Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Кафедра «Прикладная математика»

Отчет по лабораторной работе №3

по дисциплине
"Математическая статистика"

Выполнил студент
группы 5030102/00101 Нгуен Хоанг Линь

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

Санкт-Петербург
2023 г.

Содержание

Список иллюстраций

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 $U(x, -\sqrt{3}, \sqrt{3})$

Необходимо:

1) Сгенерировать выборки размером 20 и 100 элементов
2) Построить для них боксплот Тьюки
3) Для каждого распределения определить долю выбросов экспериментально (сгенерировав выборку, соответствующую распределению 1000раз, и вычислив среднюю долю выбросов) и сравнить с результатами, полученными теоретически

2. Теория

2.1. Рассматриваемые распределения

Плотности:

- Нормальное распределение
$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{1}$$
- Распределение Коши
$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \tag{2}$$
- Распределение Лапласа
$$L(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{3}$$
- Распределение Пуассона
$$P(k, 10) = \frac{10^k}{k!} e^{-10} \tag{4}$$
- Равномерное распределение
$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, |x| \leq \sqrt{3} \\ 0, |x| > \sqrt{3} \end{cases} \tag{5}$$

2.2. Боксплот Тьюки

Боксплот Тьюки - график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей. С помощью такого графика можно в удобной форме показать множество характеристик распределения, как медиану, нижний и верхний квартили, минимальное и максимальное значение и выбросы.

2.3. Построение

При построении боксплота границами выступают первый и третий квартили, серединой ящика выступает медиана. Концы "усов края статистически значимой выборки (без выбросов). Длина определяется по формуле:

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1) \tag{6}$$

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

Выбросы выходят за границы усов и отображаются на графике в виде кружков

2.4. Теоретическая вероятность выбросов

Можно вычислить теоретически первый и третий квартили Q_1^T и Q_3^T и нижнюю и верхнюю границы уса X_1^T и X_2^T по формуле (6). После этого можно определить выбросы x :

$$\left[\begin{array}{l} x < X_1^T \\ x > X_2^T \end{array}\right. \tag{7}$$

Теоретическая вероятность выбросов:

- для непрерывных распределений:

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)) \tag{8}$$

- для дискретных распределений:

$$P_B^T = P(x < X_1^T) + P(x > x_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)) \tag{9}$$

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

3. Реализация

Лабораторная работа выполнена на языке Python в виртуальной среде Anaconda с интерпретатором версии 3.9 в среде разработки Visual Studio Code. Дополнительные зависимости:

