

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ
ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ
КАФЕДРА "ПРИКЛАДНАЯ МАТЕМАТИКА"

ОТЧЁТ
ЛАБОРАТОРНАЯ РАБОТА №4
ПО ДИСЦИПЛИНЕ
"МАТЕМАТИЧЕСКАЯ СТАТИСТИКА"

ВЫПОЛНИЛ СТУДЕНТ:
САЛИХОВ С.Р.
ГРУППА: 3630102/70401

ПРОВЕРИЛ:
К.Ф-М.Н., ДОЦЕНТ
БАЖЕНОВ АЛЕКСАНДР НИКОЛАВИЧ

САНКТ-ПЕТЕРБУРГ
2020 г.

Содержание

Стр.

1. Постановка задачи	5
2. Теория	5
2.1. Эмпирическая функция распределения	5
2.1.1 Статистический ряд	5
2.1.2 Определение	5
2.1.3 Описание	5
2.2. Оценки плотности вероятности	5
2.2.1 Определение	5
2.2.2 Ядерные оценки	6
3. Реализация	6
4. Результаты	7
4.1. Эмпирическая функция распределений	7
4.2. Ядерные оценки плотности распределений	12
5. Обсуждение	27
6. Литература	27
7. Приложения	27

Список иллюстраций

1	Нормальное распределение	7
2	Распределение Коши	8
3	Распределение Лапласа.....	9
4	Распределение Пуассона	10
5	Равномерное распределение.....	11
6	Нормальное распределение $n = 20$	12
7	Нормальное распределение $n = 60$	13
8	Нормальное распределение $n = 100$	14
9	Распределение Коши $n = 20$	15
10	Распределение Коши $n = 60$	16
11	Распределение Коши $n = 100$	17
12	Распределение Лапласа $n = 20$	18
13	Распределение Лапласа $n = 60$	19
14	Распределение Лапласа $n = 100$	20
15	Распределение Пуассона $n = 20$	21
16	Распределение Пуассона $n = 60$	22
17	Распределение Пуассона $n = 100$	23
18	Равномерное распределение $n = 60$	24
19	Равномерное распределение $n = 20$	25
20	Равномерное распределение $n = 100$	26

Список таблиц

1 Постановка задачи

Для 5-ти 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.

2 Теория

2.1 Эмпирическая функция распределения

2.1.1 Статистический ряд

Статистическим рядом называется последовательность различных элементов выборки z_1, \dots, z_k , расположенных в возрастающем порядке с указанием частот n_1, \dots, n_k , с которыми эти элементы содержатся в выборке. Статистический ряд обычно записывается в виде таблицы.

2.1.2 Определение

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

$$F_n^*(x) = P^*(X < x)$$

2.1.3 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

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

$F_n^*(x)$ — функция распределения дискретной случайной величины X^* , заданной таблицей распределения. Эмпирическая функция распределения является оценкой, т. е. приближённым значением, генеральной функции распределения.

$$F_n^*(x) \approx F_X(x)$$

2.2 Оценки плотности вероятности

2.2.1 Определение

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

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

2.2.2 Ядерные оценки

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

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right)$$

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

1) при $n \rightarrow \infty$ $h_n \rightarrow 0$

2) $\frac{h_n}{n^{-1}} \rightarrow \infty$, когда $n \rightarrow \infty$.

Такие оценки называются непрерывными ядерными.

Замечание. Свойство, означающее сближение оценки с оцениваемой величиной при $n \rightarrow \infty$ 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

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

Правило Сильвермана

$$h_n = 1.06 \hat{\sigma} n^{-0.2}$$

, где $\hat{\sigma}$ - выборочное среднее отклонение.

3 Реализация

Для генерации выборки был использован *Python 3.7*: модуль *random* библиотеки *numpy* для генерации случайных чисел с различными распределениями. Графики построены средствами *matplotlib*.

4 Результаты

4.1 Эмпирическая функция распределений

Красный график - эталонная функция. Синий график - эмпирическая соответственно.

