Санкт-Петербургский политехнический университет
Петра Великого

Физико-механический институт

Кафедра «Прикладная математика»

Отчет по лабораторной работе №4

по дисциплине
"Математическая статистика"

Выполнил студент
группы 5030102/00101 Нгуен Хоанг Линь

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

Санкт-Петербург
2023 г.

Содержание

Список иллюстраций

1. 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 $U(x, -\sqrt{3}, \sqrt{3})$

Необходимо:

1) Сгенерировать выборки размером 20, 60 и 100 элементов
2) 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

Статистический ряд - последовательность различных элементов выборки $\{z_i\}_{i=1}^{k}$, расположенных по возрастанию, с указанием частот $\{n_i\}_{i=1}^{k}$, с которыми эти элементы содержатся в выборке. Обычно записывается в виде таблицы.

Эмпирическая (выборочная) функция распределения - относительная частота события $X < x$, полученная по данной выборке:

$$F_n^*(x) = P^*(X < x) \tag{1}$$

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:

$$F^*(x) = \frac{1}{n}\sum_{z_i<x} n_i \tag{2}$$

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 1. Таблица распределения

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

$$F_n^*(x) \approx F_X(x) \tag{3}$$

2.2. Ядерные оценки плотности вероятности

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

$$\widehat{f}(x) \approx f(x) \tag{4}$$

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

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{x - x_i}{h_n}) \tag{5}$$

Здесь функция $K(u)$ - ядро (ядерная функция), непрерывна и является плотностью вероятности, $x_1, ..., x_n$ - элементы выборки, любая последовательность положительных чисел $\{h_n\}$, обладающая свойствами:

$$h_n \xrightarrow[n \to \infty]{} 0; nh_n \xrightarrow[n \to \infty]{} \infty. \tag{6}$$

Такие оценки называются непрерывными ядерными

Гауссово (нормальное) ядро:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{7}$$

Правило Сильвермана:

$$h_n = 1.06 \hat{\sigma} n^{-1/5} \tag{8}$$

где $\hat{\sigma}$ - выборочное стандартное отклонение.

3. Реализация

Лабораторная работа выполнена на языке Python в виртуальной среде Anaconda с интерпретатором версии 3.9 в среде разработки Visual Studio Code. Дополнительные зависимости:

- scipy
- numpy
- matplotlib
- seaborn
- statsmodels

4. Результаты

4.1. Эмпирическая функция распределения

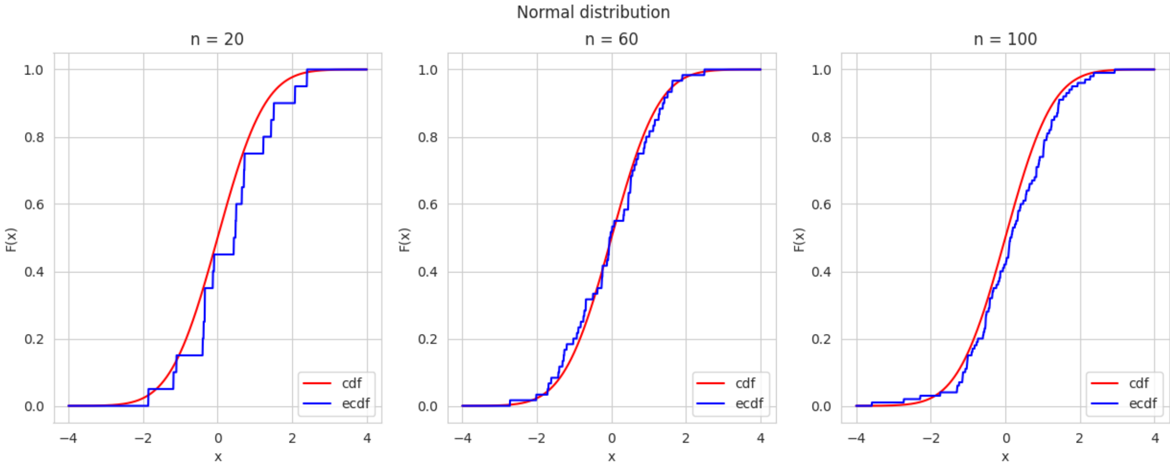

Рис. 1. Нормальное распределение

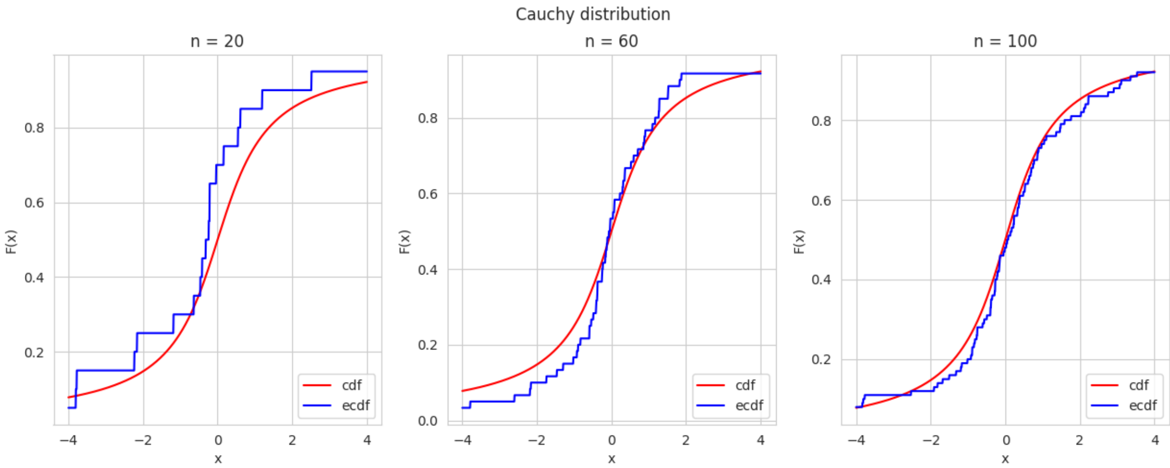

Рис. 2. Распределение Коши

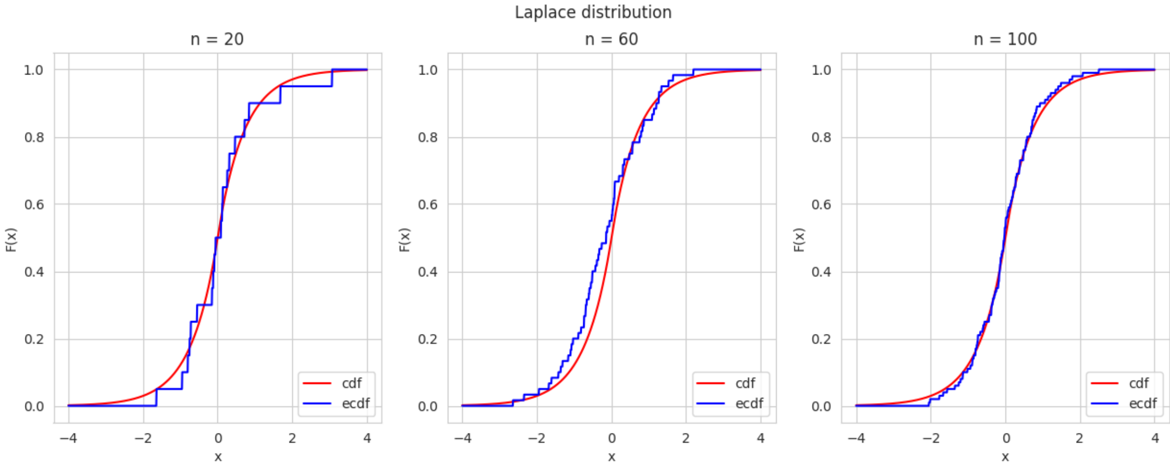

Рис. 3. Распределение Лапласа

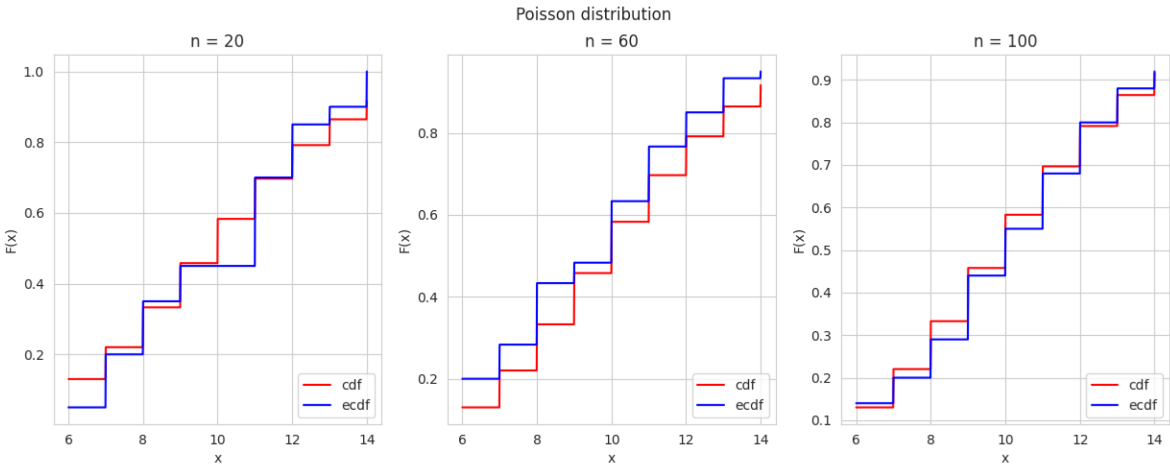

Рис. 4. Распределение Пуассона

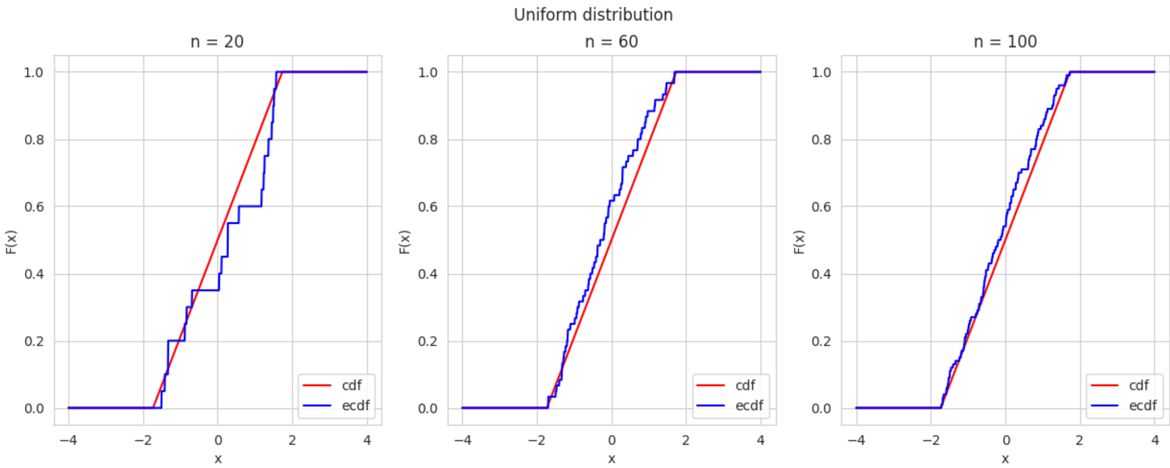