- scipy
- numpy
- matplotlib
- seaborn

4. Результаты

4.1. Графики

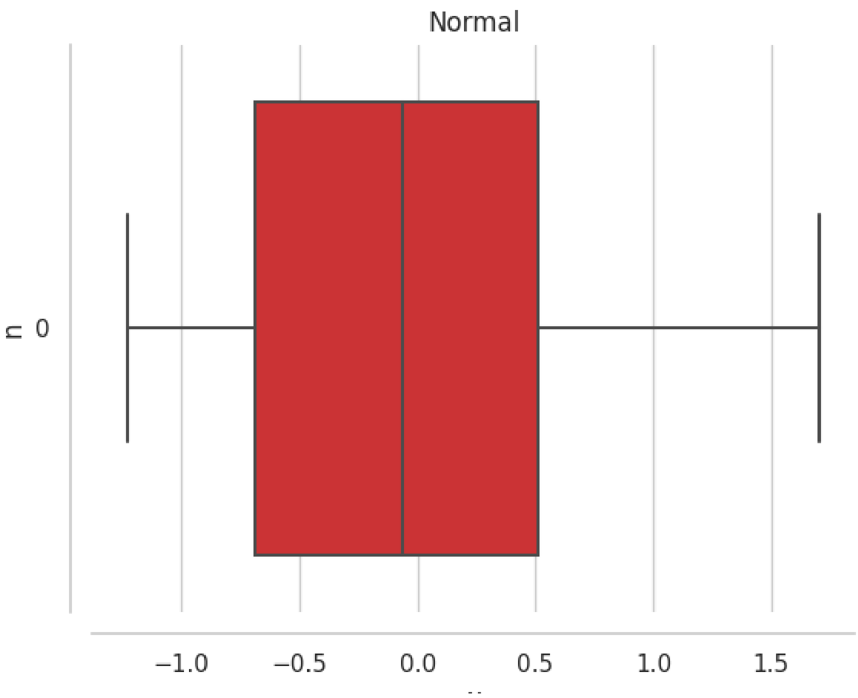

Рис. 1. Нормальное распределение с размером выборки 20

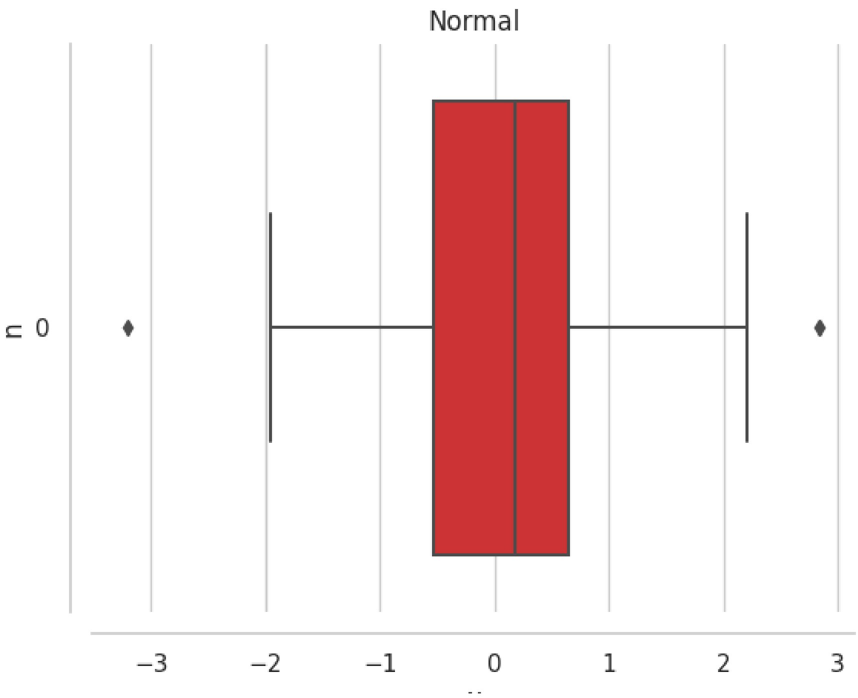

Рис. 2. Нормальное распределение с размером выборки 100

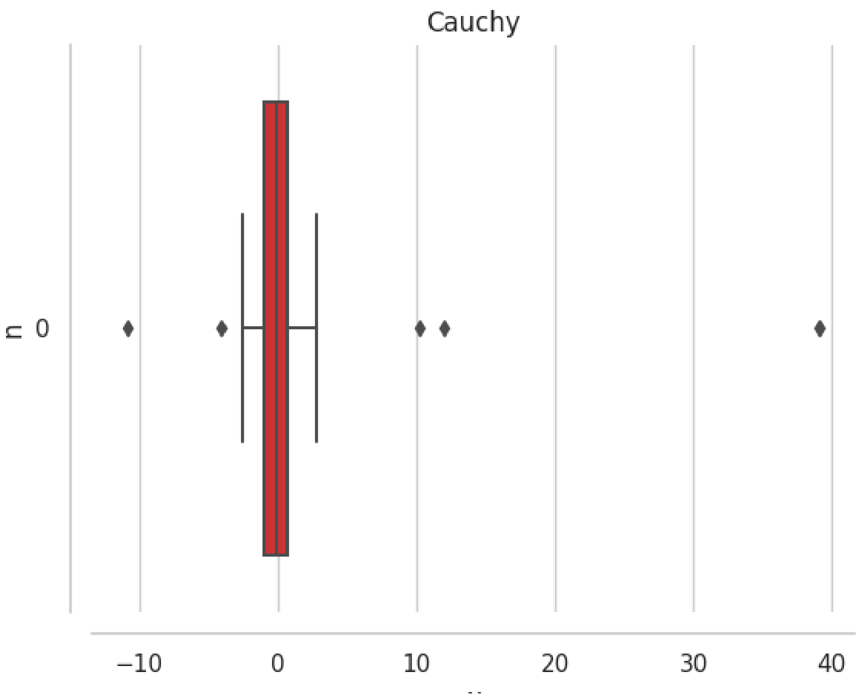

Рис. 3. Распределение Коши с размером выборки 20

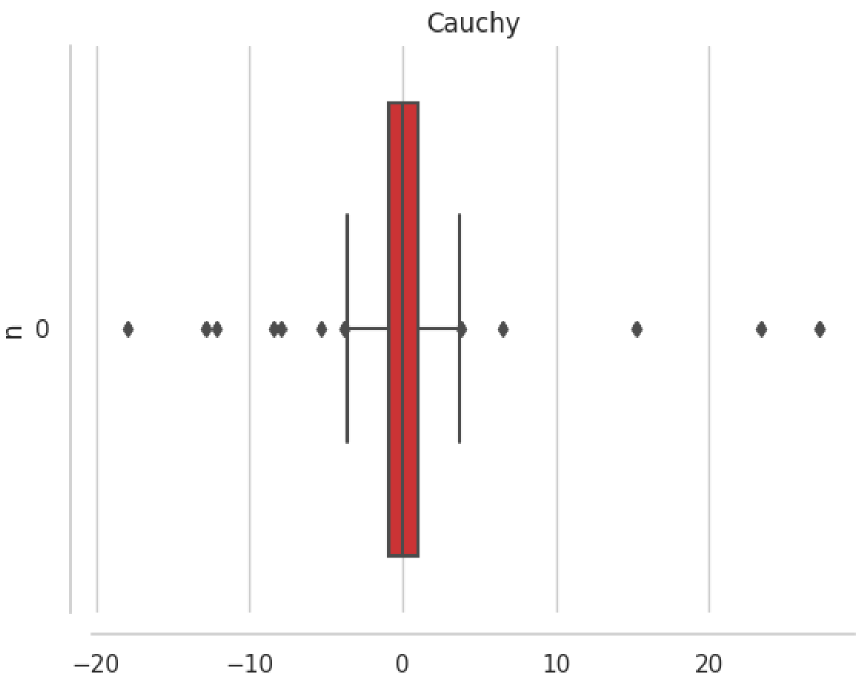