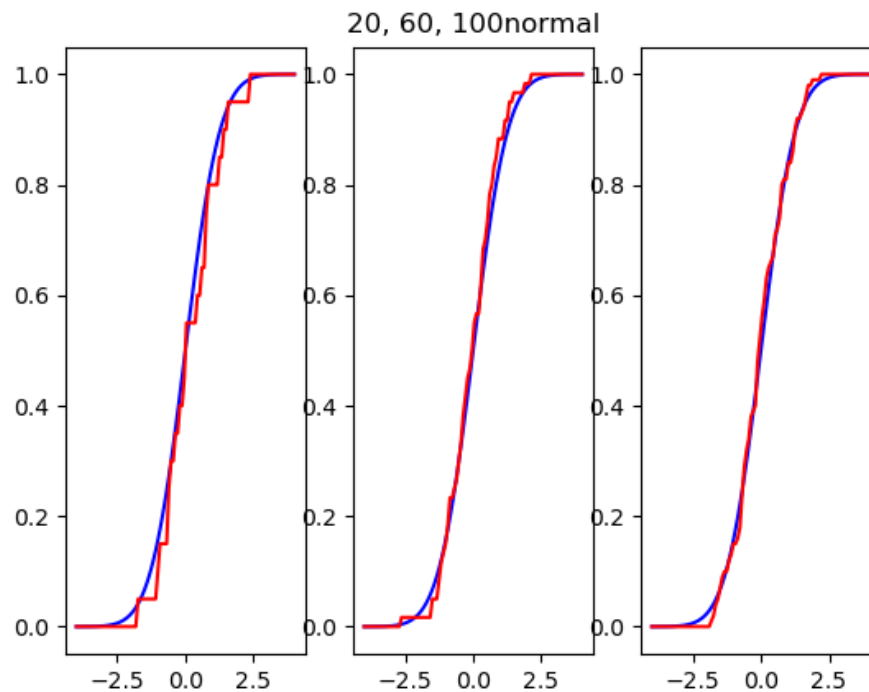


Рис. 1: Нормальное распределение

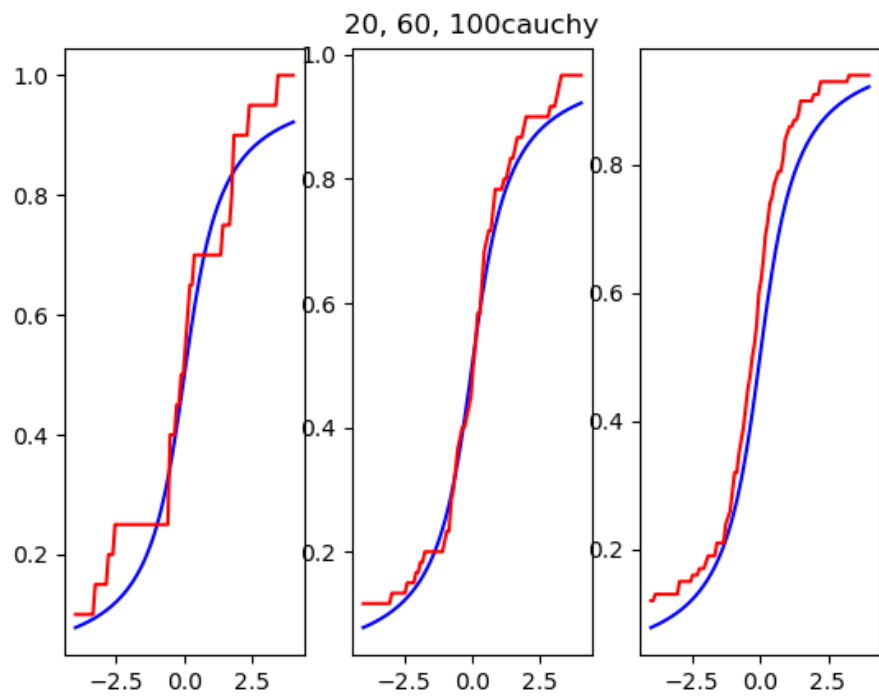


Рис. 2: Распределение Коши

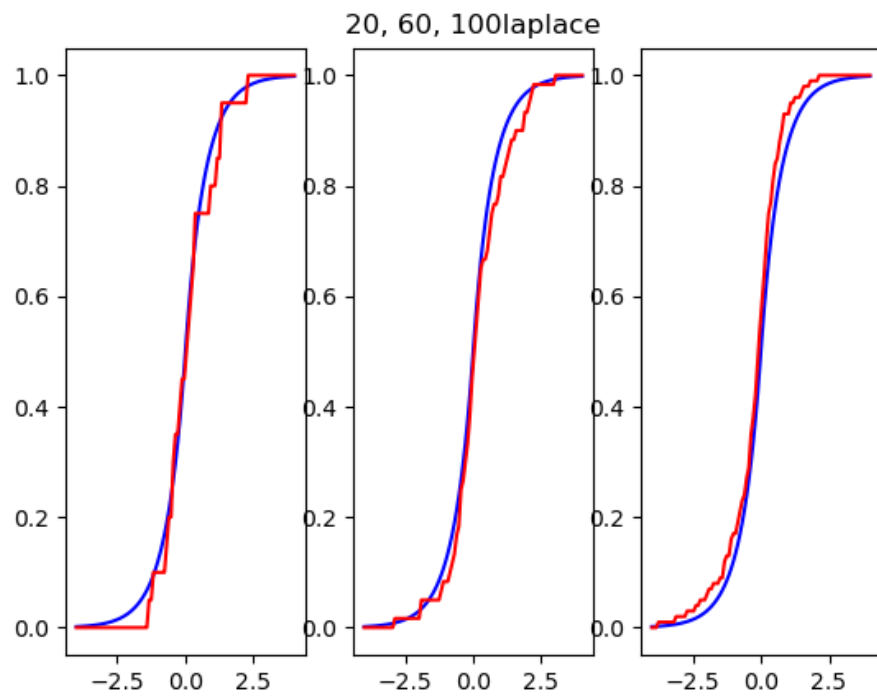


Рис. 3: Распределение Лапласа

