Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа прикладной математики и вычислительной физики

Математическая статистика

Отчёт по лабораторной работе №3-4

Работу выполнил:
П. П. Филиппов
Группа:
5030102/10101
Преподаватель:
А. Н. Баженов

Санкт-Петербург
2024

Содержание

1. Постановка задачи

Даны 5 распределений:

- Нормальное распределение:
$$N(x, 0, 1)$$
- Распрделение Коши:
$$C(x, 0, 1)$$
- Распределение Стьюдента:
$$t(x, 0, 3)$$
- Распределение Пуассона:
$$P(k, 10)$$
- Нормальное распрделение:
$$U(x, -\sqrt{3}, \sqrt{3})$$

1. 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.

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 Стьюдента
$$t(x, 0, 3) = \frac{Y_0}{\sqrt{\sum\limits_{i=0}^{3} Y_i^2}}, Y_i \sim N(0, 1)$$
- Распределение Пуассона
$$P(k, 10) = \frac{10^k}{k!} e^{-10}$$
- Нормальное распределение
$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, |x| \leq \sqrt{3} \\ 0, |x| > \sqrt{3} \end{cases}$$

2.2. Боксплот Тьюки

2.2.1. Определение

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

2.2.2. Описание

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

2.2.3. Построение

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

$$X_1 = Q_1 - \ (Q_3 - Q_1), X_2 = Q_3 - \ (Q_3 - Q_1)$$

,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

