Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчет по лабораторной работе № 4
по дисциплине: Математическая статика.

Выполнила студентка:
Заболотских Екатерина Дмитриевна
группа: 3630102/70301

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

Санкт-Петербург
2020 г.

Оглавление

Постановка задачи 1
Теория 3
1. Эмпирическая функция распределения 3
2. Ядерная оценка плотности распределения 3
Реализация 4
Результаты 5
1. Эмпирические функции распределения 5
2. Ядерные оценки плотности распределения 8
Обсуждение 11
Список литературы 11
Ссылка на github 11

Список иллюстраций

Рисунок 1: Нормальное распределение 5
Рисунок 2: Распределение Коши 5
Рисунок 3: Распределение Лапласа 6
Рисунок 4: Распределение Пуассона 6
Рисунок 5: Равномерное распределение 7
Рисунок 6: Нормальное распределение 8
Рисунок 7: Распределение Коши 8
Рисунок 8: Распределение Лапласа 9
Рисунок 9: Распределение Пуассона 9
Рисунок 10: Равномерное распределение 10

Постановка задачи

Для каждого из 5 распределений:

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

Сгенерировать выборки размеров: 20, 60, 100. Построить на них эмпирические функции распределения и ядерные оценки плотности распределения на отрезке [-4, 4] для непрерывных распределений и на отрезке [6, 14] для распределения Пуассона.

Теория

1. Эмпирическая функция распределения

Эмпирической функцией распределения, построенной на выборке $(x_1, \ldots, x_n)$ объема n, называется случайная величина

$$F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} I(x_i < y) \tag{1}$$

где I – индикатор события $x_i < y$.

2. Ядерная оценка плотности распределения

Если имеется выборка, полученная по распределению с некоторой плотностью f, то ядерной оценкой плотности этой функции называется [2]:

$$\hat{f}_h = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{2}$$

где K – ядро (неотрицательная функция), $h > 0$ – сглаживающий параметр (ширина полосы).

Чаще всего используется нормальное (гауссово) ядро, в силу его удобных математических свойств:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{3}$$

И если используется гауссово ядро, и оцениваемая плотность является гауссовой, оптимальный выбор для h определяется правилом Сильвермана [2]:

$$h_n = \left(\frac{4s_n^5}{3n}\right)^{\frac{1}{5}} \approx 1.06 s_n n^{-\frac{1}{5}} \tag{4}$$

где s_n – выборочное среднеквадратичное отклонение (корень из выборочной дисперсии).

Реализация

Код программы, реализующий данную задачу, был написан на языке Python в интегрированной среде разработке PyCharm.

Были использованы библиотеки:

- **Numpy** – библиотека для работы с данными.
- **Matplotlib** – вывод графиков.