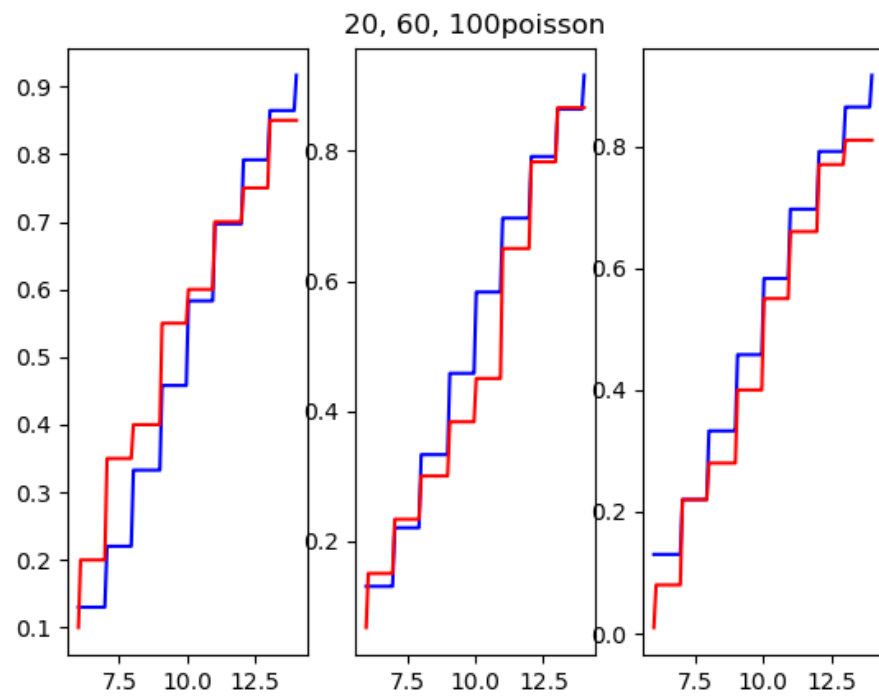


Рис. 4: Распределение Пуассона

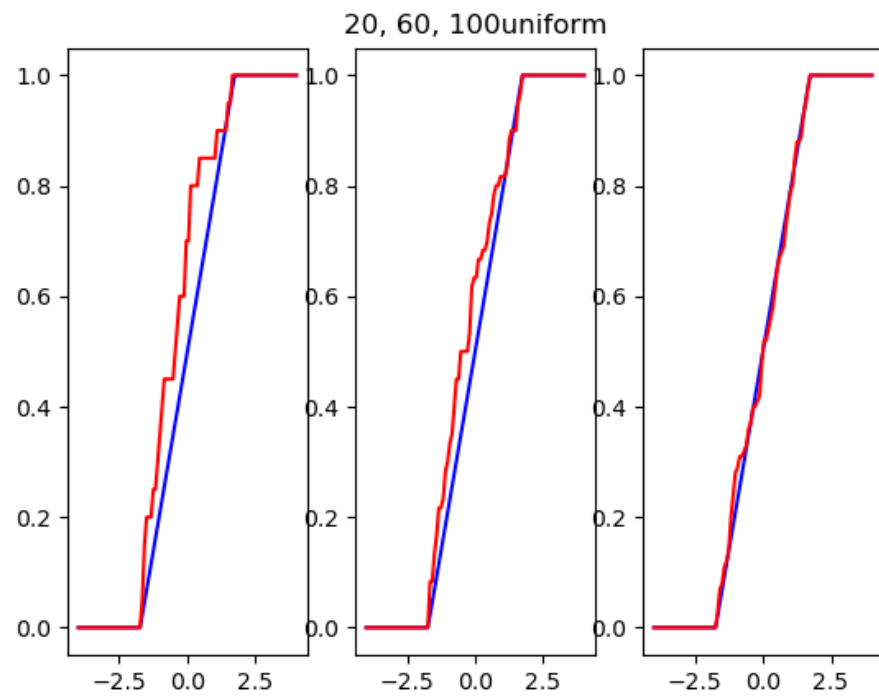


Рис. 5: Равномерное распределение

4.2 Ядерные оценки плотности распределений

Черный цвет - эталонная функция. Красный цвет - эмпирическая.

