Санкт-Петербургский Политехнический университет

Петра Великого

Институт Прикладной Математики и Механики

Кафедра «Прикладная Математика и Информатика»

Отчет

По лабораторной работе № 4

По Дисциплине «Математическая статистика»

Выполнил:

Студент Селянкин Федор

Группа 3630102/70301

Проверил:

к.ф. – м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

Список Иллюстраций:

Постановка задачи
Для 5 распределений:
- Нормальное распределение $N(x, 0, 1)$
- Распределение Коши $C(x, 0, 1)$
- Распределение Лапласа $L(x, 0, \frac{1}{\sqrt{2}})$
- Распределение Пуассона $P(k, 10)$
- Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$

Сгенерировать выборки размером 20, 60 и 100 элементов.
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
Распределения
- Нормальное распределение

1

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

- Распределение Коши

2

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

- Распределение Лапласа

3

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

- Распределение Пуассона

4

$$\boldsymbol{P(k, 10) = \frac{10^k}{k!} e^{-10}} \tag{4}$$

- Равномерное распределение

5

$$\mathbf{U}(\mathbf{x}, -\sqrt{\mathbf{3}}, \sqrt{\mathbf{3}}) = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{2}\sqrt{\mathbf{3}}} \text{ при } |\boldsymbol{x}| \le \sqrt{\mathbf{3}} \\ \mathbf{0} \text{ при } |\boldsymbol{x}| > \sqrt{\mathbf{3}} \end{cases} \tag{5}$$

Эмпирическая функция распределения
Статистический ряд
Статистическим рядом называется последовательность различных элементов выборки z_1, z_2, …, z_k, расположенных в возрастающем порядке с указанием частот n_1, n_1, …,n_k, с которыми эти элементы содержаться в выборке. Статический ряд обычно записывается в виде таблицы

z	z_1	z_2	…	z_k
n	n_1	n_2	…	n_k

Таблица 1 Статистический ряд

Определение
Эмпирической (выборочной) функцией распределения (э. ф. р.) называется относительная частота события X < x, полученная по данной выборке:

6

$$\mathrm{F}_\mathrm{n}^*(\mathrm{x}) = \mathrm{P}^*(\mathrm{X} < \mathrm{x}) \tag{6}$$

Описание

Для получения относительной частоты $P^*(X < x)$ просуммируем в статистическом ряде, построенном по данной выборке, все частоты n_i, для которых элементы z_i статистического ряда меньше x. Тогда $P^*(X < x) = \frac{1}{n}\sum_{z_i<x} n_i$. Получаем

7

$$F^*(x) = \frac{1}{n}\sum_{z_i<x} n_i \tag{7}$$

Где $F^*(x)$ – функция распределения дискретной случайной величины X^*, заданной таблицей распределения

X^*	z_1	z_2	…	z_k
P	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_k}{n}$

Таблица 2Таблица распределений

Эмпирическая функция распределения является оценкой, т.е. приближённым значением, генеральной функции распределения

8

$$F^*(x) = \frac{1}{n}\sum_{z_i<x} n_i \tag{8}$$

Оценки плотности вероятности

Определение

Оценкой плотности вероятности $f(x)$ называется функция $\hat{f}(x)$, построенная на основе выборки, приближенно равная $f(x)$

9

$$\hat{f}(x) \approx f(x) \tag{9}$$

Ядерные оценки

Представим оценку в виде суммы с числом слагаемых, равным объему выборки:

10

$$\widehat{f_n} = \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{x-x_i}{h_n}) \tag{10}$$

Здесь функция K(u), называемая ядерной (ядром), непрерывна и является плотностью вероятности, $x_1,\ldots, x_n$ – элементы выборки, $\{h_n\}$ – любая последовательность положительных чисел, обладающая свойствами

11

$$h_n \xrightarrow[n\to\infty]{} 0;\ \frac{h_n}{n^{-1}} \xrightarrow[n\to\infty]{} \infty \tag{11}$$

Такие оценки называются непрерывными ядерными.

Замечание. Свойство, означающее сближение оценки с оцениваемой величиной при $n \to \infty$ в каком-либо смысле, называется состоятельностью оценки.

