

Санкт-Петербургский политехнический университет
Петра Великого
Институт прикладной математики и механики
Кафедра “Прикладная математика”

Отсчет
По лабораторным работам №1–4
По дисциплине
“Математическая статистика”

Выполнил студент;
Золотухин Илья Сергеевич
Группа:
5030102/90101

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

Санкт-Петербург
2022 г.

1. ПОСТАНОВКА ЗАДАЧИ	3
2. ТЕОРИЯ	4
2.1 РАСПРЕДЕЛЕНИЯ	4
2.2 ХАРАКТЕРИСТИКИ ПОЛОЖЕНИЯ	4
2.2.1 <i>Характеристики рассеяния</i>	5
2.3 БОКСПЛОТ ТЬЮРИ	5
1.3.1 <i>Построение</i>	5
2.4 ТЕОРЕТИЧЕСКАЯ ВЕРОЯТНОСТЬ ВЫБРОСОВ	5
2.5 ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ	6
2.5.1 <i>Статистический ряд</i>	6
2.5.2 <i>Эмпирическая функция распределения</i>	6
2.5.3 <i>Нахождение э. ф. р.</i>	6
2.6 ОЦЕНКА ПЛОТНОСТИ ВЕРОЯТНОСТИ	6
2.6.1 <i>Определение</i>	6
2.6.2 <i>Ядерные оценки</i>	6
3. РЕАЛИЗАЦИЯ	7
3.1 ГИСТОГРАММА И ГРАФИК ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ	7
3.2. ХАРАКТЕРИСТИКИ ПОЛОЖЕНИЯ И РАССЕЯНИЯ	9
3.3 БОКСПЛОТ ТЬЮРИ	11
3.4 ДОЛЯ ВЫБРОСОВ	14
3.5 ТЕОРЕТИЧЕСКАЯ ВЕРОЯТНОСТЬ ВЫБРОСОВ	14
3.6 ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ	15
3.7 ЯДЕРНЫЕ ОЦЕНКИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ	16
4 ОБСУЖДЕНИЕ	21
4.1 ГИСТОГРАММА И ГРАФИК ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ	21
4.2 ХАРАКТЕРИСТИКИ ПОЛОЖЕНИЯ И РАССЕЯНИЯ	22
4.3 ДОЛЯ И ТЕОРЕТИЧЕСКАЯ ВЕРОЯТНОСТЬ ВЫБРОСОВ	22
4.4 ЭМПИРИЧЕСКАЯ ФУНКЦИЯ И ЯДЕРНЫЕ ОЦЕНКИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ	22

Рисунок 1 НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ	7
Рисунок 2 РАСПРЕДЕЛЕНИЕ КОШИ	7
Рисунок 3 РАСПРЕДЕЛЕНИЕ ЛАПЛАСА	8
Рисунок 4 РАСПРЕДЕЛЕНИЕ ПУАССОНА	8
Рисунок 5 РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ.....	8
Рисунок 6 НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ	11
Рисунок 7 РАСПРЕДЕЛЕНИЕ КОШИ	12
Рисунок 8 РАСПРЕДЕЛЕНИЕ ЛАПЛАСА	13
Рисунок 9 РАСПРЕДЕЛЕНИЕ ПУАССОНА	13
Рисунок 10 РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ	14
Рисунок 11 НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ	15
Рисунок 12 РАСПРЕДЕЛЕНИЕ КОШИ	15
Рисунок 13 РАСПРЕДЕЛЕНИЕ ЛАПЛАСА	16
Рисунок 14 РАСПРЕДЕЛЕНИЕ ПУАССОНА	16
Рисунок 15 РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ.....	16
Рисунок 16 НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ $n = 20$	17
Рисунок 17 НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ $n = 60$	17
Рисунок 18 НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ $n = 100$	17
Рисунок 19 РАСПРЕДЕЛЕНИЕ КОШИ $n = 20$	18
Рисунок 20 РАСПРЕДЕЛЕНИЕ КОШИ $n = 60$	18
Рисунок 21 РАСПРЕДЕЛЕНИЕ КОШИ $n = 100$	18
Рисунок 22 РАСПРЕДЕЛЕНИЕ ЛАПЛАСА $n = 20$	19
Рисунок 23 РАСПРЕДЕЛЕНИЕ ЛАПЛАСА $n = 60$	19
Рисунок 24 РАСПРЕДЕЛЕНИЕ ЛАПЛАСА $n = 100$	19
Рисунок 25 РАСПРЕДЕЛЕНИЕ ПУАССОНА $n = 20$	20
Рисунок 26 РАСПРЕДЕЛЕНИЕ ПУАССОНА $n = 60$	20
Рисунок 27 РАСПРЕДЕЛЕНИЕ ПУАССОНА $n = 100$	20
Рисунок 28 РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ $n = 20$	21
Рисунок 29 РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ $n = 60$	21
Рисунок 30 РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ $n = 100$	21
Таблица 1 НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ	9
Таблица 2 РАСПРЕДЕЛЕНИЕ КОШИ	9
Таблица 3 РАСПРЕДЕЛЕНИЕ ЛАПЛАСА	10
Таблица 4 РАСПРЕДЕЛЕНИЕ ПУАССОНА	10
Таблица 5 РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ.....	11
Таблица 6 ДОЛЯ ВЫБРОСОВ	14
Таблица 7 ТЕОРЕТИЧЕСКАЯ ВЕРОЯТНОСТЬ ВЫБРОСОВ.....	15

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 $U(x, -\sqrt{3}, \sqrt{3})$
- Сгенерировать выборки размером 10, 100 и 1000 элементов. Построить на одном рисунке гистограмму и график плотности распределения.
- Сгенерировать выборки размером 10, 100 и 1000 элементов. Для каждой выборки вычислить следующие статистические характеристики положения данных: $\hat{x}, medx, z_R, z_Q, z_{tr}$. Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

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

1. Вычислить оценку дисперсии по формуле:

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

2. Представить полученные данные в виде таблиц
3. Сгенерировать выборки размером 20 и 100 элементов. Построить для них боксплот Тьюри. Для каждого распределения определить долю выбросов экспериментально (сгенерировав выборку, соответствующую распределению 1000 раз, и вычислив 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

Плотности классических распределений:

Нормальное распределение

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

Распределение Коши

$$C(x, x_0, \gamma) = \frac{1}{\pi \cdot \gamma} \frac{1}{1 + \frac{(x - x_0)^2}{\gamma^2}} \quad (2)$$

Распределение Лапласа

$$L(x, \beta, \alpha) = \frac{\alpha}{2} e^{-\alpha|x-\beta|} \quad (3)$$

Распределение Пуассона

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

Равномерное распределение

$$U(x, a, b) = \begin{cases} \frac{1}{b-a} & \text{при } x \in [a, b] \\ 0 & \text{при } x \notin [a, b] \end{cases} \quad (5)$$

2.2 Характеристики положения

Выборочное среднее

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

Выборочная медиана

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

Полу сумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \quad (8)$$

Полу сумма квантилей

Выборочная квантиль z_p порядка p определяется формулой

$$z_p = \begin{cases} x_{([np]+1)} & \text{при } np \text{ дробном,} \\ x_{(np)} & \text{при } np \text{ целом} \end{cases} \quad (9)$$

Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \quad (10)$$

Усечённое среднее

$$z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_{(i)}, \quad r \approx \frac{n}{4} \quad (11)$$

2.2.1 Характеристики рассеяния

Выборочная дисперсия

$$D = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (12)$$

2.3 Боксплот Тьюри

1.3.1 Построение

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

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1) \quad (13)$$

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 кружочков

2.4 теоретическая вероятность выбросов

Можно вычислить теоретические первый и третий квартили распределений - Q_1^T и Q_3^T . По ф-ле (12) - теоретические нижнюю и верхнюю границы уса - X_1^T и X_2^T . Выбросы – величины x :

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases} \quad (14)$$

Теоретическая вероятность выбросов:

Для непрерывных распределений

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)) \quad (15)$$

Для дискретных распределений

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

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

2.5 Эмпирическая функция распределения

2.5.1 Статистический ряд

Статистический ряд – 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. Обычно записывается в виде таблицы.

2.5.2 Эмпирическая функция распределения

Эмпирическая (выборочная) функция распределения (э. ф. р.) – относительная частота события $X < x$, полученная по данной выборке:

$$F_n^*(x) = P^*(X < x) \quad (17)$$

2.5.3 Нахождение э. ф. р.

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

$$F^*(x) = \frac{1}{n} \sum_{z_i < x} n_i \quad (18)$$

$F^*(x)$ – 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 0: Таблица распределения

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

$$F_n^*(x) \approx F_X(x) \quad (19)$$

2.6 Оценка плотности вероятности

2.6.1 Определение

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

$$\hat{f}(x) \approx f(x) \quad (20)$$

2.6.2 Ядерные оценки

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

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) \quad (21)$$

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

$$\lim_{n \rightarrow \infty} h_n = 0; \lim_{n \rightarrow \infty} \frac{h_n}{n^{-1}} = \infty. \quad (22)$$

Такие оценки называются непрерывными ядерными.