Рис. 4. Распределение Коши с размером выборки 100

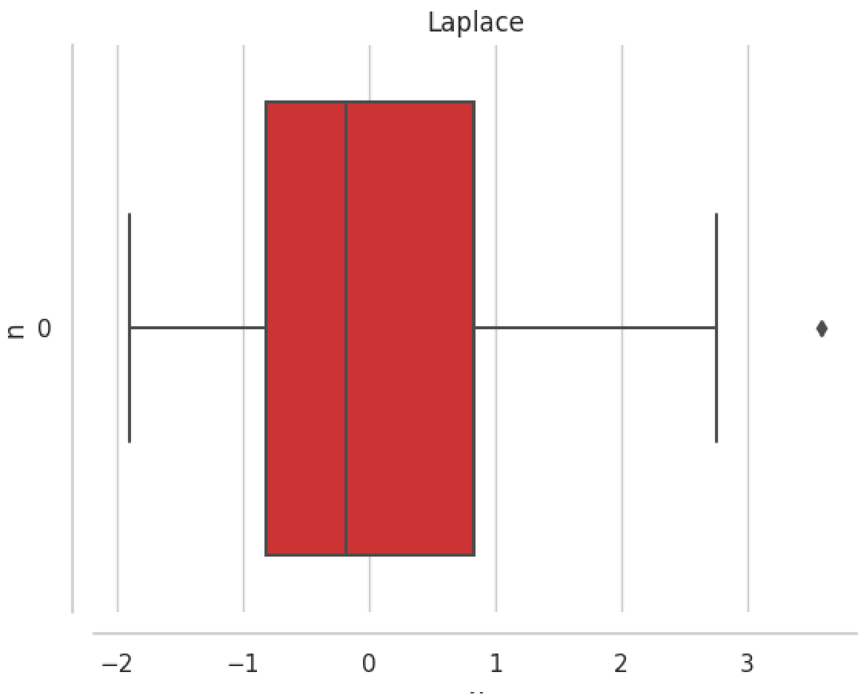

Рис. 5. Распределение Лапласа с размером выборки 20

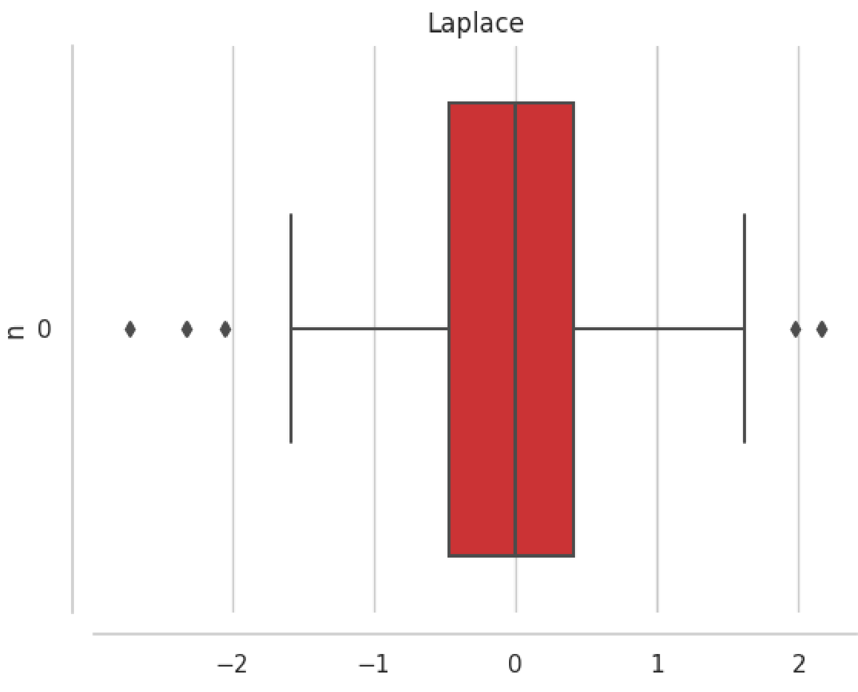

Рис. 6. Распределение Лапласа с размером выборки 100

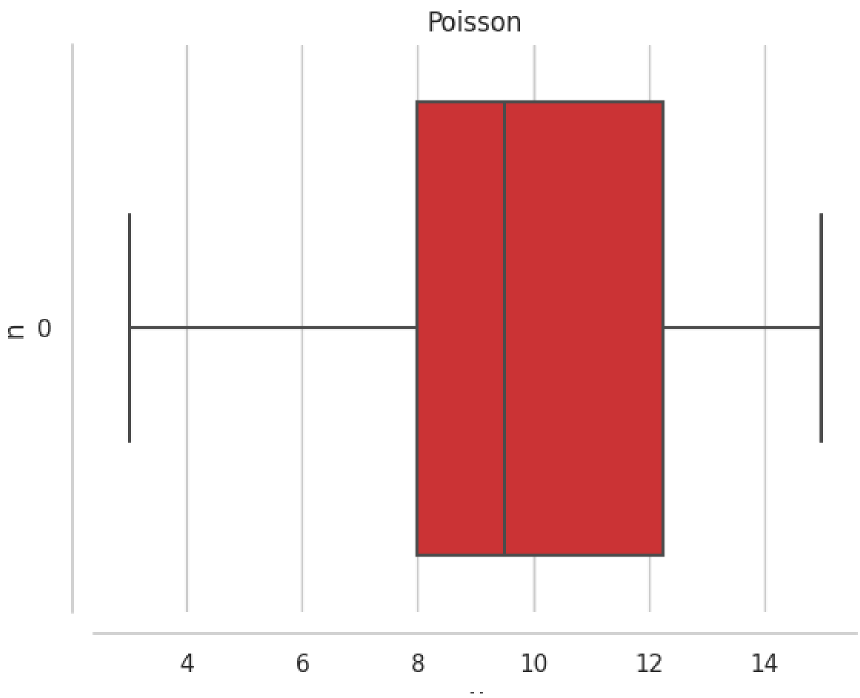

Рис. 7. распределение Пуассона с размером выборки 20

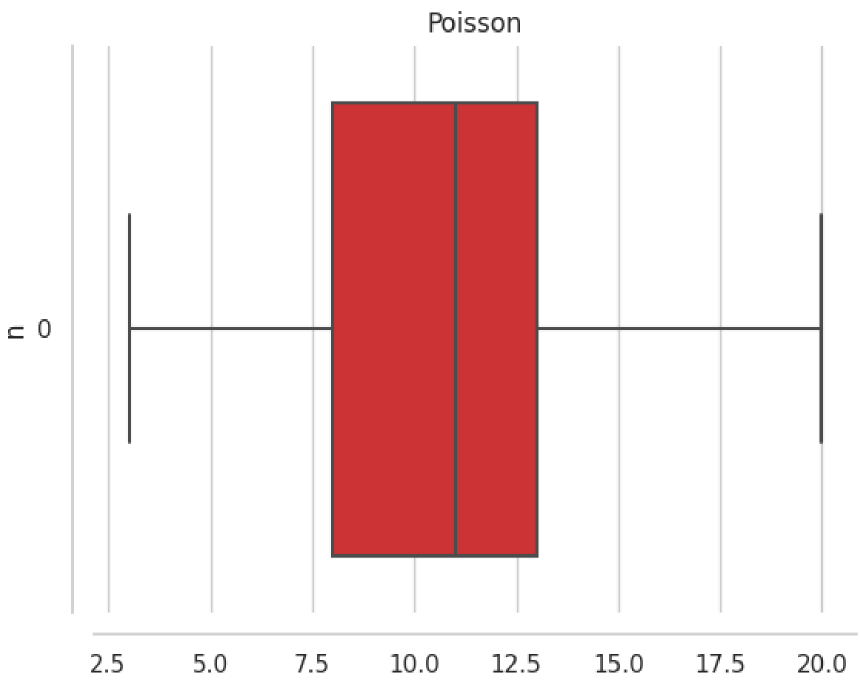

Рис. 8. Распределение Пуассона с размером выборки 100

