Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики

Кафедра «Прикладная математика»

# Отчет по лабораторной работе №4

по дисциплине
"Математическая статистика"

Выполнила студентка
группы 3630102/80201 Деркаченко Анна Олеговна

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

Санкт-Петербург
2021 г.

# Содержание

# Список иллюстраций

# 1. Постановка задачи

Даны распределения:

- 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 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, суказанием частот $\{n_i\}_{i=1}^k$, с которыми эти элементы содержатся в выборке.

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

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

Ее можно найти по формуле:

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

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

### 2.2. Ядерные оценки плотности вероятности

*Оценка плотности вероятности* $f(x)$ - построенная на основе выборки функция $\widehat{f}(x)$ : $\widehat{f}(x) \approx f(x)$

Непрерывная ядерная оценка задается формулой:

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{x - x_i}{h_n}) \tag{3}$$

где $K(u)$ - ядро, т. е. непрерывная функция, являющаяся плотностью вероятности, $x_1, ..., x_n$ - элементы выборки, последовательность $\{h_n\}$ :

$$h_n \xrightarrow[n \to \infty]{} 0; nh_n \xrightarrow[n \to \infty]{} \infty. \tag{4}$$

Используется Гауссово ядро:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{5}$$

А также правило Сильвермана:

$$h_n = 1.06\hat{\sigma}n^{-1/5} \tag{6}$$

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

## 3. Реализация

Реализация лабораторной работы проводилась на языке Python в среде разработки PyCharm с использованием дополнительных библиотек:

- scipy
- numpy
- math
- matplotlib
- seaborn
- statsmodels

Исходный код лабораторной работы размещен в GitHub-репозитории.

