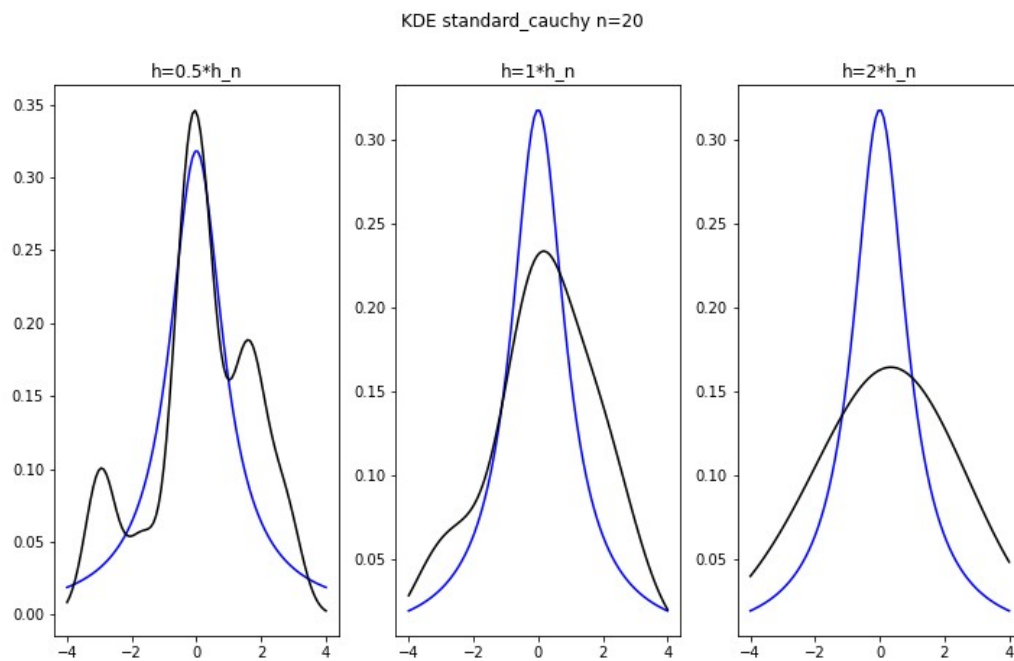


Рис. 35 ЯФП для Стандартного распределения Коши при  $n=60$

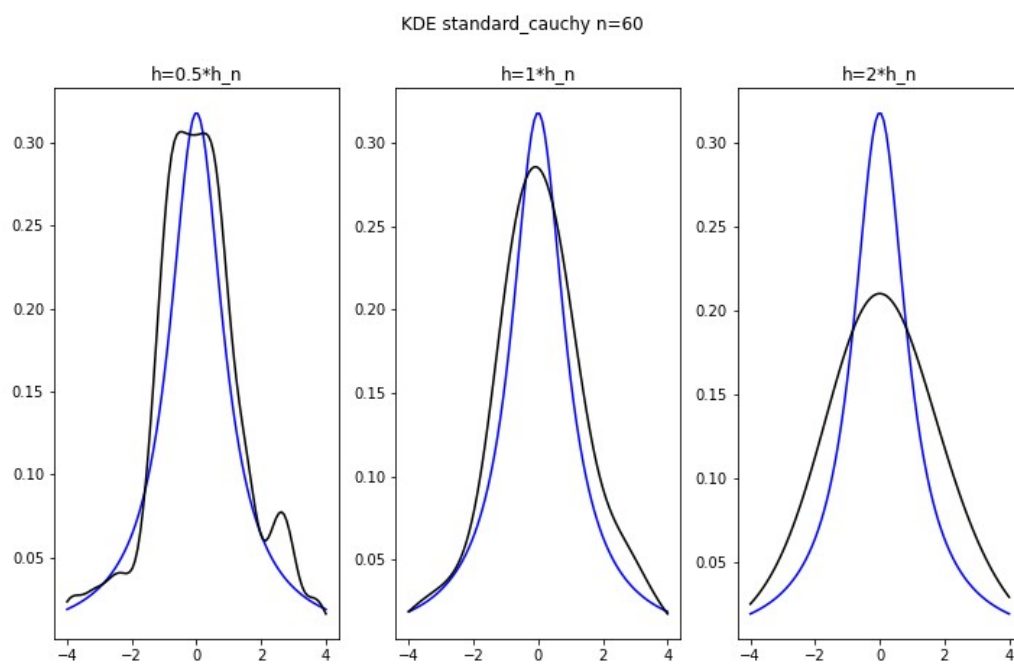


Рис. 36 ЯФП для Стандартного