Рис. 5. Равномерное распределение

4.2. Ядерные оценки плотностей распределения

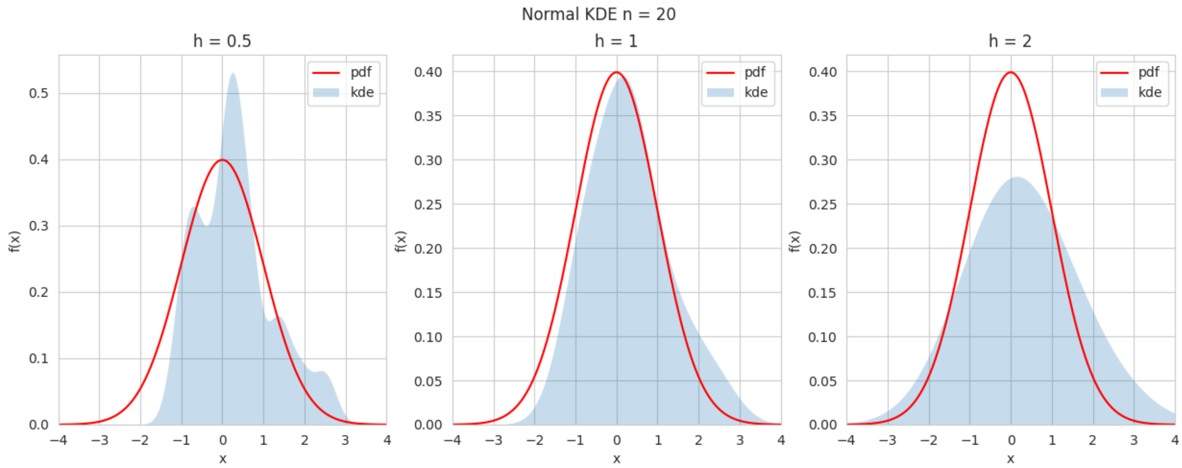

Рис. 6. Нормальное распределение размерностью 20

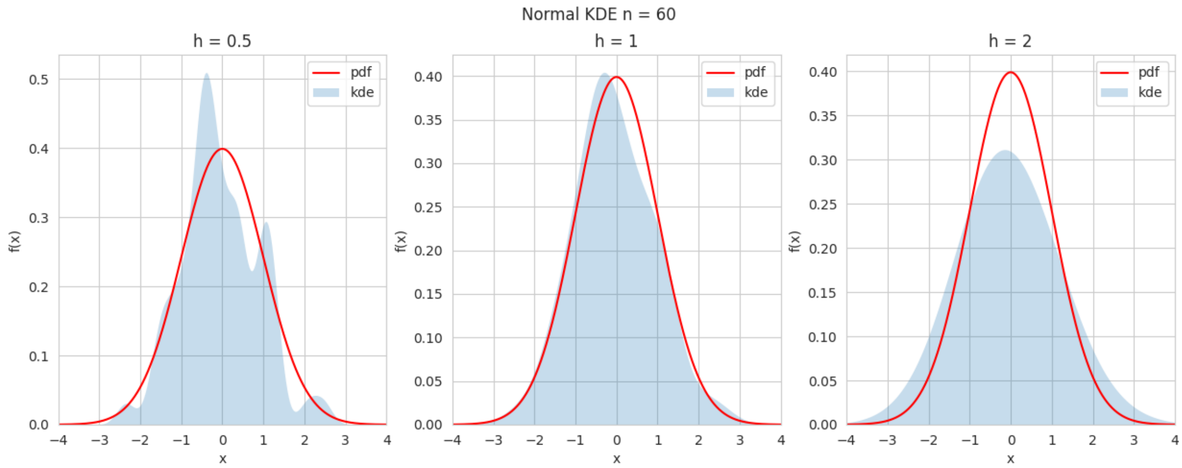

Рис. 7. Нормальное распределение размерностью 60

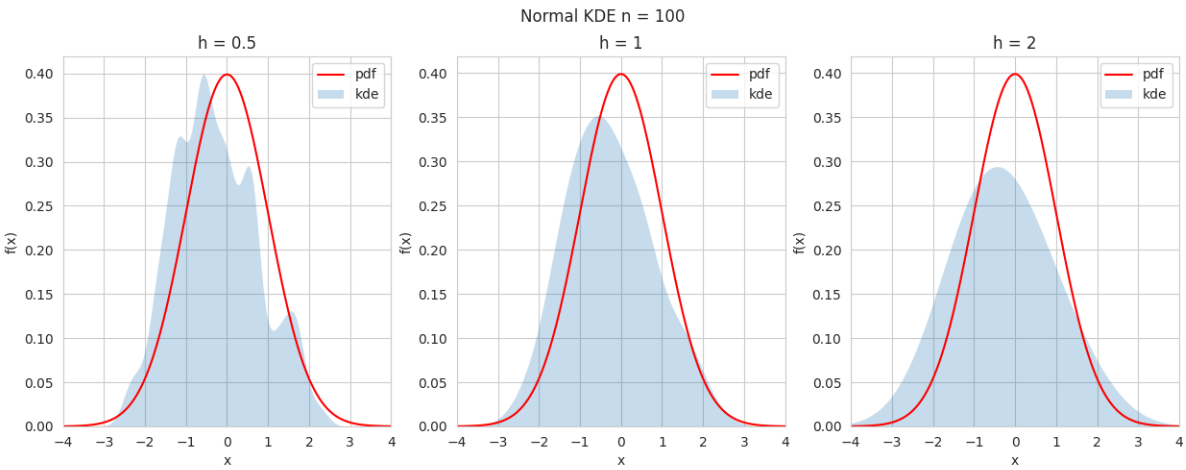

Рис. 8. Нормальное распределение размерностью 100

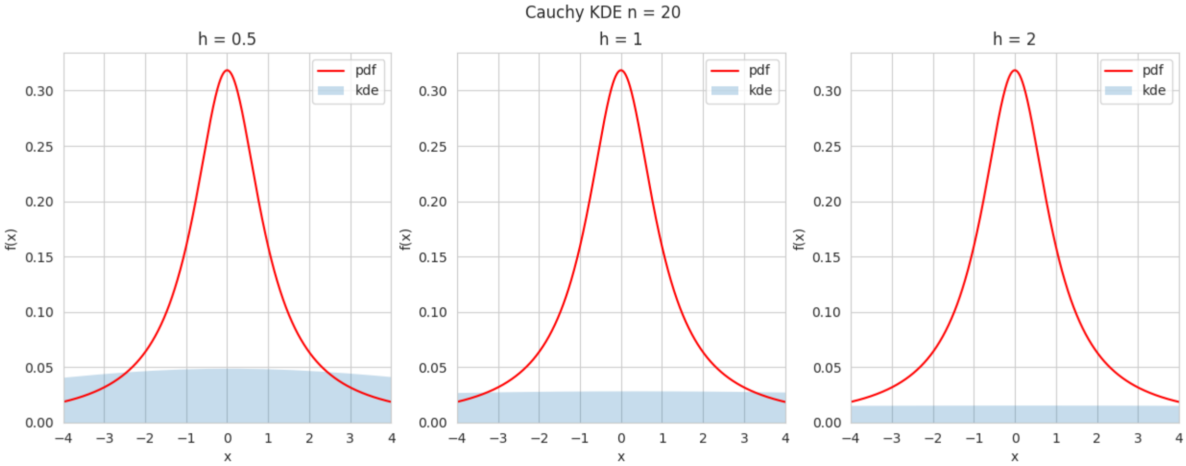

Рис. 9. Распределение Коши размерностью 20