Если плотность $f(x)$ кусочно-непрерывная, то ядерная оценка плотности является состоятельной при соблюдении условий, накладываемых на параметр сглаживания h_n, а также на ядро *K(u)*.

Гауссово (нормальное) ядро

12

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{12}$$

Правило Сильвермана

13

$$\mathrm{h_n} \ = \ 1.06\hat{\sigma}\mathrm{n}^{-\frac{1}{5}} \tag{13}$$

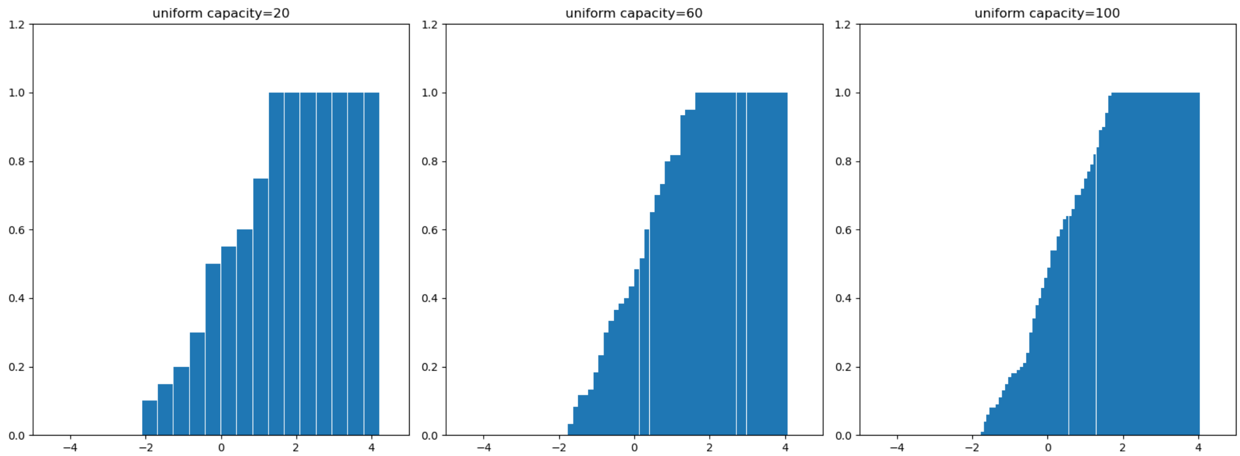
Где – $\hat{\sigma}$ выборочное стандартное отклонение.

Реализация

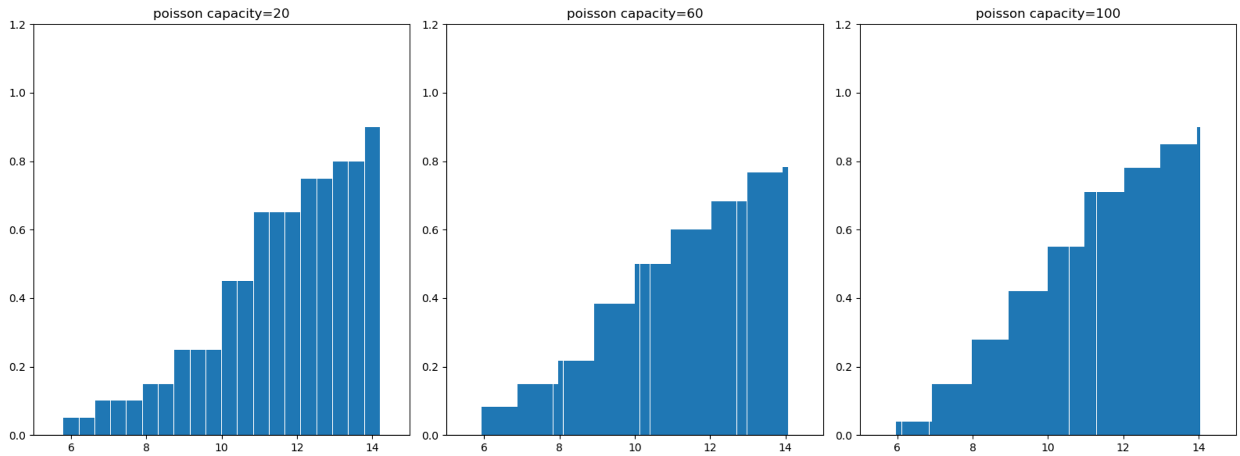
Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm, с использованием дополнительных библиотек для отображения и расчетов. Исходный код лабораторной выложен на веб-сервисе GitHub [2].

