

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ПЕТРА ВЕЛИКОГО

ФИЗИКО-МЕХАНИЧЕСКИЙ ИНСТИТУТ

ВЫСШАЯ ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ВЫЧИСЛИТЕЛЬНОЙ ФИЗИКИ

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

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

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

Санкт-Петербург
2023 г.

Содержание

1	Постановка задачи	2
2	Теория	3
2.1	Рассматриваемые распределения	3
2.2	Гистограмма	3
2.3	Вариационный ряд	3
2.4	Выборочные числовые характеристики	4
2.4.1	Характеристики положения	4
2.4.2	Характеристики рассеяния	4
2.5	Боксплот Тьюки	5
2.5.1	Построение	5
2.6	Теоретическая вероятность выбросов	5
3	Реализация	6
4	Результаты	7
4.1	Гистограмма и график плотности распределения	7
4.2	Характеристики положения и рассеяния	8
4.3	Боксплот Тьюки	10
4.4	Доля выбросов	13
4.5	Теоретическая вероятность выбросов	13
5	Обсуждение	14
	Литература	17

1 Постановка задачи

Для 5 распределений:

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

1. Сгенерировать выборки размером 10, 50 и 1000 элементов. Построить на одном рисунке гистограмму и график плотности распределения.
2. Сгенерировать выборки размером 10, 100 и 1000 элементов. Для каждой выборки вычислить следующие статистические характеристики положения данных: \bar{x} , $med\ x$, z_R , z_Q , z_{tr} . Повторить такие вычисления 1000 раз для каждой выборки и найти среднее значение характеристик положения 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. Сгенерировать выборки размером 20 и 100 элементов.
Построить для них боксплот Тьюки.
Для каждого распределения определить долю выбросов экспериментально (сгенерировав выборку, соответствующую распределению 1000 раз, и вычислив среднюю долю выбросов) и сравнить с результатами, полученными теоретически.
4. Сгенерировать выборки размером 20, 60 и 100 элементов.
Построить на них эмпирические функции распределения и ядерные оценки плотности распределения на отрезке $[-4; 4]$ для непрерывных распределений и на отрезке $[6; 14]$ для распределения Пуассона.

2 Теория

2.1 Рассматриваемые распределения

Плотности вероятности рассматриваемых распределений:

- Нормальное распределение

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

- Распределение Коши

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

- Распределение Лапласа

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

- Распределение Пуассона

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

- Равномерное распределение

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

2.2 Гистограмма

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

2.3 Вариационный ряд

Вариационный ряд – последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются.

2.4 Выборочные числовые характеристики

2.4.1 Характеристики положения

- Выборочное среднее

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

- Выборочная медиана

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

- Полусумма экстремальных выборочных элементов

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

- Полусумма квартилей

Выборочная квартиль z_p порядка p определяется формулой

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

Полусумма квартилей

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

- Усечённое среднее

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

2.4.2 Характеристики рассеяния

Выборочная дисперсия

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

2.5 Боксплот Тьюки

2.5.1 Построение

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

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

где X_1 — нижняя граница уса, X_2 — верхняя граница уса, Q_1 — первый квартиль, Q_3 — третий квартиль.

Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков.