Гауссово (нормальное) ядро

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (23)$$

Правило Сильвермана

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

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

3.Реализация

3.1 Гистограмма и график плотности распределения.

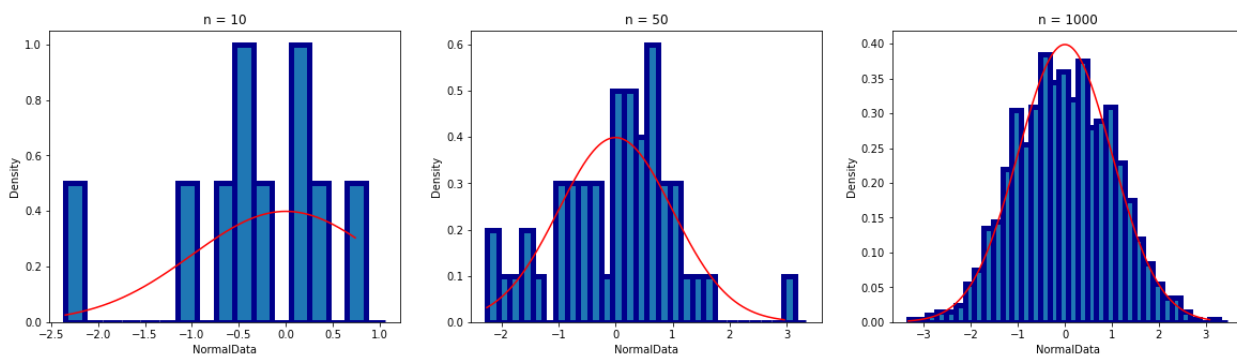


Рисунок 1 Нормальное распределение

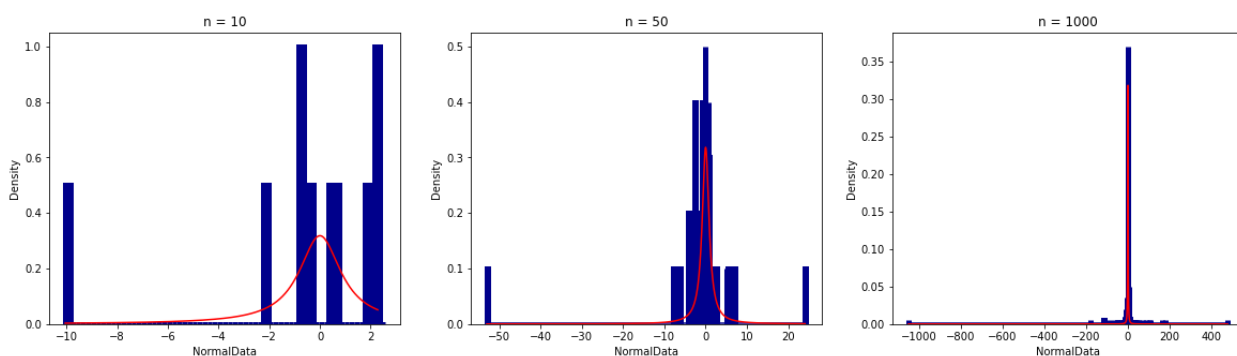


Рисунок 2 Распределение Коши

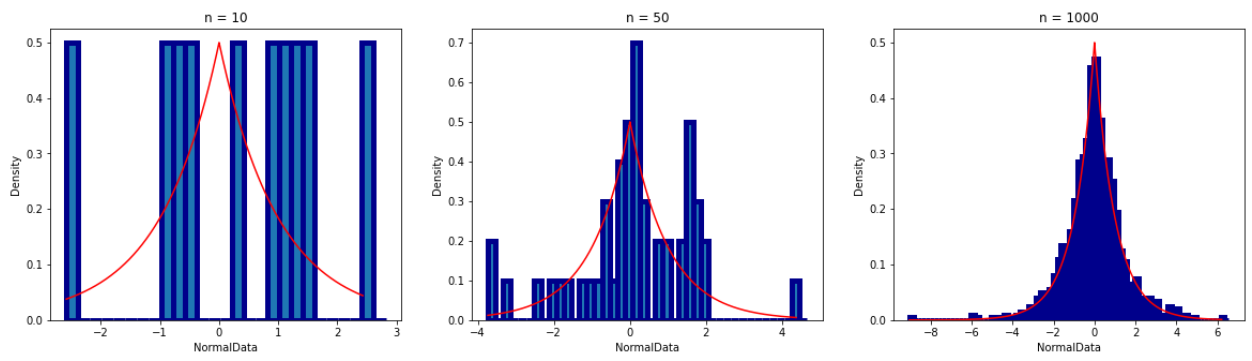


Рисунок 3 Распределение Лапласа

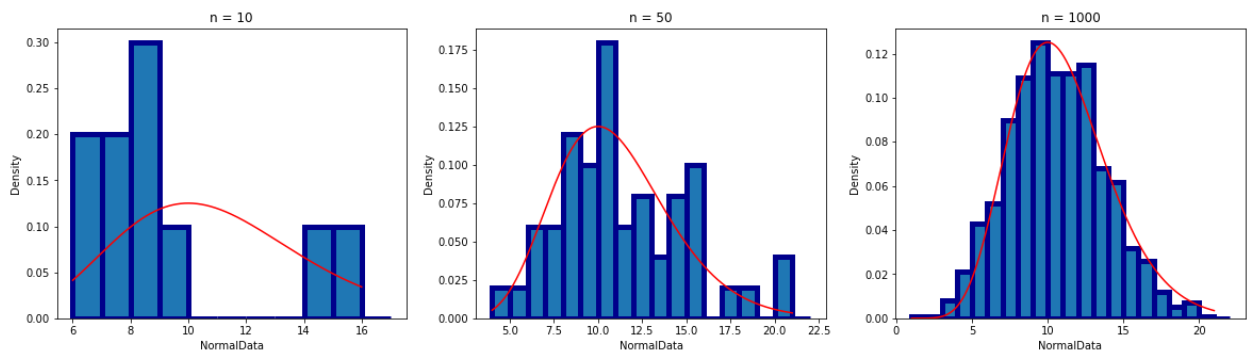


Рисунок 4 Распределение Пуассона

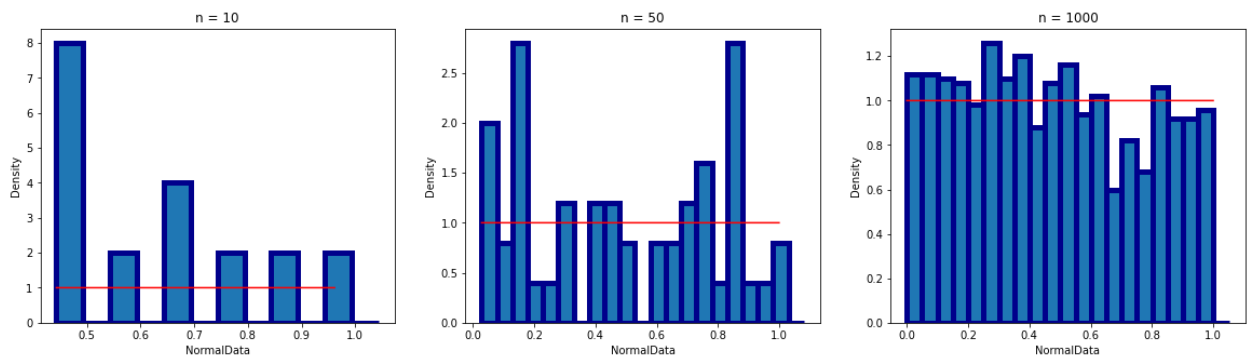


Рисунок 5 Равномерное распределение

3.2. Характеристики положения и рассеяния

Таблица 1 Нормальное распределение

normal n = 10					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0001	0.0001	-0.004	0.0021	-0.0017
$D(z)$	0.0012	0.0020	0.060	0.0022	0.0013
normal n = 100					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.00011	0.0000	-0.0143	0.0000	-0.0020
$D(z)$	0.00097	0.0016	0.0571	0.0012	0.0013
normal n = 1000					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0038	0.0030	0.013	0.0047	0.0019
$D(z)$	0.0010	0.0015	0.062	0.0013	0.0012

Таблица 2 Распределение Коши

cauchy n = 10					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$ (1)	0.0004	-0.0003	0.002	0.0036	-0.0015
$D(z)$ (2)	0.0011	0.0016	0.060	0.0050	0.0013
cauchy n = 100					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.0029	-0.0012	0.0	0.0003	-0.0028
$D(z)$	0.0028	0.0015	2.4	0.0014	0.0012
cauchy n = 1000					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0005	0.0008	0.0	0.0021	-0.0008
$D(z)$	0.0014	0.0015	191.0	0.0012	0.0011

