

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ им. ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ

КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ

ЛАБОРАТОРНАЯ РАБОТА №3 БОКСПЛОТ ТЬЮКИ

Студент группы 3630102/70301

Камянский Д.В.

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	11
8. Список литературы	11
9. Приложения	11

1 Список иллюстраций

1	Boxplot распределений для размера выборки $N=20$	5
2	Boxplot стандартное нормальное распределение	6
3	Boxplot стандартное распределение Лапласа	7
4	Boxplot стандартное распределение Коши	8
5	Boxplot распределение Пуассона	9
6	Boxplot равномерное распределение	10

2 Список таблиц

1	Теоретические оценки доли выбросов	10
2	Экспериментальные оценки доли выбросов	11

3 Постановка задачи

Для, приведённых ниже, пяти распределений сгенерировать выборки объёмом 20, 100, для каждой выборки построить боксплот Тьюки. Для каждого распределения экспериментально определить процент выбросов, сгенерировав выборку, соответствующую распределению 1000 раз и, вычислив среднюю долю выбросов. Сравнить полученные оценки с результатами, полученными теоретически.

Распределения:

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (1)$$

$$C(x, 0, 1) = \frac{1}{\pi(1+x^2)} \quad (2)$$

$$L\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \quad (3)$$

$$P(5, k) = \frac{5^k}{k!} e^{-5} \quad (4)$$

$$M(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & |x| \leq \sqrt{3} \\ 0 & |x| > \sqrt{3} \end{cases} \quad (5)$$

4 Теория

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

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.

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

Правая и левая границы: $X_1 = Q_1 - 1.5(Q_3 - Q_1)$, $X_2 = Q_3 + 1.5(Q_3 - Q_1)$

Теоретическая вероятность выбросов:

Для непрерывных распределений: $P_e^t = F(X_1^t) + (1 - F(X_2^t))$

Для дискретных распределений: $P_e^t = (F(X_1^t) - P(x = X_1^t)) + (1 - F(X_2^t))$

5 Реализация

Для генерации выборки был использован *Python 3.8.2*: модуль *stats* библиотеки *scipy* для генерации выборок различных распределений.

Боксплот Тьюки был построен средствами библиотеки *seaborn*.

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 распределений для размера выборки $N=20$

