Санкт-Петербургский Политехнический Университет им. Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Отчёт по лабораторной работе №3 по дисциплине “Математическая статистика”

Боксплот Тьюки

Выполнил студент:

Мишутин Д. В.

Группа:

3630102/70301

Проверил:

К.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2020 г.

Оглавление

Список иллюстраций и таблиц

1 Постановка задачи

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.

Распределения:

- Стандартное нормальное распределение:

$$N(x,0,1)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \tag{1.1}$$

- Стандартное распределение Коши:

$$C(x,0,1)=\frac{1}{\pi(1+x^2)} \tag{1.2}$$

- Распределение Лапласа:

$$L\left(x,0,\frac{1}{\sqrt{2}}\right)=\frac{1}{\sqrt{2}}e^{-\sqrt{2}|x|} \tag{1.3}$$

- Распределение Пуассона:

$$P(k,10)=\frac{10^k}{k!}e^{-10} \tag{1.4}$$

- Равномерное распределение:

$$U(x,-\sqrt{3},\sqrt{3})=\begin{cases}\frac{1}{2\sqrt{3}}, при\,|x|\le\sqrt{3}\\ 0, при\,|x|>\sqrt{3}\end{cases} \tag{1.5}$$

2 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.

Выброс – результат, сильно выделяющийся из общей выборки.

Характеристики положения:

- Выборочное среднее:

$$avrg=\frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2.1)$$

- Выборочная медиана:

$$med\,x=\begin{cases} x_{k+1}, n=2k+1 \\ \frac{1}{2}\left(x_k+x_{k+1}\right), n=2k \end{cases} \qquad (2.2)$$

- Квартиль:

$$Z_{[p]}=\begin{cases} x_{[np]}, np\in Z \\ x_{[np]+1}, np\notin Z \end{cases} \qquad (2.3)$$

3 Реализация

Был использован язык *Python 3.8.2*: модуль *numpy* для генерации выборок с различными распределениями и математических расчётов, модуль *matplotlib* для построения и отображения боксплотов, модуль *pandas* для оптимального хранения статистических данных и функция *display* из модуля *IPython.display* для их корректного отображения в таблицах.