Таблица 3 Распределение Лапласа

laplace n = 10					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.0004	-0.0004	0.010	0.0012	-0.0016
$D(z)$	0.0010	0.0016	0.065	0.0014	0.0012
laplace n = 100					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.00168	0.0023	-0.003	0.0001	-0.0039
$D(z)$	0.00093	0.0015	0.061	0.0012	0.0013
laplace n = 1000					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0010	0.0008	-0.003	0.0028	-0.0006
$D(z)$	0.0010	0.0016	0.060	0.0013	0.0012

Таблица 4 Распределение Пуассона

pois n = 10					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$ (1)	0.0008	0.0005	0.002	0.0022	-0.0008
$D(z)$ (2)	0.0011	0.0015	0.064	0.0013	0.0012
pois n = 100					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.0015	-0.0028	-0.003	0.0006	-0.0031
$D(z)$	0.0011	0.0016	0.060	0.0013	0.0013
pois n = 1000					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.0006	-0.0013	-0.0077	0.0004	-0.0022
$D(z)$	0.0011	0.0016	0.0589	0.0013	0.0013

Таблица 5 Равномерное распределение

pois n = 10					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$ (1)	0.0015	0.0010	0.012	0.0030	-0.0003
$D(z)$ (2)	0.0012	0.0017	0.063	0.0015	0.0012
pois n = 100					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	-0.0003	0.0010	-0.004	0.0002	-0.0013
$D(z)$	0.0012	0.0018	0.066	0.0015	0.0014
pois n = 1000					
	\bar{x}	med x	z_R	z_Q	z_{tr}
$E(z)$	0.0011	0.0024	0.007	0.0023	0.0002
$D(z)$	0.0013	0.0020	0.062	0.0015	0.0015