Результаты

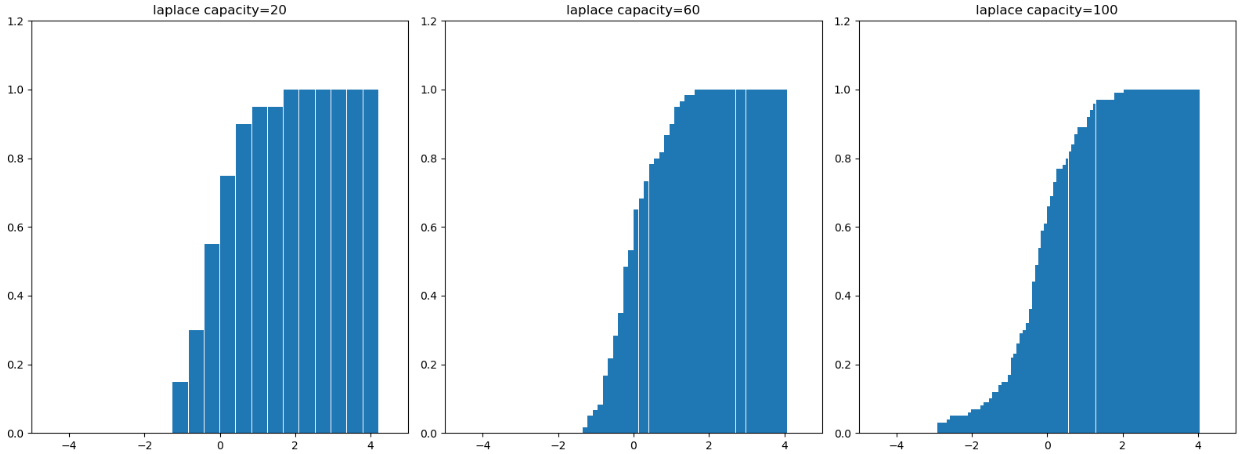
Эмпирическая функция распределения



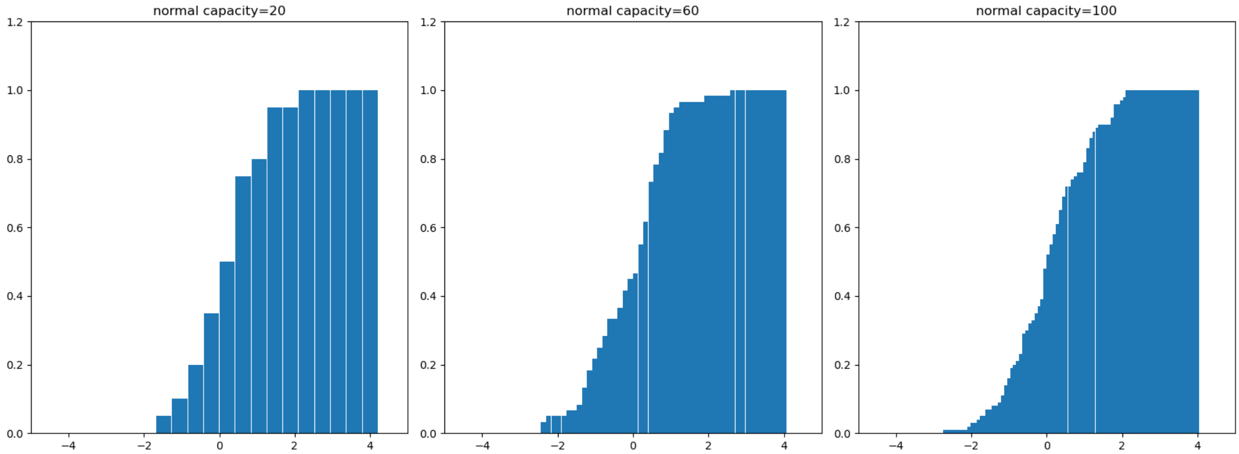
1 ЭФР: Равномерное распределение



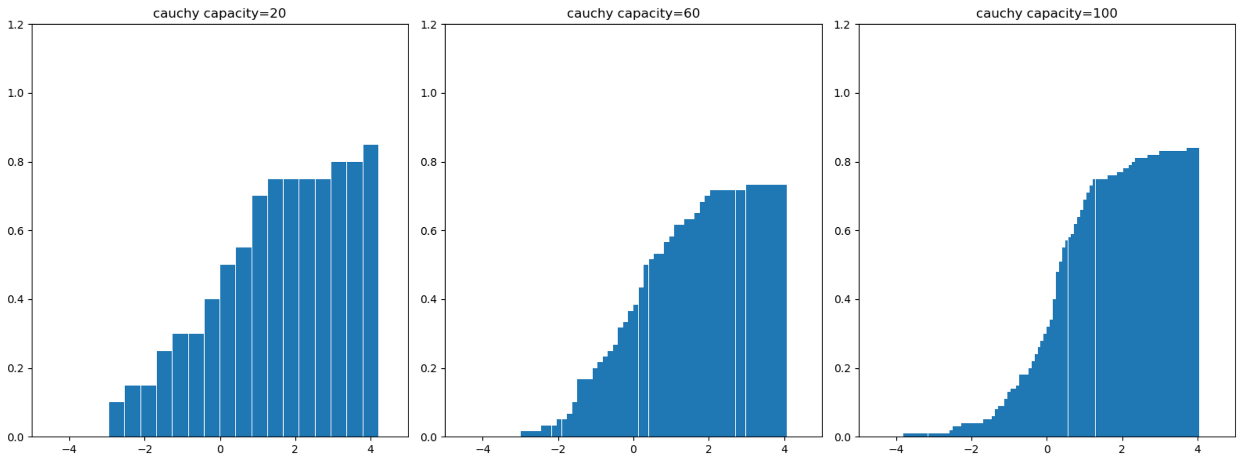
2 ЭФР: Распределение Пуассона



3 ЭФР: Распределение Лапласа

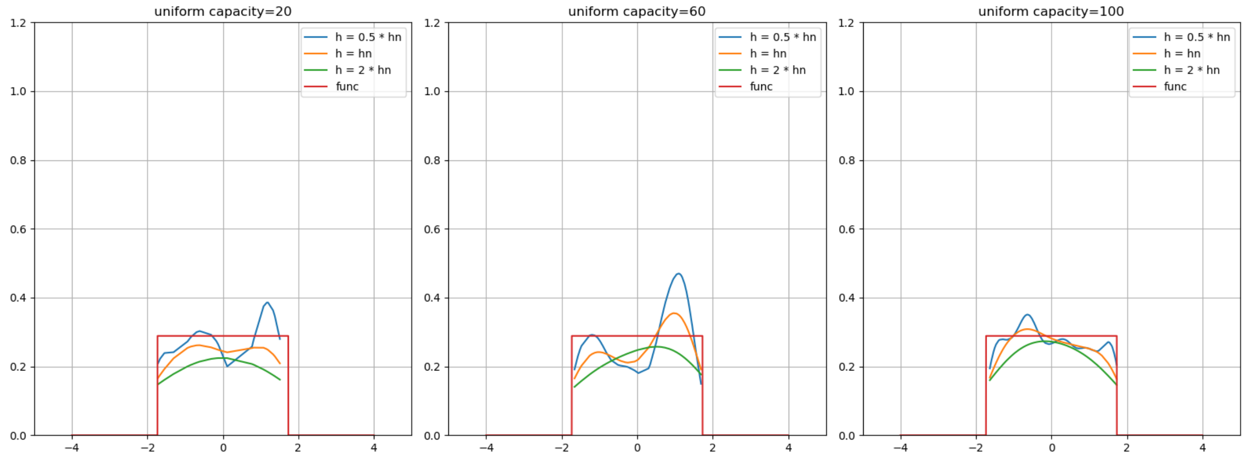


4 ЭФР: Нормальное распределение