URL: https://github.com/derkanw/Mathstat/tree/main/lab4

# 4. Результаты

## 4.1. Эмпирическая функция распределения

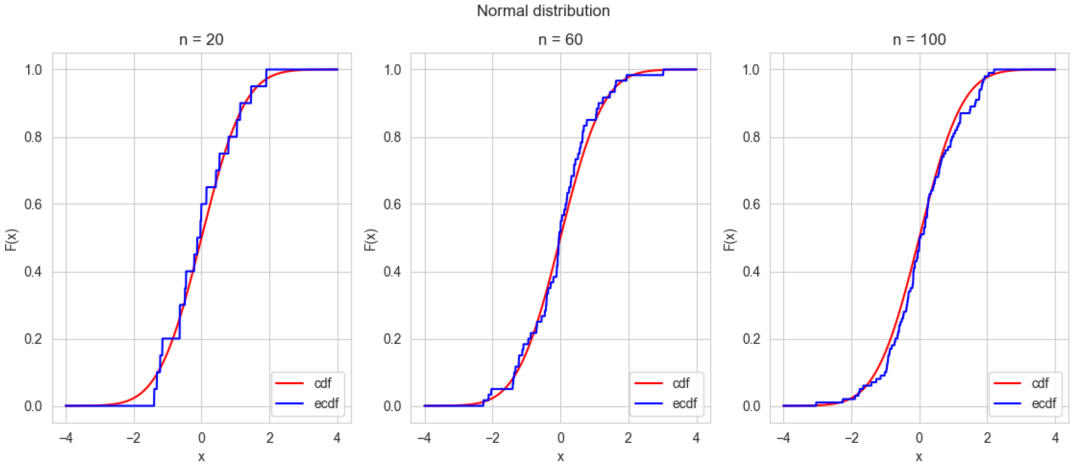


Рис. 1. Нормальное распределение

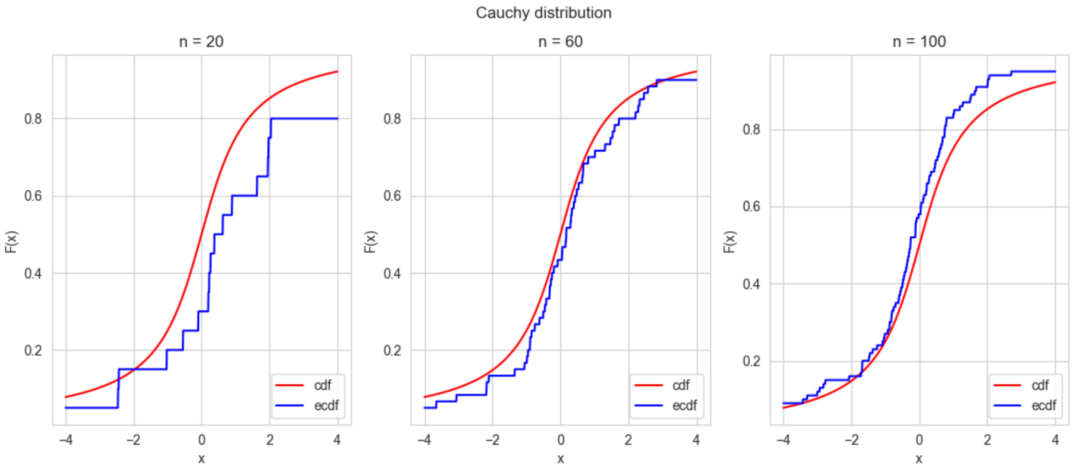


Рис. 2. Распределение Коши

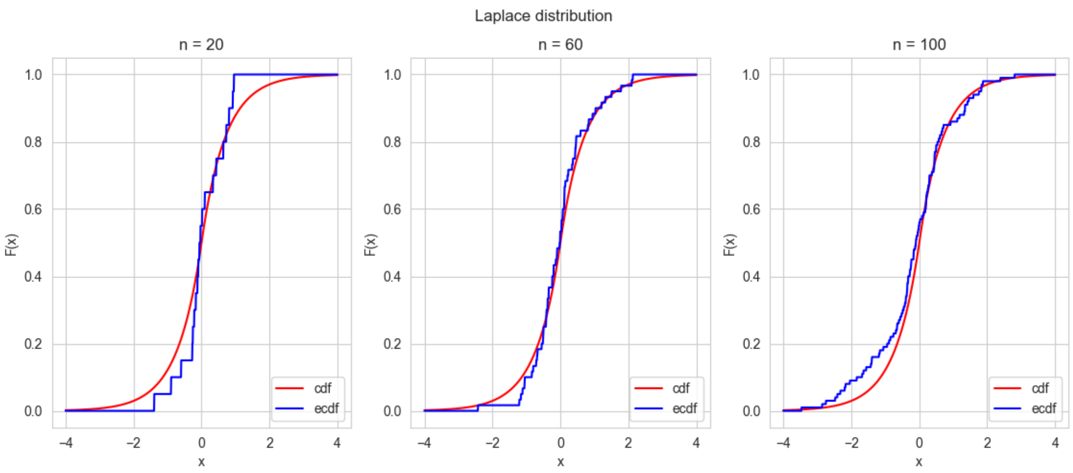


Рис. 3. Распределение Лапласа

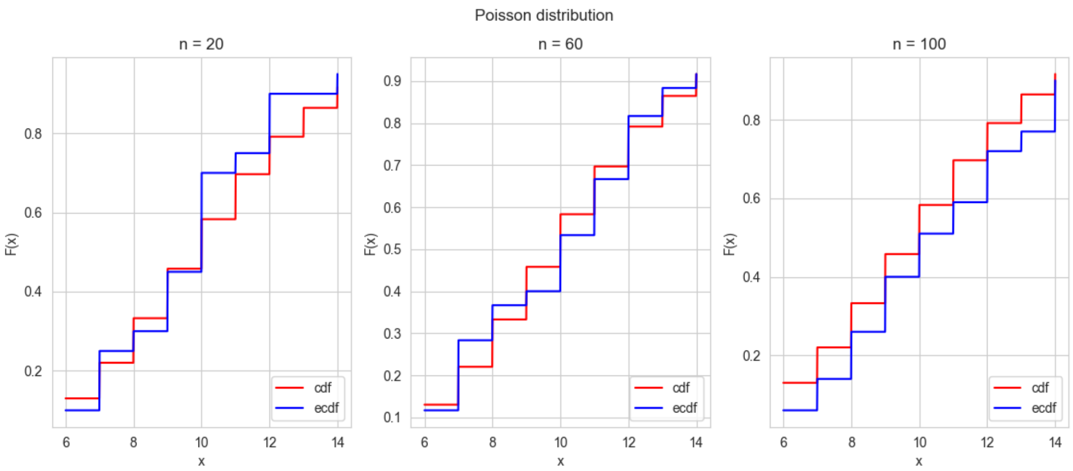


Рис. 4. Распределение Пуассона

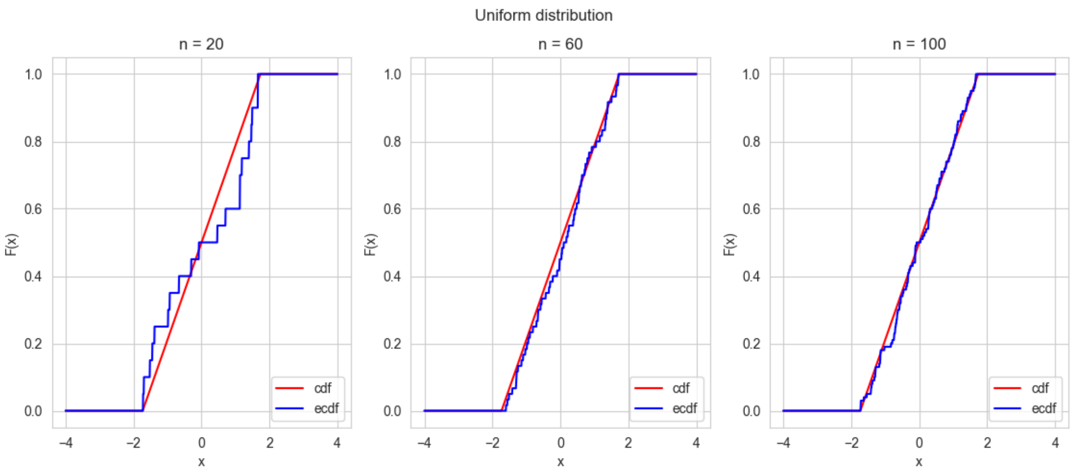