Результаты

1. Эмпирические функции распределения

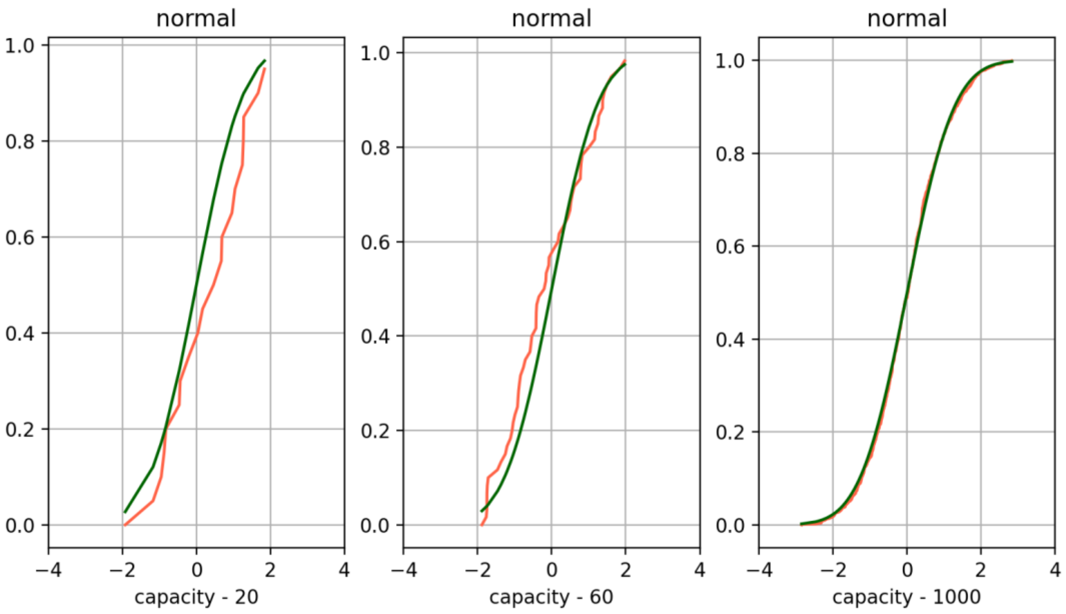

Рисунок 1: Нормальное распределение

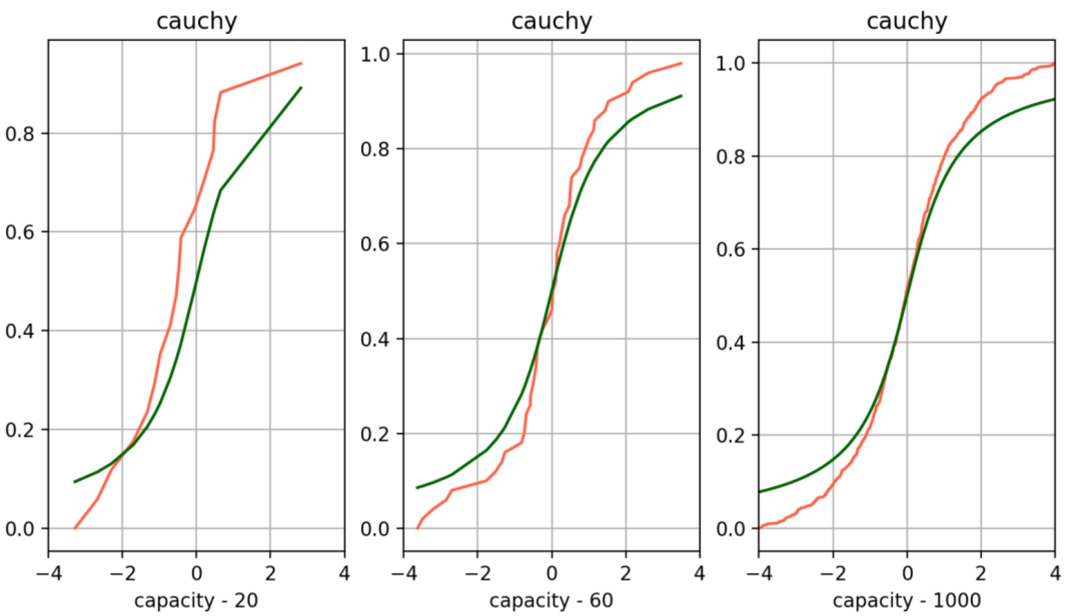

Рисунок 2: Распределение Коши

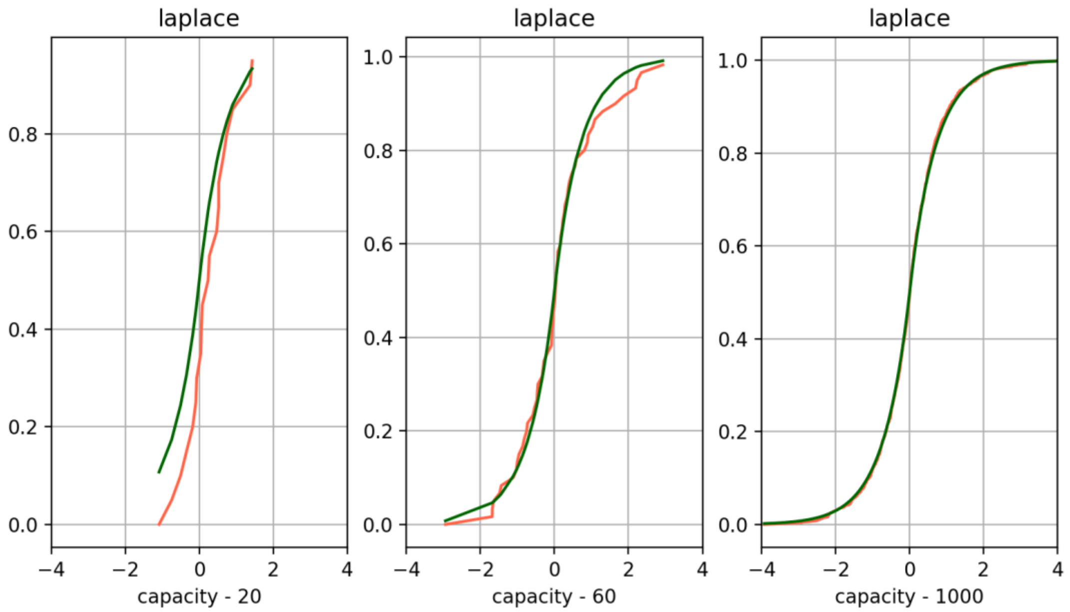

