

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ им. ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ

КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ

ЛАБОРАТОРНАЯ РАБОТА №3 БОКСПЛОТ ТЬЮКИ

3 КУРС, ГРУППА 3630102/70301

Студент

Лебедев К.С.

Преподаватель

Баженов А. Н.

САНКТ-ПЕТЕРБУРГ
2020 г.

Содержание

1. Список иллюстраций	3
2. Список таблиц	3
3. Постановка задачи	4
4. Теория.....	4
5. Реализация.....	4
6. Результаты	5
7. Выводы	10
8. Список литературы	10
9. Приложения	10

1 Список иллюстраций

1	Boxplot нормальное распределение	5
2	Boxplot стандартное распределение Лапласа	6
3	Boxplot стандартное распределение Коши	7
4	Boxplot распределение Пуассона	8
5	Boxplot равномерное распределение	9

2 Список таблиц

1	Зависимость выбросов от размера выборки	9
---	---	---

3 Постановка задачи

Для, приведённых ниже, пяти распределений сгенерировать выборки объёмом 20, 100, для каждой выборки построить боксплот Тьюки. Для каждого распределения определить процент выбросов экспериментально. Сгенерировать выборку, соответствующую распределению 1000 раз и, вычислив средний процент, сравнить его с результатами, полученными теоретически.

Распределения:

- Стандартное нормальное распределение
- Стандартное распределение Коши
- Распределение Лапласа
- Равномерное распределение
- Распределение Пуассона

4 Теория

Боксплот Тьюки - график, использующийся в описательной статистике, изображающий одномерное распределение вероятностей.

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

1. Выборочная медиана [1]:

$$\text{med } x = \begin{cases} x_{k+1}, & n = 2k + 1 \\ \frac{1}{2} (x_k + x_{k+1}), & n = 2k \end{cases} \quad (1)$$

2. Квартиль [2]:

$$z_{[p]} = \begin{cases} x_{np}, & np \in \mathbb{Z} \\ x_{[np]+1}, & np \notin \mathbb{Z} \end{cases} \quad (2)$$

Выбросом в статистике называют результат измерения, выделяющийся из общей выборки.

5 Реализация

Для генерации выборки был использован *Python 3.6*: модуль *random* библиотеки *numpy* [4].

Боксплот Тьюки был построен средствами библиотеки *matplotlib* [5].

Правая и левые границы: $R = LQ + l(UQ - LQ)$, $L = UQ - k(UQ - LQ)$, где k обычно полагают равным 1.5 [3]

Число выбросов определялось таким образом: если значение из выборки находится вне установленных левой и правых границ, то оно является выбросом.