Рис. 5. Равномерное распределение

## 4.2. Ядерные оценки плотностей распределения

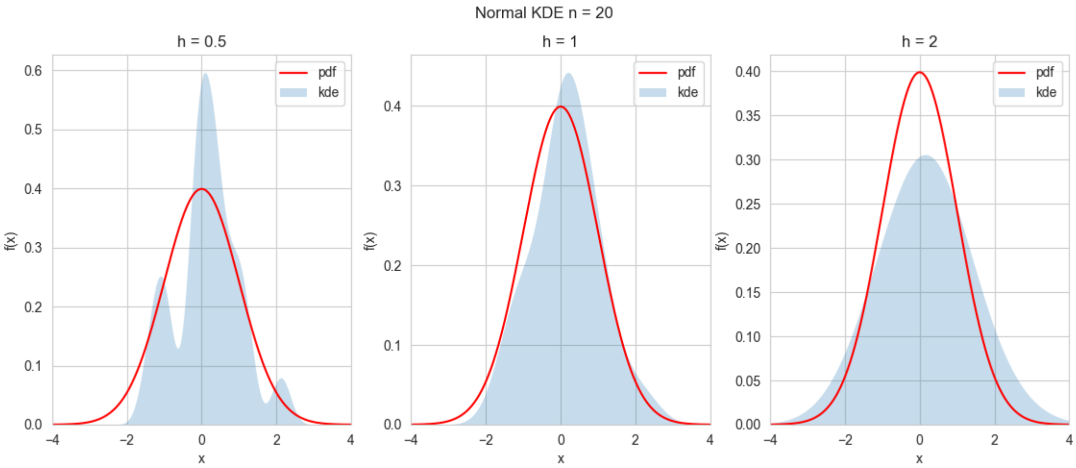


Рис. 6. Нормальное распределение размерностью 20

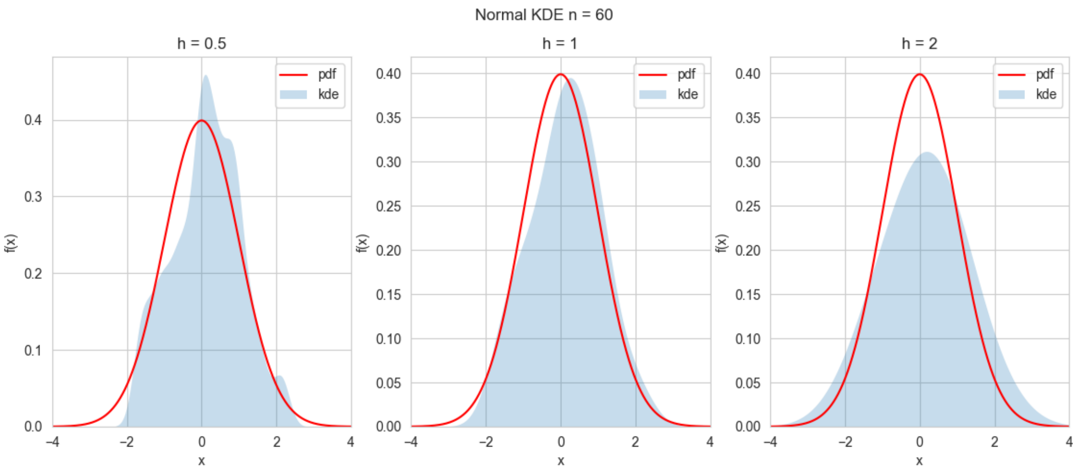


Рис. 7. Нормальное распределение размерностью 60

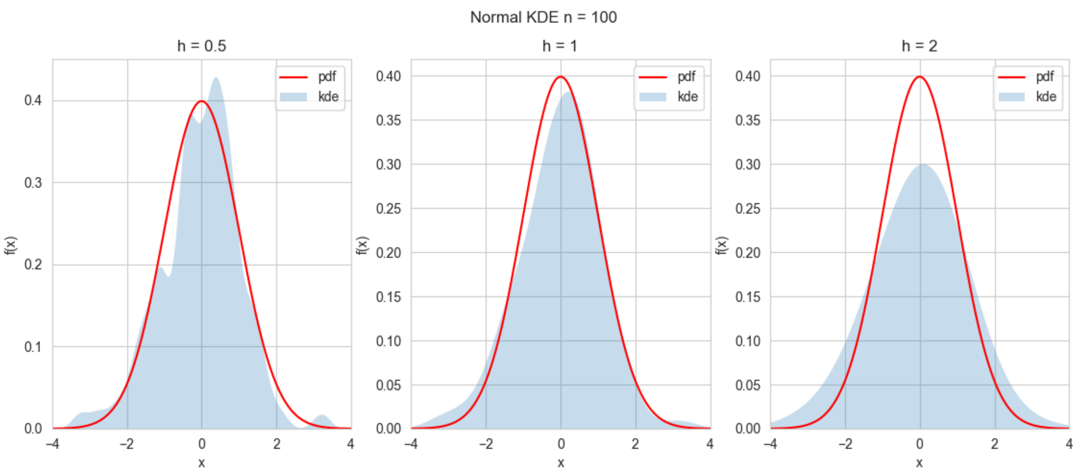


Рис. 8. Нормальное распределение размерностью 100

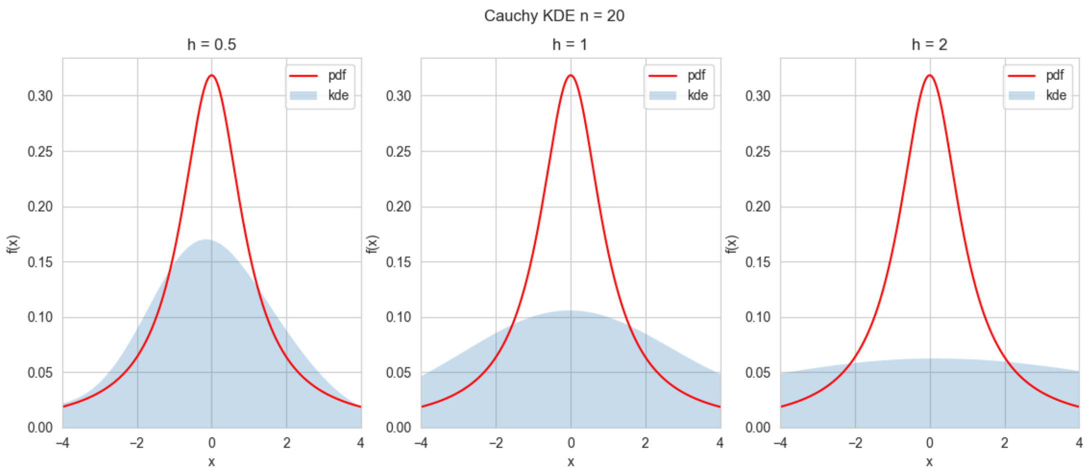


Рис. 9. Распределение Коши размерностью 20