3.3 Боксплот Тьюри

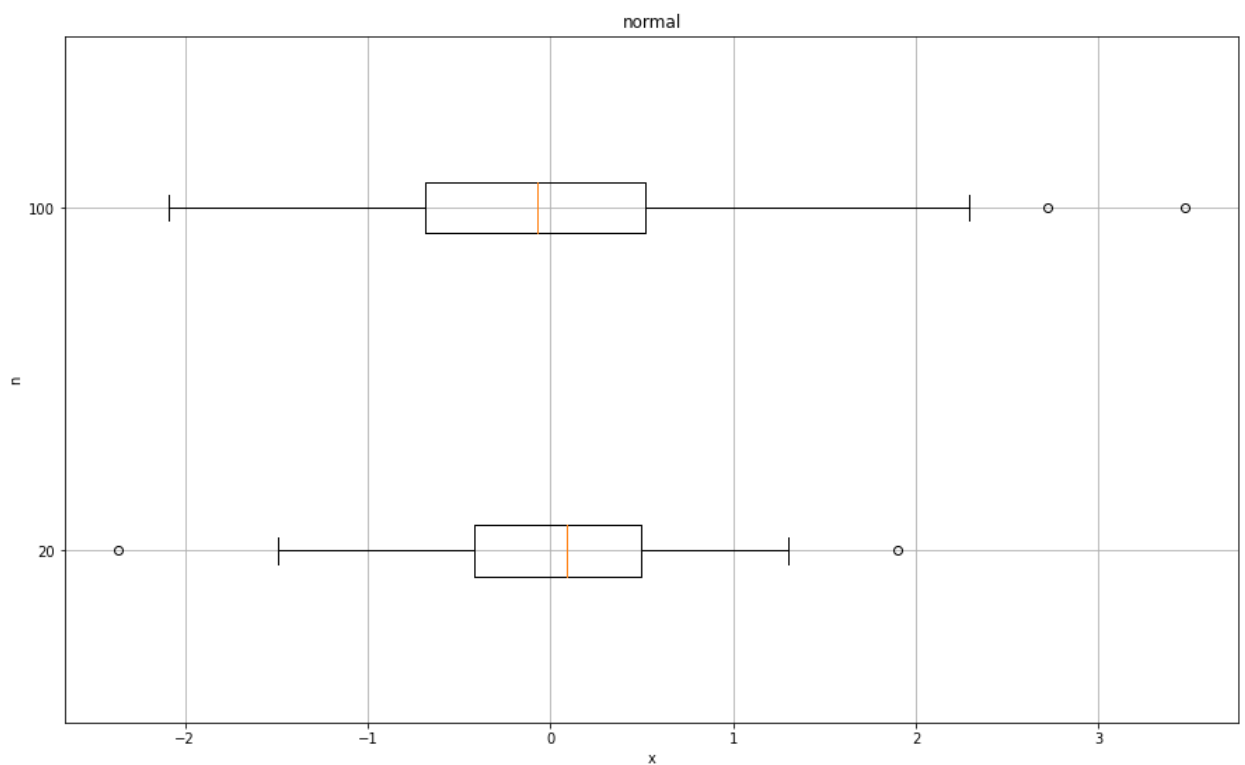


Рисунок 6 Нормальное распределение

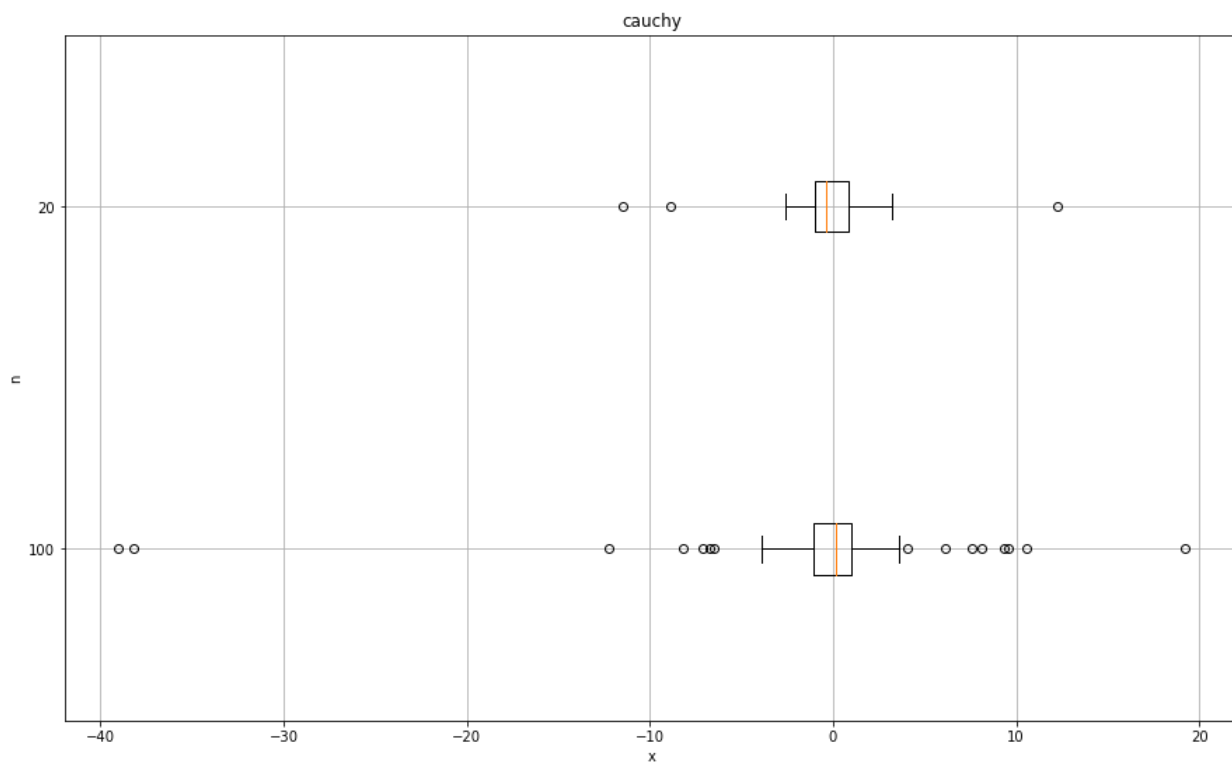


Рисунок 7 Распределение Коши

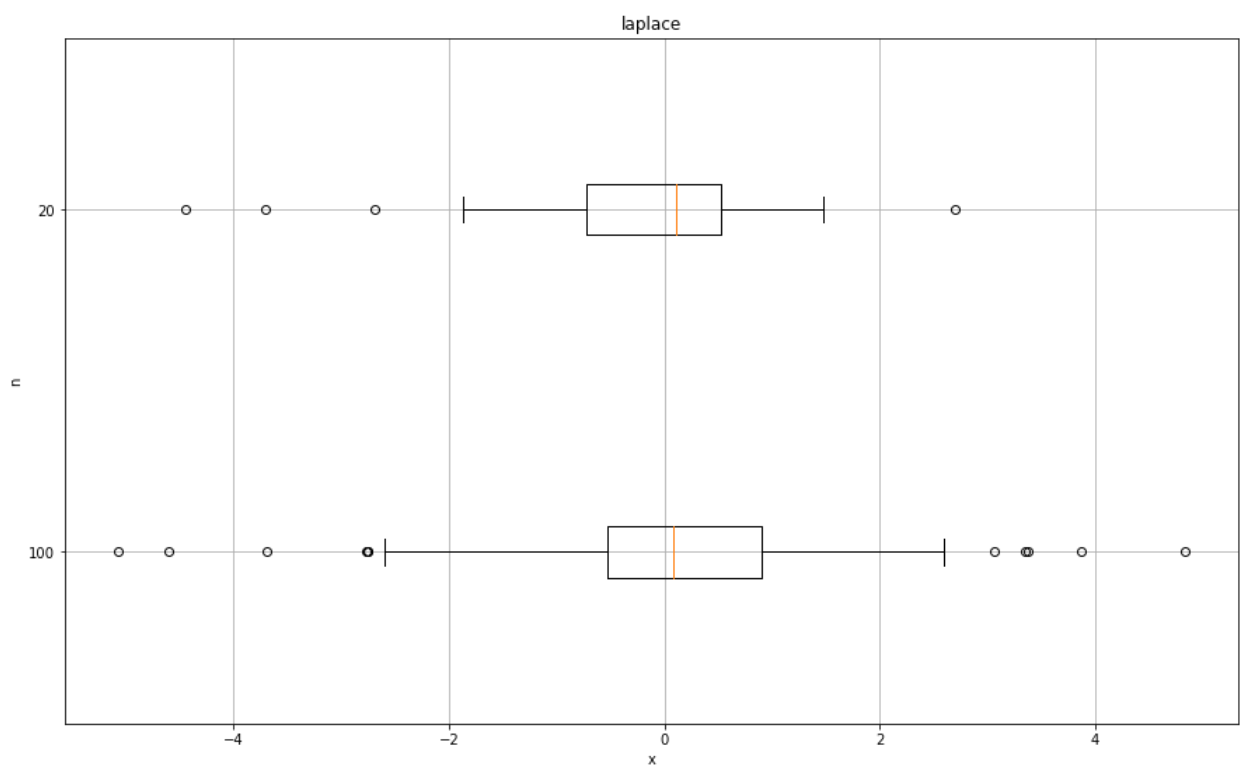


Рисунок 8 Распределение Лапласа

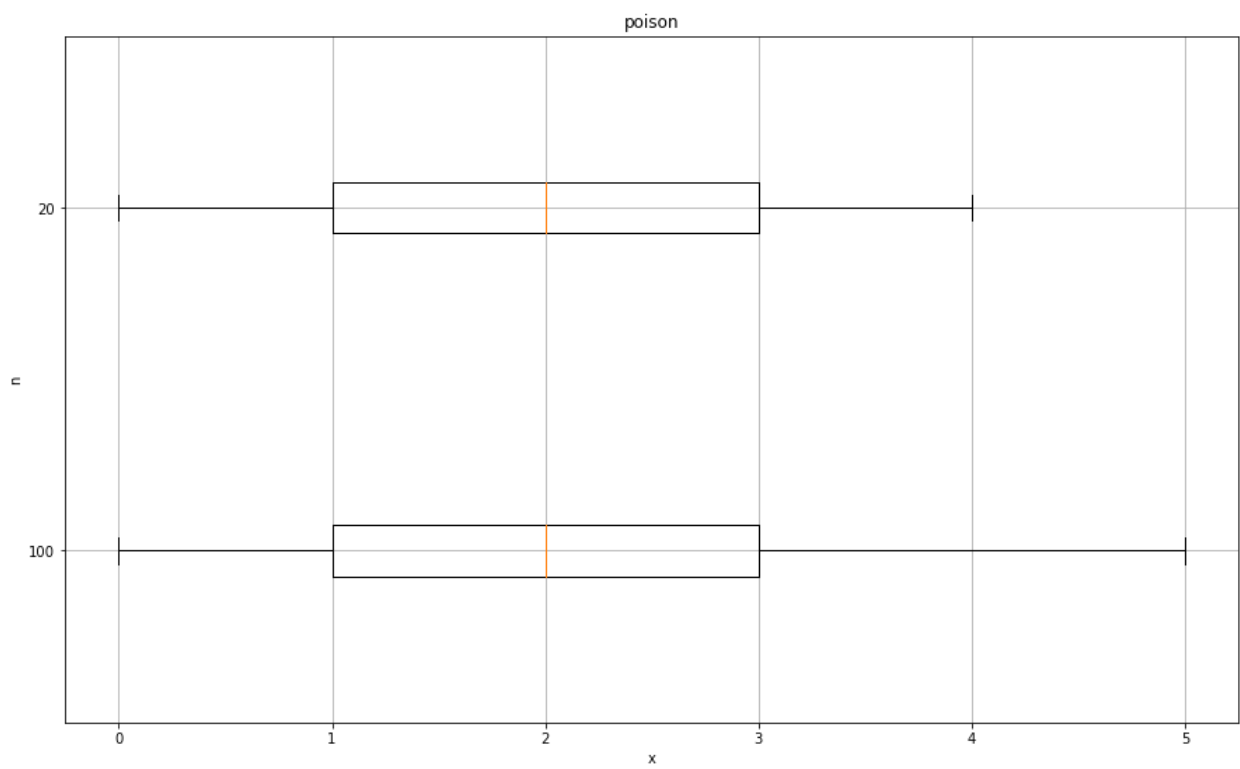


Рисунок 9 Распределение Пуассона

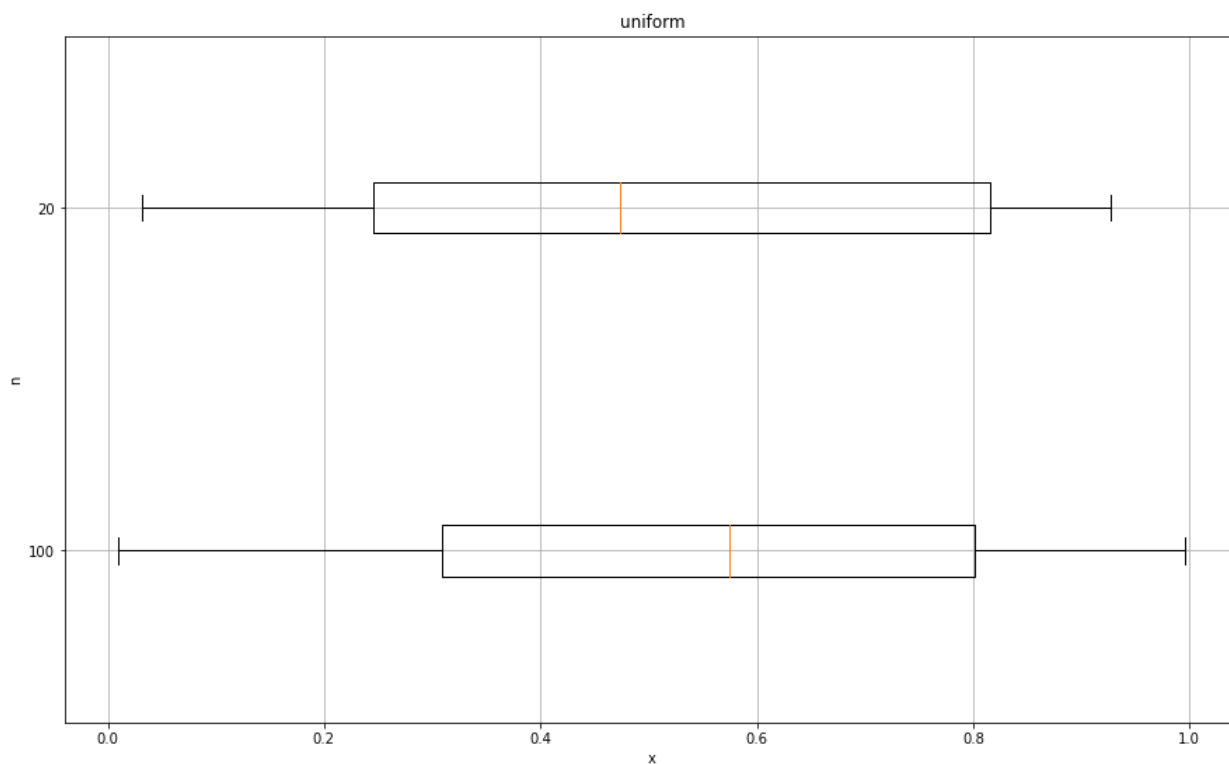


Рисунок 10 равномерное распределение

3.4 Доля выбросов

Таблица 6 Доля выбросов

Выборка	Доля выбросов
Normal n = 20	0.10
Normal n = 100	0.03
Cauchy n = 20	0.05
Cauchy n = 100	0.15
Laplace n = 20	0.05
Laplace n = 100	0.06
Pois n = 20	0.10
Pois n = 100	0.08
Unif n = 20	0.0
Unif n = 100	0.0