2.6 Теоретическая вероятность выбросов

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

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

Теоретическая вероятность выбросов для непрерывных распределений

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

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

Теоретическая вероятность выбросов для дискретных распределений

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

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

3 Реализация

4 Результаты

4.1 Гистограмма и график плотности распределения

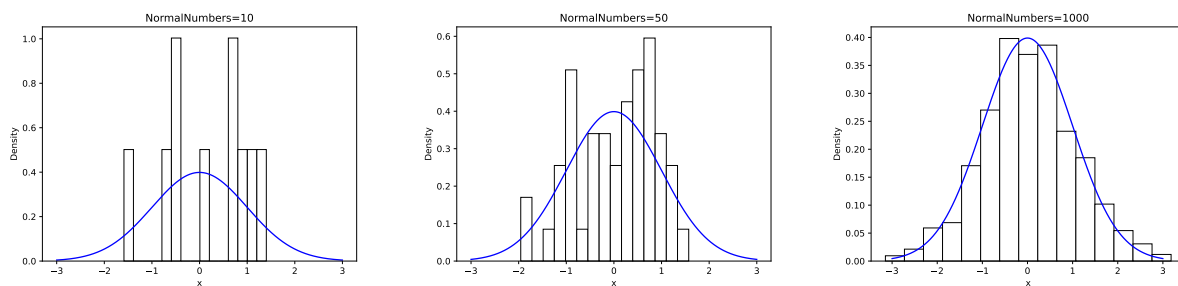


Рис. 1: Нормальное распределение

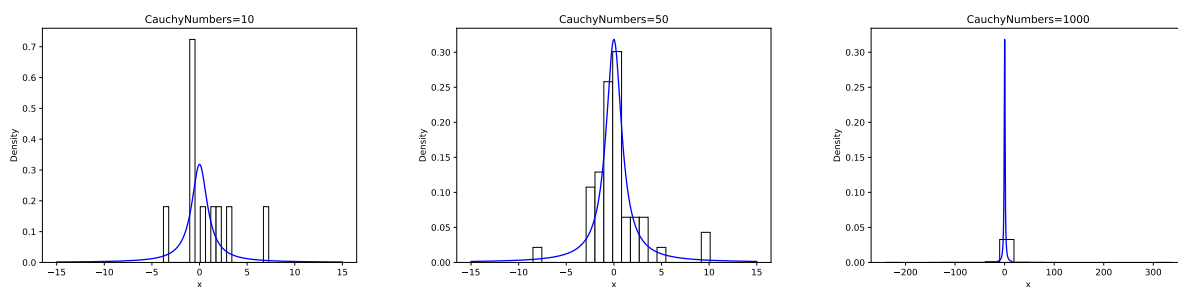


Рис. 2: Распределение Коши

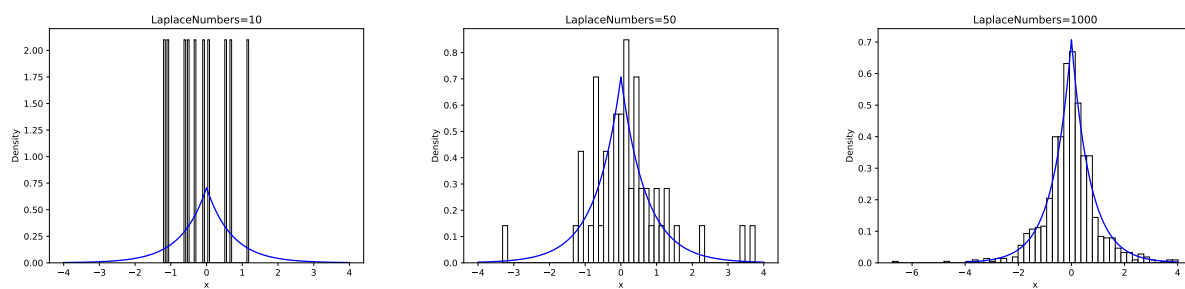


Рис. 3: Распределение Лапласа

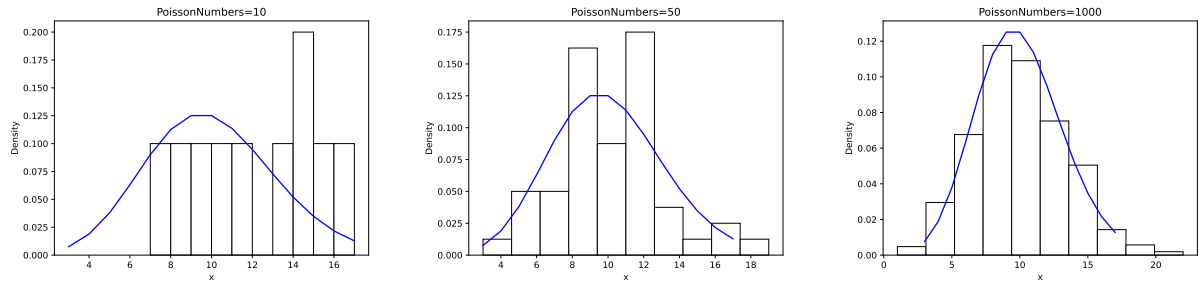


Рис. 4: Распределение Пуассона

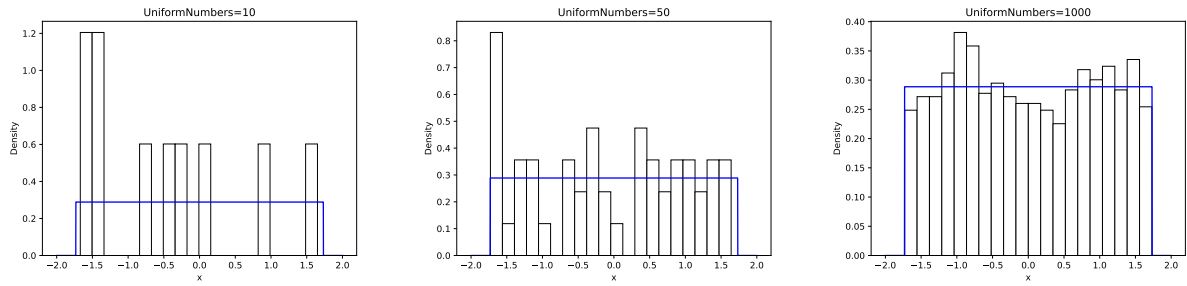


Рис. 5: Равномерное распределение

4.2 Характеристики положения и рассеяния

Normal n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00369	0.00140	0.00253	0.00218	0.00621
$D(z)$	0.09933	0.13162	0.18904	0.12373	0.11933
Normal n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.00061	0.00180	0.00675	-0.01403	-0.00010
$D(z)$	0.00950	0.01575	0.08953	0.01197	0.01167
Normal n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-2.21952	-0.00053	-0.01664	-0.00134	-0.00042
$D(z)$	0.00101	0.00164	0.06311	0.00124	0.00122

Таблица 1: Характеристики положения и рассеяния нормального распределения

Cauchy n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.60631	-0.01793	3.07935	0.00118	-0.02171
$D(z)$	3825.12746	0.36398	95430.01133	1.47888	0.38230
Cauchy n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.68621	-0.00218	33.379	-0.02695	0.00145