3. Изображения

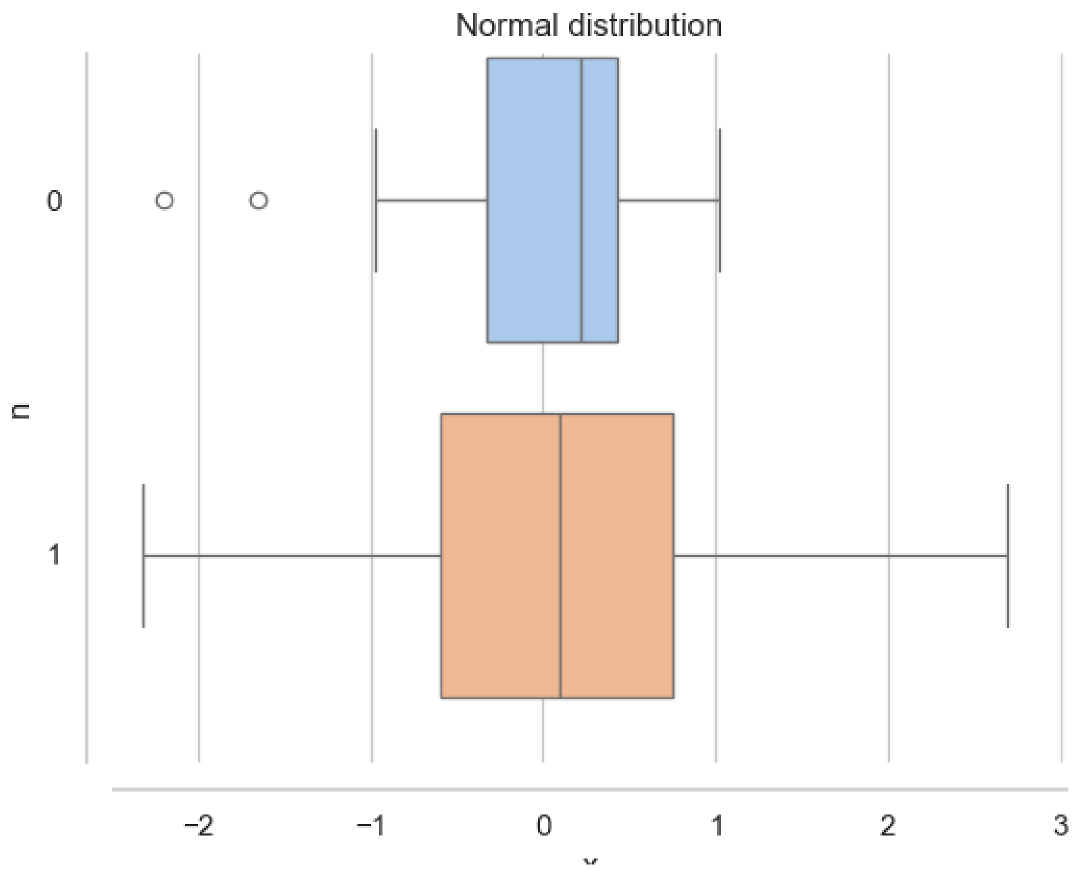

Рисунок 3.1. Нормальное распределение

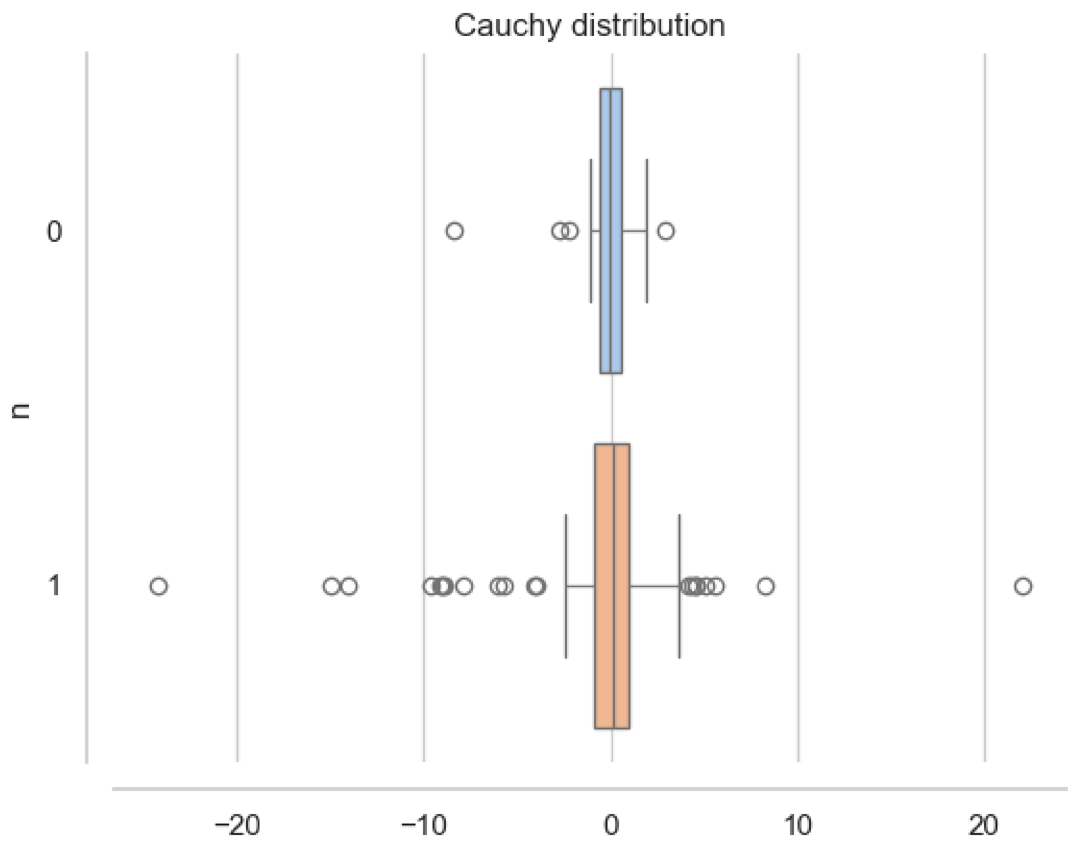

Рисунок 3.2. Распределение Коши

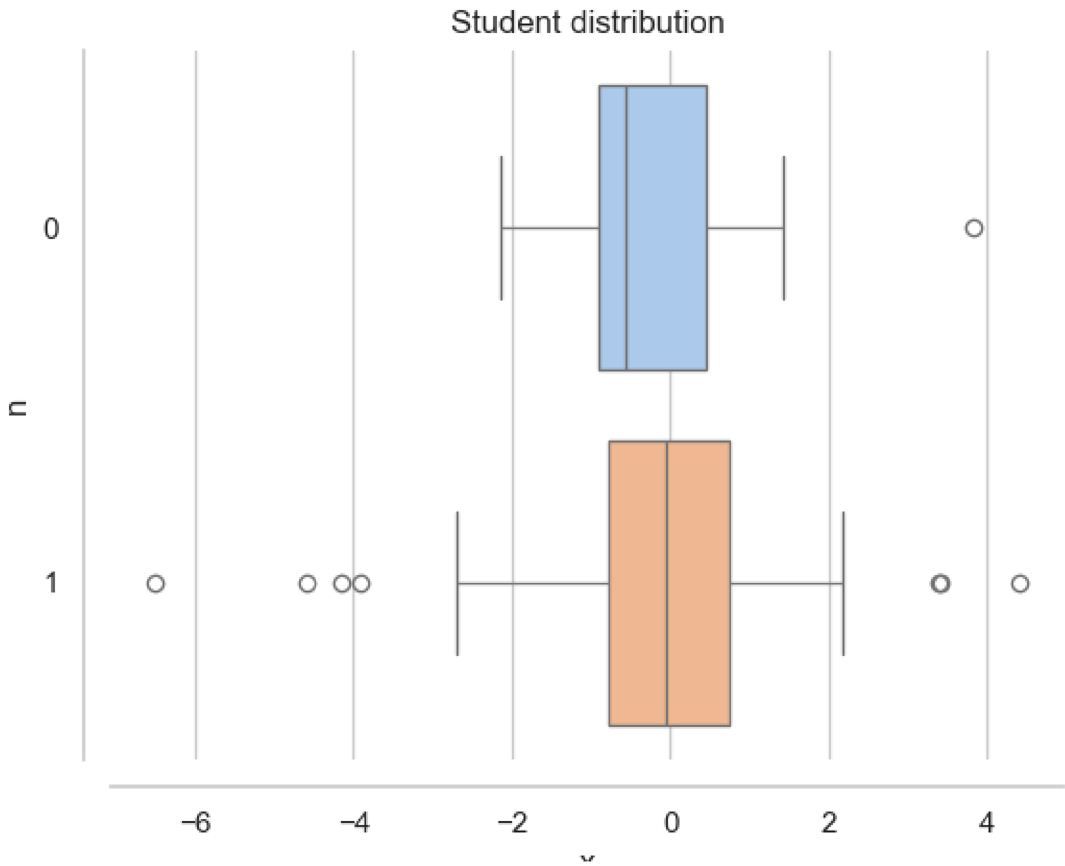

Рисунок 3.3. Распределение Стьюдента

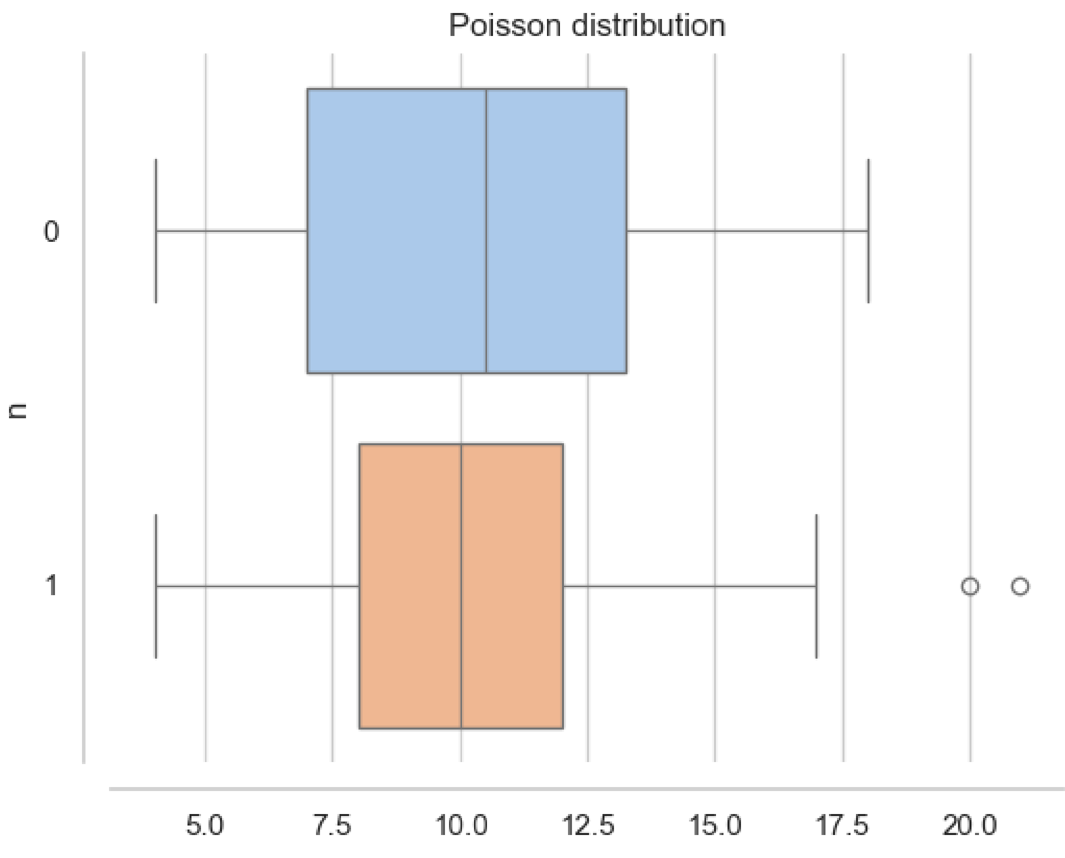

Рисунок 3.4. Распределение Пуассона

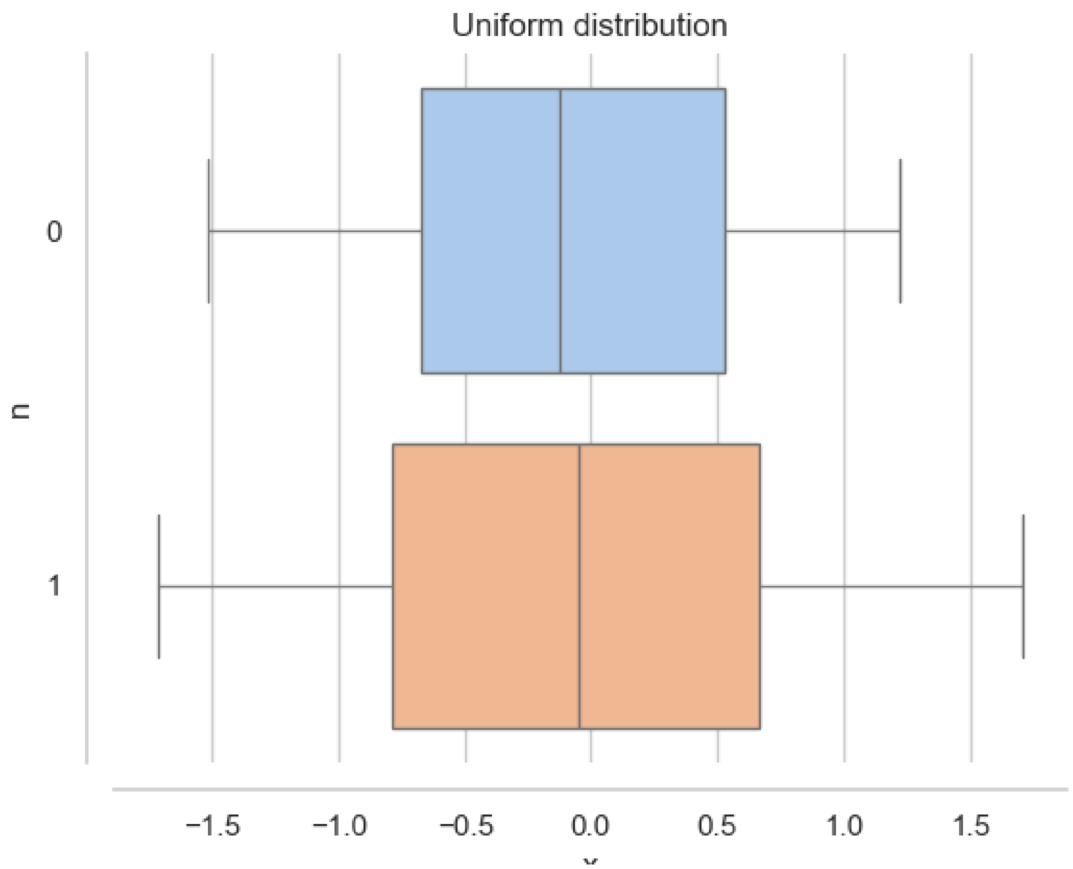

Рисунок 3.5. Равномерное распределение