

Санкт-Петербургский Политехнический Университет Петра
Великого
Физико-механический институт

Высшая школа прикладной математики и вычислительной
физики

Лабораторная работа №2
по дисциплине “Математическая статистика”

Обучающаяся:

А.Д. Балакшина
(группа 5030102/20101)

Преподаватель:

А.Н. Баженов

Санкт-Петербург

2025

Содержание

1	Формулировка задания	3
2	Формализация	4
3	Выполнение работы	4
4	Результаты	5
4.1	Бокс-плоты	5
4.2	Количество выбросов	9
5	Оценка характера выбросов и вида боксплотов	9

1 Формулировка задания

Для 4 распределений:

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

1. Сгенерировать выборки размером 20, 100 и 1000 элементов.
2. Построить бокс-плоты Тьюки
3. Определить число выбросов, занести в таблицу
4. Оценить вид бокс-плотов и относительное число выбросов при изменении мощности выборки

2 Формализация

В данной работе вычисление числа выбросов основано на межквартильном размахе (IQR). Это значение определялось следующими величинами:

1. Квартили:

Q_1 (первый квартиль): значение, ниже которого находится 25% данных.

Q_3 (третий квартиль): значение, ниже которого находится 75% данных.

2. Межквартильный размах: разница между третьим и первым квартилями:

$$IQR = Q_3 - Q_1$$

3. Границы выбросов:

$$Upper = Q_3 + 1.5 * IQR$$

$$Lower = Q_1 - 1.5 * IQR$$

В итоге, любое значение, которое оказалось меньше нижней границы или больше верхней границы, считалось выбросом.

3 Выполнение работы

Лабораторная работа выполнена на языке программирования Python 3.12 с использованием библиотек `numpy`, `seaborn`, `pandas`, `matplotlib`.

Были сгенерированы выборки, построены бокс-плоты Тьюки (сохранялись в виде файлов `png`), определено число выбросов (выводилось в консоль в формате таблиц LATEX). Программа отработала корректно.