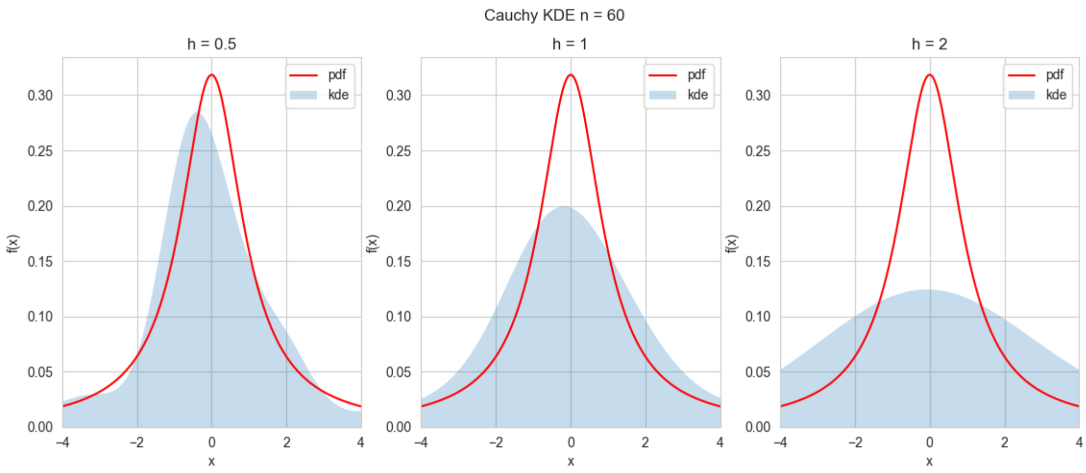


Рис. 10. Распределение Коши размерностью 60

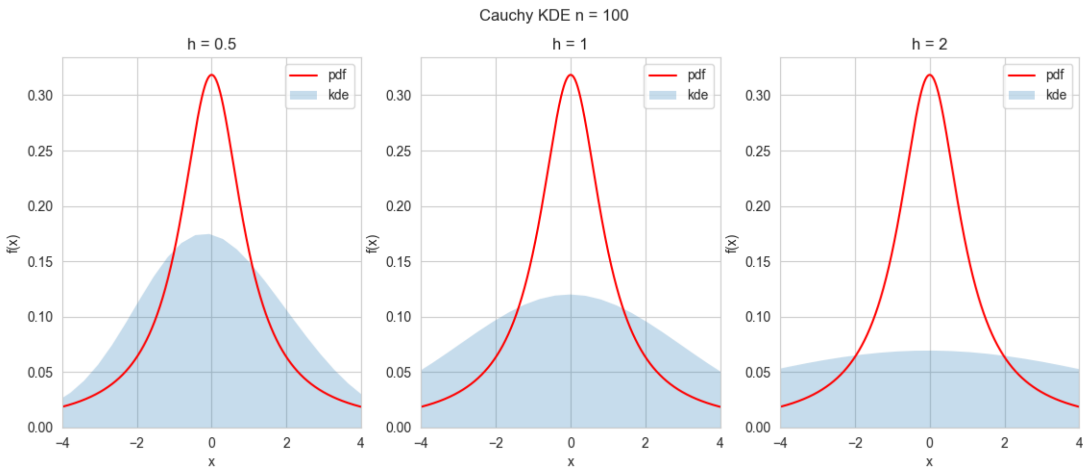


Рис. 11. Распределение Коши размерностью 100

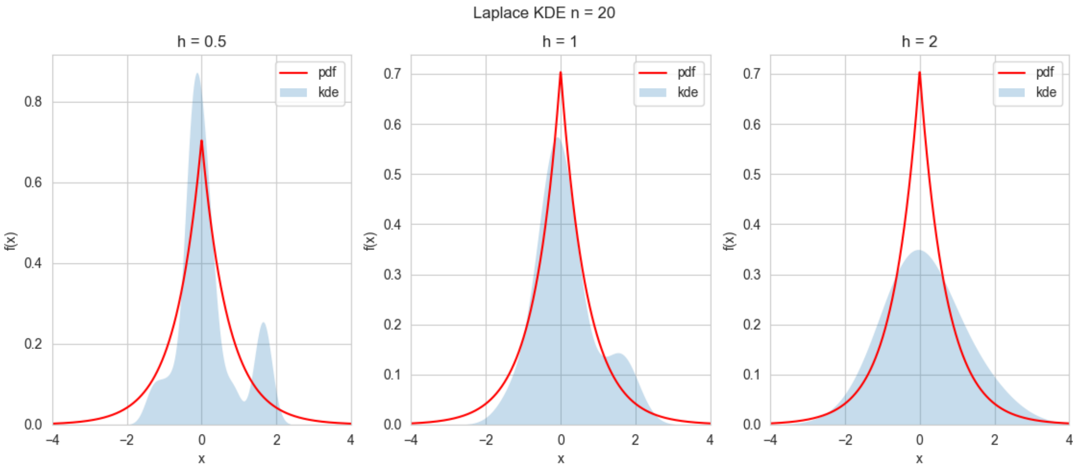


Рис. 12. Распределение Лапласа размерностью 20

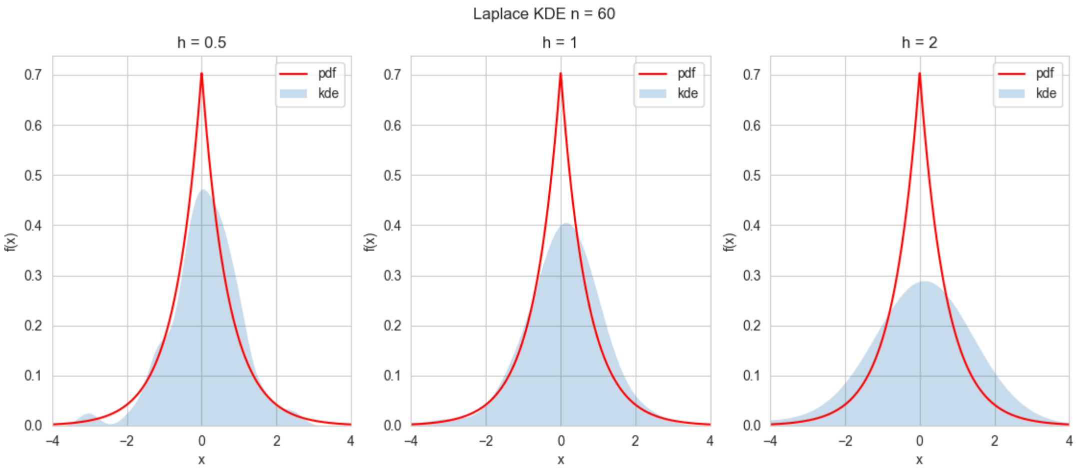


Рис. 13. Распределение Лапласа размерностью 60

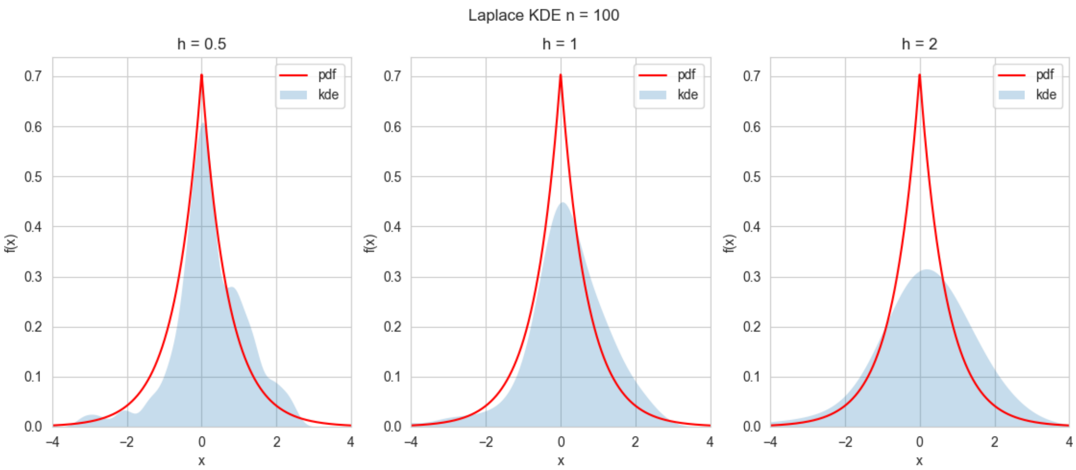


Рис. 14. Распределение Лапласа размерностью 100

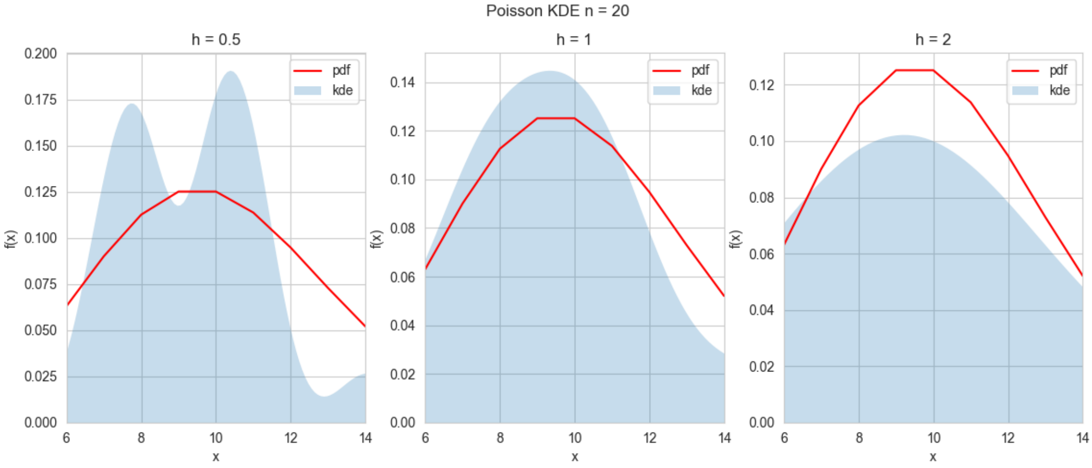