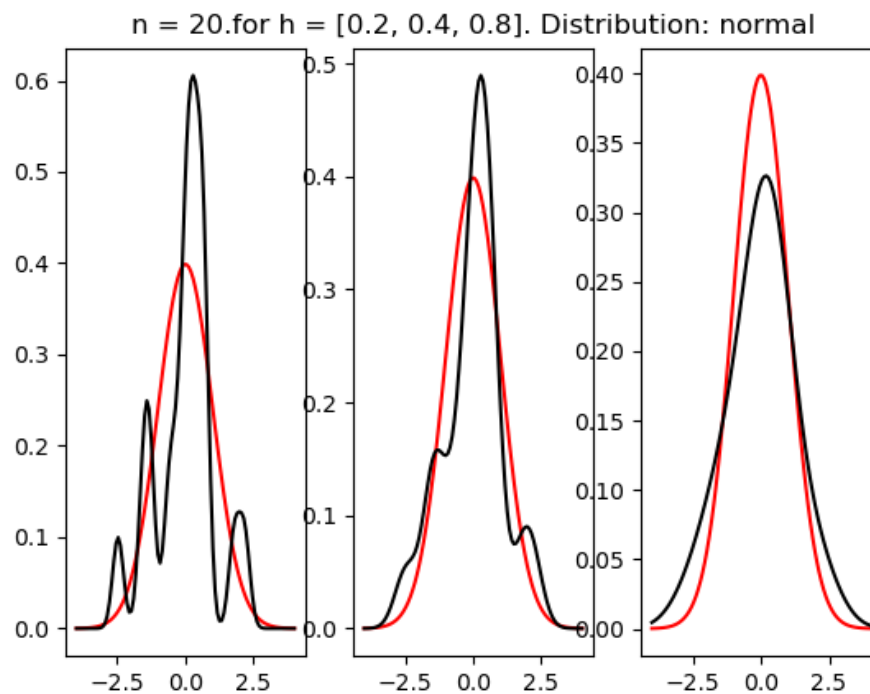


Рис. 6: Нормальное распределение $n = 20$

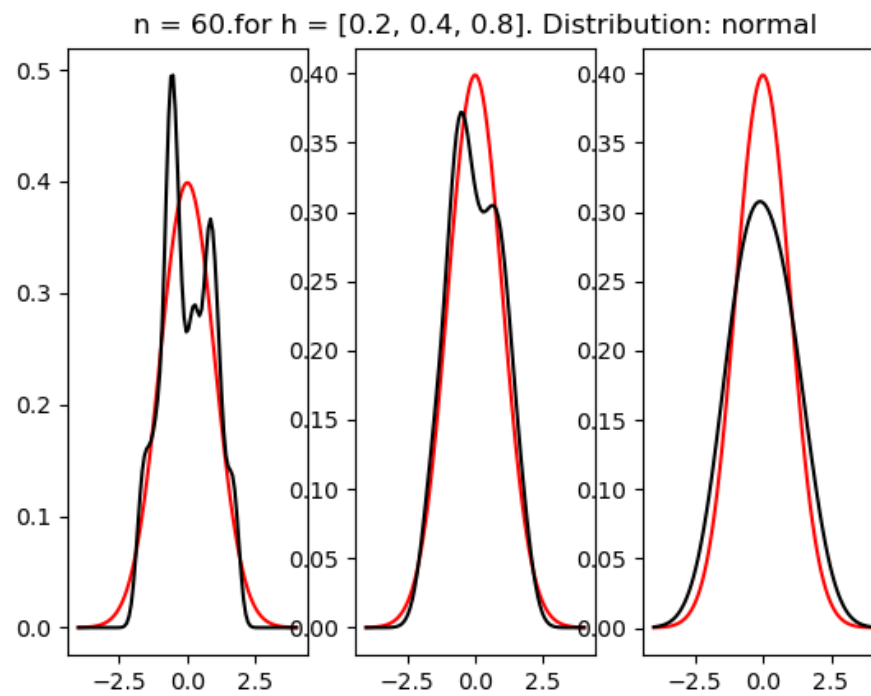


Рис. 7: Нормальное распределение $n = 60$

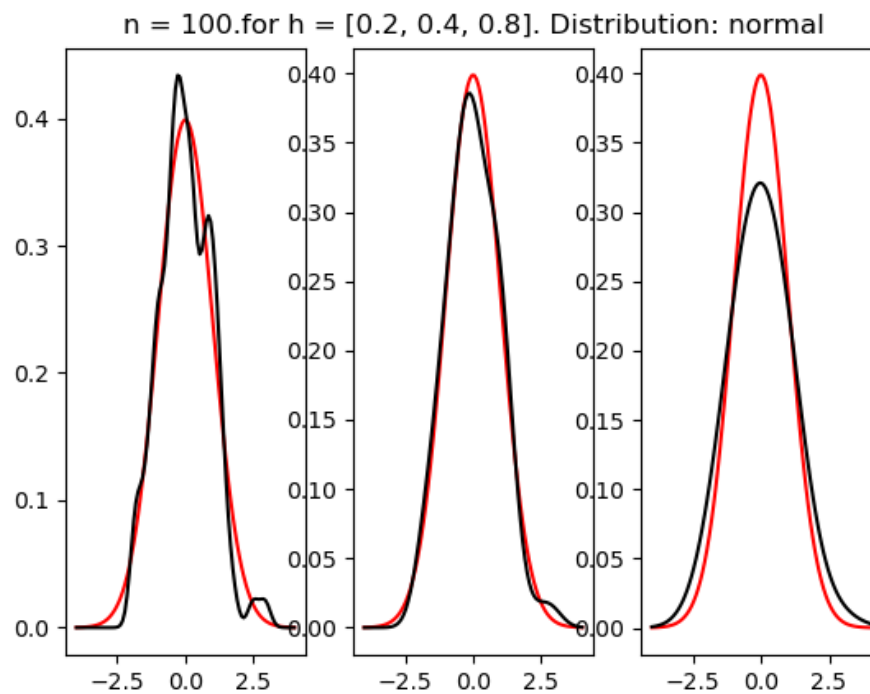


Рис. 8: Нормальное распределение $n = 100$

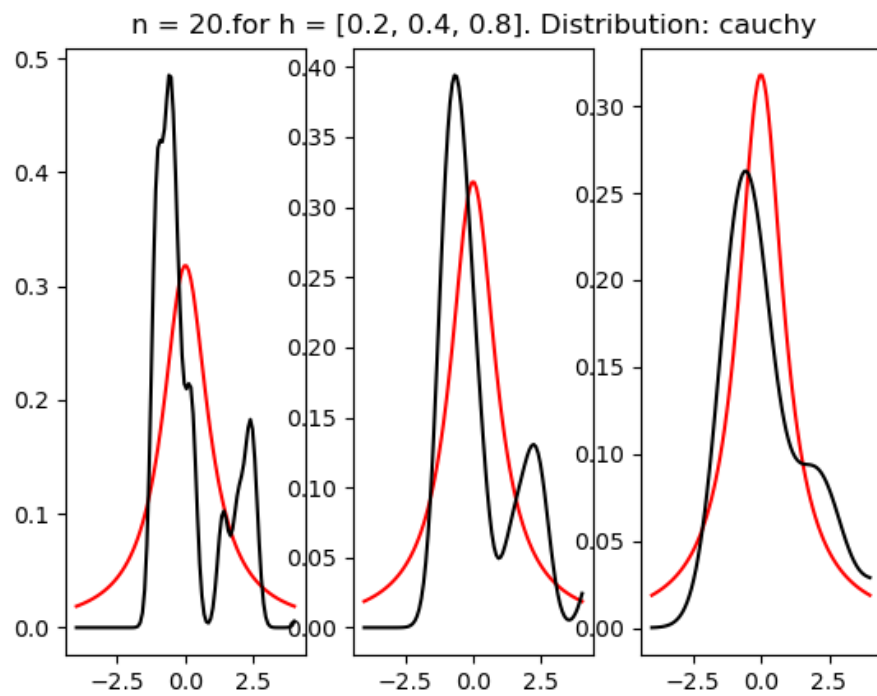


Рис. 9: Распределение Коши $n = 20$

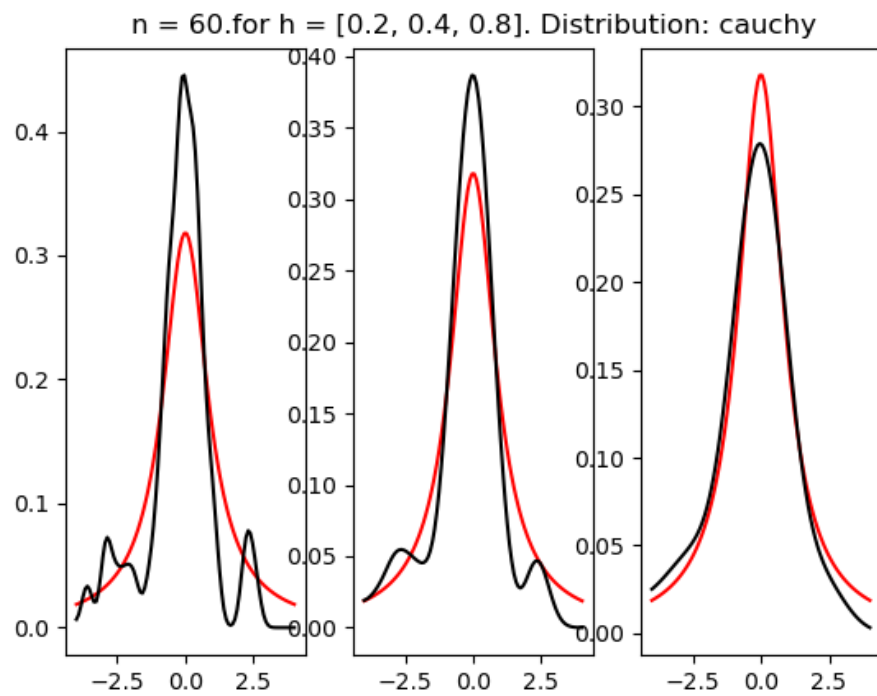