4 Результаты

4.1 Бокс-плоты

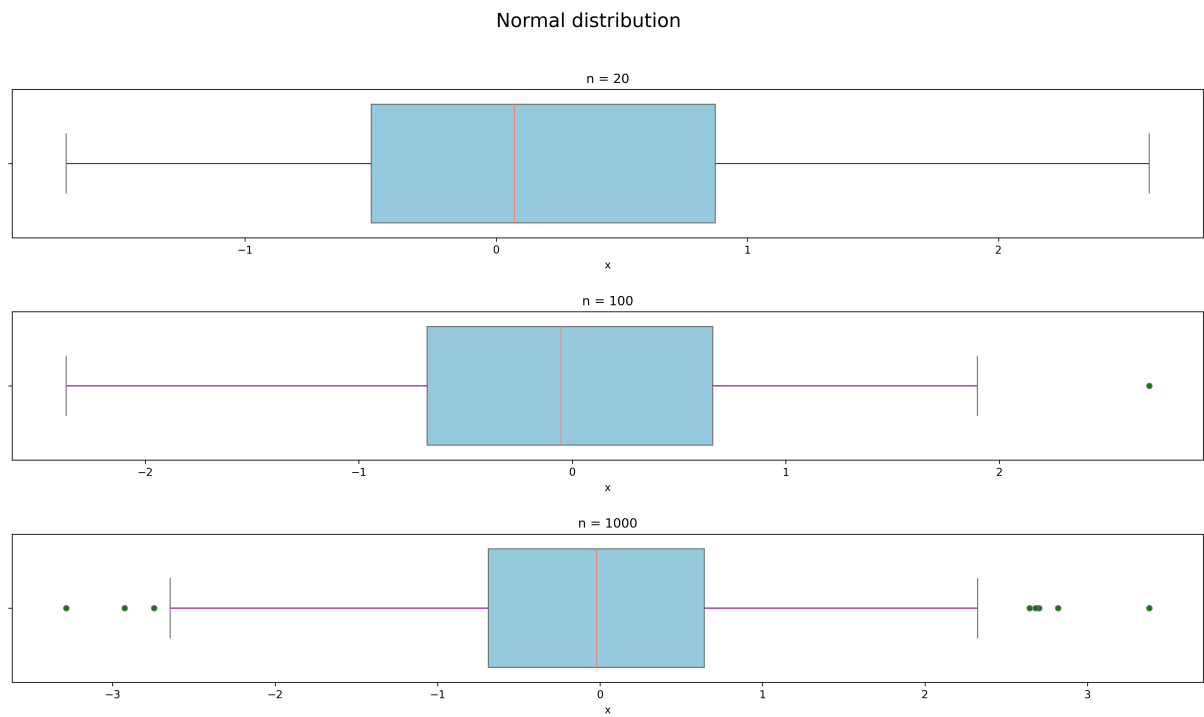


Рис. 1: Нормальное распределение.

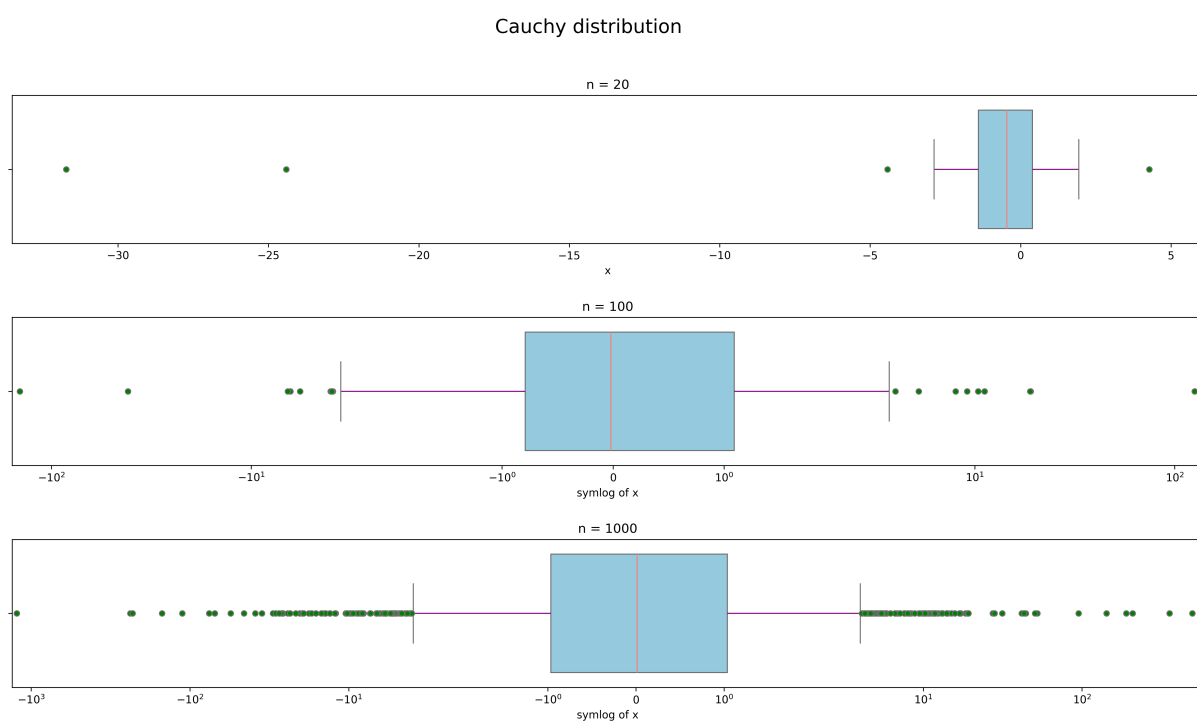


Рис. 2: Распределение Коши.