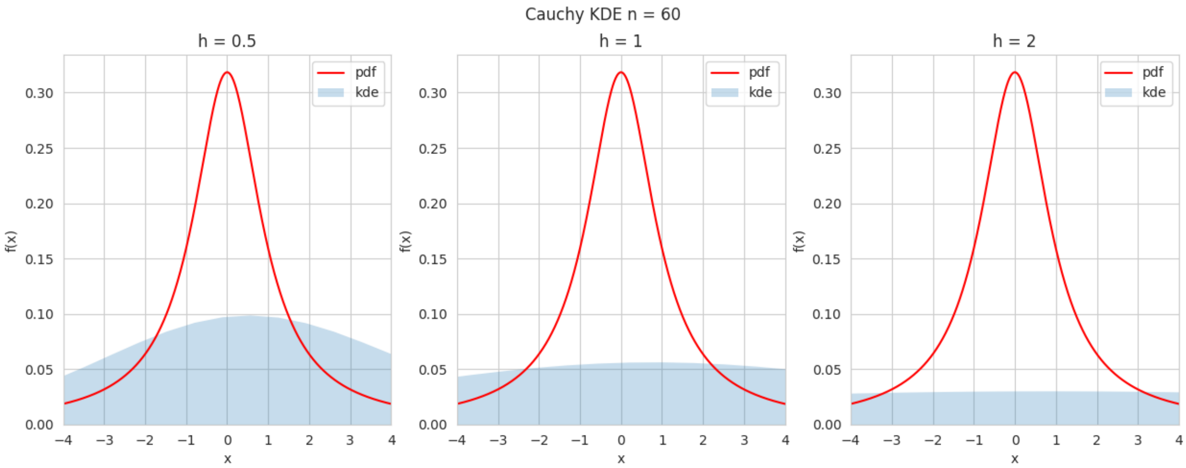

Рис. 10. Распределение Коши размерностью 60

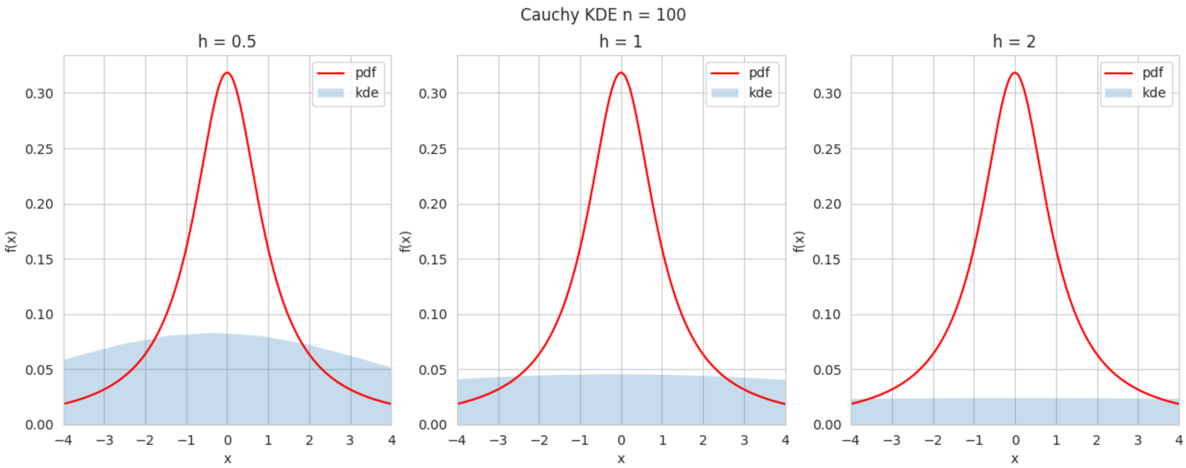

Рис. 11. Распределение Коши размерностью 100

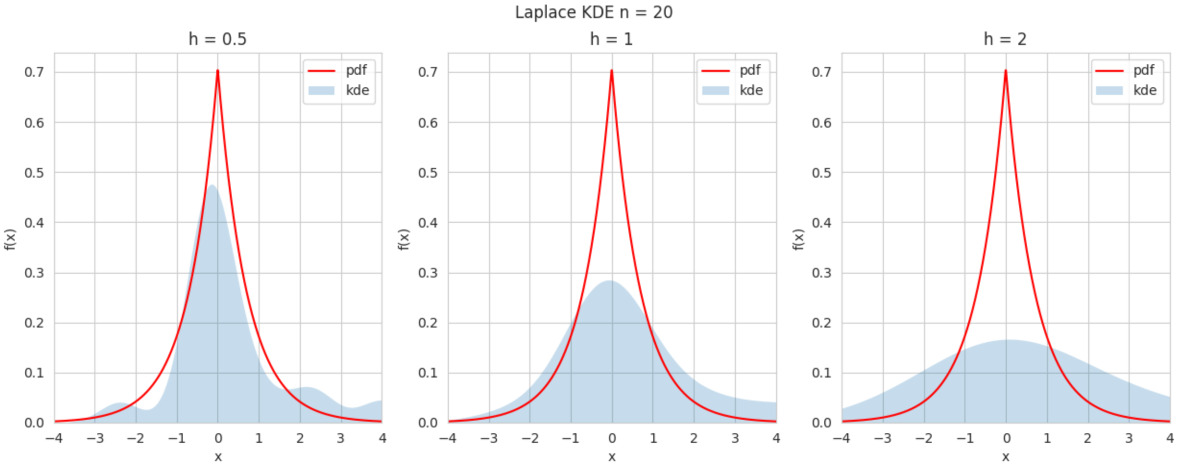

Рис. 12. Распределение Лапласа размерностью 20

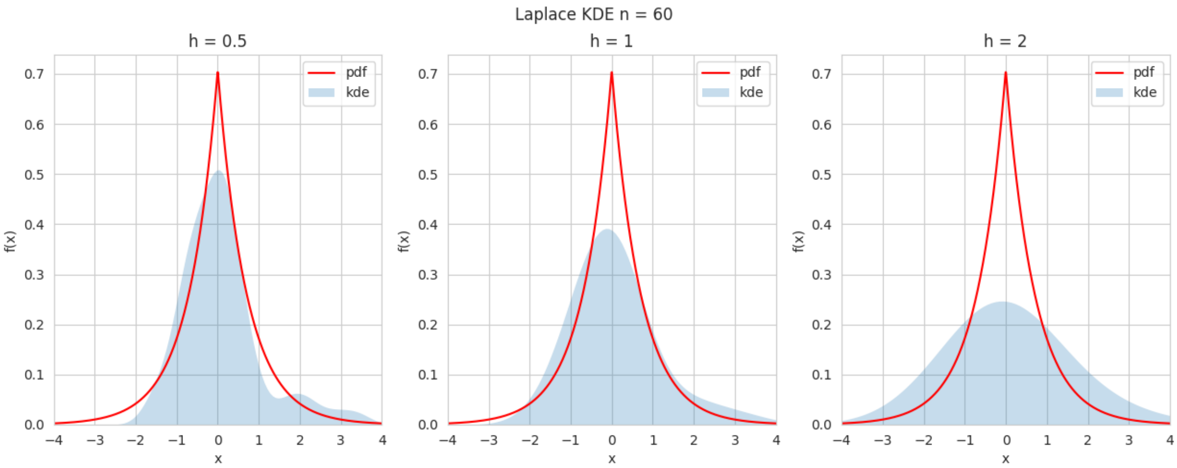

Рис. 13. Распределение Лапласа размерностью 60

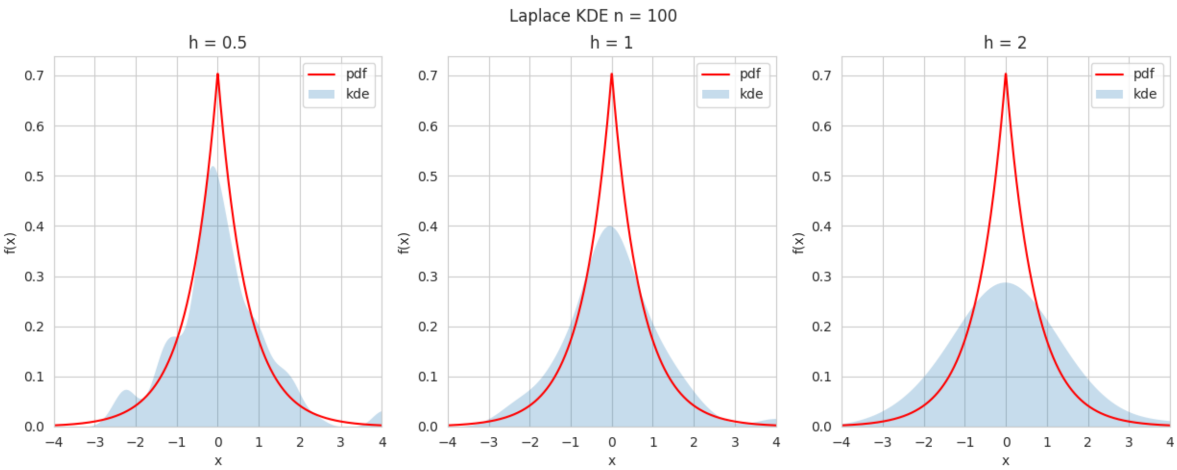