распределения Коши при  $n=100$

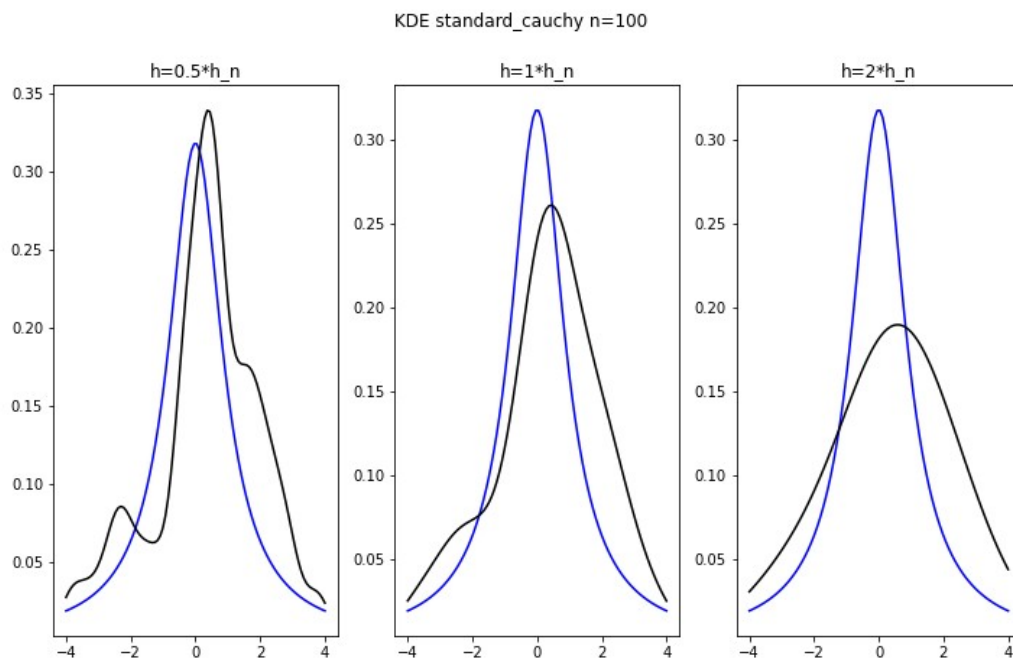


Рис. 37 ЯФП для распределения Лапласа при  $n=20$

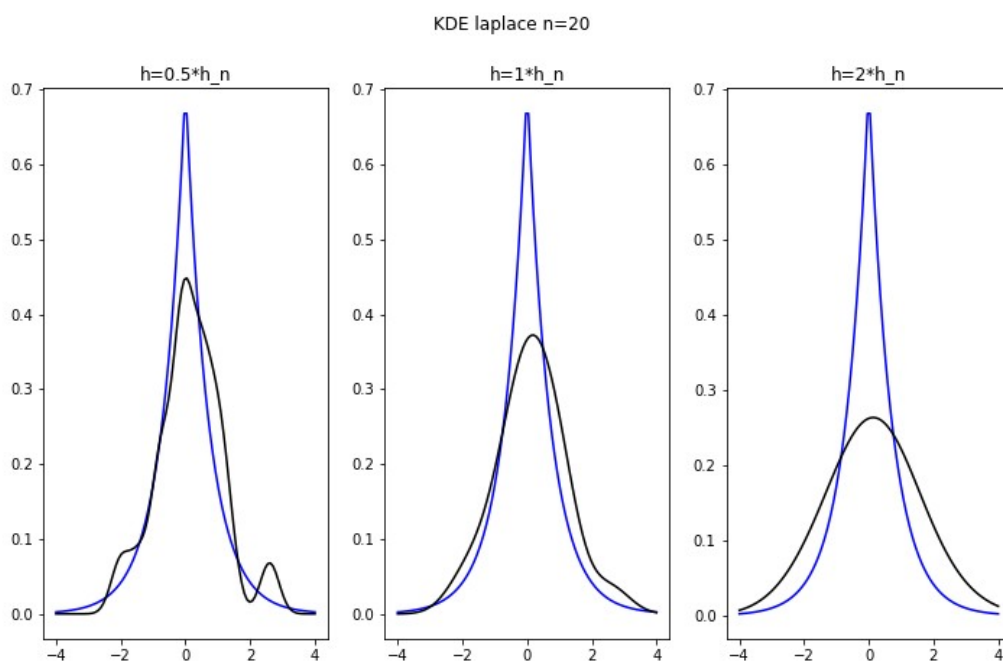