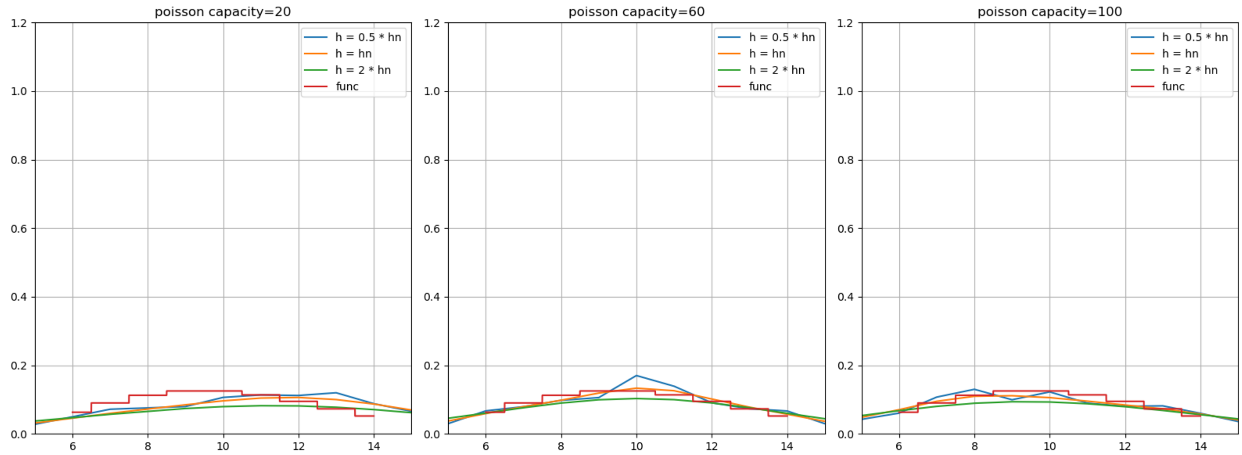


5 ЭФР: Распределение Коши

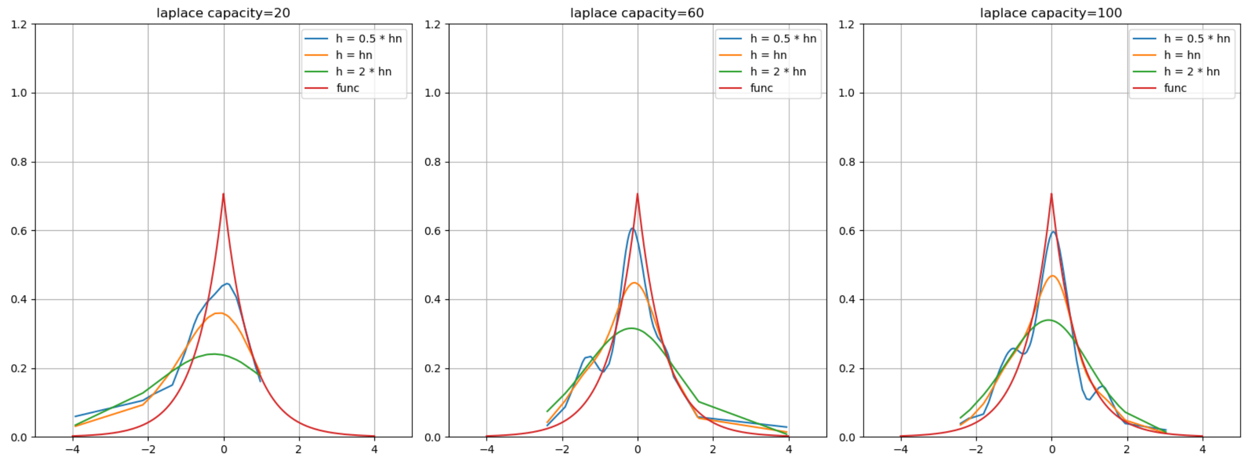
Ядерные плотности распределения



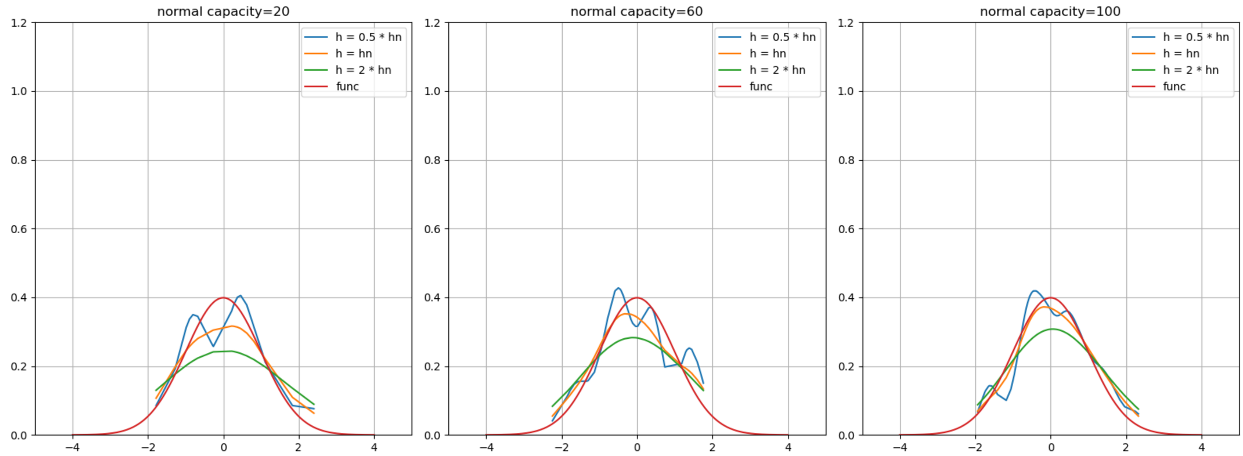
6 ЯПР: Равномерное распределение



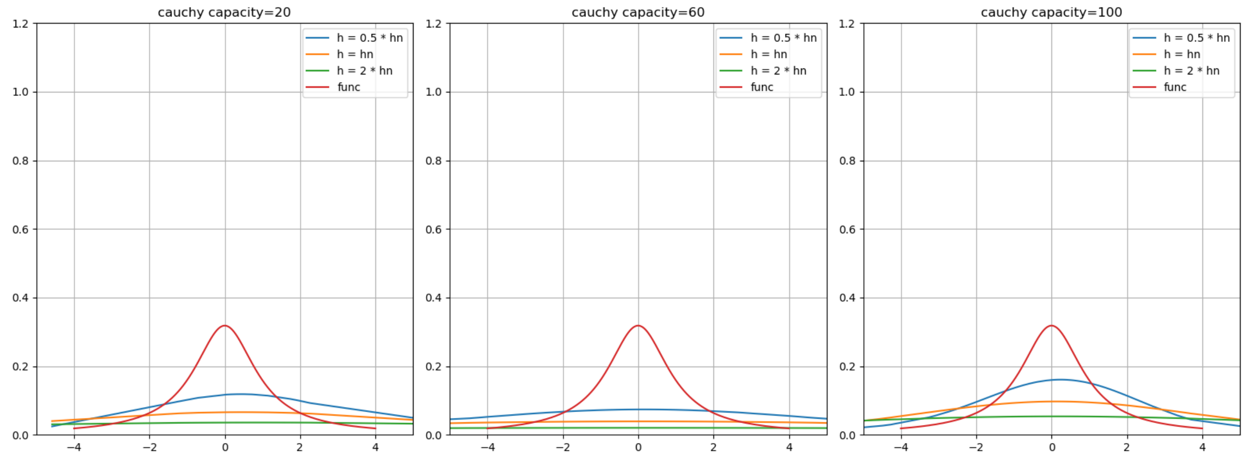
7 ЯПР: Распределение Пуассона



8 ЯПР: Распределение Лапласа



9 ЯПР: Нормальное распределение



10 ЯПР: Распределение Коши

Литература

1. Выборочная функция распределения
   https://en.wikipedia.org/wiki/Empirical_distribution_function
2. Ссылка на репозиторий GitHub
   https://github.com/SelyankinFyodor/math-statistics/tree/master/Lab4

Обсуждения

С увеличением мощности выборки Эмпирическая функция распределения становится более походящей на функцию распределения.

С увеличением мощности выборки график ядерной плотности распределения все больше становится похож на график плотности распределения. Причем на многих иллюстрациях график с h выбранном согласно правилу Сильвермана без удваивания или уполовинивания приближен к графику плотности распределения лучше, чем остальные.