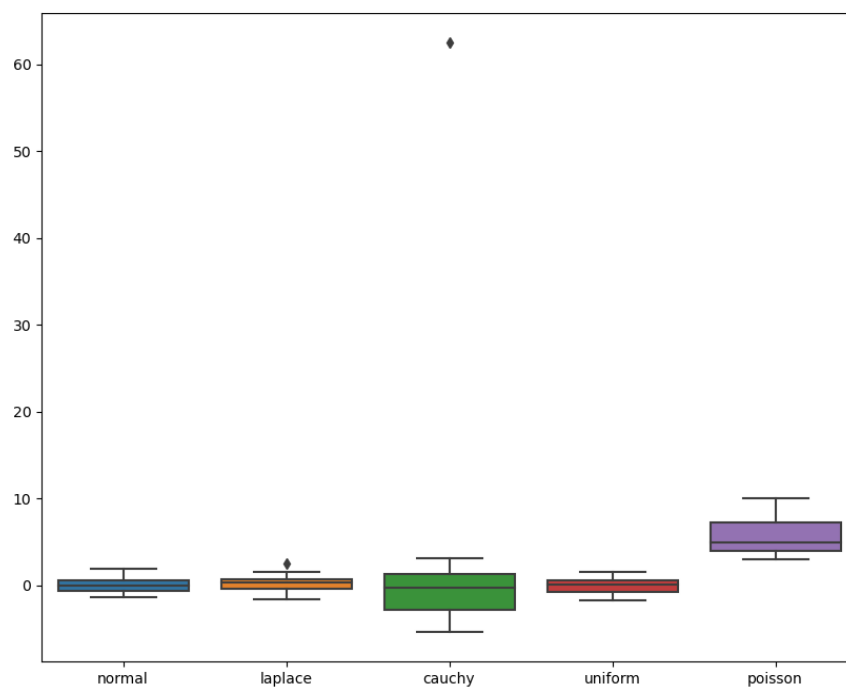


Рис. 2: Voxplot стандартное нормальное распределение

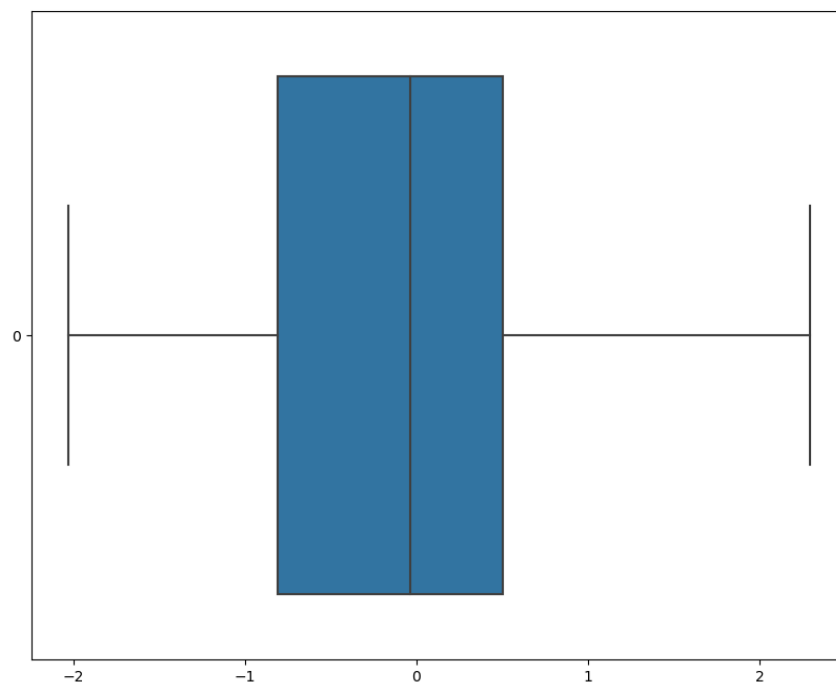


Рис. 3: Boxplot стандартное распределение Лапласа

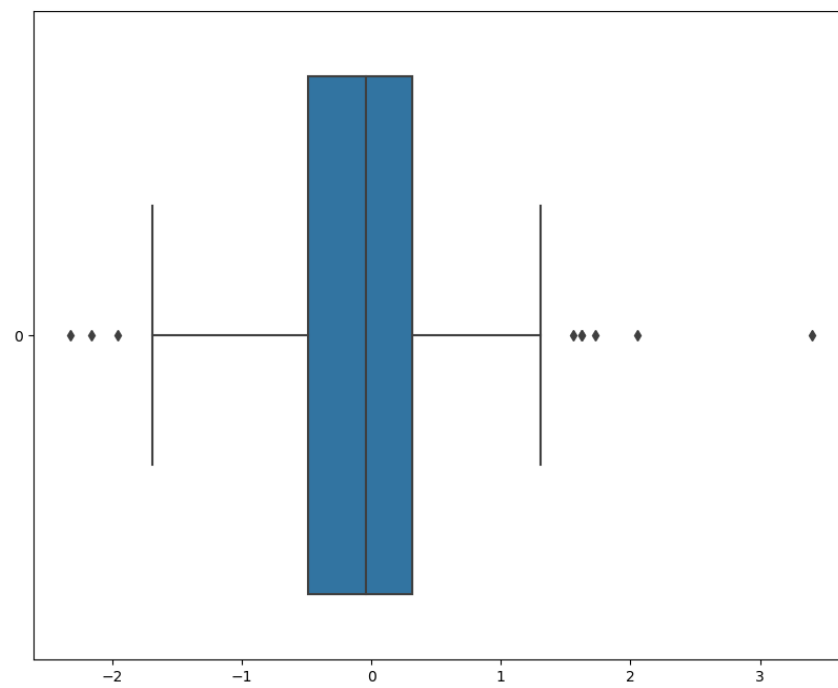


Рис. 4: Boxplot стандартное распределение Коши

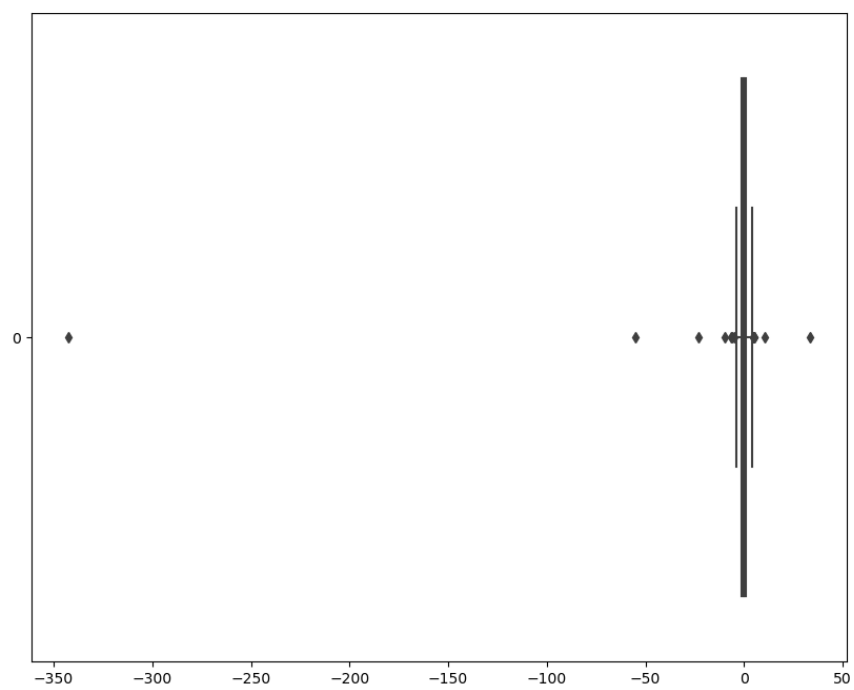


Рис. 5: Voxplot распределение Пуассона

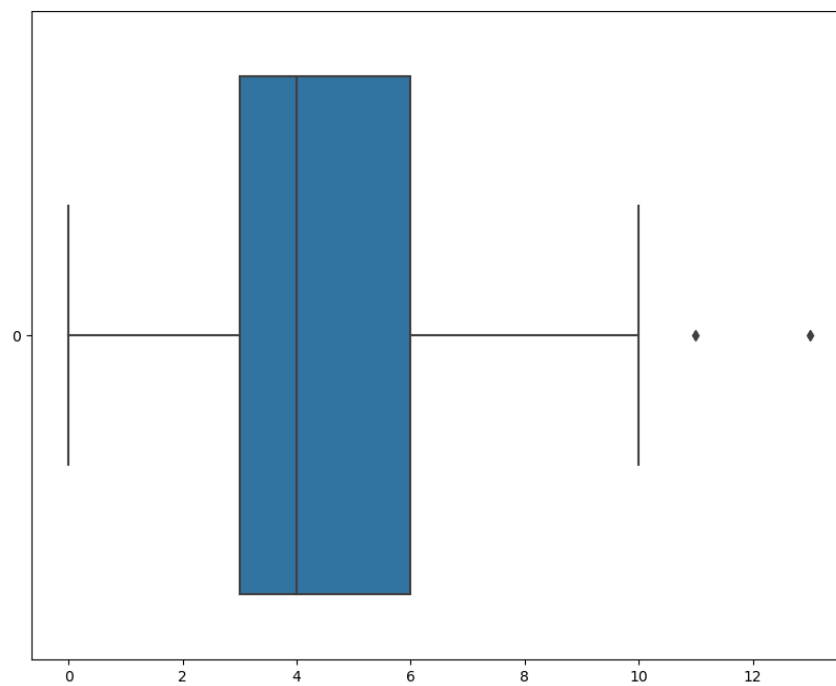


Рис. 6: Вохplot равномерное распределение

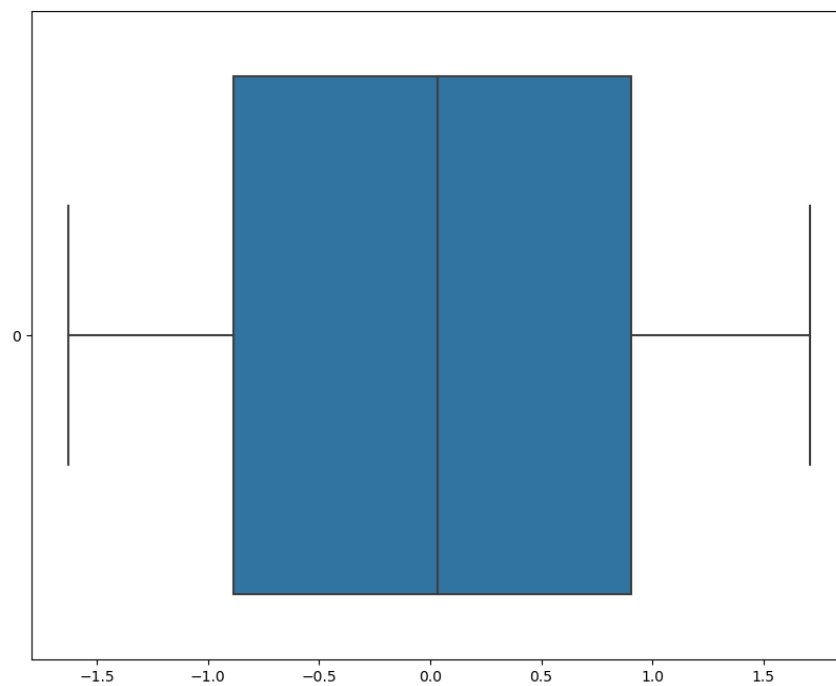


Таблица 1: Теоретические оценки доли выбросов

Распределение	Доля выбросов
normal	0.006977