Рис. 38 ЯФП для распределения Лапласа при  $n=60$

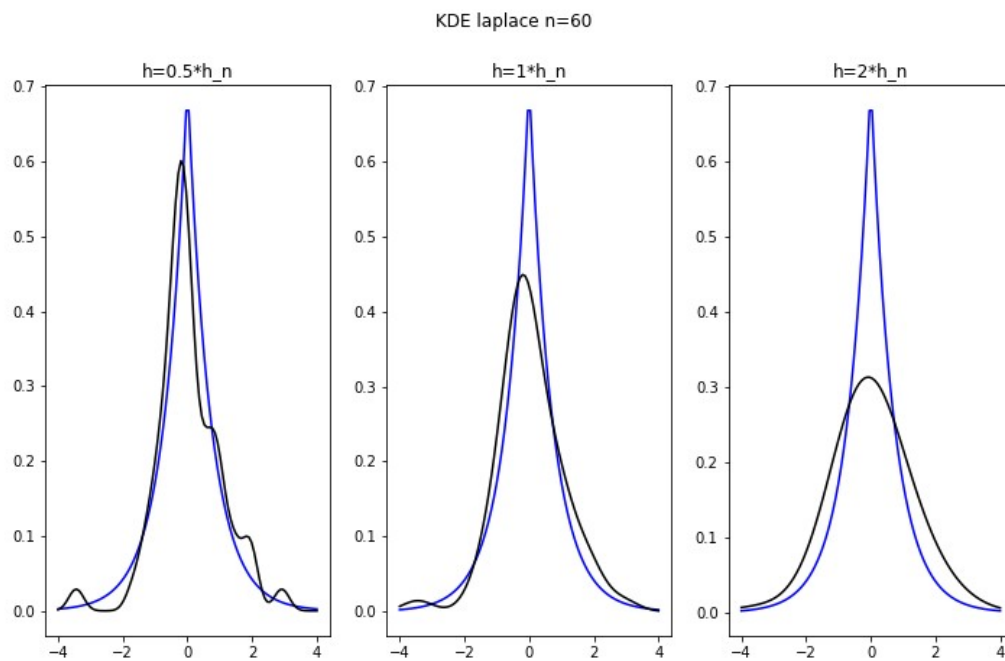


Рис. 39 ЯФП для распределения Лапласа при  $n=100$

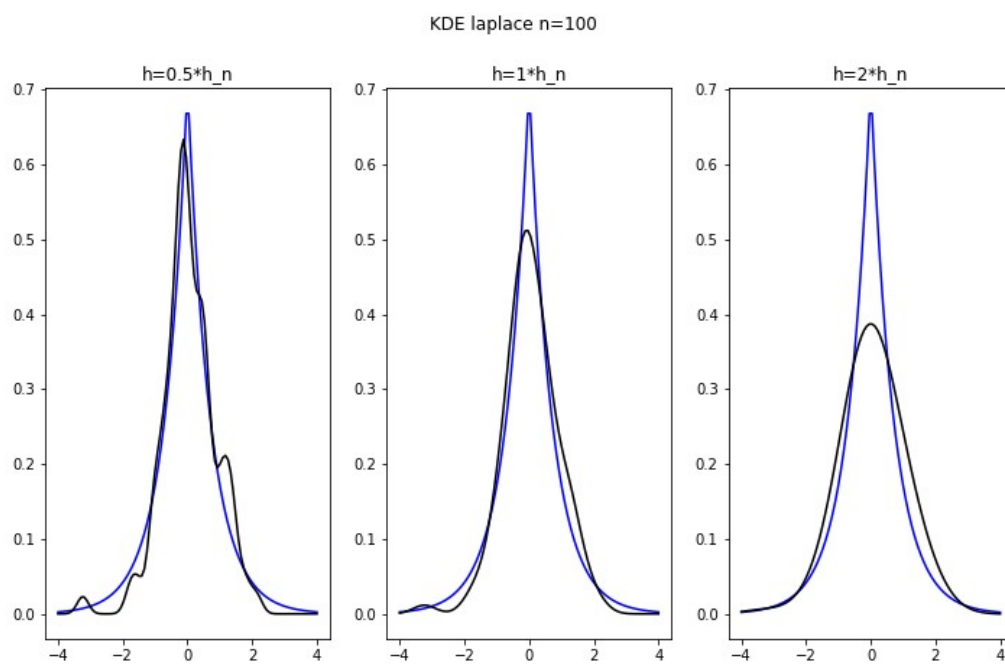