4 Результаты

Рис. 1 Стандартное нормальное распределение

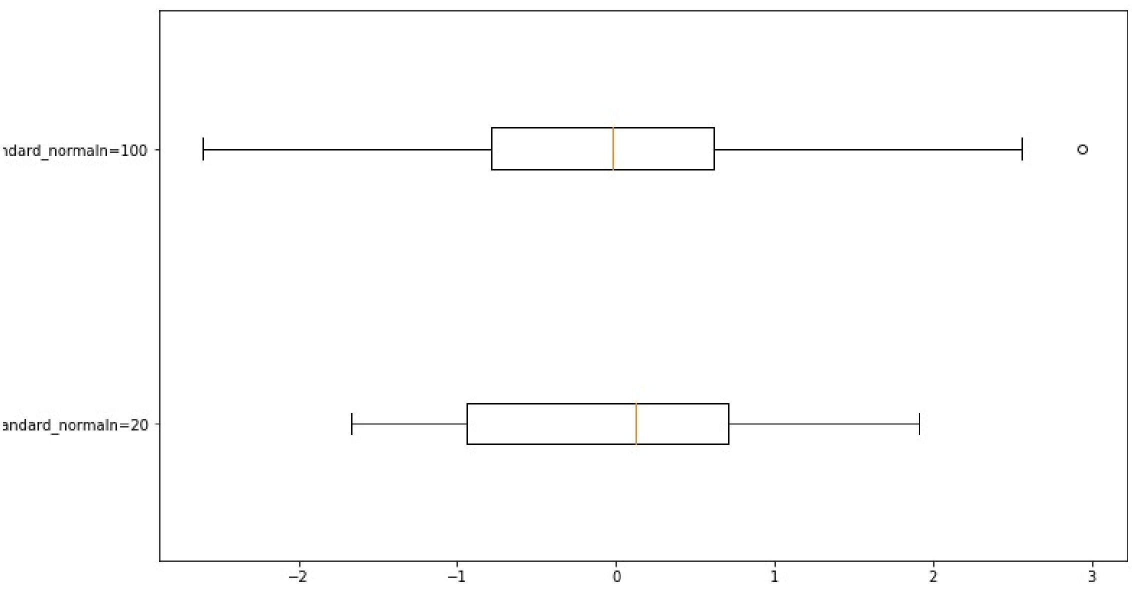

Рис. 2 Стандартное распределение Коши

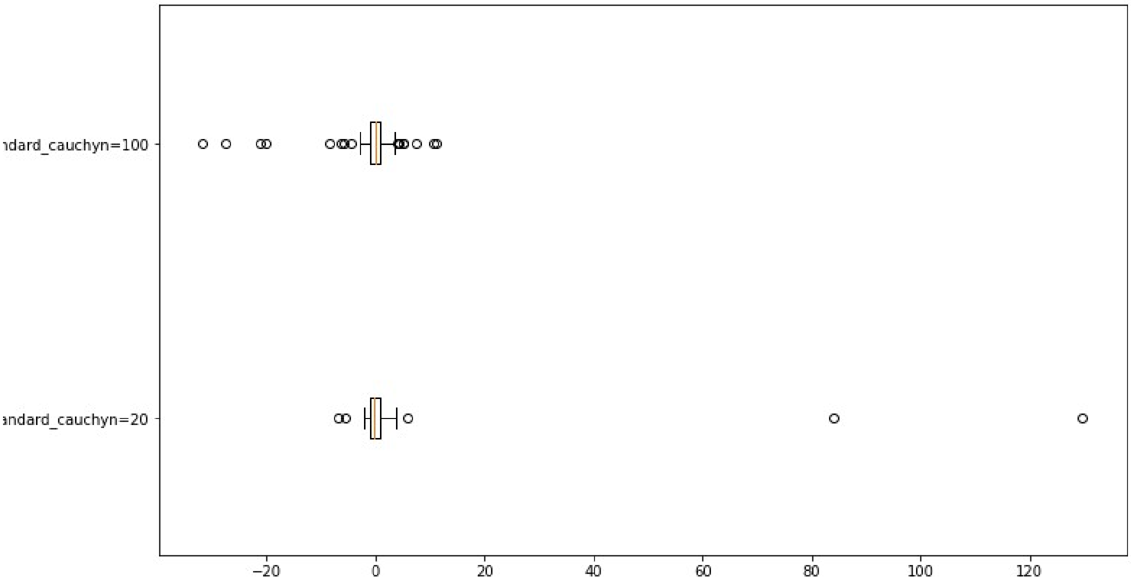

Рис. 3 Распределение Лапласа

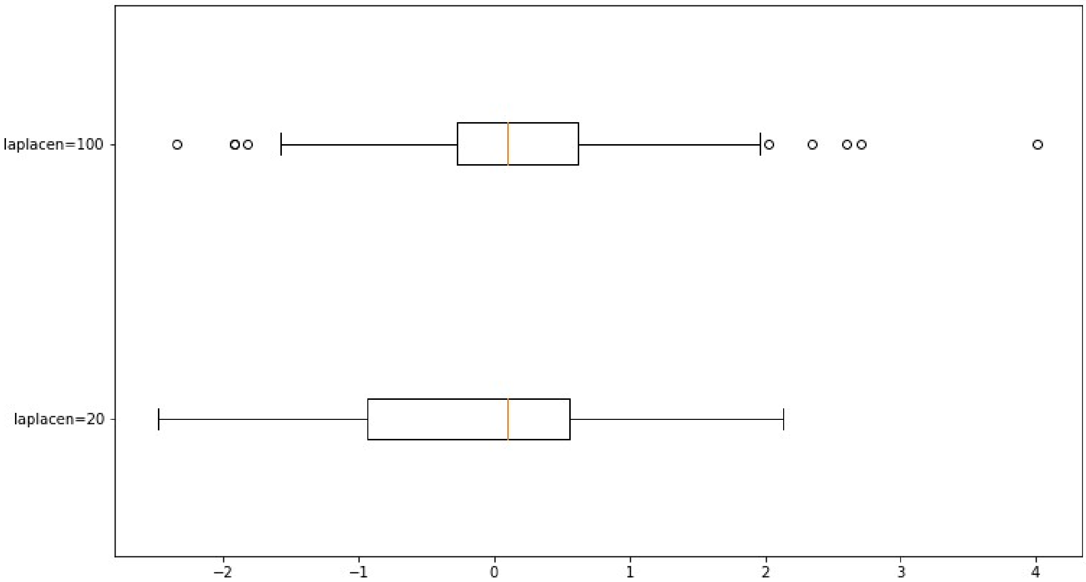

Рис. 4 Распределение Пуассона

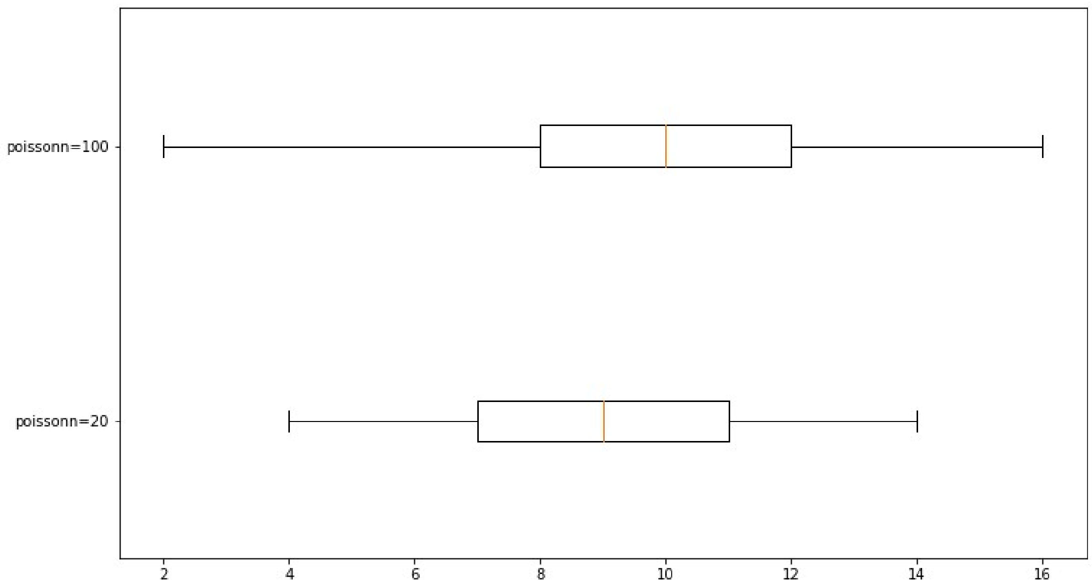

Рис. 5 Равномерное распределение

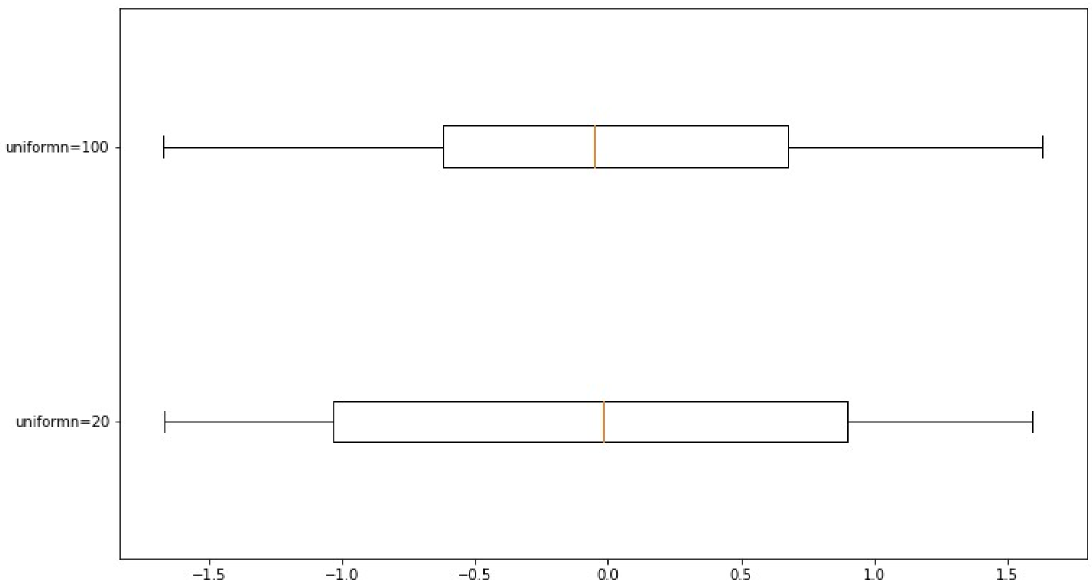

Таблица 1 Средние проценты выбросов

Выборка	Средний процент выбросов
Стандартное нормальное распределение (1.1)	
N = 20	2.67
N = 100	1.02
Стандартное распределение Коши (1.2)	
N = 20	15.19
N = 100	16.38
Распределение Лапласа (1.3)	
N = 20	8.76
N = 100	6.26
Распределение Пуассона (1.4)	
N = 20	1.62
N = 100	0.81
Равномерное распределение (1.5)	
N = 20	0.05
N = 100	0

5 Выводы

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

$$(1.5) < (1.4) < (1.1) < (1.3) < (1.2)$$

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

6 Литература

Основы работы с *numpy* (отдельная глава курса)

Pandas обзор

7 Приложения

Код лабораторной