cauchy	0.155958
laplace	0.0625
uniform	0.0
poisson	0.013695

Таблица 2: Экспериментальные оценки доли выбросов

Распределение	Средняя доля выбросов	Дисперсия доли выбросов
normal		
n = 20	0.02	0.00152
n = 100	0.01	0.000161
cauchy		
n = 20	0.15	0.004827
n = 100	0.155	0.001067
laplace		
n = 20	0.07	0.004469
n = 100	0.065	0.000995
uniform		
n = 20	0.002	0.000212
n = 100	0.0	0.0
poisson		
n = 20	0.02	0.002284
n = 100	0.014	0.000285

7 Выводы

Экспериментально полученные оценки доли выбросов стремятся к теоретическим с ростом размера выборки, Значение для средней доли выбросов было ограничено первым значащим разрядом в значении дисперсии. Можно вывести соотношение между процентами выбросов:

$$uniform < normal < poisson < laplace < cauchy \quad (6)$$

По полученным данным видно, что наименьший процент выбросов у равномерного распределения, а наибольший процент выбросов у распределения Коши, при чем значения этих выбросов могут отклоняться от выборочного среднего на порядки.

8 Список литературы

1. [Боксплот](#)
2. [Модуль scipy.stats](#)
3. [Модуль seaborn](#)

9 Приложения

Код лаборатрной: <https://github.com/dkamianskii/MatStatLabs/tree/master/Lab3>