Санкт-Петербургский политехнический университет Петра Великого

Физико-механический институт

Кафедра прикладной математики и информатики

Математическая статистика

Отчет по лабораторной работе № 2

Выполнил студент гр. 5030102/20202 Тишковец С. Е.

Преподаватель Баженов А.Н.

Санкт-Петербург

2025

Оглавление

1. Постановка задачи

Даны 4 распределения:

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

Задание:

1.1. Сгенерировать выборки размером 20, 100 и 1000 элементов.
1.2. Построить бокс-плоты Тьюки.
1.3. Определить число выбросов, занести в таблицу.
1.4. Обсудить вид бокс-плотов и относительное число выбросов при изменении мощности выборки.

2. Теоретическая информация

2.1. Распределения

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

2.2. Бокс-плот Тьюти

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

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.

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 квартиля и полутора межквартильных расстояний. Формула имеет вид

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

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. Выбросами считаются такие величины, что:

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

3. Результаты исследования

3.1. Графики

3.1.1. Нормальное распределение

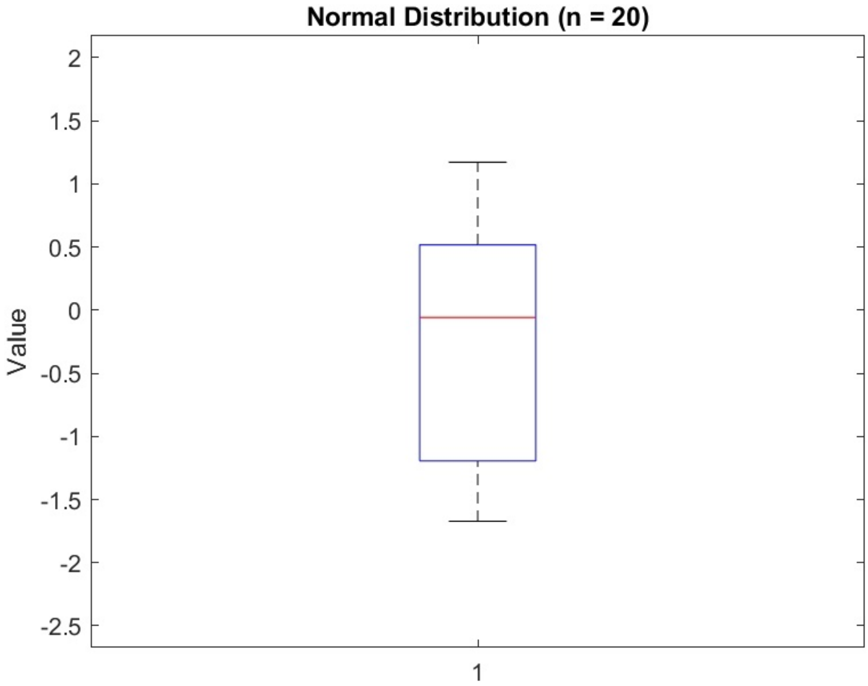

Рис. 1. Выборка размером 20 элементов

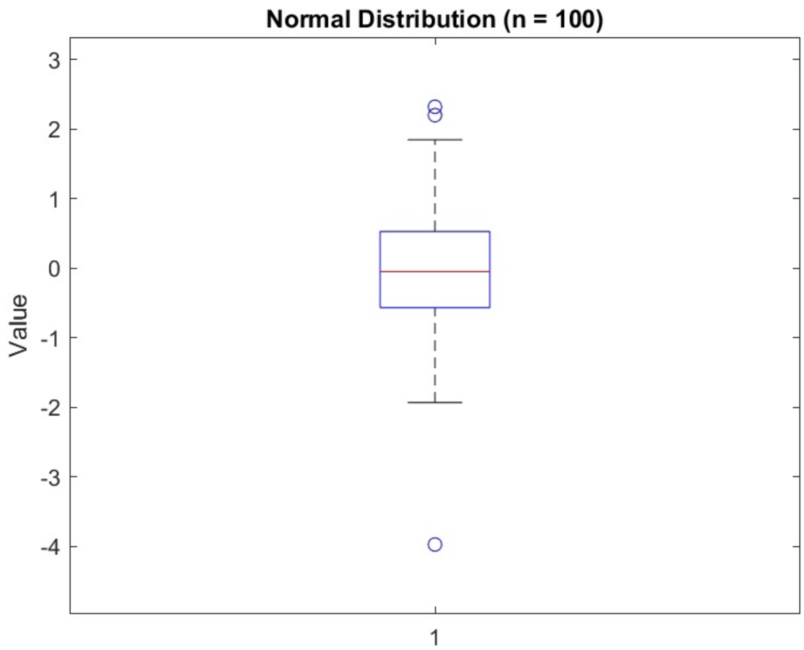

Рис. 2. Выборка размером 100 элементов

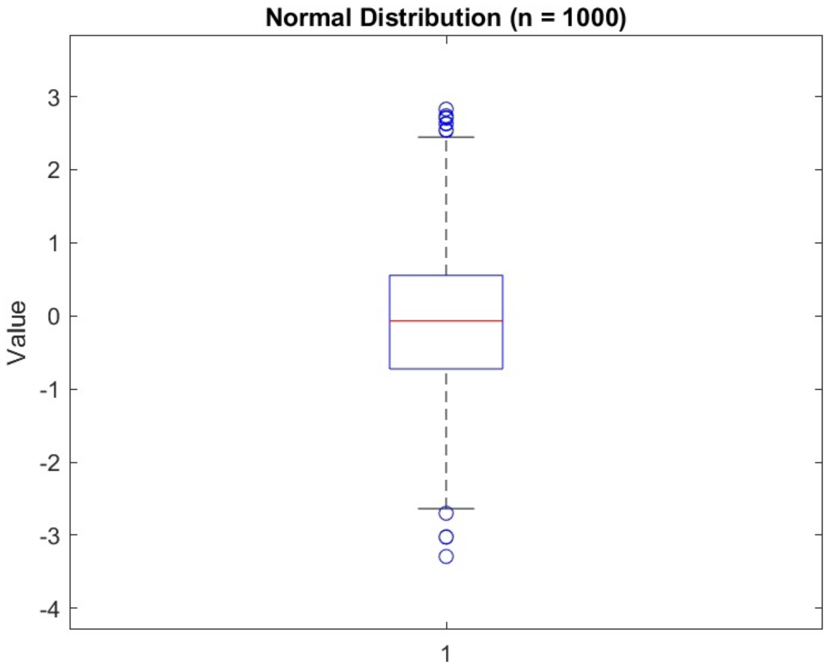

Рис. 3. Выборка размером 1000 элементов

