Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра прикладной математики

ОТЧЕТ

Тема: ***Исследование распределений вероятностей***

Направление: 01.03.02 Прикладная математика и информатика

Выполнил студент гр. 33631/4 Камалетдинова Ю.

Преподаватель Баженов А.

Санкт-Петербург
2019

Содержание

Постановка задачи

В данной лабораторной работе исследуется такой способ представления набора статистических данных как гистограмма. Требуется сгенерировать выборки разного размера для заданных распределений, построить гистограммы и сделать выводы о взаимосвязи функции плотности распределения и функции гистограммы.

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

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

$$L(x,0,\frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}}e^{-\sqrt{2}|x|} \text{ – Лаплас} \tag{3}$$

$$U(x,-\sqrt{3},\sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| \leq \sqrt{3} \\ 0, & |x| > \sqrt{3} \end{cases} \text{ – равномерное} \tag{4}$$

$$P(\lambda) = \frac{e^{-\lambda}}{k!}\lambda^k \text{ – Пуассон} \tag{5}$$

Описание метода

Для построения гистограммы разобьем множество возможных значений ξ выбоки на непересекающиеся интервалы $\Delta_j = [t_j, t_{j+1}),\ j = \overline{0,m};\ t_0 = -\infty,\ t_{m+1} = +\infty$. Количество элементов X выборки из объема n, попавших в интервал Δ_j, обозначим через v_j, то есть $v_j = \sum_{i=1}^{n} I(X_i \in \Delta_j)$. По теореме Бернулли при $n \to \infty$

$$v_j/n \to P(\xi \in \Delta_j) = \int_{\Delta_j} f(u)du = |\Delta_j| f(c),\ c \in \Delta_j \tag{6}$$

Кусочно-постоянная функция, приведенная ниже, есть нормализованная гистограмма

$$f_n(t) = \frac{v_j}{n|\Delta_j|},\ t \in \Delta_j,\ j = \overline{1, m} \tag{7}$$

Итак, гистограмма является статистическим аналогом функции плотности распределения и функции распределения. Формулы (6), (7) указаны в пособии [1].

