

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторной работе №3
по дисциплине
«Математическая статистика»

Выполнил студент:
Самутичев Евгений Романович
группа: 3630102/70201

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

Санкт-Петербург
2020 г.

Содержание

| | | |
|----------|--|-----------|
| 1 | Постановка задачи | 2 |
| 2 | Теория | 3 |
| 2.1 | Выбросы | 3 |
| 2.2 | Боксплот Тьюки | 3 |
| 2.2.1 | Описание | 3 |
| 2.2.2 | Построение | 3 |
| 3 | Реализация | 4 |
| 4 | Результаты | 5 |
| 4.1 | Боксплоты | 5 |
| 4.2 | Теоретическая вероятность выбросов | 8 |
| 4.3 | Доля выбросов | 8 |
| 5 | Обсуждение | 9 |
| 6 | Приложения | 10 |

Список иллюстраций

| | | |
|---|-------------------------------------|---|
| 1 | Нормальное распределение | 5 |
| 2 | Распределение Коши | 5 |
| 3 | Распределение Лапласа | 6 |
| 4 | Распределение Пуассона | 6 |
| 5 | Равномерное распределение | 7 |

Список таблиц

| | | |
|---|--|---|
| 1 | Теоретическая вероятность выбросов | 8 |
|---|--|---|

1 Постановка задачи

Для каждого из 5 распределений:

1. Нормального $N(x, 0, 1)$
2. Коши $C(x, 0, 1)$
3. Лапласа $L(x, 0, \frac{1}{\sqrt{2}})$
4. Пуассона $P(k, 10)$
5. Равномерного $U(x, -\sqrt{3}, \sqrt{3})$

сгенерировать выборки размера 20 и 100, построить боксплот Тьюки. Определить долю выбросов экспериментально (сгенерировав выборку каждого размера 1000 раз) и сравнить с результатами полученными теоретически.

2 Теория

2.1 Выбросы

Результат измерения, выделяющийся из выборки называется *выбросом*. Простейший критерий основан на межквартильном расстоянии, выбросами считаются элементы выборки лежащие вне диапазона $[X_1, X_2]$:

$$X_1 = LQ - \frac{3}{2}(UQ - LQ), X_2 = UQ + \frac{3}{2}(UQ - LQ) \quad (1)$$

, где LQ, UQ - выборочные нижний и верхний квартили.

Теоретическая вероятность выбросов для непрерывных распределений:

$$P_{outlier} = P(x < X_1) + P(x > X_2) = F(X_1) + (1 - F(X_2)) \quad (2)$$

, а для дискретных с учетом возможного скачка

$$P_{outlier} = F(X_1) - (F(X_1+) - F(X_1)) + (1 - F(X_2)) \quad (3)$$

2.2 Боксплот Тьюки

2.2.1 Описание

Боксплот (англ. box plot) — график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей: в удобной форме показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы. [2]

2.2.2 Построение

Границами ящика служат LQ и UQ , линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выбросов): X_1 и X_2 (1).

3 Реализация

Работа выполнена с использованием языка **Python** в интегрированной среде разработки **PyCharm**, были задействованы библиотеки:

- **NumPy** - вычисление квартилей для дальнейшего подсчета выбросов
- **SciPy** - модуль **stats** для генерации данных по распределениям
- **Matplotlib** - построение боксплотов

Исходный код работы приведен в приложении.