Рис. 10: Распределение Коши $n = 60$

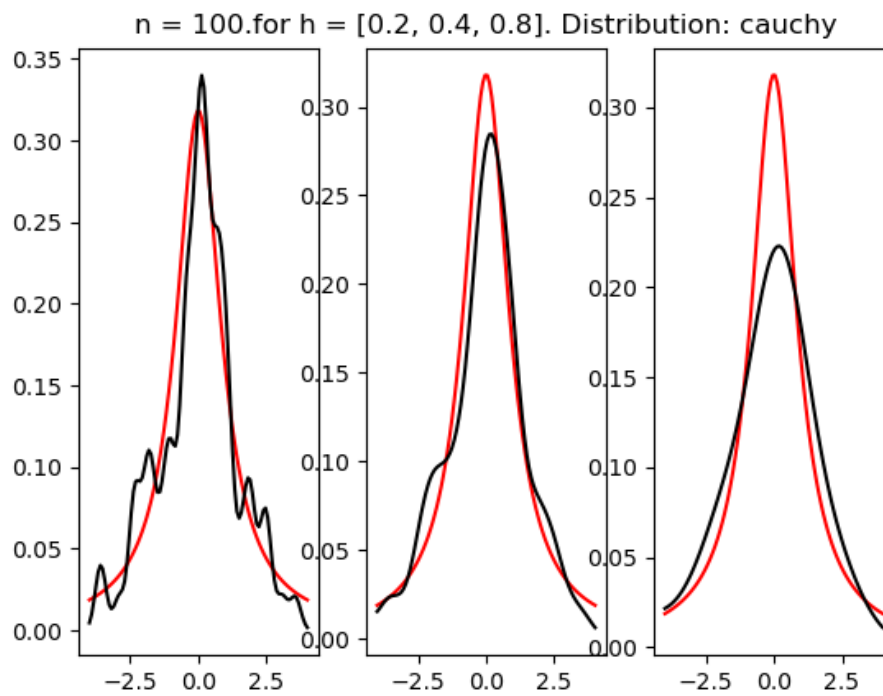


Рис. 11: Распределение Коши $n = 100$

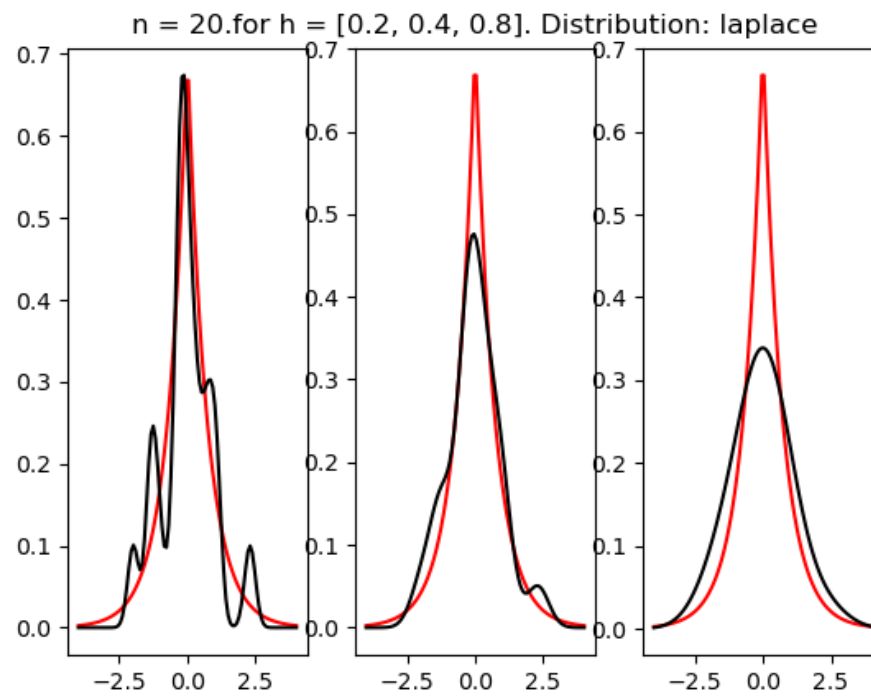


Рис. 12: Распределение Лапласа $n = 20$

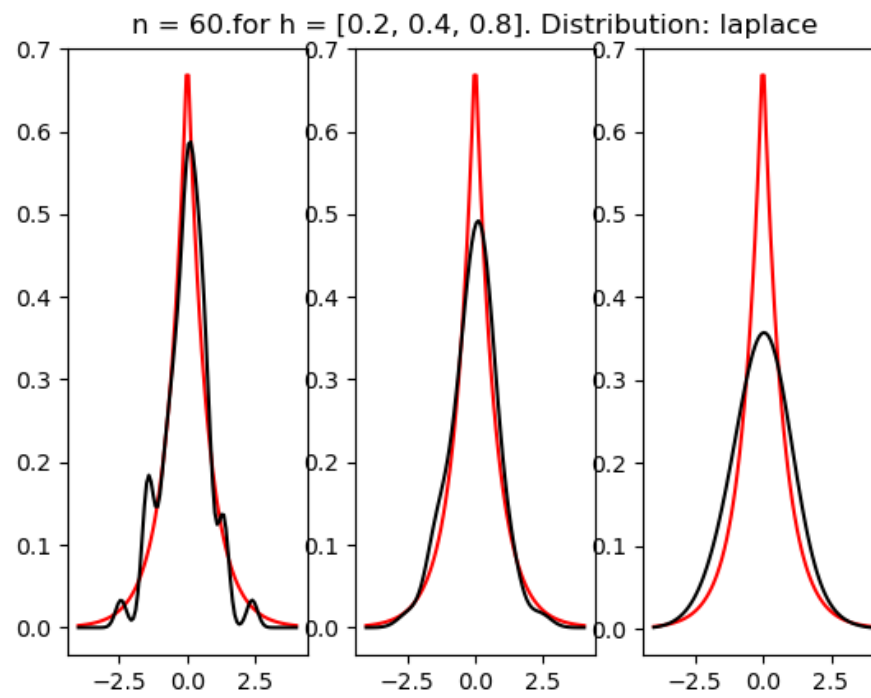


Рис. 13: Распределение Лапласа $n = 60$

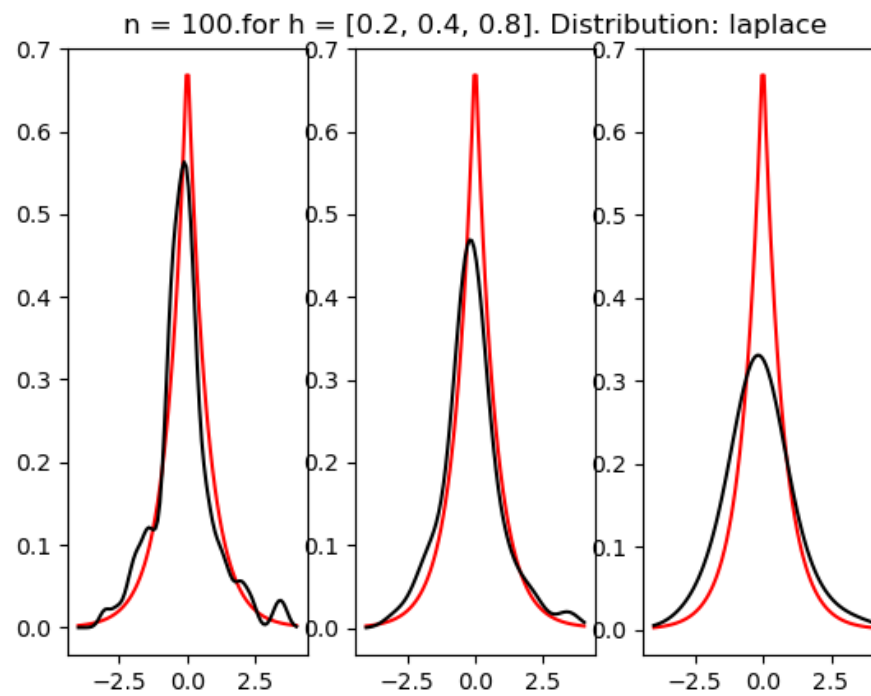


Рис. 14: Распределение Лапласа $n = 100$