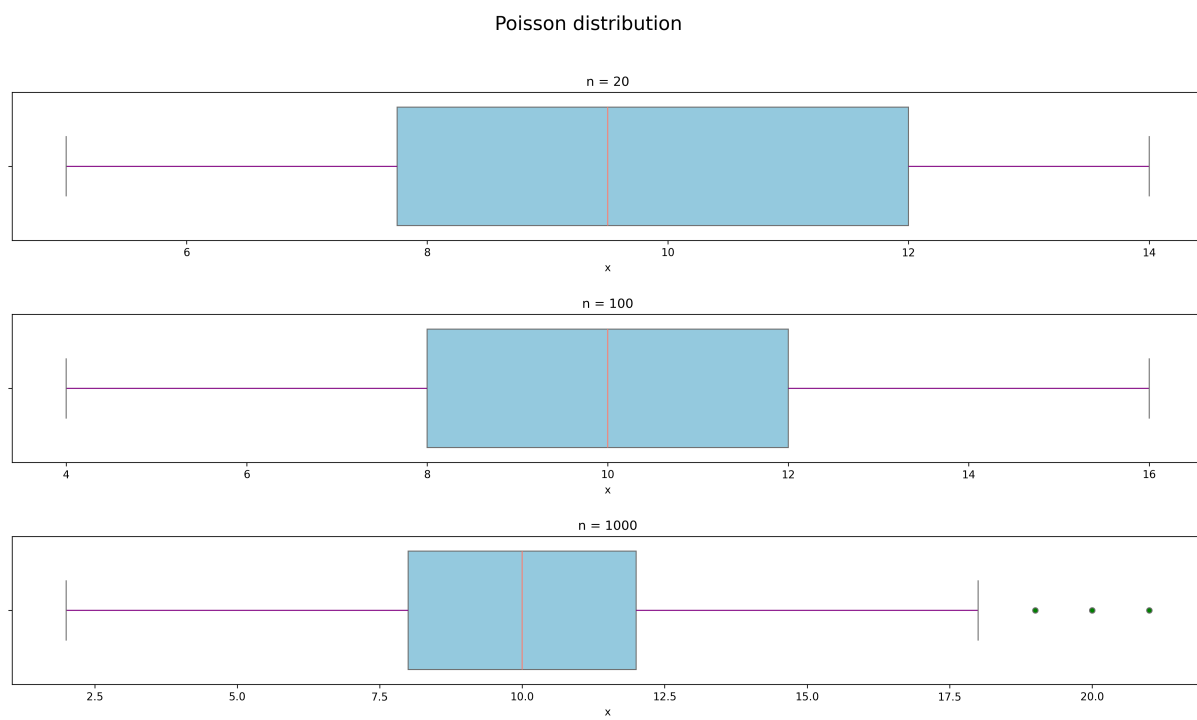


Рис. 3: Распределение Пуассона.