Рис. 40 ЯФП для Равномерного распределения при  $n=20$

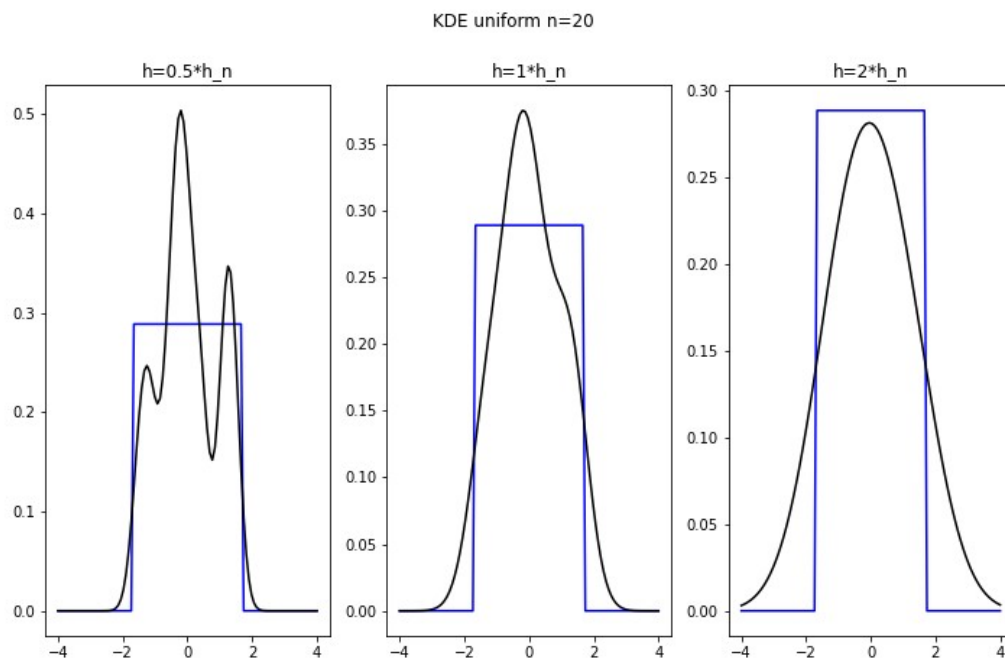


Рис. 41 ЯФП для Равномерного распределения при  $n=60$

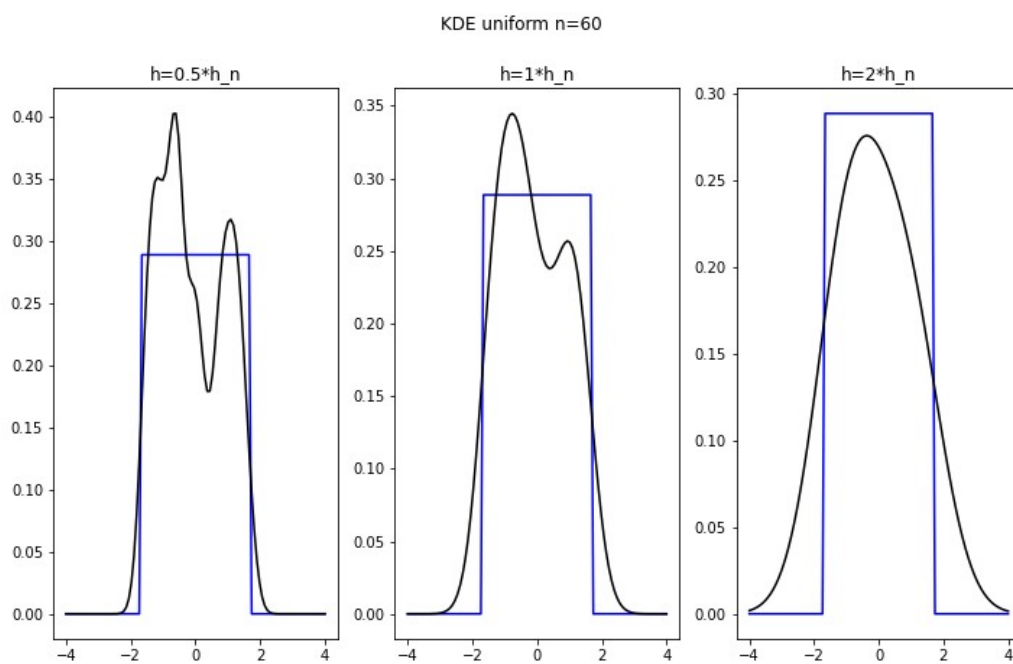