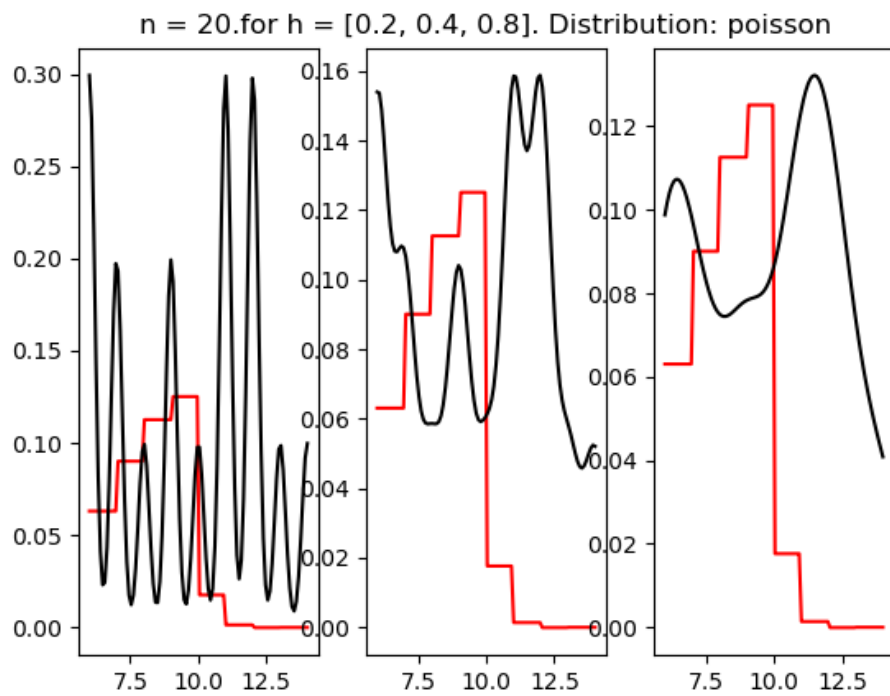


Рис. 15: Распределение Пуассона $n = 20$

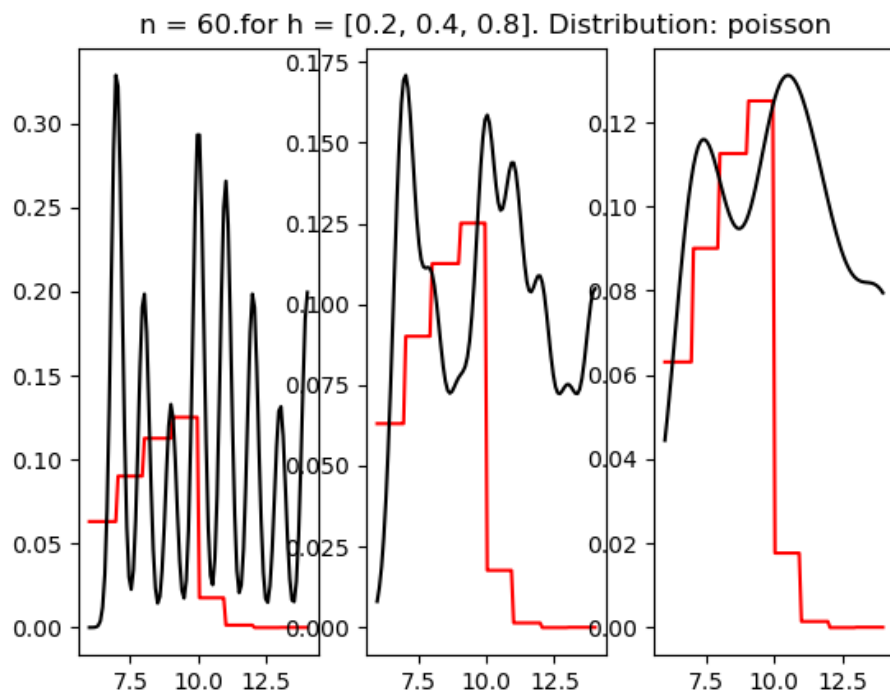


Рис. 16: Распределение Пуассона $n = 60$

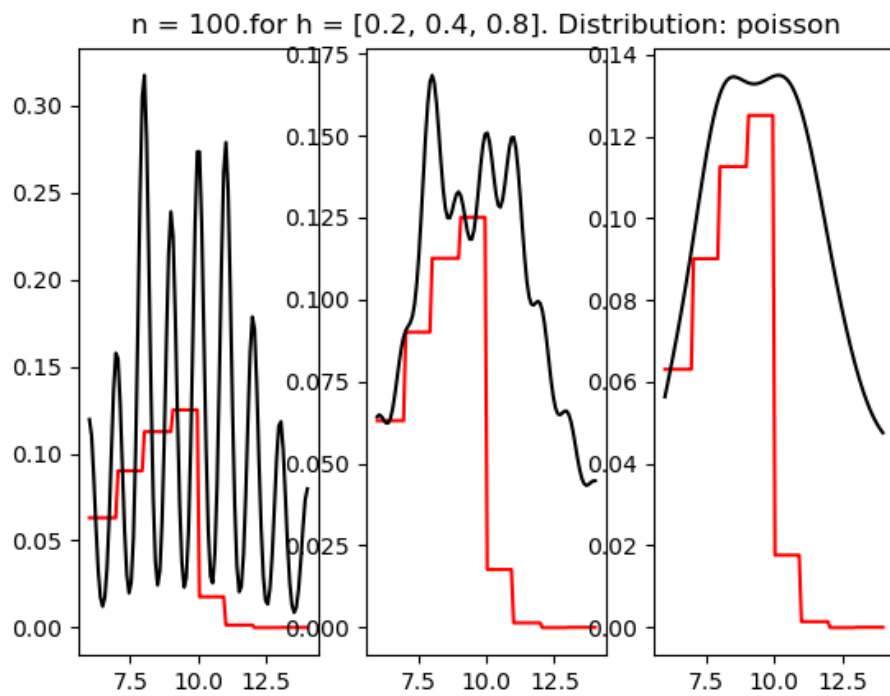


Рис. 17: Распределение Пуассона $n = 100$

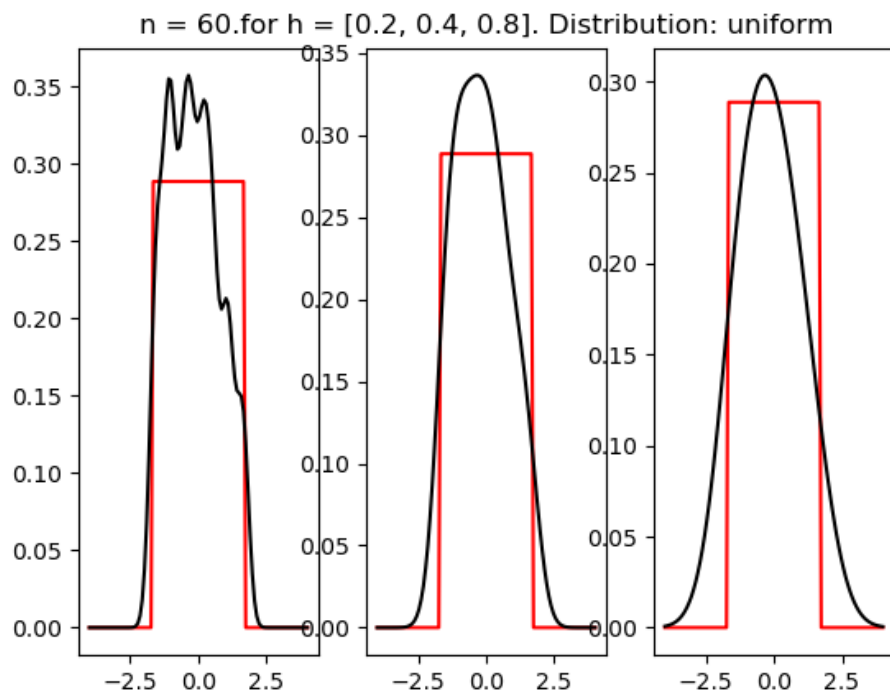


Рис. 18: Равномерное распределение $n = 60$

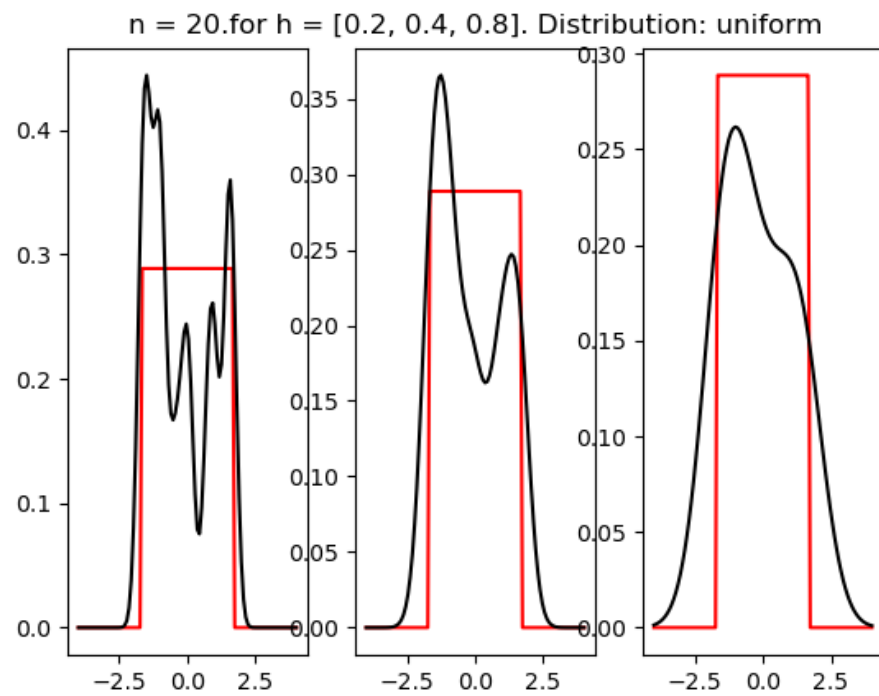