$D(z)$	415.87354	0.02449	1025245.02557	0.05200	0.02604
Cauchy n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.02740	-0.00038	-5.55196	-0.00173	7.45726
$D(z)$	118.83309	0.00243	28306188.93642	0.00501	0.00252

Таблица 2: Характеристики положения и рассеяния распределения Коши

Laplace n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.01279	0.00579	0.02646	0.01273	0.00412
$D(z)$	0.09453	0.07397	0.40262	0.09303	0.06910
Laplace n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00358	0.00157	0.00132	-0.01101	0.00213
$D(z)$	0.00990	0.00523	0.42664	0.00966	0.00588
Laplace n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.00036	1.53039	-0.01446	-0.00114	0.00018
$D(z)$	0.00102	0.00050	0.41539	0.00096	0.00058

Таблица 3: Характеристики положения и рассеяния распределения Лапласа

Poisson n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	9.97609	9.8235	10.2955	9.9105	9.8415
$D(z)$	0.97231	1.37659	1.79642	1.22023	1.21450
Poisson n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	9.98505	9.836	10.9405	9.8475	9.84008
$D(z)$	0.10400	0.22110	0.99120	0.16499	0.12406
Poisson n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	9.99791	9.9965	11.639	9.992	9.85761
$D(z)$	0.00964	0.00323	0.61767	0.00443	0.01066

Таблица 4: Характеристики положения и рассеяния распределения Пуассона

Uniform n=10	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00336	0.00843	0.00505	-0.00111	0.00502
$D(z)$	0.10159	0.22721	0.04560	0.14254	0.19700
Uniform n=100	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	0.00098	0.00587	0.00060	-0.01393	0.00199
$D(z)$	0.01036	0.03018	0.00057	0.01545	0.02068
Uniform n=1000	\bar{x}	$med\ x$	z_R	z_Q	z_{tr}
$E(z)$	-0.00111	-0.00171	-0.00003	-0.00240	-0.00139
$D(z)$	0.00100	0.00288	5.99778	0.00154	0.00198