3.5 Теоретическая вероятность выбросов

Таблица 7 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	-1.000	1.000	-4.000	4.000	0.156
Распределение Лапласа	-0.693	0.693	-2.773	2.773	0.062
Распределение Пуассона	8.000	12.00	2.000	18.00	0.008
Равномерное распределение	0.250	0.750	-0.500	1.500	0.000

3.6 Эмпирическая функция распределения

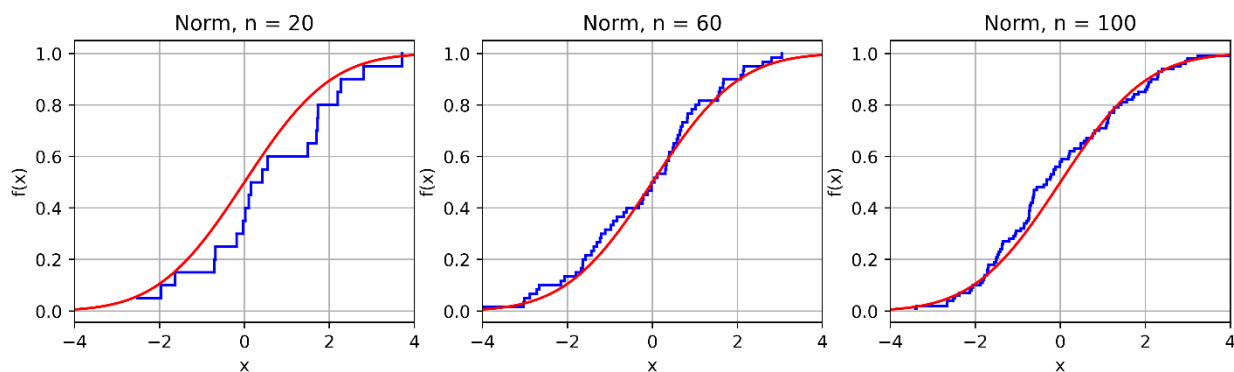


Рисунок 11 Нормальное распределение

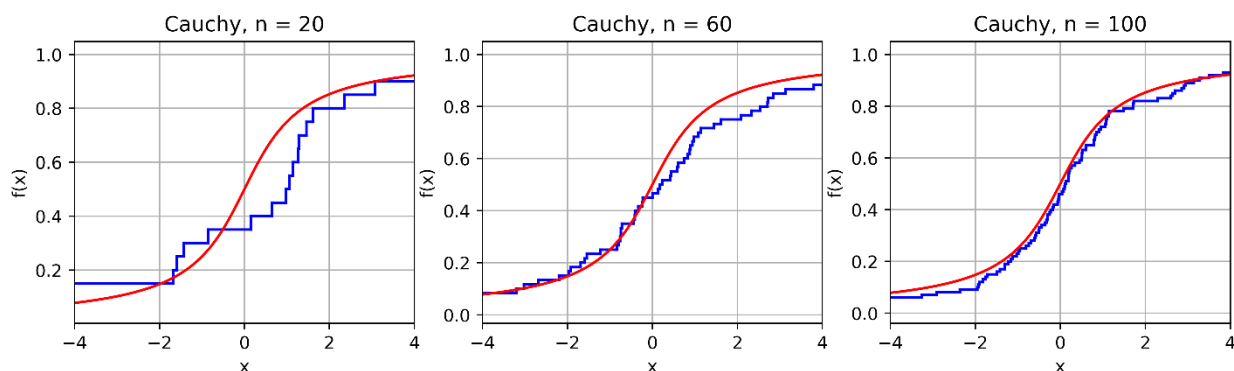


Рисунок 12 Распределение Коши

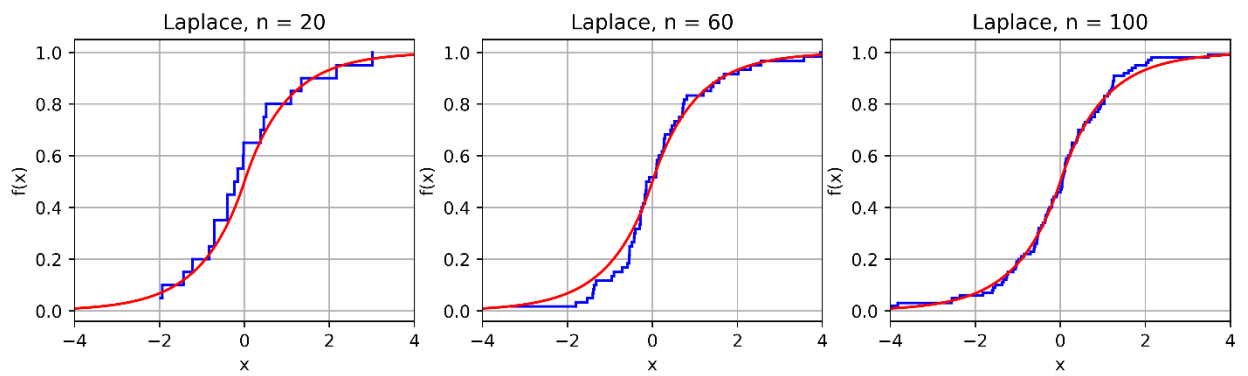


Рисунок 13 Распределение Лапласа

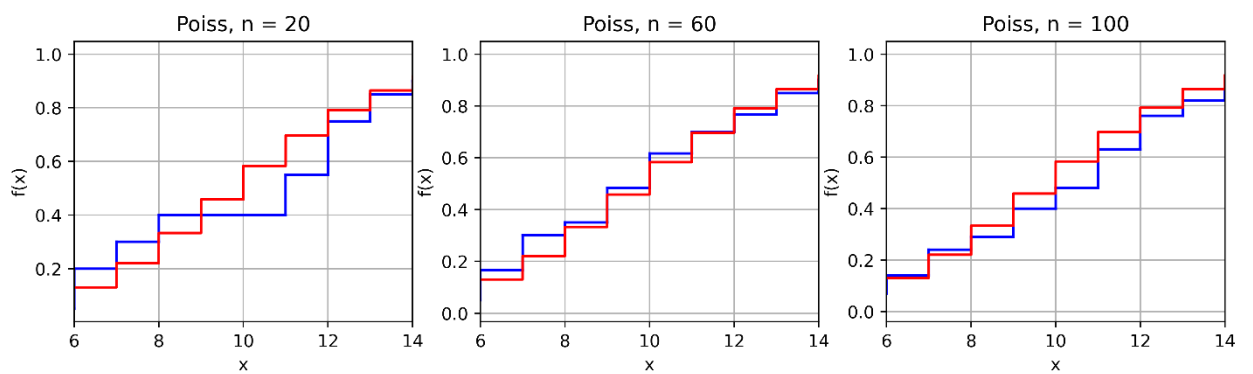


Рисунок 14 Распределение Пуассона

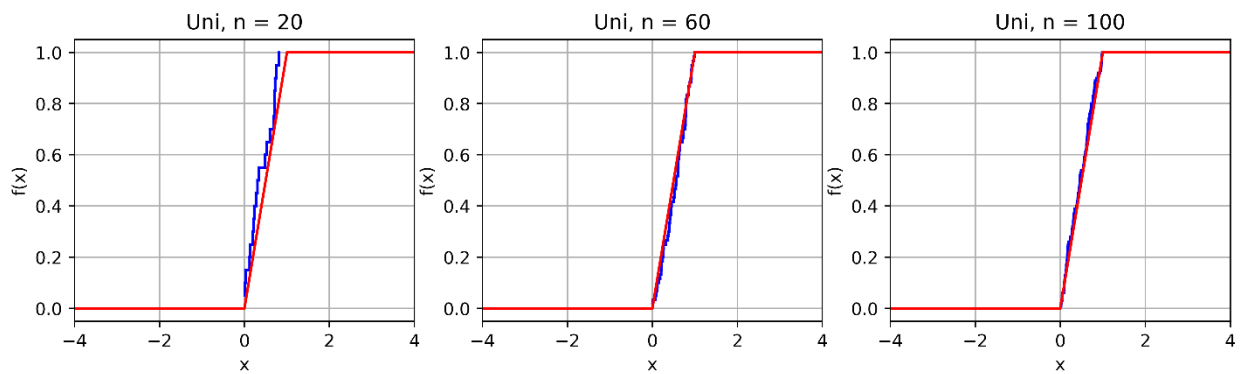


Рисунок 15 Равномерное распределение

3.7 Ядерные оценки плотности распределения

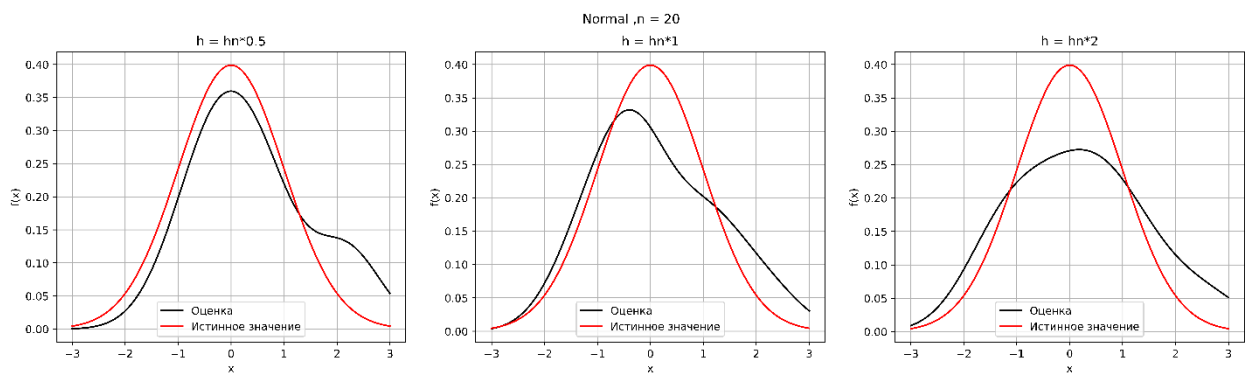


Рисунок 16 Нормальное распределение $n = 20$

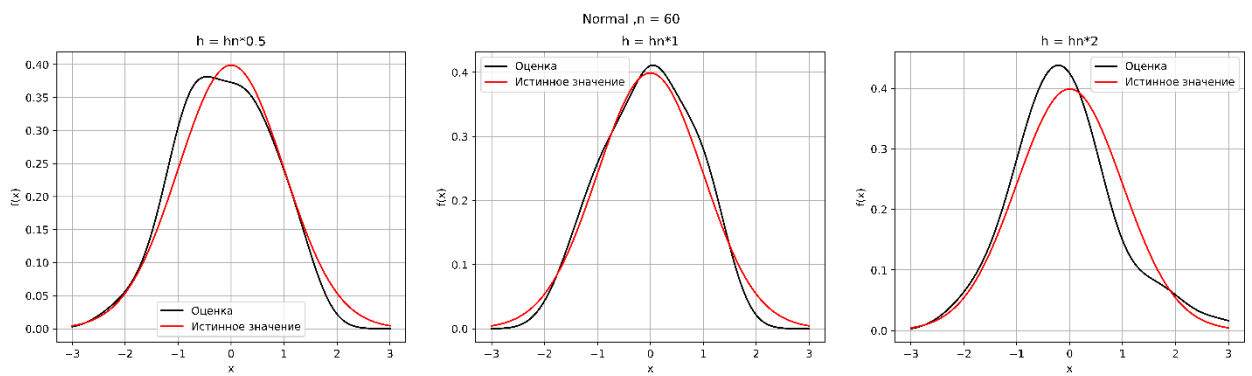


Рисунок 17 Нормальное распределение $n = 60$

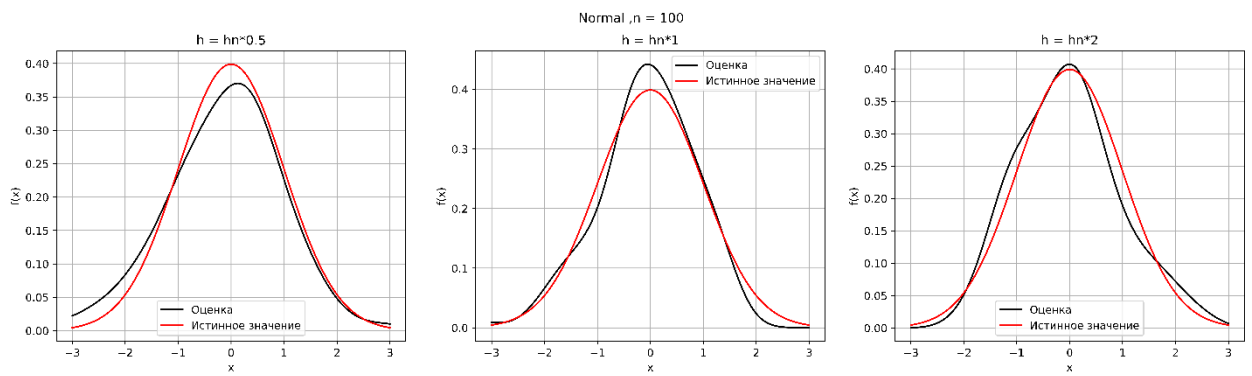