6 Результаты

Рис. 1: Voxplot нормальное распределение

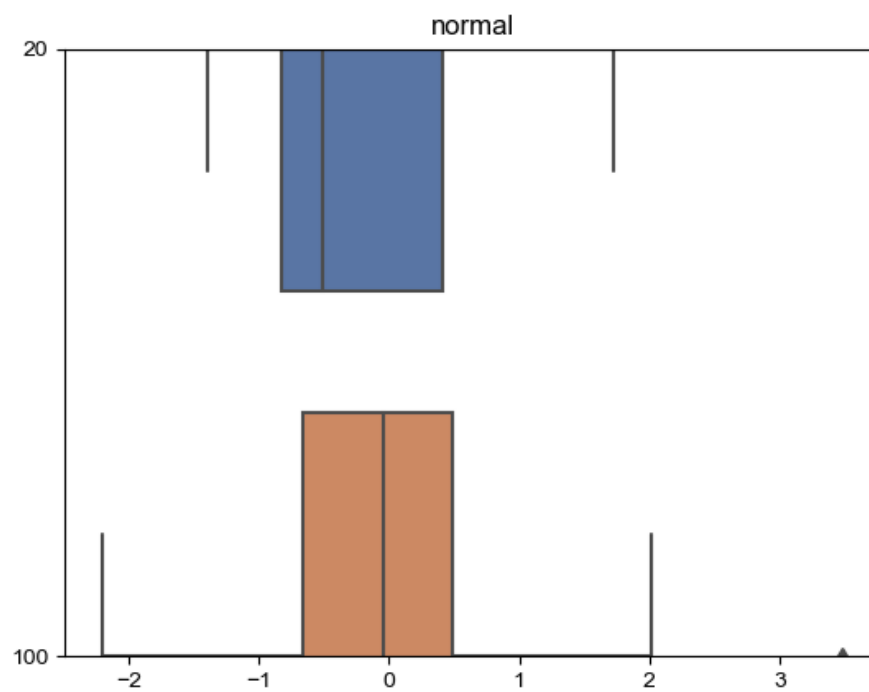


Рис. 2: Boxplot стандартное распределение Лапласа

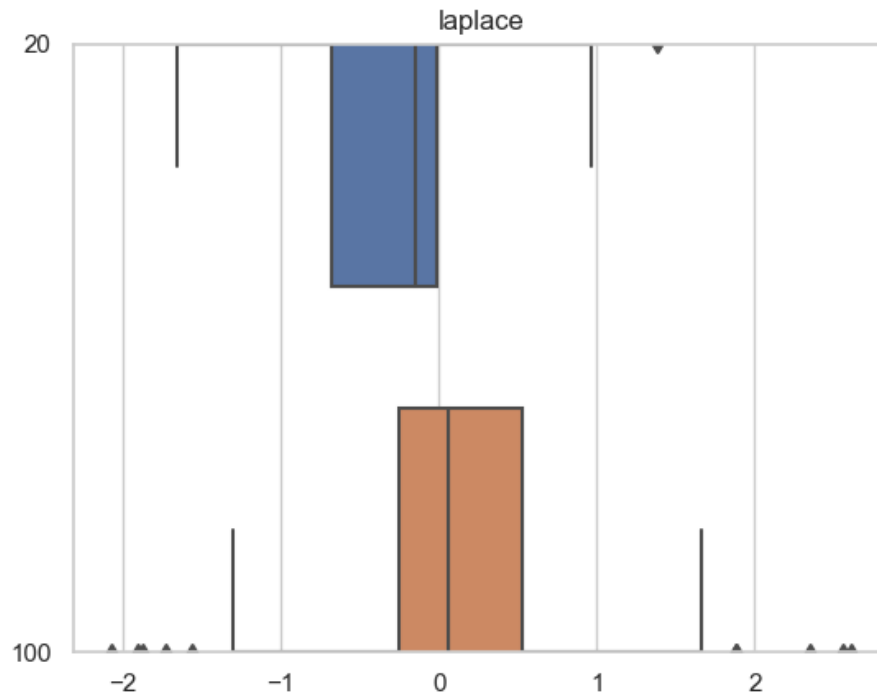


Рис. 3: Boxplot стандартное распределение Коши

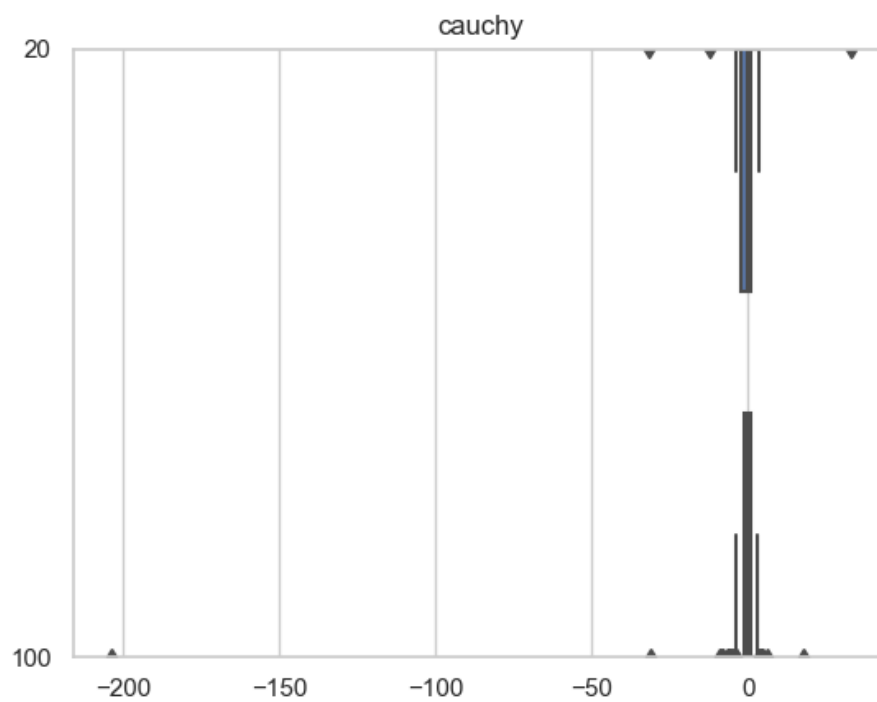


Рис. 4: Voxplot распределение Пуассона

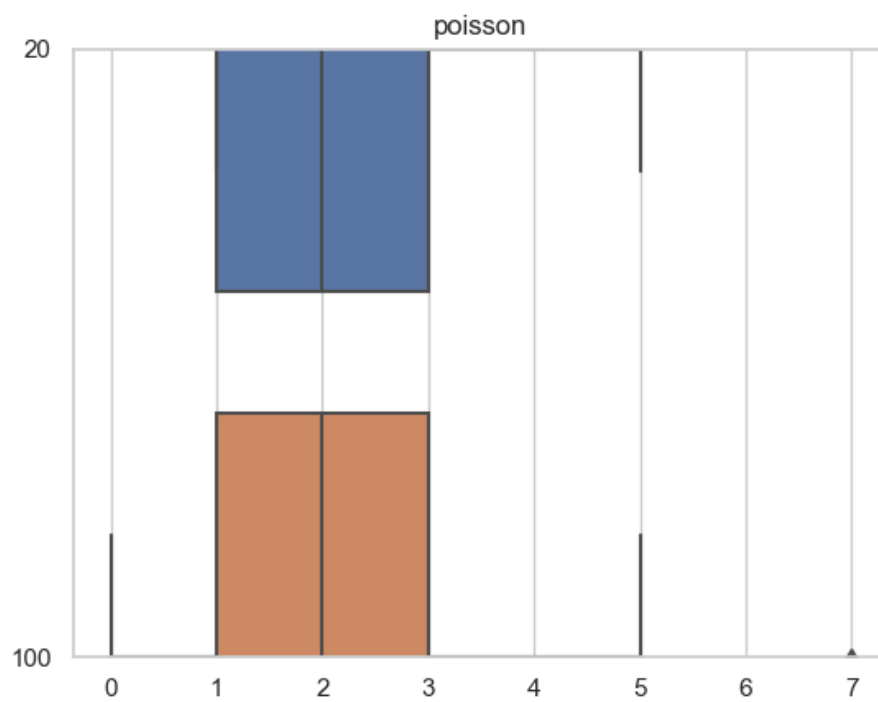


Рис. 5: Voxplot равномерное распределение

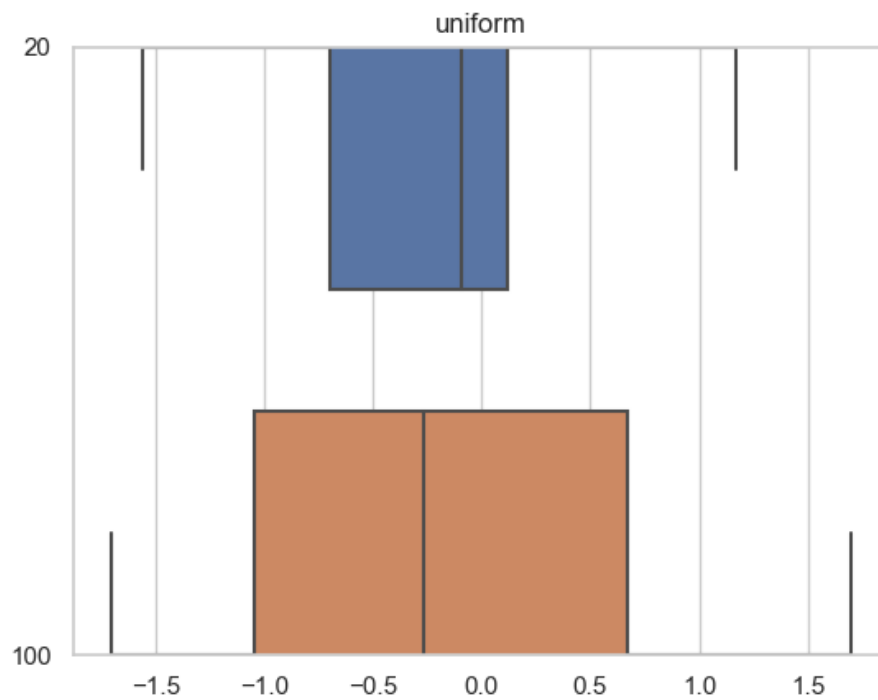


Таблица 1: Зависимость выбросов от размера выборки

Выборка	Доля выбросов
normal	
n = 20	0.022250
n = 100	0.009710
cauchy	
n = 20	0.147700
n = 100	0.156480
laplace	
n = 20	0.072300
n = 100	0.066380
uniform	
n = 20	0.002200
n = 100	0.000000
poisson	
n = 20	0.034300
n = 100	0.009650

7 Выводы

Экспериментально полученные проценты выбросов, близки к теоретическим. Можно вывести соотношение между процентами выбросов:

$$uniform < normal < poisson < laplace < cauchy \quad (3)$$

По полученным данным видно, что наименьший процент выбросов у равномерного распределения, а наибольший процент выбросов у распределения Коши.

8 Список литературы

- [1] Выборочная медиана - http://femto.com.ua/articles/part_1/2194.html
- [2] Квартили - <https://studfiles.net/preview/2438125/page:13/>
- [3] Боксплот - https://en.wikipedia.org/wiki/Box_plot
- [4] Модуль numpy - <https://physics.susu.ru/vorontsov/language/numpy.html>
- [5] Модуль matplotlib - <https://matplotlib.org/users/index.html>

9 Приложения

Код отчёта: <https://github.com/MisterProper9000/MatStatLabs/blob/master/MatStatLab3/MatStatLab3.tex>

Код лабораторной: <https://github.com/MisterProper9000/MatStatLabs/blob/master/MatStatLab3/MatStatLab3.py>