Рис. 15. Распределение Пуассона размерностью 20

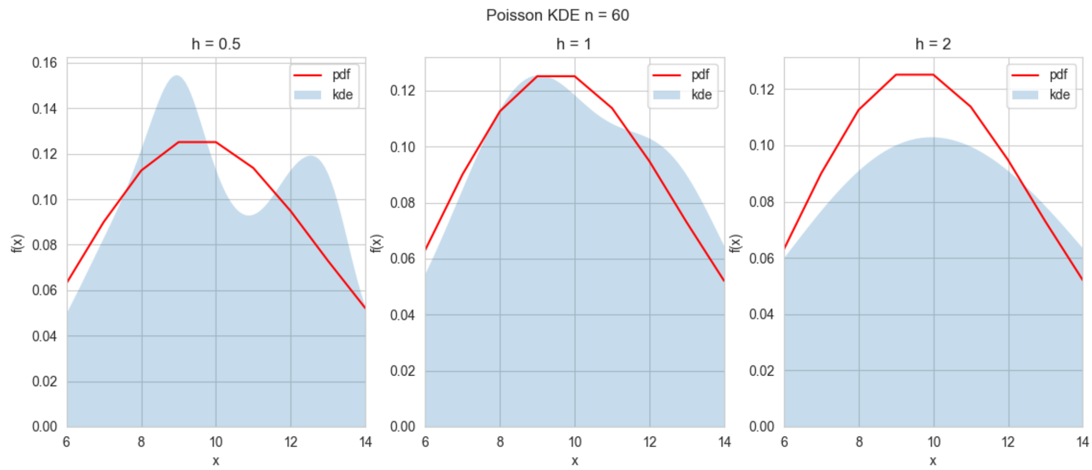


Рис. 16. Распределение Пуассона размерностью 60

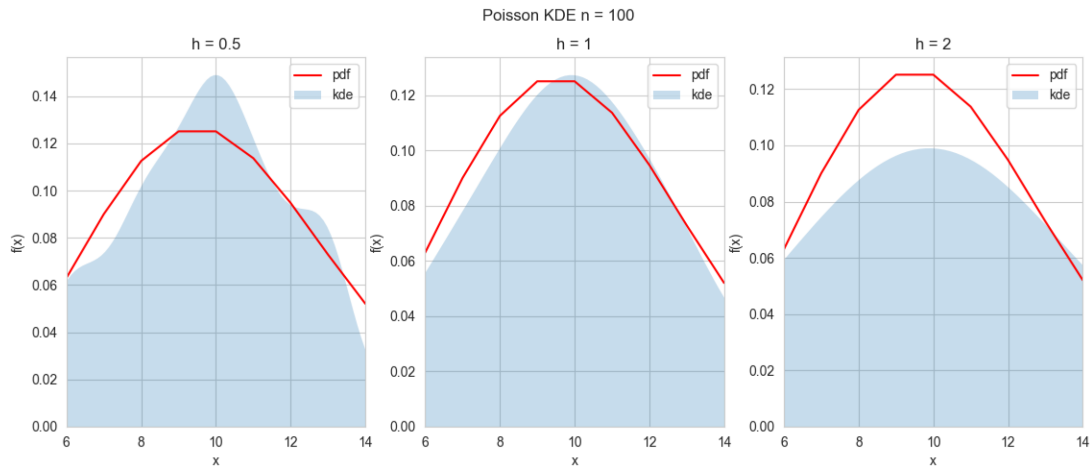


Рис. 17. Распределение Пуассона размерностью 100

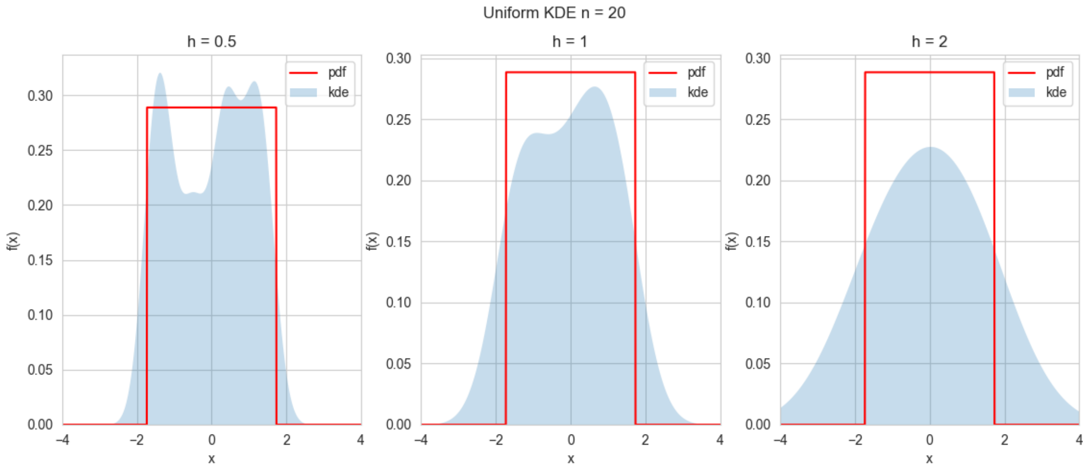


Рис. 18. Равномерное распределение размерностью 20

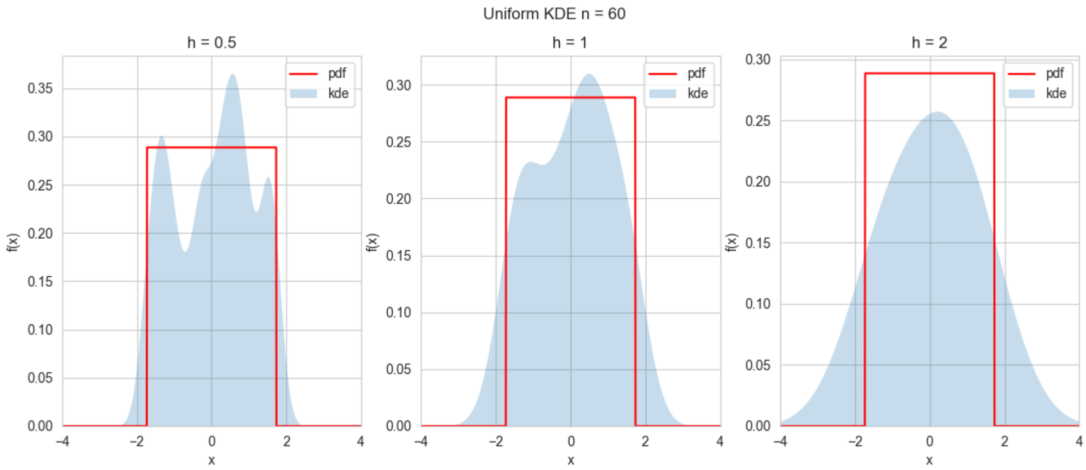


Рис. 19. Равномерное распределение размерностью 60