Рис. 42 ЯФП для Равномерного распределения при  $n=100$

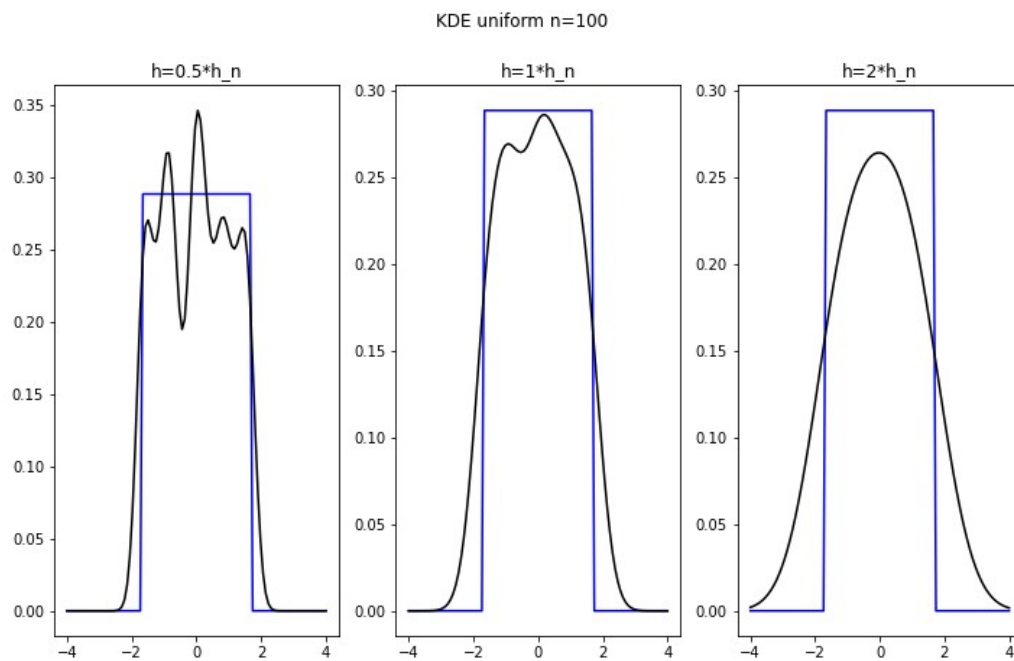


Рис. 43 ЯФП для распределения Пуассона при  $n=20$