Рисунок 3: Распределение Лапласа

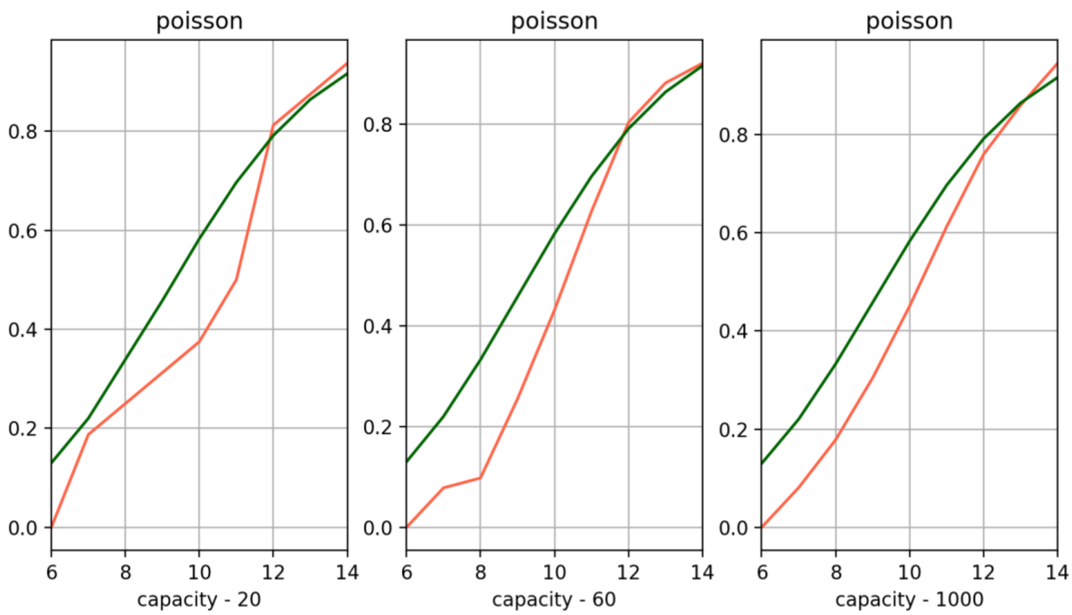

Рисунок 4: Распределение Пуассона

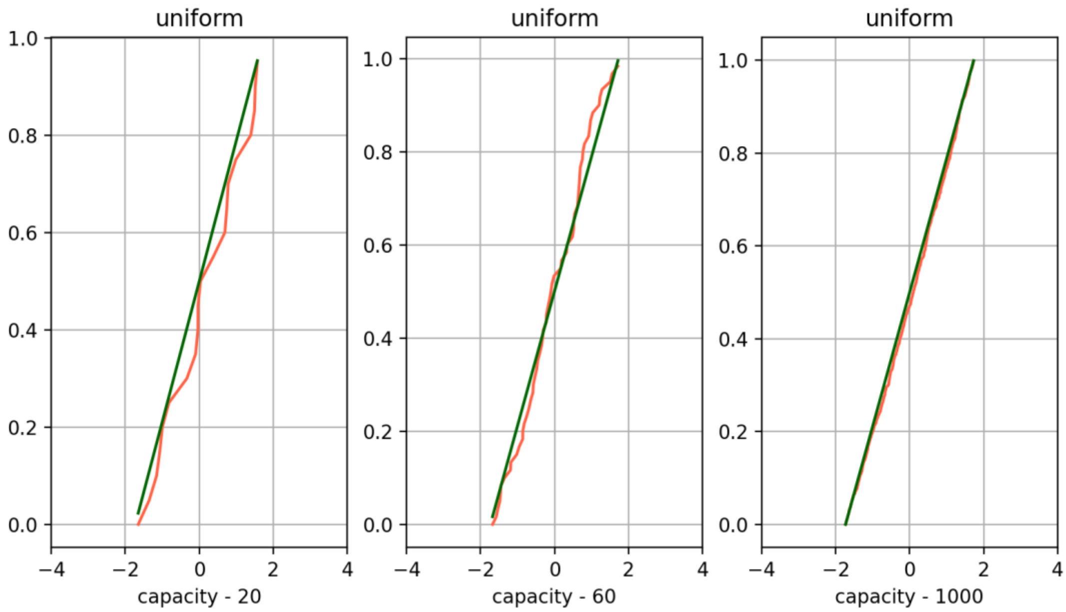

Рисунок 5: Равномерное распределение

2. Ядерные оценки плотности распределения