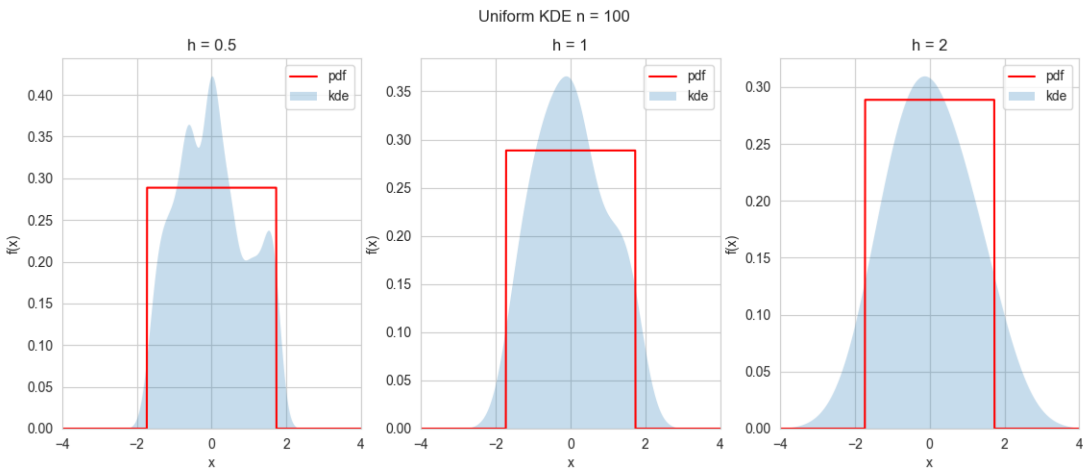


Рис. 20. Равномерное распределение размерностью 100

# 5. Обсуждение

При рассмотрении графиков эмпирических функций наблюдается следующая закономерность: чем больше размерность выборки, тем ступенчатая эмпирическая функция, построенная по ней, больше приближается к эталонной функции распределения данной величины. Стоит отметить, что для распределения Пуассона характерно наибольшее отклонение графика эмпирической функции распределения от эталонной.

Иллюстрации ядерных оценок плотностей распределения демонстрируют в большинстве случаев приближение ядерной оценки к функции плотности вероятности по всем $h$ с увеличением размерности выборки. Однако оптимальным значением для распределения Коши можно назвать размерность выборки, равной 60.

Для каждого из распределений оптимальным является свое значение параметра $h$. Параметр $h = h_n$ лучше всего приближает нормальное распределение, распределение Пуассона и равномерного распределение. Для распределения Коши и Лапласа оптимальным значением является $h = \frac{h_n}{2}$.

Также для распределения Коши и равномерного распределения при увеличении параметра $h$ характерно уменьшение значения пиковой точки и расширение основания графика ядерной оценки. А для всех распределений наблюдается преобразование графика ядерной оценки в функцию с единственным максимумом при значении параметра $h = 2h_n$.