Таблица 5: Характеристики положения и рассеяния равномерного распределения

4.3 Боксплот Тьюки

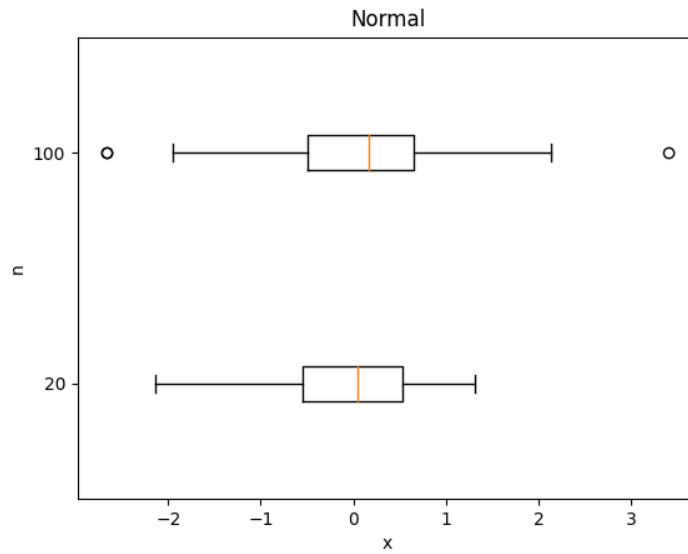


Рис. 6: Боксплот нормального распределения

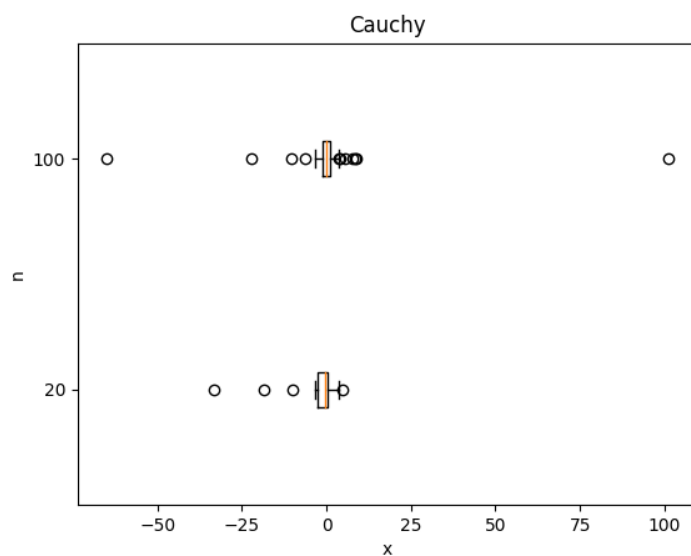


Рис. 7: Боксплот распределения Коши

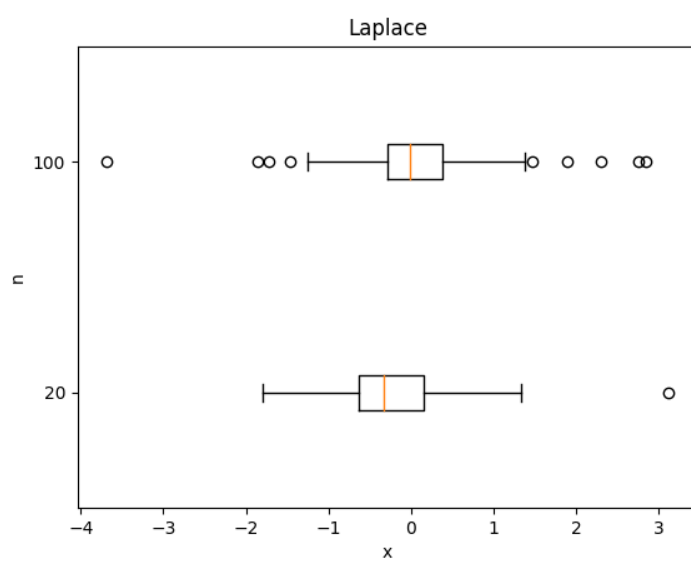


Рис. 8: Боксплот распределения Лапласа

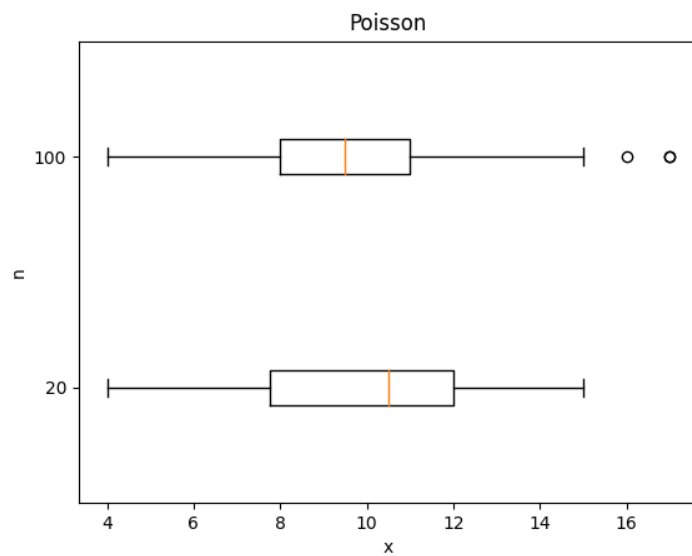


Рис. 9: Боксплот распределения Пуассона

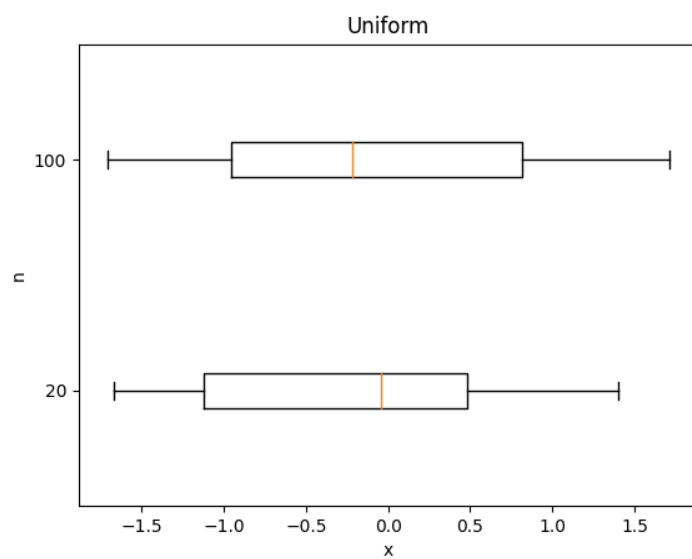


Рис. 10: Боксплот равномерного распределения

4.4 Доля выбросов

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

Таблица 6: Доля выбросов

4.5 Теоретическая вероятность выбросов

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

Таблица 7: Теоретическая вероятность выбросов

5 Обсуждение

Список иллюстраций

1	Нормальное распределение	7
2	Распределение Коши	7
3	Распределение Лапласа	7
4	Распределение Пуассона	8
5	Равномерное распределение	8
6	Боксплот нормального распределения	10
7	Боксплот распределения Коши	11
8	Боксплот распределения Лапласа	11
9	Боксплот распределения Пуассона	12
10	Боксплот равномерного распределения	12

Список таблиц

1	Характеристики положения и рассеяния нормального распределения	8
2	Характеристики положения и рассеяния распределения Коши	9
3	Характеристики положения и рассеяния распределения Лапласа	9
4	Характеристики положения и рассеяния распределения Пуассона	9
5	Характеристики положения и рассеяния равномерного распределения	10
6	Доля выбросов	13
7	Теоретическая вероятность выбросов	13

Литература

- [1] Histogram. URL: <https://en.wikipedia.org/wiki/Histogram>
- [2] Вероятностные разделы математики. Учебник для бакалавров технических направлений. //Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.
- [3] Box plot. URL: https://en.wikipedia.org/wiki/Box_plot
- [4] Анатольев, Станислав (2009) «Непараметрическая регрессия», Квантиль, №7, стр. 37-52.