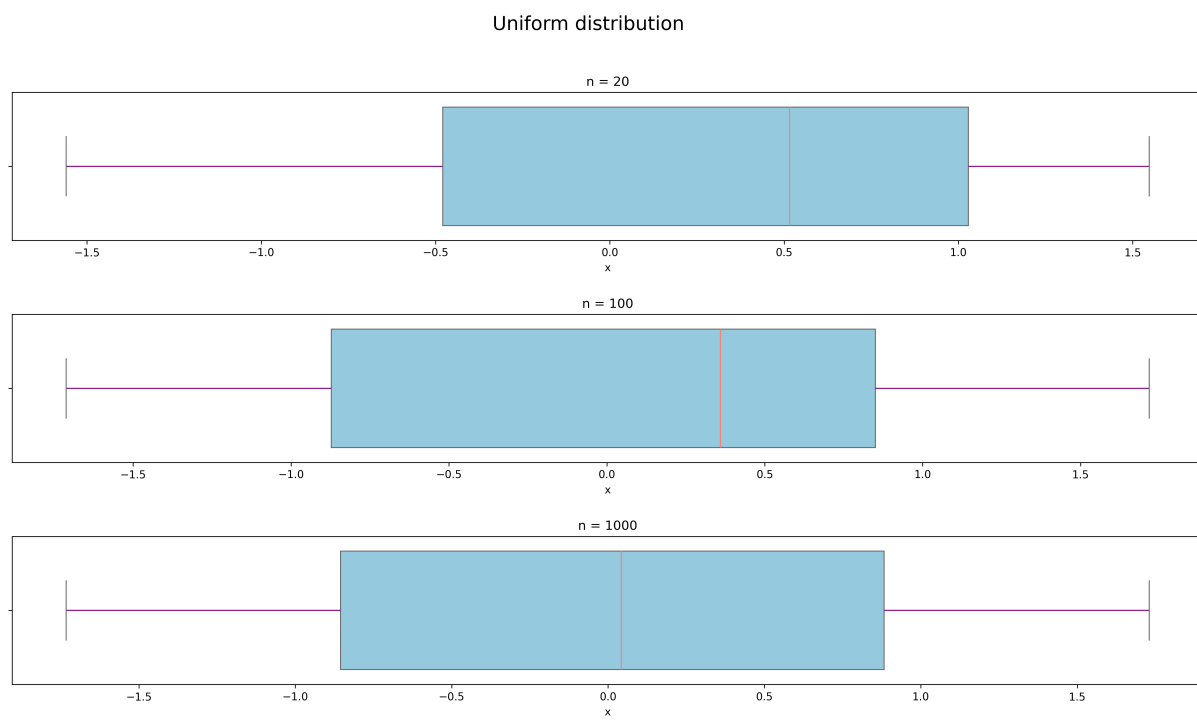


Рис. 4: Равномерное распределение.

4.2 Количество выбросов

Cauchy distribution

Sample size	Outliers
20	3
100	19
1000	170

Normal distribution

Sample size	Outliers
20	0
100	0
1000	9

Poisson distribution

Sample size	Outliers
20	0
100	1
1000	10

Uniform distribution

Sample size	Outliers
20	0
100	0
1000	0

5 Оценка характера выбросов и вида боксплотов

1. Нормальное распределение

Боксплот: При увеличении размера выборки стремится к симметричной форме, с медианой, близкой к 0.

Выбросы: В нормальном распределении выбросы встречаются редко, так как вероятность появления значений, удалённых от среднего, убывает экспоненциально быстро. Однако, так как это рас-

пределение определено на бесконечном носителе, при увеличении выборки количество выбросов увеличивается.

2. Распределение Коши

Боксплот: Медиана близка к 0, но из-за тяжёлых хвостов распределения Коши межквартильный размах (IQR) очень широк. Усы на боксплоте длинные, так как распределение Коши имеет бесконечную дисперсию и часто порождает экстремальные выбросы.

Выбросы: У распределения Коши "хвосты" убывают медленно (по степенному закону), поэтому выбросы встречаются очень часто. При увеличении выборки количество выбросов будет расти пропорционально объёму данных.

3. Распределение Пуассона

Боксплот: Асимметричный, с медианой около 10 (λ).

Выбросы: В распределении Пуассона выбросы возможны, но не являются такими экстремальными, как в распределении Коши. Выбросы появляются реже, чем у Коши, но чаще, чем у нормального распределения. Математически это следует из асимптотики $P(k) = \frac{10^k * e^{-10}}{k!}$, что убывает быстрее, чем степенной закон, но медленнее, чем экспоненциальный. При увеличении выборки количество выбросов растёт (так как распределение определено на бесконечном носителе), но их доля остаётся небольшой.

4. Равномерное распределение

Боксплот: Медиана, близка к 0. Усы простираются до границ распределения.

Выбросы: В равномерном распределении выбросы отсутствуют, так как все значения ограничены интервалом $[a, b]$. При увеличении выборки выбросы не появятся.