Рис. 19: Равномерное распределение $n = 20$

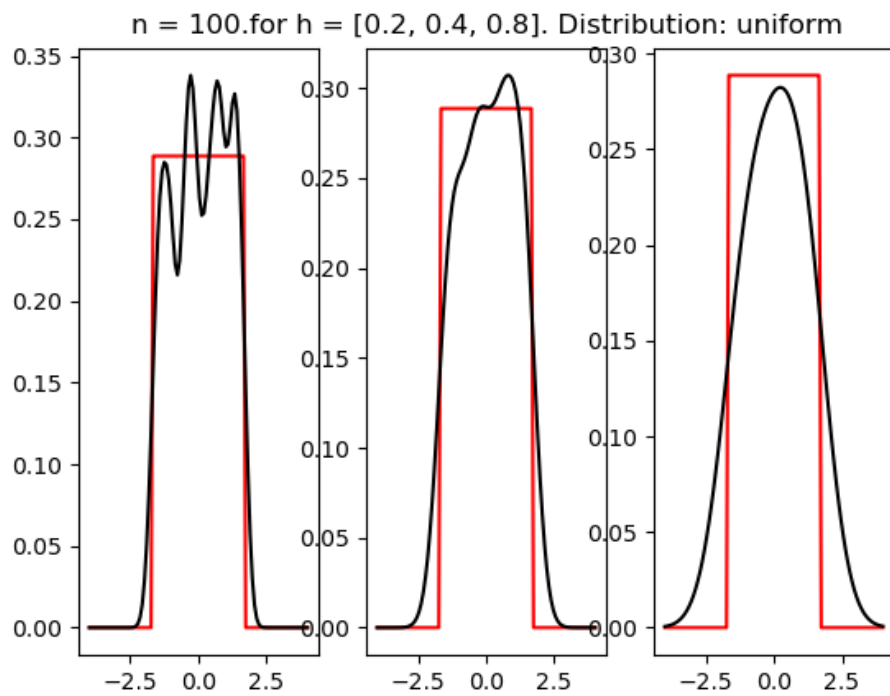


Рис. 20: Равномерное распределение $n = 100$

5 Обсуждение

Видно, что эмпирическая функция лучше приближает эталонную функцию на больших выборках.

При фиксированной ширине окна точнее приблизить функцию распределения позволяет увеличение выборки.

Наилучшее приближение функции распределения ядерной функции для распределения Лапласа, нормального распределения и распределения Коши достигается при $h = h_n$, в свою очередь для равномерного распределения наилучшее приближение достигается при $h = \frac{h_n}{2}$ и h_n . Для распределения Пуассона при $h = 2h_n$.

6 Литература

Модуль `numpy`

Модуль `matplotlib`

Модуль `scipy`

Ядерная оценки плотности

7 Приложения

Код лабораторной

Код отчёта