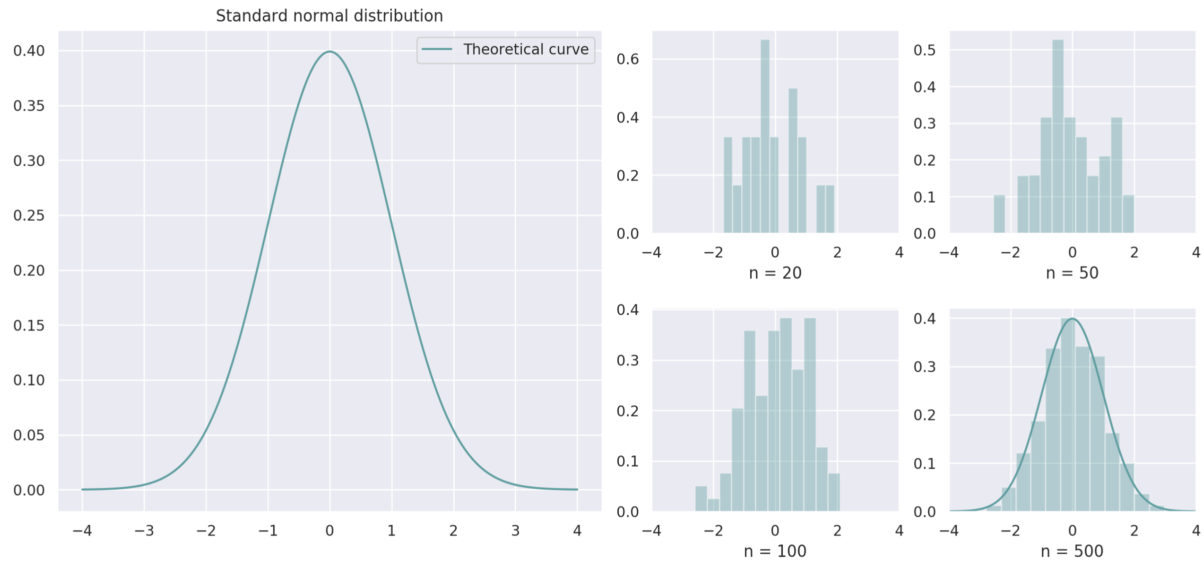
Реализация

Для выполнения поставленной задачи будем пользоваться библиотеками для языка Python: *numpy, scipy* – расчеты, законы распределения вероятностей; *matplotlib, seaborn* – визуализация результатов. Ход работы:

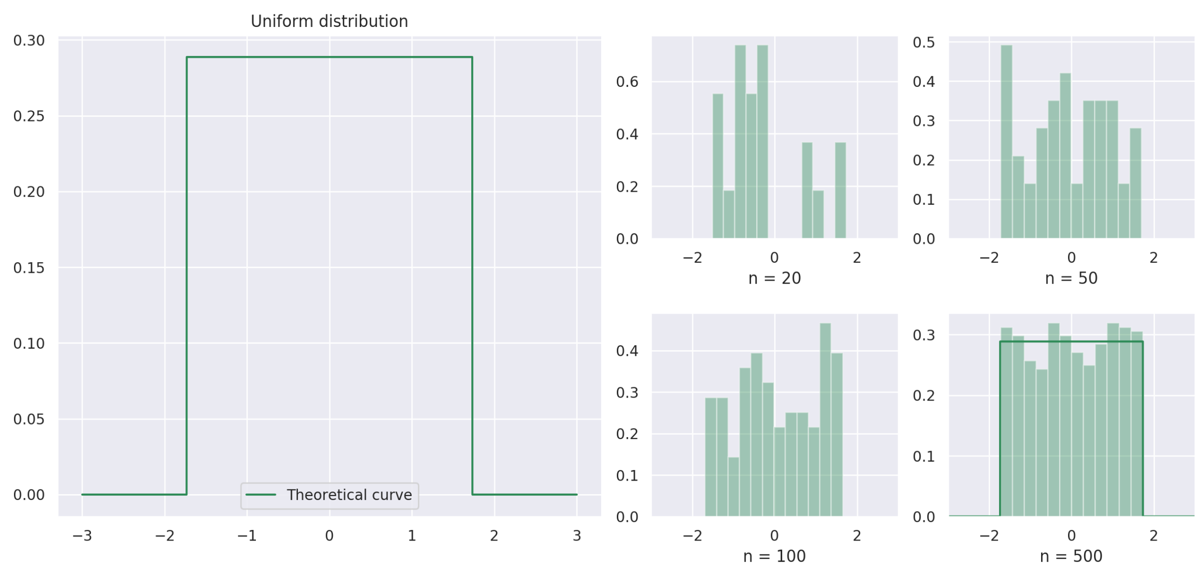
- Производим построение графика теоретической функции распределения
- При помощи метода **.rvs** берется выборка размера n из распределения с заданными параметрами
- Определяем границы области интересующих нас значений и разделяем на равные интервалы (*bins*)
- Строим гистограммы, представляющие собой столбцы высотой, пропорциональной числу попавших в их ширину значений
- Нормируем гистограммы и отрисовываем теоретическую функцию распределения в одних осях