Рис. 14. Распределение Лапласа размерностью 100

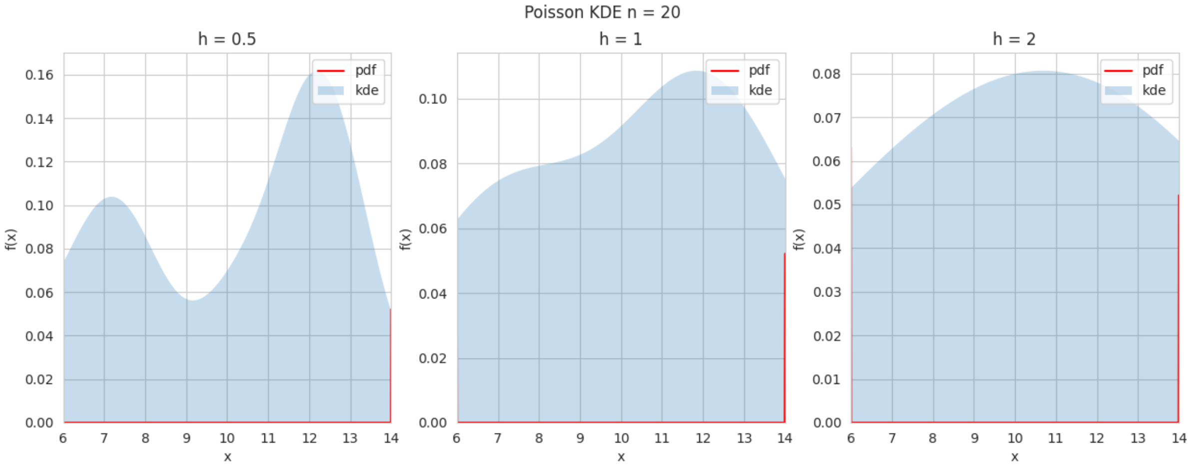

Рис. 15. Распределение Пуассона размерностью 20

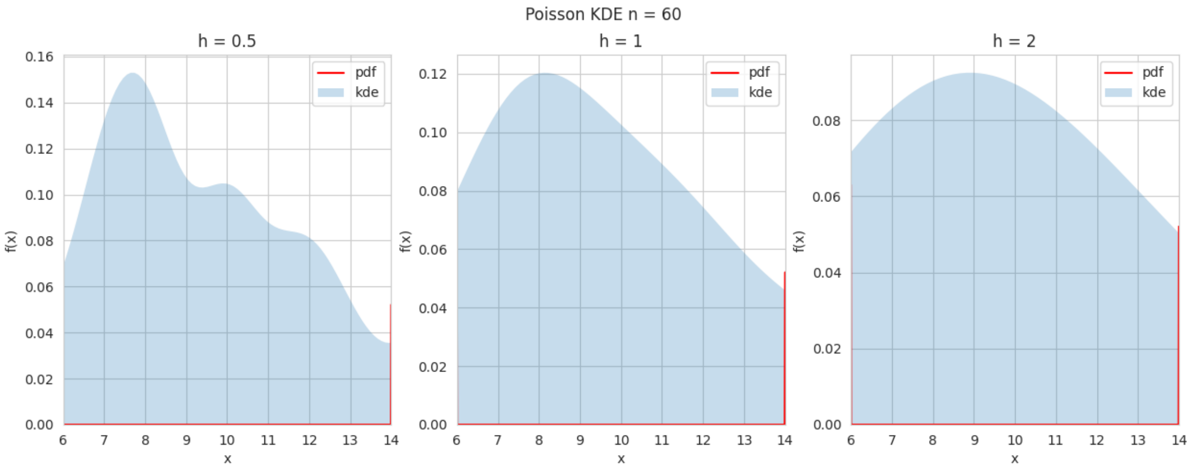

Рис. 16. Распределение Пуассона размерностью 60

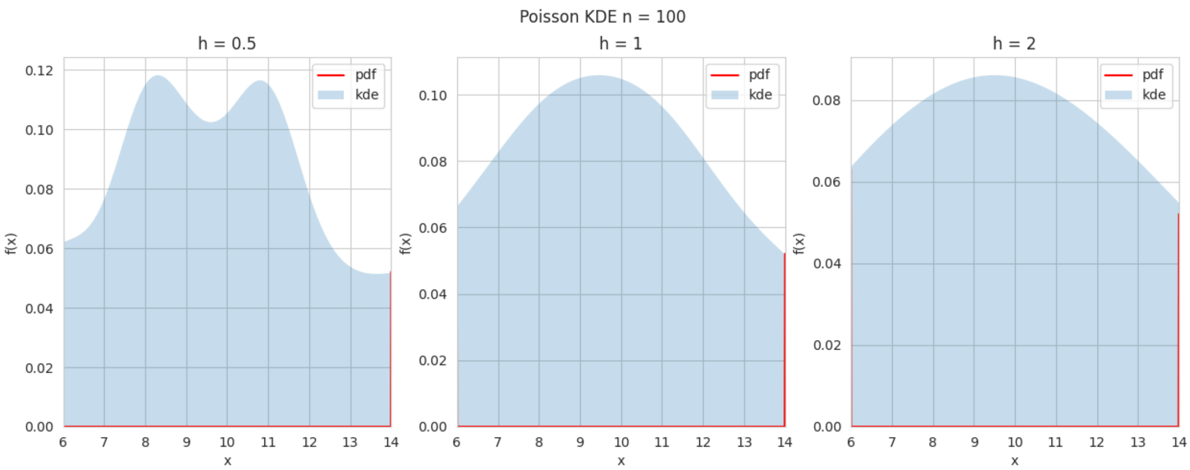

Рис. 17. Распределение Пуассона размерностью 100

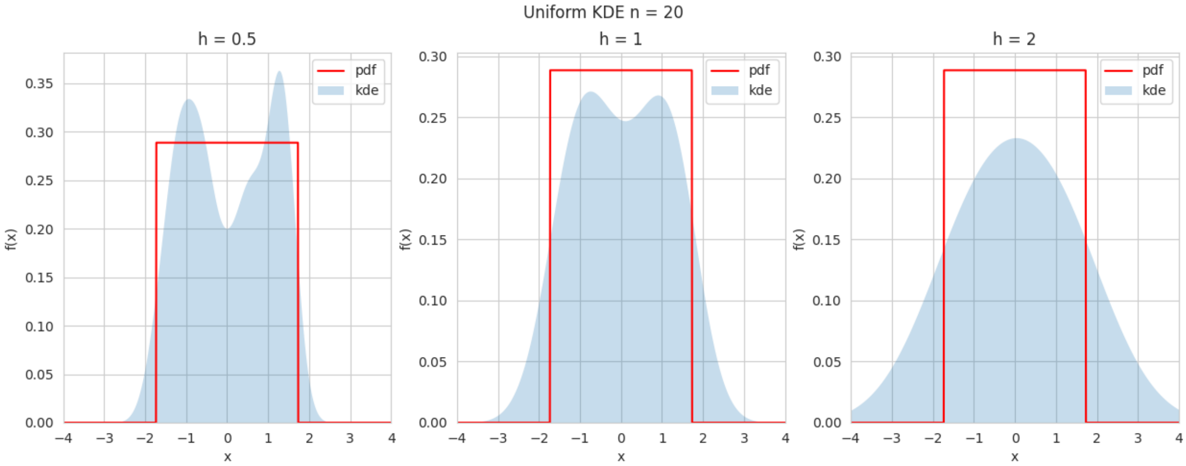

Рис. 18. Равномерное распределение размерностью 20

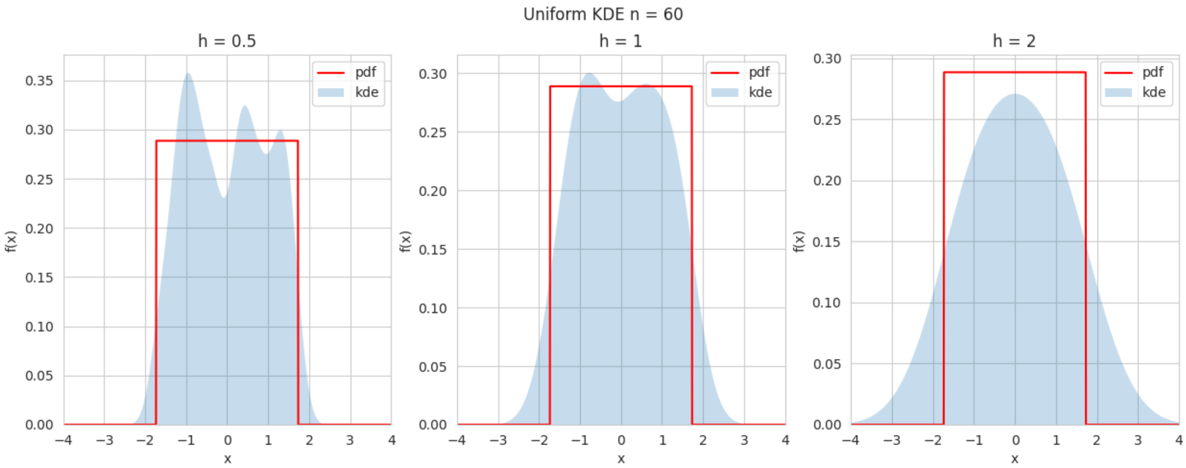