3.1.2. Распределение Коши

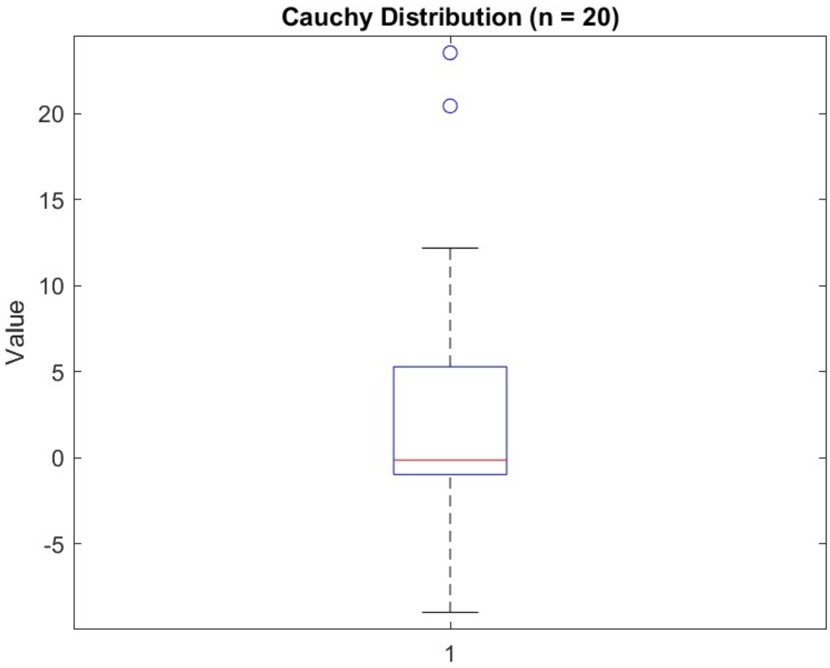

Рис. 4. Выборка размером 20 элементов

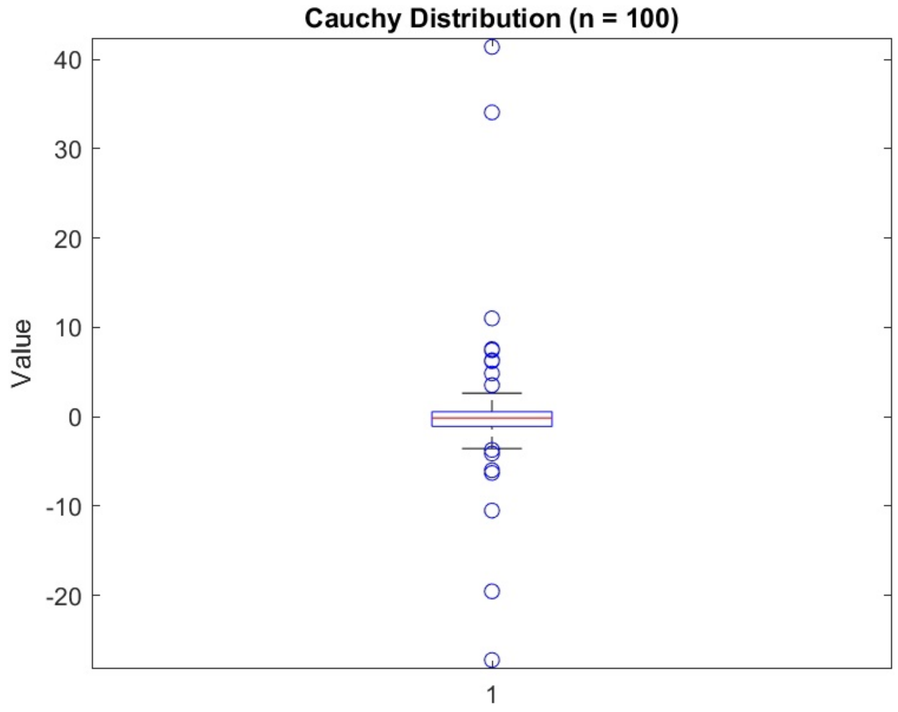

Рис. 5. Выборка размером 100 элементов

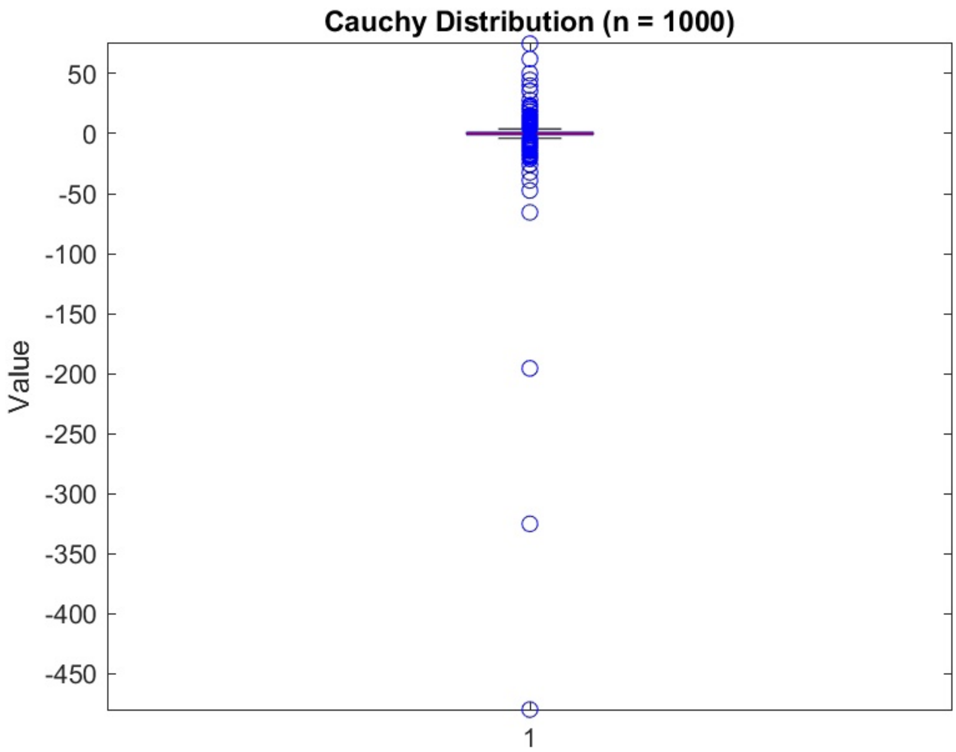

Рис. 6. Выборка размером 1000 элементов

3.1.3. Распределение Пуассона

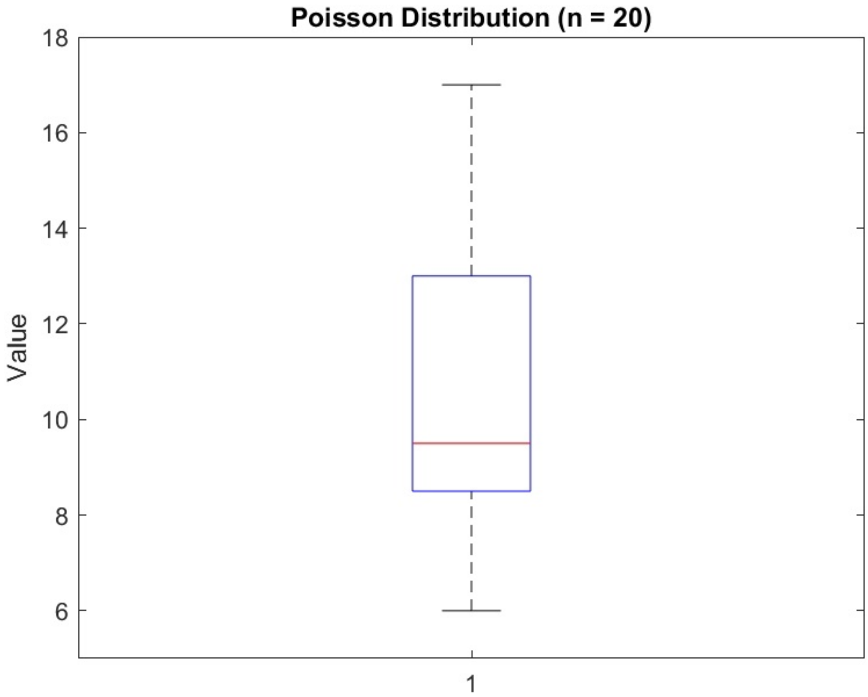

Рис. 7. Выборка размером 20 элементов

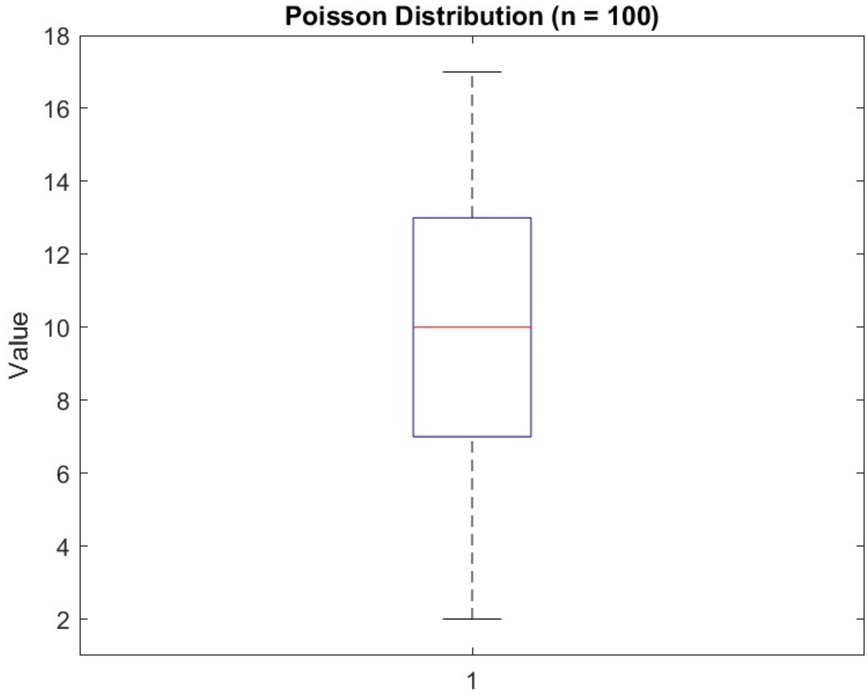

Рис. 8. Выборка размером 100 элементов

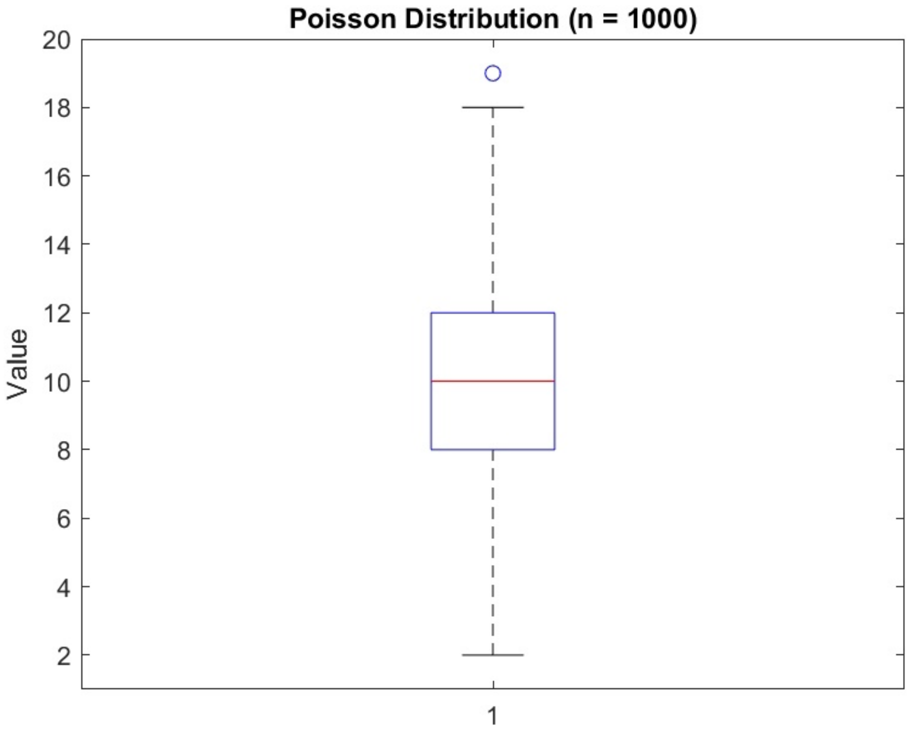