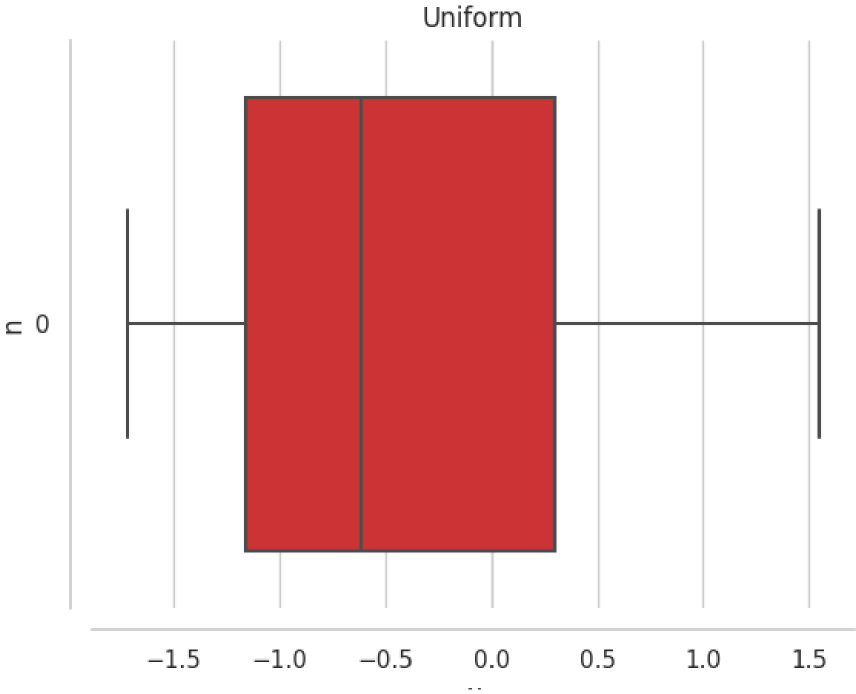

Рис. 9. Равномерное распределение с размером выборки 20

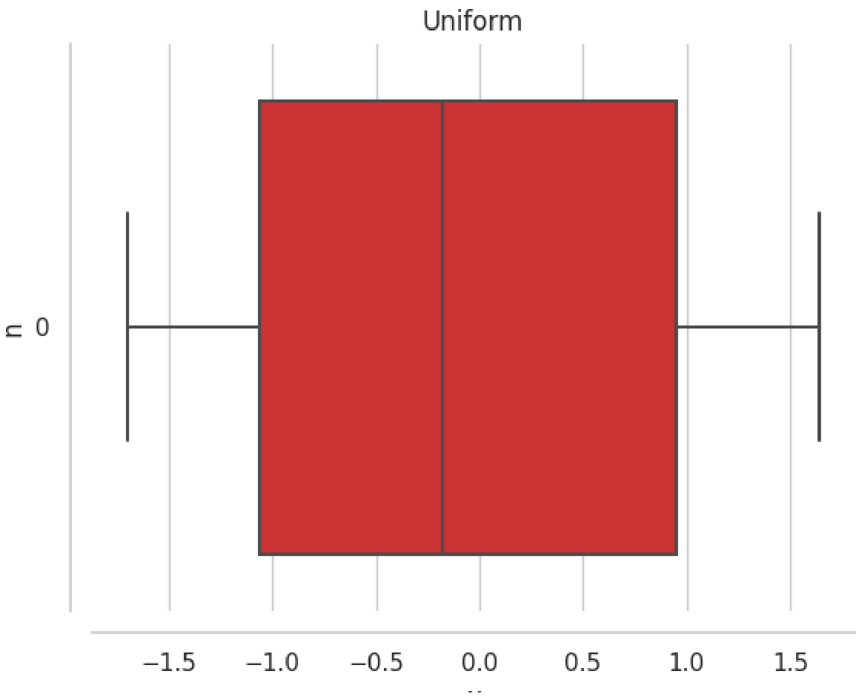

Рис. 10. Равномерное распределение с размером выборки 100

4.2. Доля выбросов

Выборка	Доля выбросов	P_B^T
Нормальное n = 20	0.0229	0.007
Нормальное n = 100	0.0096	0.007
Коши n = 20	0.1526	0.156
Коши n = 100	0.1555	0.156
Лапласа n = 20	0.0744	0.063
Лапласа n = 100	0.0663	0.063
Пуассона n = 20	0.0238	0.008
Пуассона n = 100	0.0096	0.008
Равномерное n = 20	0.0027	0
Равномерное n = 100	0	0

Таблица 1. Практическая доля выбросов

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

Таблица 2. Теоретическая вероятность выбросов

5. Обсуждение

Боксплот Тьюки позволяет удобно продемонстрировать характеристики заданного распределения: медиану, первый и третий квартили, выбросы, максимальные и минимальные значения. Анализировать такой график проще и удобнее, чем аналитические расчёты.

Таблицы показывают, что для всех распределений чем больше размерность выборки, тем ближе найденная доля выбросов к теоретической оценке. В распределении Коши доля выбросов значительно выше, чем во всех остальных распределениях, а у равномерного выбросы почти не наблюдаются.