Рисунок 18 Нормальное распределение $n = 100$

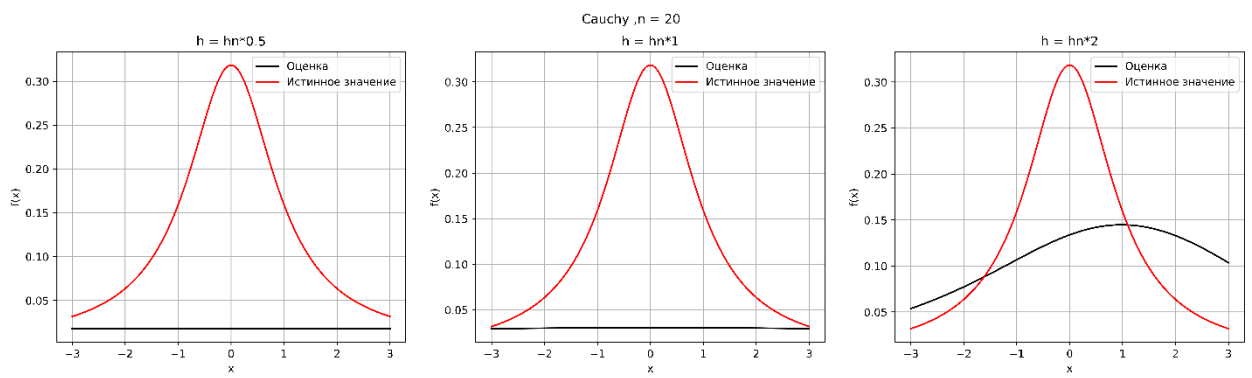


Рисунок 19 Распределение Коши $n = 20$

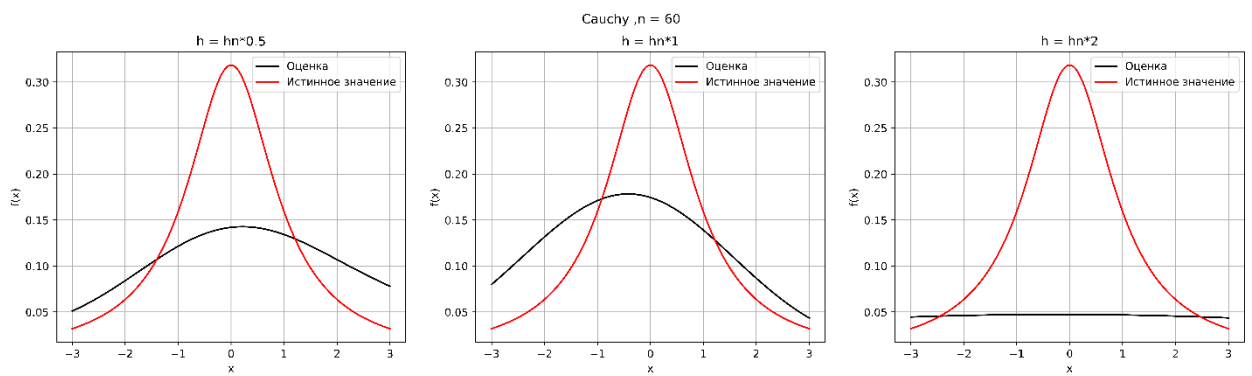


Рисунок 20 Распределение Коши $n = 60$

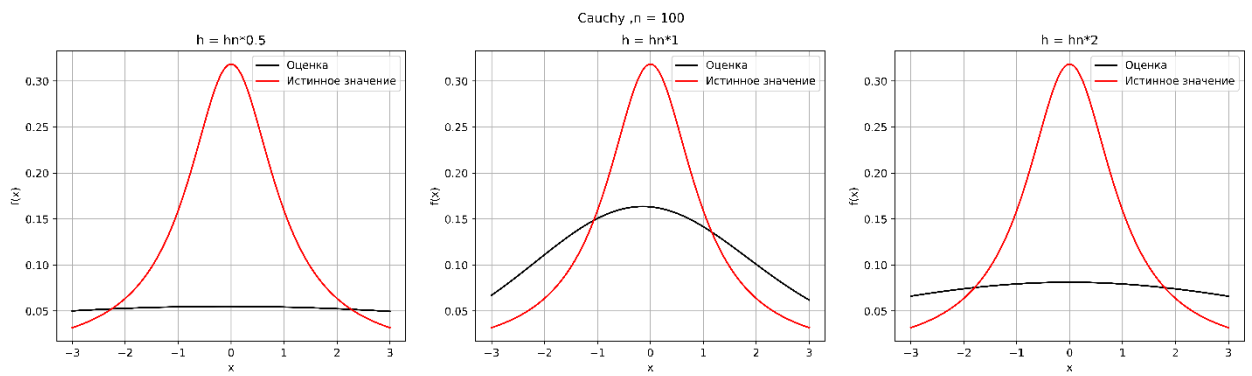


Рисунок 21 Распределение Коши $n = 100$

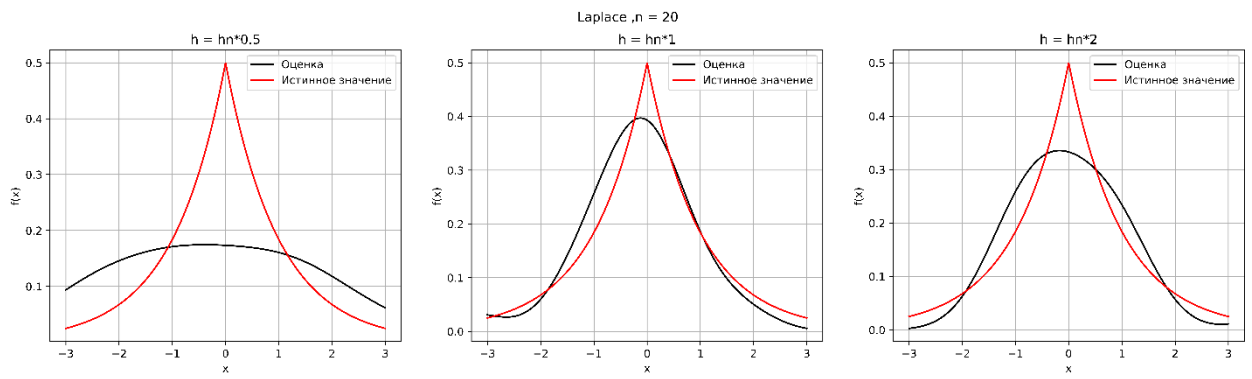


Рисунок 22 Распределение Лапласа $n = 20$

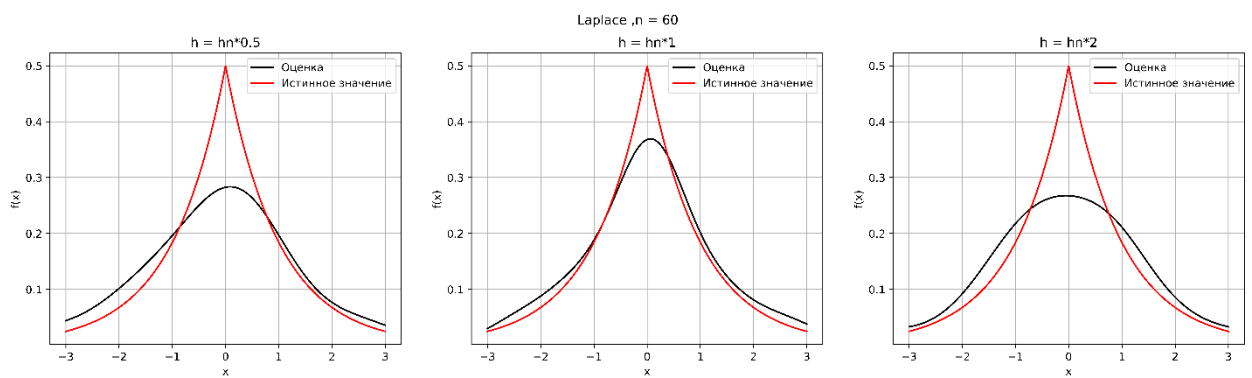


Рисунок 23 Распределение Лапласа $n = 60$

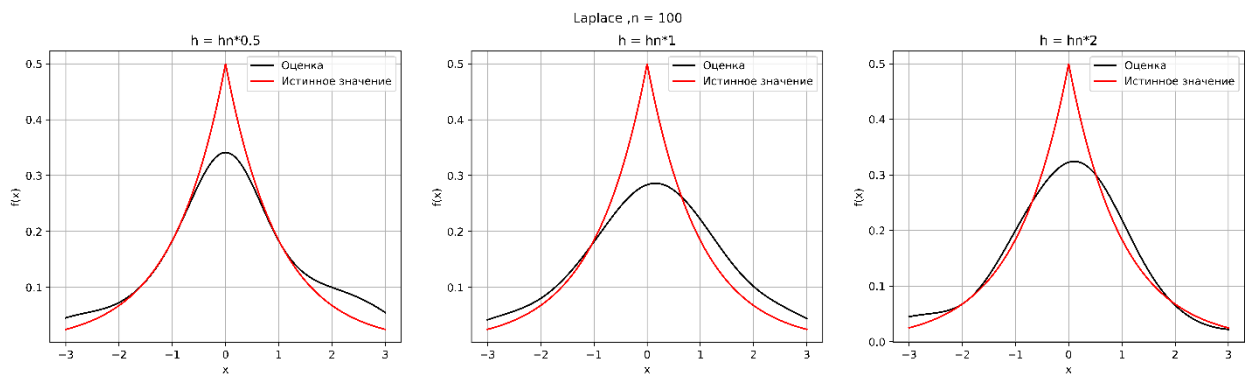


Рисунок 24 Распределение Лапласа $n = 100$

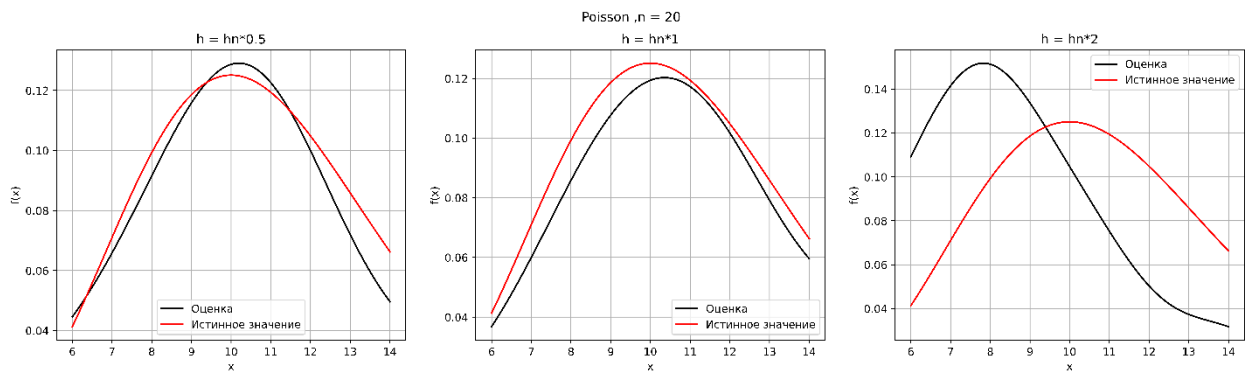


Рисунок 25 Распределение Пуассона $n = 20$

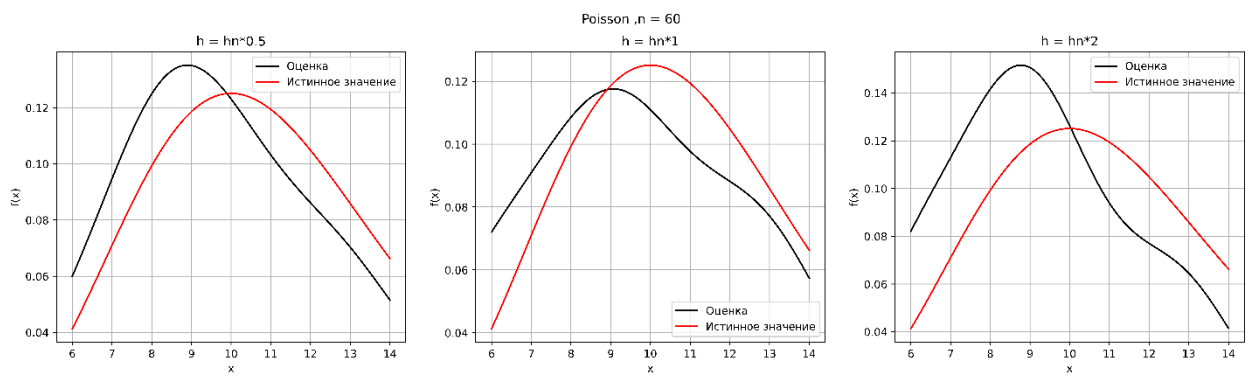


Рисунок 26 Распределение Пуассона $n = 60$

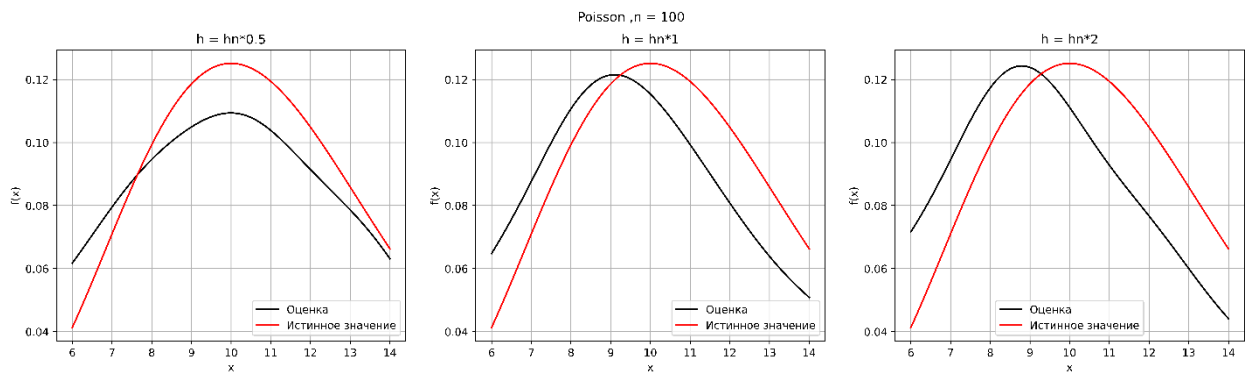


Рисунок 27 Распределение Пуассона $n = 100$

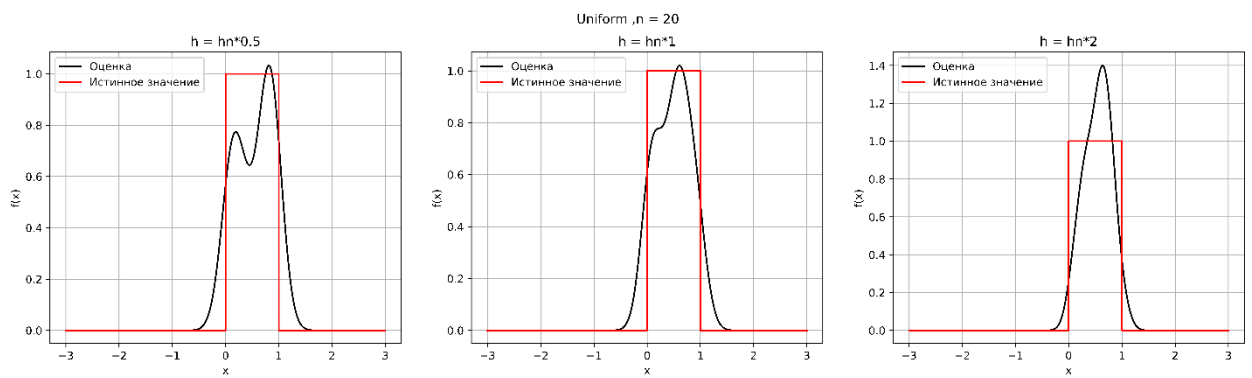


Рисунок 28 Равномерное распределение $n = 20$

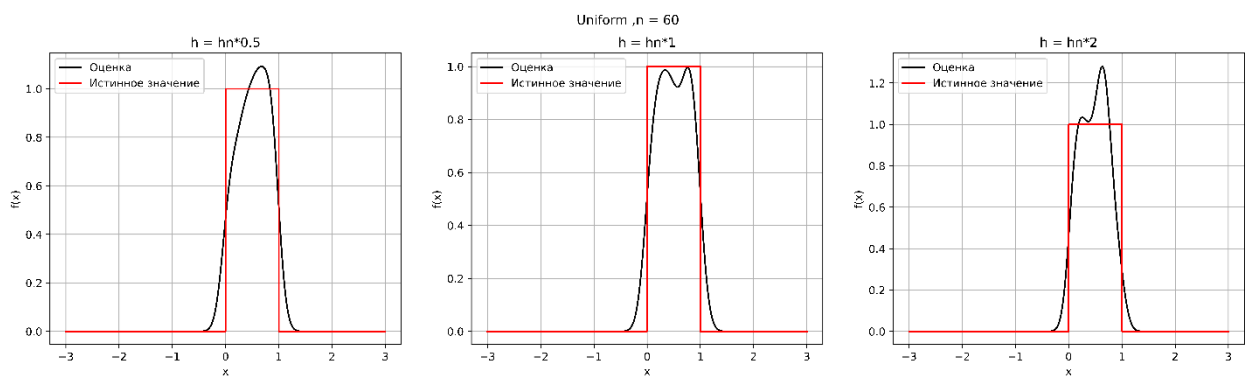


Рисунок 29 Равномерное распределение $n = 60$