Рис. 19. Равномерное распределение размерностью 60

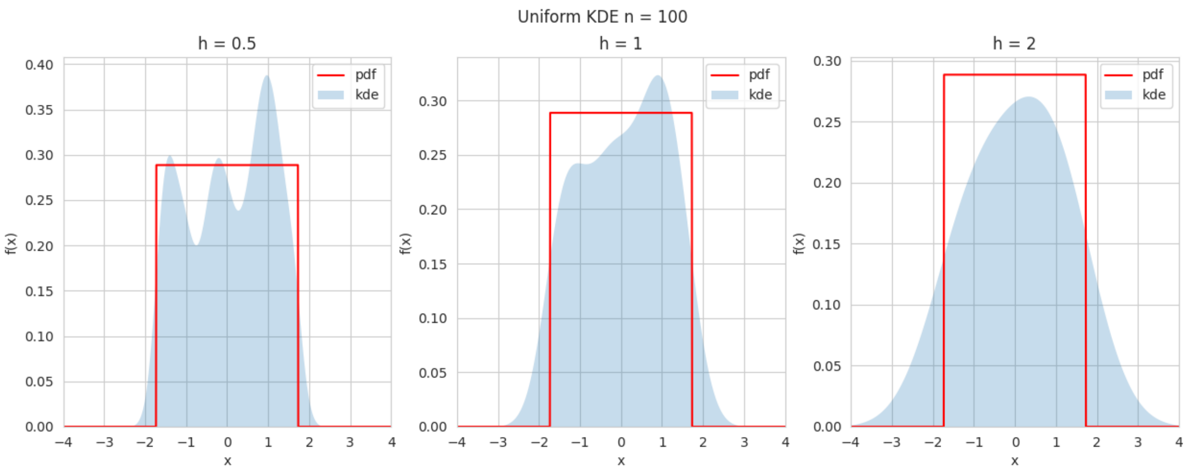

Рис. 20. Равномерное распределение размерностью 100

5. Обсуждение

На графиках эмпирических функций можно заметить, что с увеличением размерности выборки соответствующая ступенчатая эмпирическая функция становится ближе к эталонной функции распределения. Из представленных распределений график эмпирической функции распределения Пуассона больше всего отклоняется от эталонной функции.

Графики ядерных оценок плотностей распределения тоже показывают приближение ядерной оценки к функции плотности вероятности по всем h при увеличении размерности выборки. При этом для распределения Коши лучшим значением размерности выборки можно считать 60.

Для разных распределений лучшее значение параметра h различается. $h = h_n$ лучше всего приближает нормальное распределение, распределение Пуассона и равномерное распределение. Для распределения Коши и Лапласа оптимально значение $h = \frac{h_n}{2}$. Помимо этого при увеличении параметра h уменьшается значение пиковой точки и расширяется основание графика ядерной оценки для распределения Коши и равномерного. И для всех распределений при значении параметра $h = 2h_n$ замечается преобразование графика ядерной оценки в функцию с единственным максимумом.