4 Результаты

4.1 Боксплоты

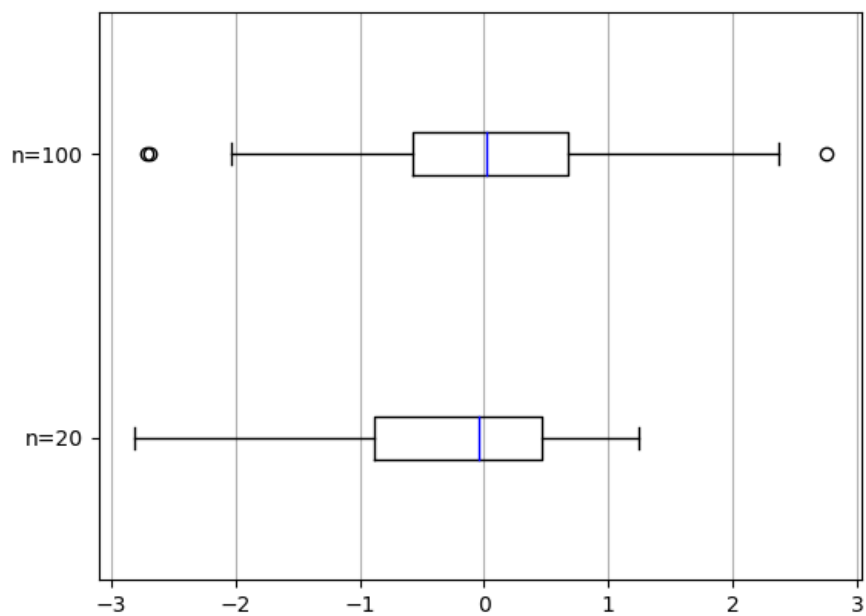


Рис. 1: Нормальное распределение

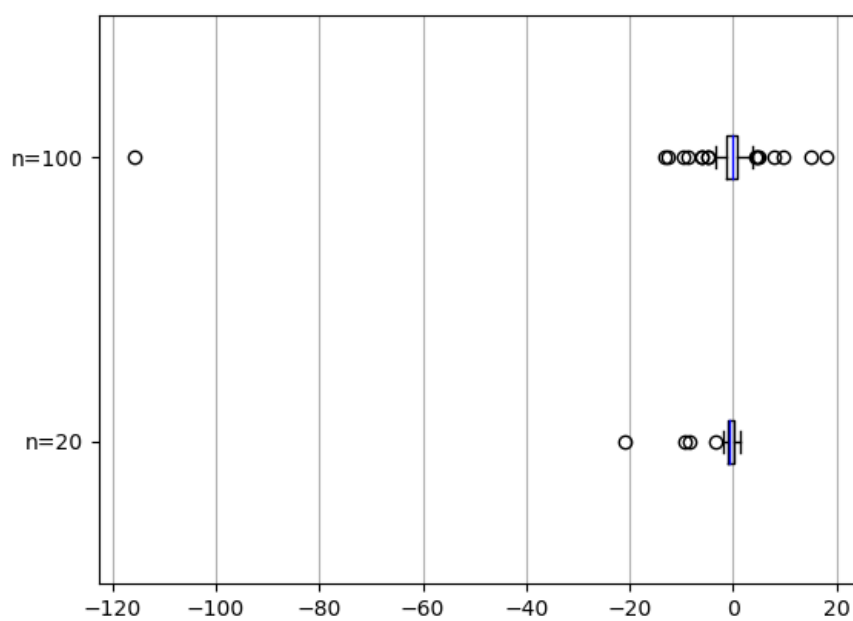


Рис. 2: Распределение Коши

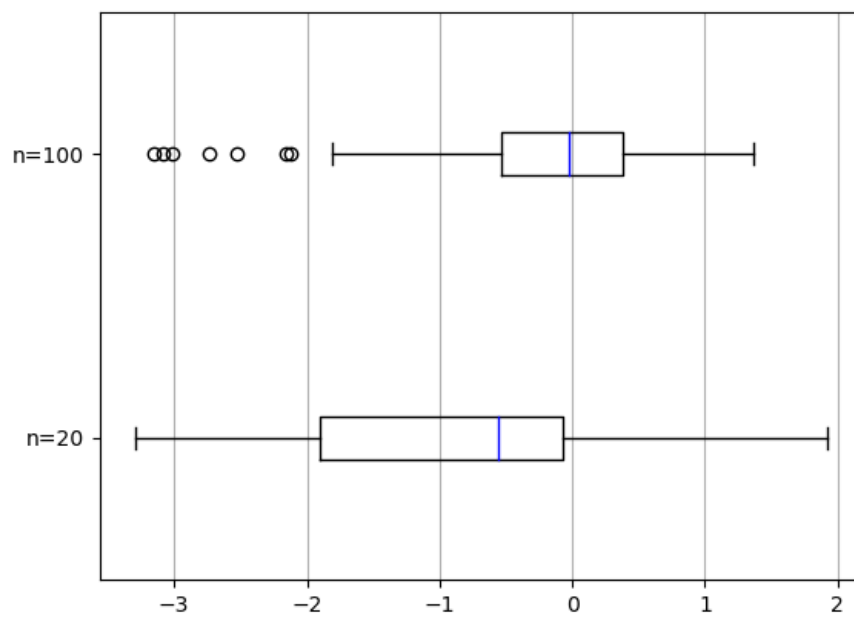


Рис. 3: Распределение Лапласа

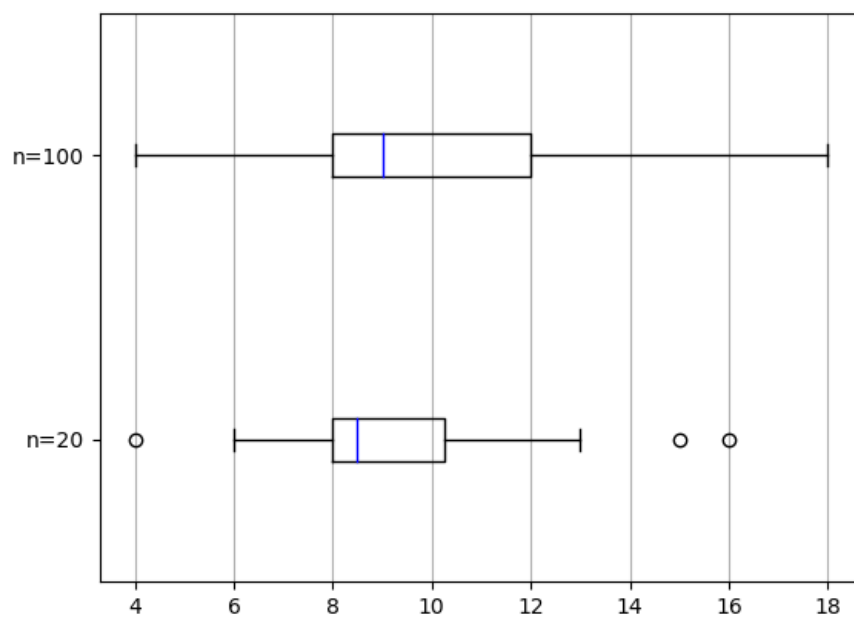


Рис. 4: Распределение Пуассона

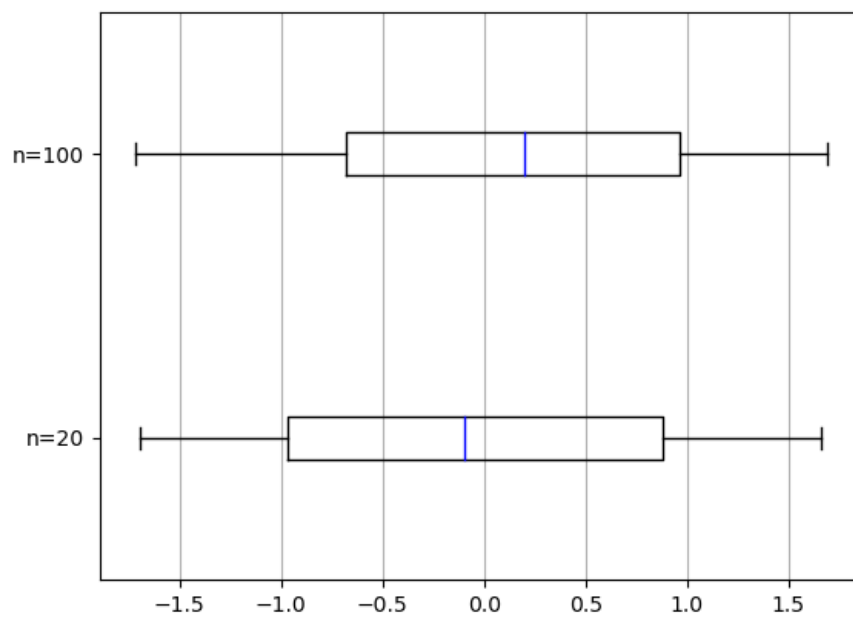


Рис. 5: Равномерное распределение

4.2 Теоретическая вероятность выбросов

Подсчитана для каждого распределения при помощи модуля `stats` библиотеки SciPy (см. Реализация):

| Распределение | normal | cauchy | laplace | poisson | uniform |
|-----------------------|--------|--------|---------|---------|---------|
| $P_{outlier}(2), (3)$ | 0.007 | 0.156 | 0.0625 | 0.008 | 0.0 |

Таблица 1: Теоретическая вероятность выбросов

4.3 Доля выбросов

5 Обсуждение

6 Приложения

1. Исходный код лабораторной <https://github.com/zhenyatos/statlabs/tree/master/Lab3>

Список литературы

- [1] Н. И. Чернова, *Математическая статистика: Учеб. пособие*. Новосиб. гос. ун-т. Новосибирск, 2007. 148 стр.
- [2] Ящик с усами // Википедия. [2020—2020]. Дата обновления: 12.01.2020. URL: <https://ru.wikipedia.org/?oldid=104502300> (дата обращения: 12.01.2020)