Рис. 9. Выборка размером 1000 элементов

3.1.4. Равномерное распределение

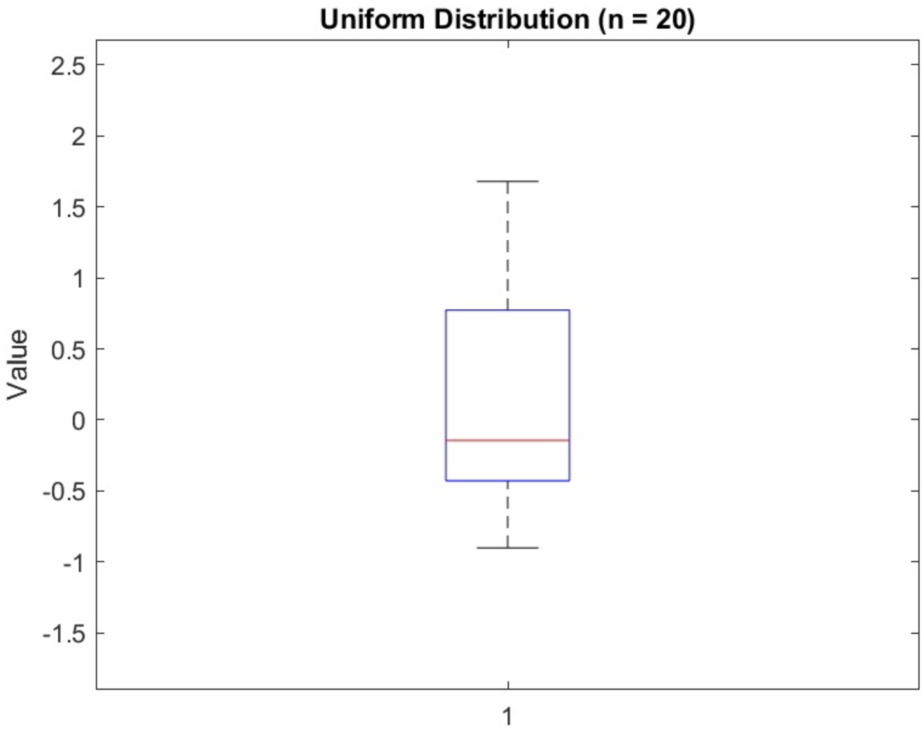

Рис. 10. Выборка размером 20 элементов

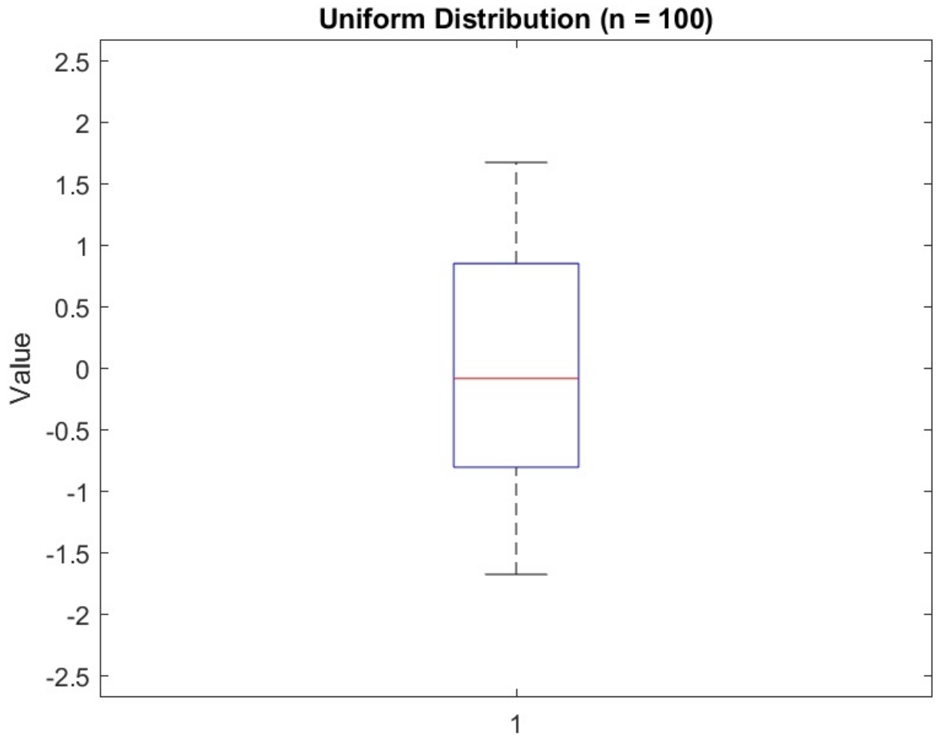

Рис. 11. Выборка размером 100 элементов

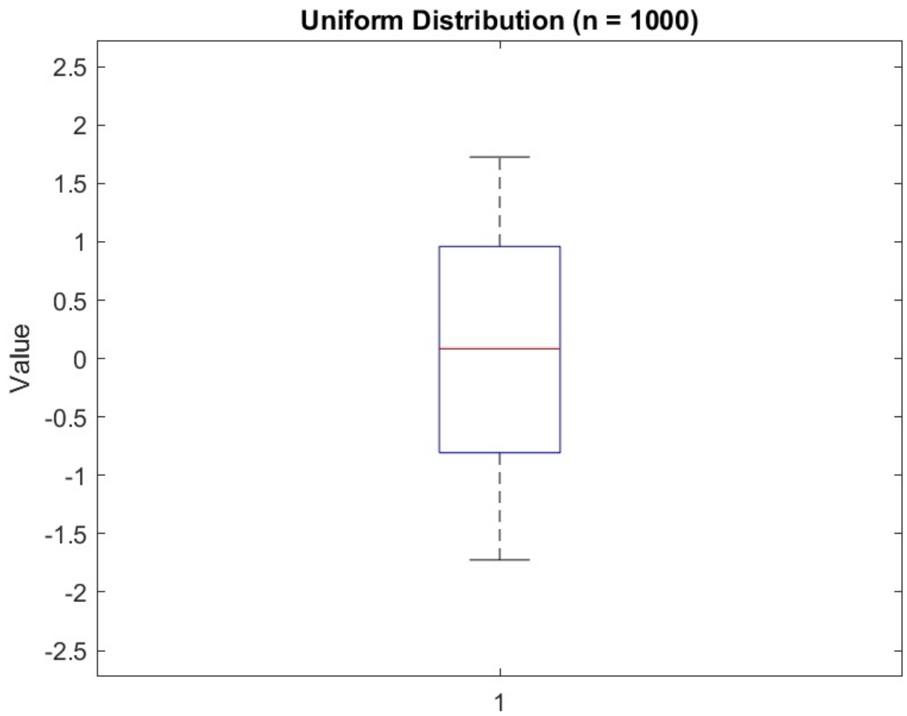

Рис. 12. Выборка размером 1000 элементов

3.2. Таблицы

Размер выборки	Распределение			
	Нормальное	Коши	Пуассона	Равномерное
20	0	4	0	0
100	3	14	0	0
1000	13	172	6	0

Таблица 1. Количество выбросов для каждого распределения в зависимости от размера выборки

4. Выводы

4.1. С увеличением выборки нормальное распределение демонстрирует достаточно слабое увеличение числа выбросов, но их относительная доля остаётся низкой.

4.2. Распределение Коши при увеличении выборки показывает значительное увеличение числа выбросов. С увеличением выборки наблюдается растущий уровень выбросов, что подтверждает характерное для Коши свойство (длинные хвосты). Это указывает на влияние особенно крайних значений в больших объёмах данных.

4.3. Показатель выбросов для распределения Пуассона остаётся практически неизменным даже при увеличении выборки, что свидетельствует о том, что выборки поддерживают кумулятивные значения последовательно. Вероятность появления более высоких значений оказалась одинаковой на больших объёмах данных.

4.4. Равномерное распределение: полное отсутствие выбросов как в малых, так и в больших выборках подтверждает равномерность данных и отсутствие крайних значений. Это иллюстрирует идеальные границы, в которых распределены значения.