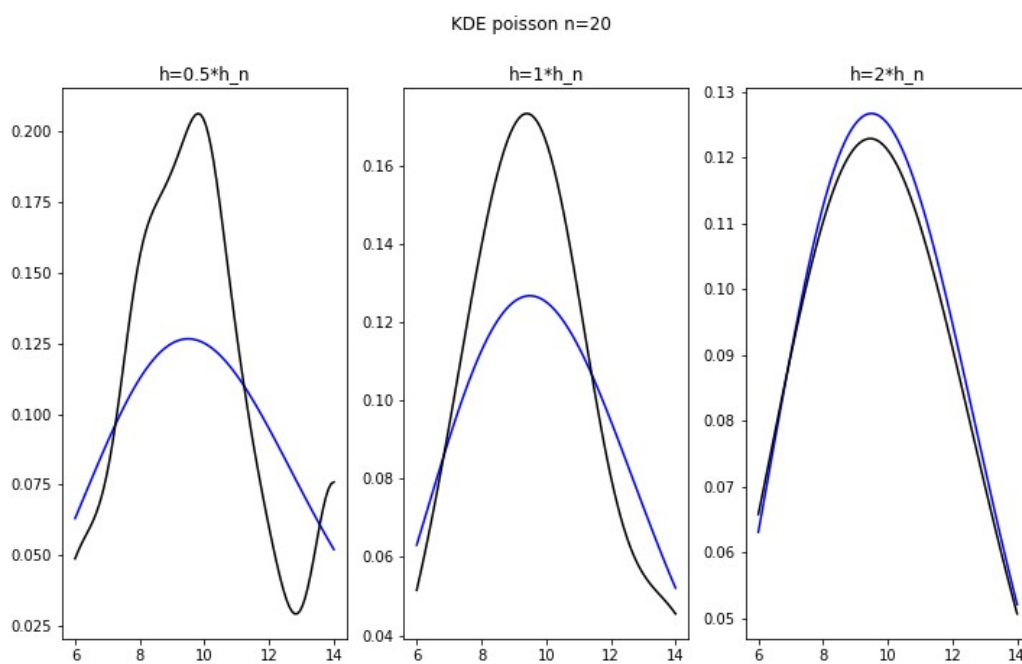


Рис. 44 ЯФП для распределения Пуассона при  $n=60$

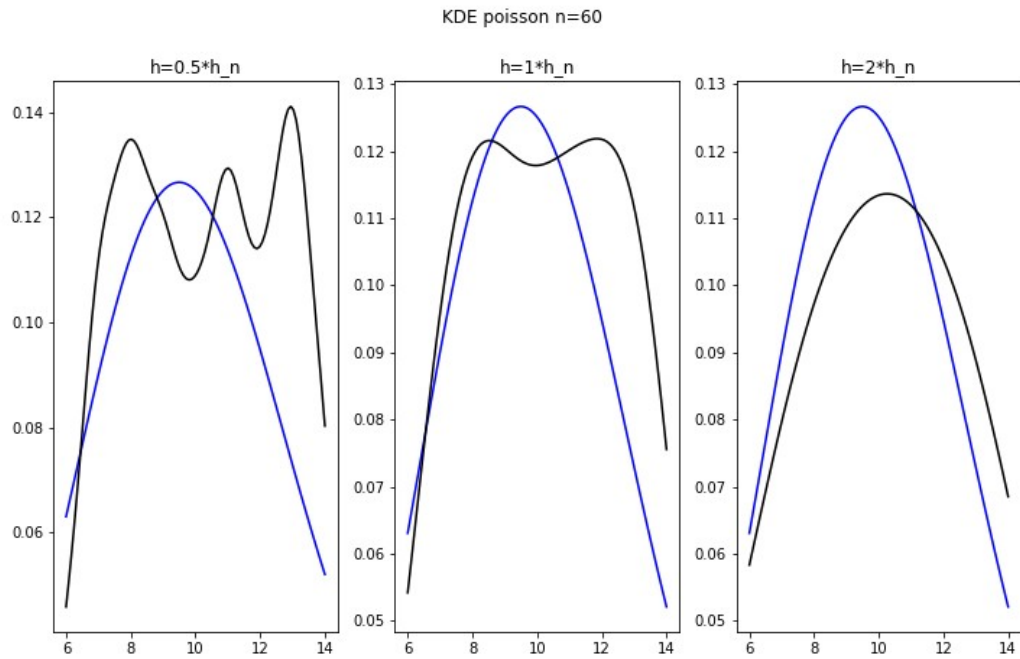
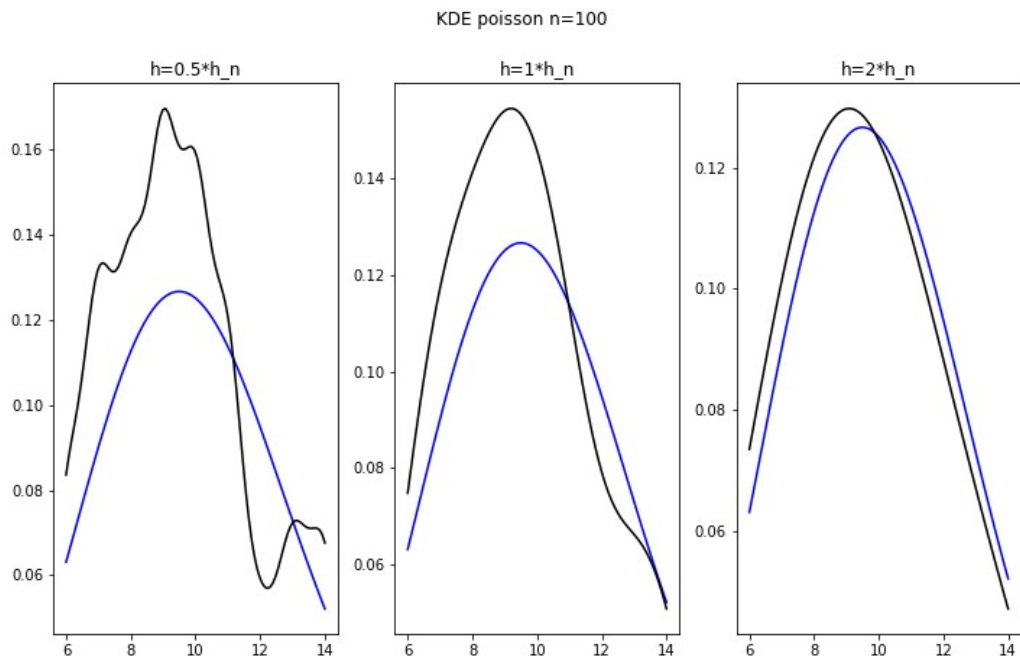


Рис. 45 ЯФП для распределения Пуассона при  $n=100$



#### 4.6 Таблицы. Средние модули ошибок

Строки – мощности выборок, столбцы – коэффициенты при  $h_n$ , маркером помечены минимальные значения.