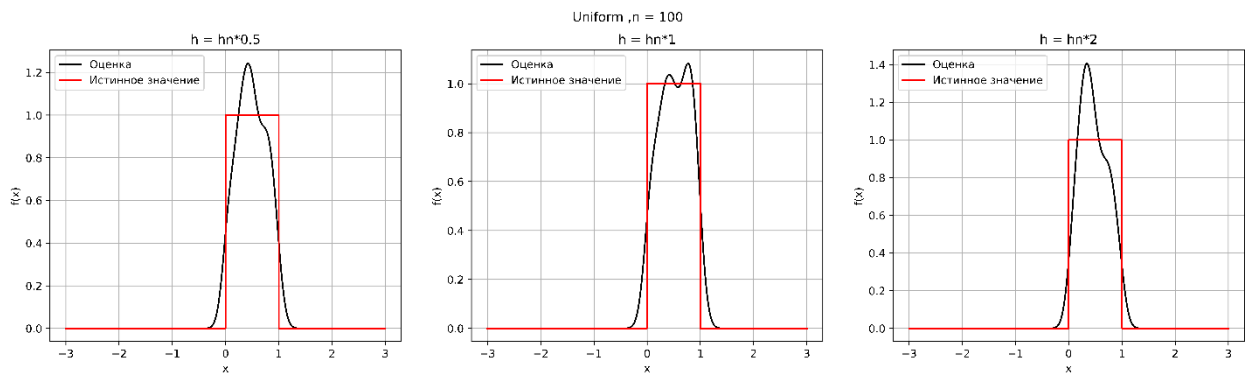


Рисунок 30 Равномерное распределение $n = 100$

4 Обсуждение

4.1 Гистограмма и график плотности распределения

Как видно из графиков, чем больше выборка для каждого из распределений, тем ближе её гистограмма к графику плотности вероятности того закона, по которому значения распределены.

Чем меньше выборка, тем соответственно хуже по ней определяется характер распределения величины.

Максимумы и минимумы гистограммы и плотностей распределения не совпадают нигде. Также можно увидеть всплески гистограмм, что видно из распределения коши.

4.2 Характеристики положения и рассеяния

Из данных, что получили можем отметить, что значения дисперсии для распределения Коши имеют аномальные значения, что подтверждает теоретический результат о том, что у распределения Коши нет дисперсии

4.3 Доля и теоретическая вероятность выбросов

По данным, приведенным в таблице, видно, что чем больше выборка, тем ближе доля выбросов к теоретической оценке. Причем доля выбросов для распределения Коши выше, чем для остальных распределений. Доля выбросов равномерного распределения повторяет в точности теоретическую оценку, причем она равняется нулю.

Боксплоты Тьюри наглядно показывают важные характеристики.

4.4 Эмпирическая функция и ядерные оценки плотности распределения

На иллюстрациях с э. ф. р. Видно, что ступенчатая эмпирическая функция распределения тем лучше приближает функцию распределения реальной выборки, чем мощнее эта выборка. Причем наибольшее отклонение наблюдается у распределения Пуассона.

Для нормального распределения при любых значениях функции практически совпадают

Для распределения Коши оценка плотности и реальная плотность распределения сильно отличаются, но хорошие параметры $n = 100$ с $h = hn*1$

Для распределения Лапласа наилучшее совпадение функций при $n = 60$ и шаге $h = hn*1$

Для распределения Пуассона хорошо себя показывают малые значения n , а то есть оптимальные значения $n = 20$ $h = hn/2$