Результат

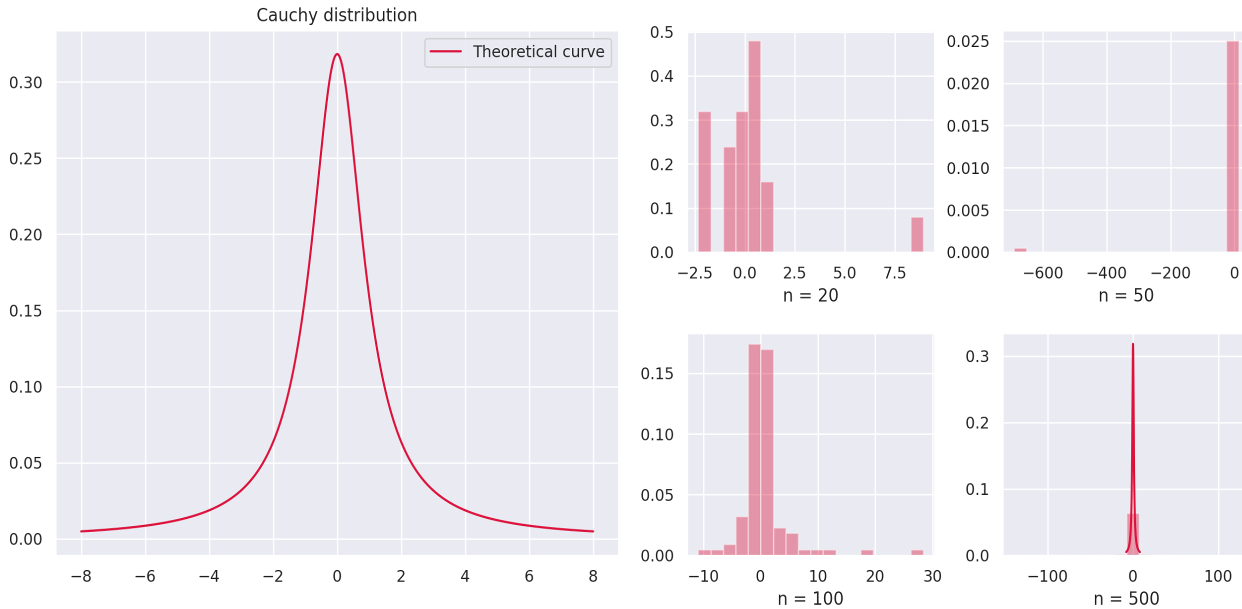
Нормальное распределение с параметрами 0, 1



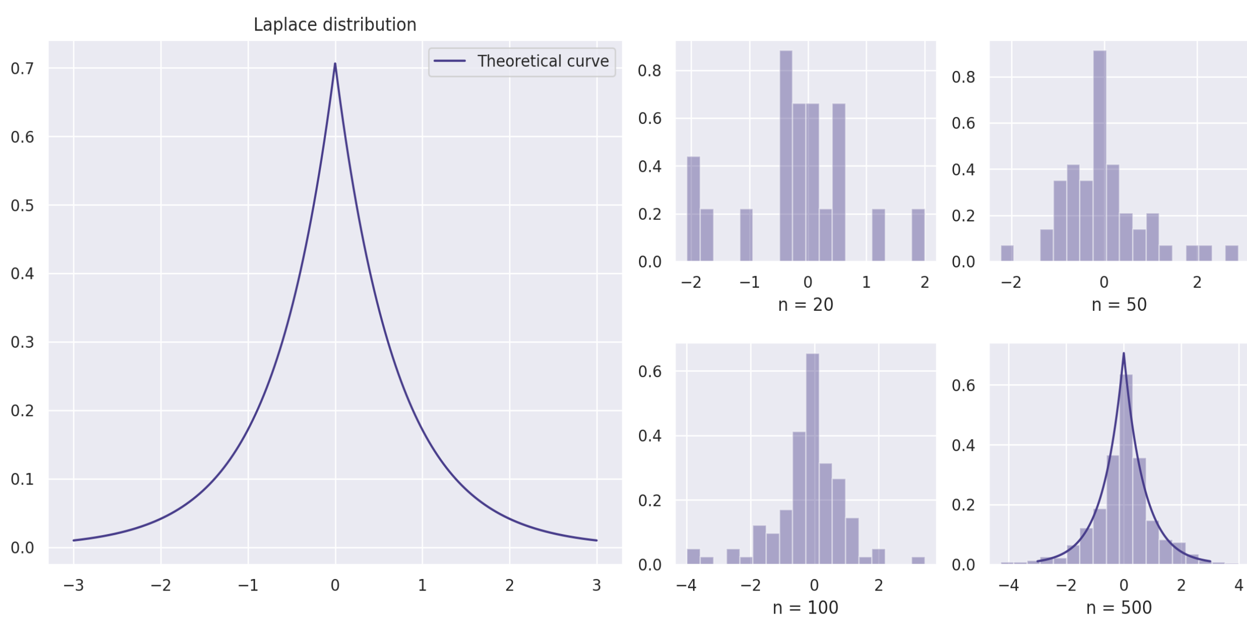
Равномерное распределение на отрезке $[-\sqrt{3}, \sqrt{3}]$