Таблица 7 Стандартное нормальное распределение

$E(z)$	0.5	1.0	2.0
$n=20$	0.061887	0.051959	0.052813

<b>n=60</b>	0.028175	0.026624	0.039276
<b>n=100</b>	0.035016	0.028130	0.037715

Таблица 8 Стандартное распределение Коши

<b>E(z)</b>	<b>0.5</b>	<b>1.0</b>	<b>2.0</b>
<b>n=20</b>	0.036591	0.037148	0.052669
<b>n=60</b>	0.022031	0.021232	0.039087
<b>n=100</b>	0.046387	0.042639	0.048923

Таблица 9 Распределение Лапласа

<b>E(z)</b>	<b>0.5</b>	<b>1.0</b>	<b>2.0</b>
<b>n=20</b>	0.035431	0.040874	0.067724
<b>n=60</b>	0.032295	0.030991	0.052448
<b>n=100</b>	0.024103	0.023139	0.034833

Таблица 10 Равномерное распределение

<b>E(z)</b>	<b>0.5</b>	<b>1.0</b>	<b>2.0</b>
<b>n=20</b>	0.041468	0.039627	0.052521
<b>n=60</b>	0.030078	0.033001	0.051592
<b>n=100</b>	0.025169	0.030812	0.054162

Таблица 11 Распределение Пуассона

<b>E(z)</b>	<b>0.5</b>	<b>1.0</b>	<b>2.0</b>
<b>n=20</b>	0.036118	0.020863	0.002852
<b>n=60</b>	0.022694	0.013810	0.011530
<b>n=100</b>	0.025138	0.017411	0.006402

## 5 Выводы

### 5.1 Функции плотности вероятности и гистограммы

При увеличении мощности выборки построенная гистограмма точнее приближает график функции плотности соответствующего распределения.

### 5.2 Характеристики положения выборки

В процессе работы вычислены значения характеристик положения для каждого из 5 распределений на выборках фиксированных мощностей и получены следующее ранжирование характеристик положения:

1. Стандартное нормальное распределение:  
 $Z_R < Z_{tr} < med\ x < avrg < Z_Q$
2. Стандартное распределение Коши:  
 $Z_R < avrg < Z_Q < med\ x < Z_{tr}$
3. Распределение Лапласа:  
 $med\ x < avrg < Z_Q < Z_{tr} < Z_R$
4. Распределение Пуассона:

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

5. Равномерное распределение:

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

### 5.3 Боксплот Тьюки

Экспериментально полученные проценты выбросов, близки к теоретическим. Можно вывести соотношение между процентами выбросов у конкретных распределений:

$$(2.1.5) < (2.1.4) < (2.1.1) < (2.1.3) < (2.1.2)$$

По полученным данным видно, что наименьший процент выбросов у равномерного распределения, а наибольший у стандартного распределения Коши.

### 5.4 Эмпирические функции и ядерные оценки

Чем больше выборка, тем лучше ЭФР приближает эталонную функцию распределения.

Исходя из данных в таблицах, для распределений

1. Стандартного нормального: лучшие оценки при  $h=h_n$
2. Стандартного Коши: для малых выборок лучшая оценка будет при  $h=\frac{1}{2}h_n$ , при больших –  $h=h_n$
3. Лапласа: для малых будет  $h=\frac{1}{2}h_n$ , для больших –  $h=h_n$
4. Равномерного: для малых –  $h=h_n$ , для больших –  $h=\frac{1}{2}h_n$
5. Пуассона: лучшие оценки при  $h=2h_n$

## 6 Литература

[Основы работы с \*numpy\* \(отдельная глава курса\)](#)

[Matplotlib. Уроки](#)

[Документация по \*scipy\*](#)

[Pandas обзор](#)

[Модуль \*math\*](#)

## 7 Приложения

[Репозиторий на GitHub с лабораторными работами](#)