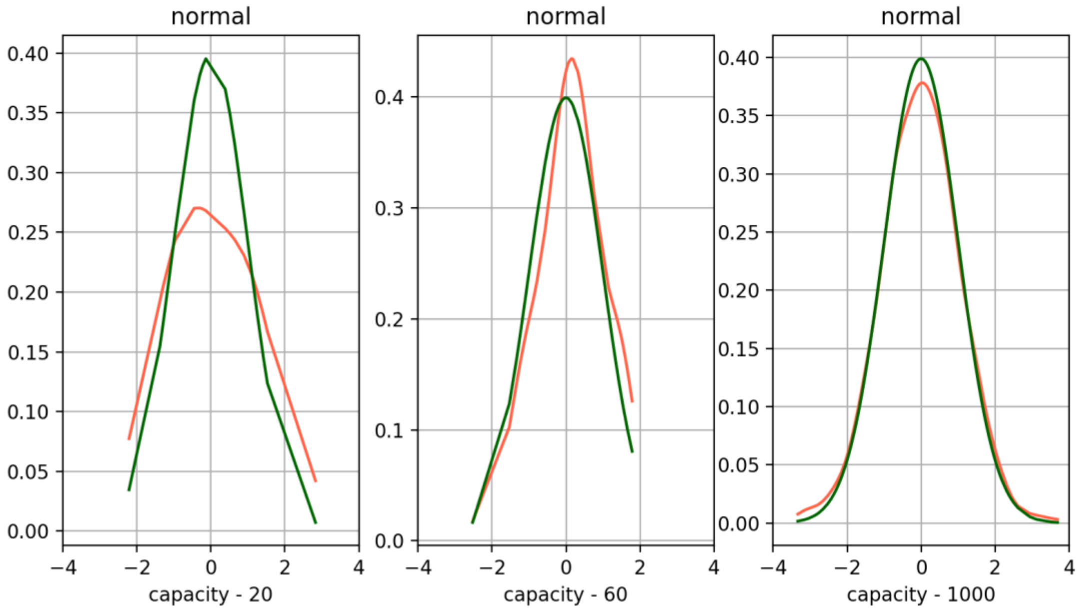

Рисунок 6: Нормальное распределение

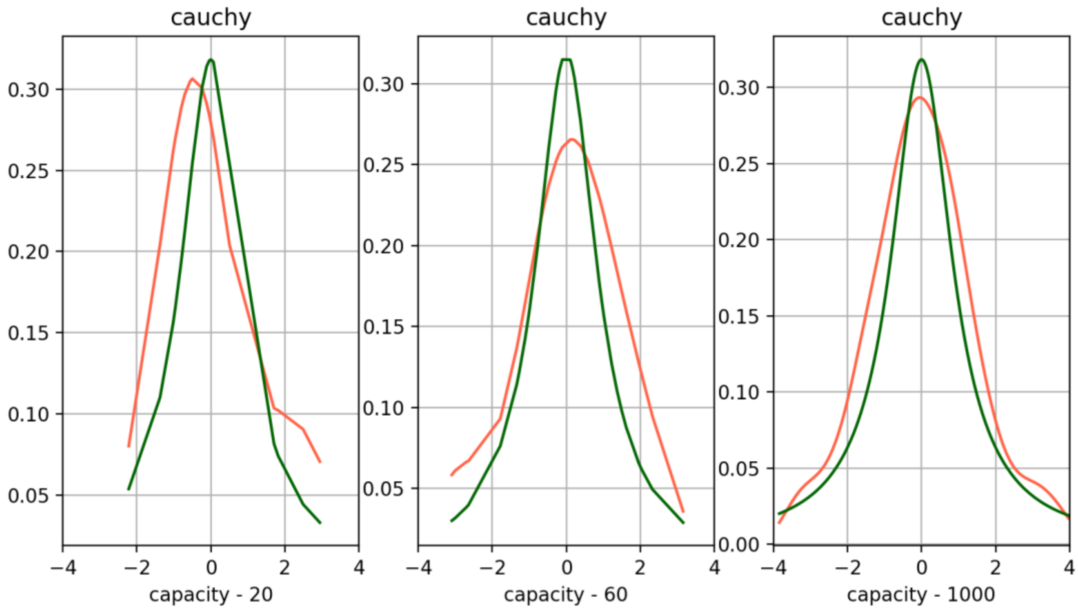

Рисунок 7: Распределение Коши

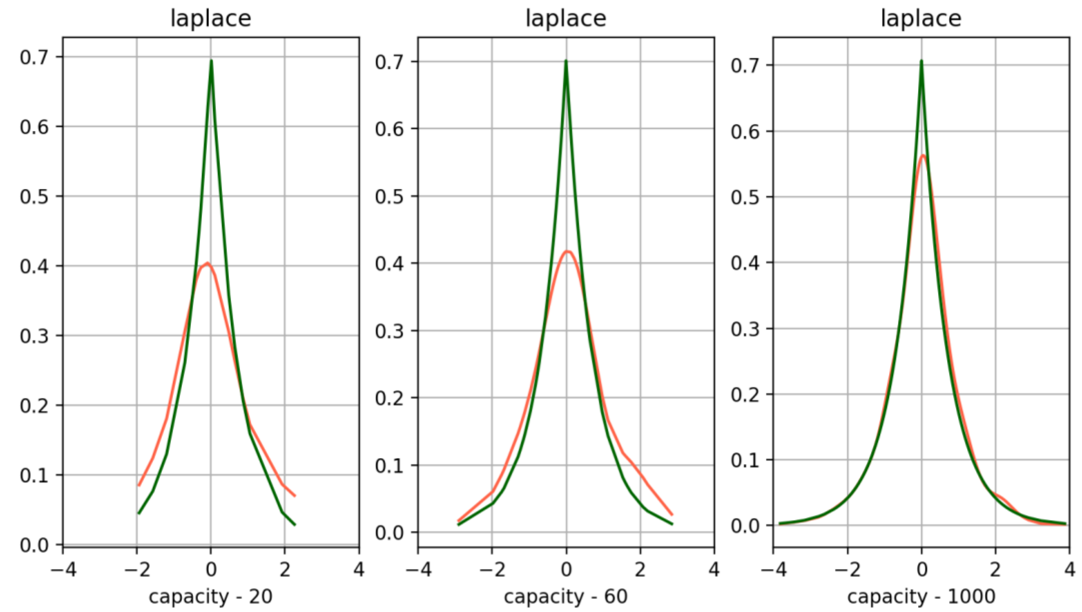

Рисунок 8: Распределение Лапласа

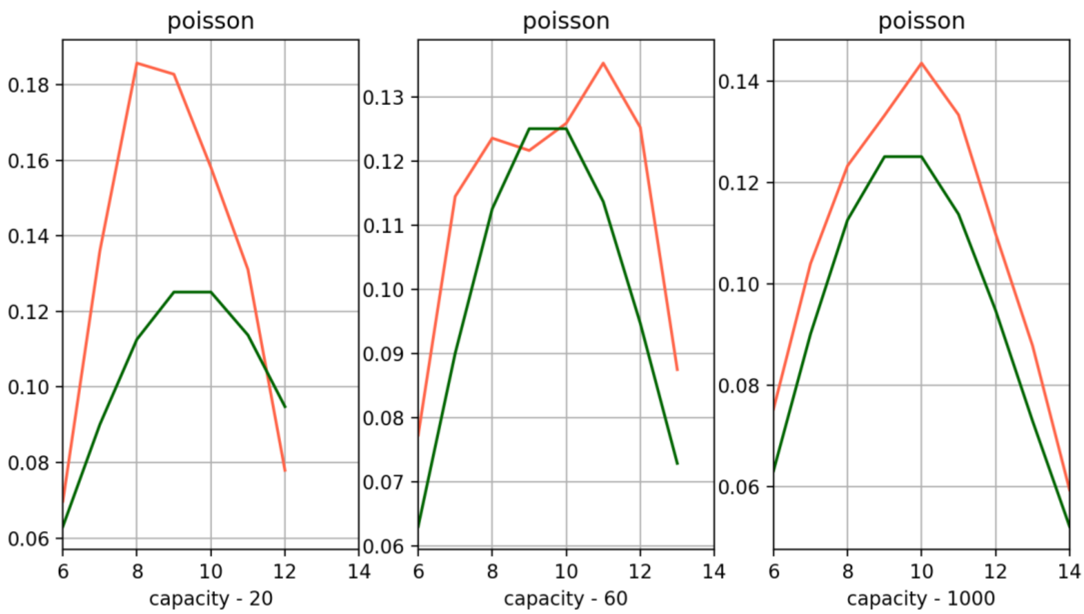