Распределение Коши с параметрами 0, 1

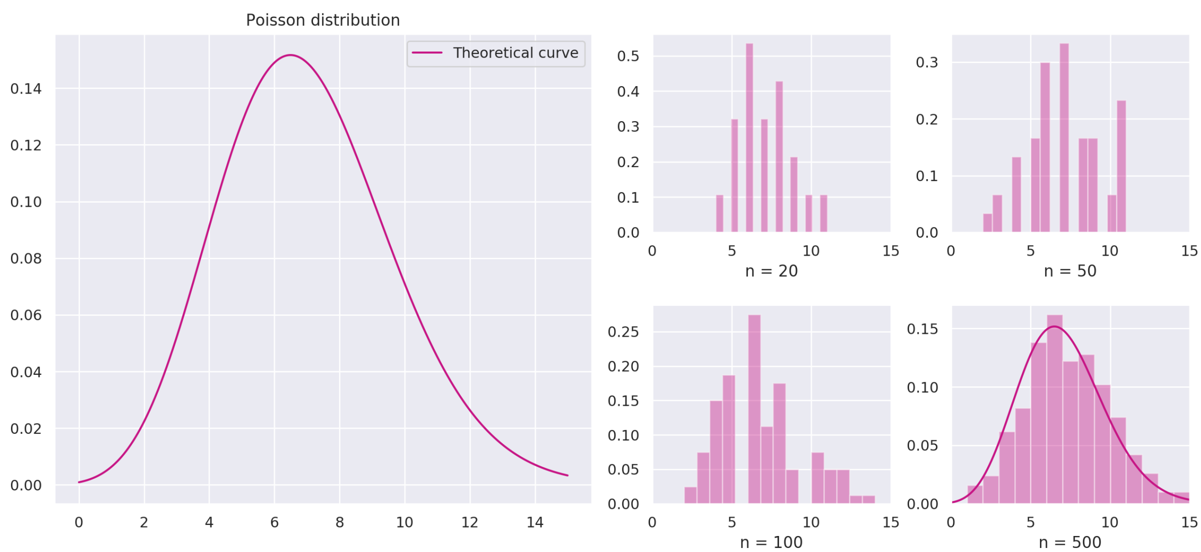


Распределение Лапласа с параметрами 0, $\frac{1}{\sqrt{2}}$



Распределение Пуассона с параметром $\lambda = 7$

Выбор параметра обоснован стремлением распределения к нормальному при увеличении λ. Для наглядной визуализации подходит число 7, к тому же являющееся числом Миллера, или предельной порцией информации, обрабатываемой человеком за раз (оригинальная статья - [2]). Этот факт является довольно символичным, учитывая область применения распределения Пуассона.



При достаточной мощности выборки из распределений (n = 100) гистограмму можно рассматривать как аналог плотности распределения непрерывной случайной величины.

Список литературы

[1] *Кадырова Н. О.* Теория вероятностей и математическая статистика. Статистический анализ данных: учеб. пособие / *Н. О. Кадырова, Л. В. Павлова, И. Е. Ануфриев.* - СПб.: Изд-во Политехн. ун-та, 2010. -54с.

[2] *Miller, G. A.* The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review. 1956; 63:81–97.