Рисунок 9: Распределение Пуассона

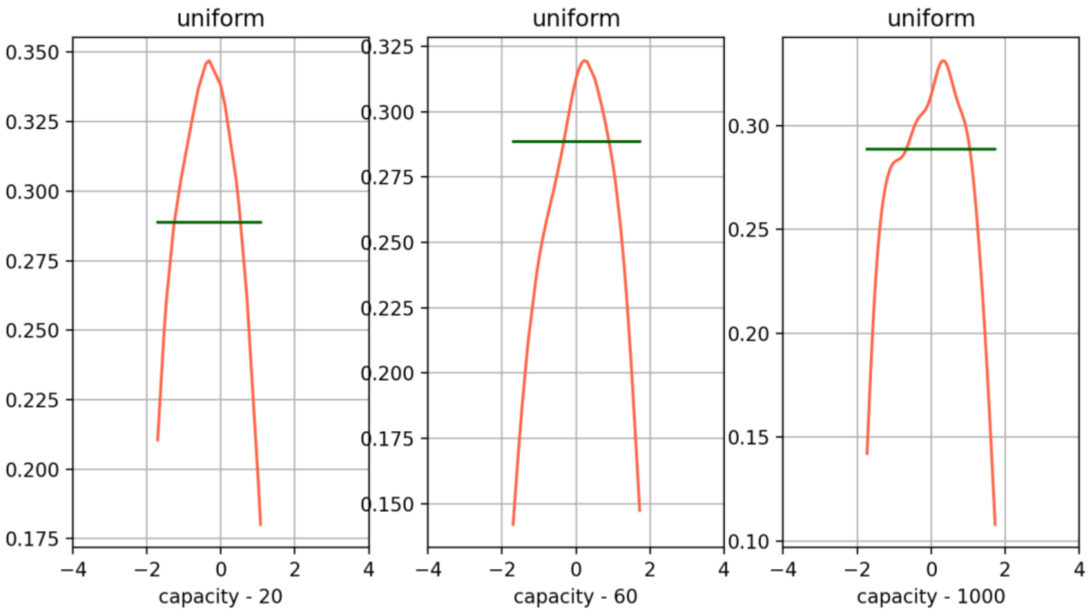

Рисунок 10: Равномерное распределение

Обсуждение

Ожидаемо, увеличение размера выборки улучшает приближение к теоретическим значениям. Так же видно, что равномерное распределение, в силу своей разрывности плохо приближается эмпирически.

Список литературы

1. Конспекты лекции
2. Википедия: https://ru.wikipedia.org/wiki

Ссылка на github: https://github.com/KateZabolotskih/MathStatLabs