

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №4

по дисциплине
"Математическая статистика"

Выполнил студент
Группы 3630102/80101

Шао Цзяци

Проверил
доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург
2021 г.

Содержание

1. Постановка задачи	4
2. Теория	4
2.1. Рассматриваемые распределения	4
2.2. Эмпирическая функция распределение	4
2.3. Ядерные оценки плотности вероятности	5
3. Реализация	5
4. Результаты	6
4.1. Эмпирическая функция распределения	6
4.2. Ядерные оценки плотности распределения	8
5. Обсуждение	15

Список иллюстраций

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

1. Постановка задачи

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. Рассматриваемые распределения

Плотности распределений

- Нормальное распределение

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

- Распределение Коши

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

- Распределение Лапласа

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

- Распределение Пуассона

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

- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| \leq \sqrt{3} \\ 0, & |x| \geq \sqrt{3} \end{cases} \quad (5)$$

2.2. Эмпирическая функция распределение

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.

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

$$F_n^*(x) = P^*(X < x) \quad (6)$$

Его можно найти по формуле

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n 1_{x_i < x} \quad (7)$$

где, 1 - это индикаторная функция. Эмпирическая функция распределения является оценкой, т. е. приближённым значением, генеральной функции распределения $F_n^*(x) \approx F_X(x)$.

2.3. Ядерные оценки плотности вероятности

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

Непрерывная ядерная оценка задается формулой:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) \quad (8)$$

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

$$h_n \xrightarrow{n \rightarrow \infty} 0; nh_n \xrightarrow{n \rightarrow \infty} \infty. \quad (9)$$

Используется Гауссово ядро:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (10)$$

А также правило Сильвермана:

$$h_n = 1.06\hat{\sigma}n^{-1/5} \quad (11)$$

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

3. Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования python в среде разработки Pycharm с дополнительными библиотеками.

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

Исходный код лабораторной работы размещен в Github-репозитории.

URL: <https://github.com/ShaoTs/shaoMathStatistic/tree/master/Lab4>

4. Результаты

4.1. Эмпирическая функция распределения

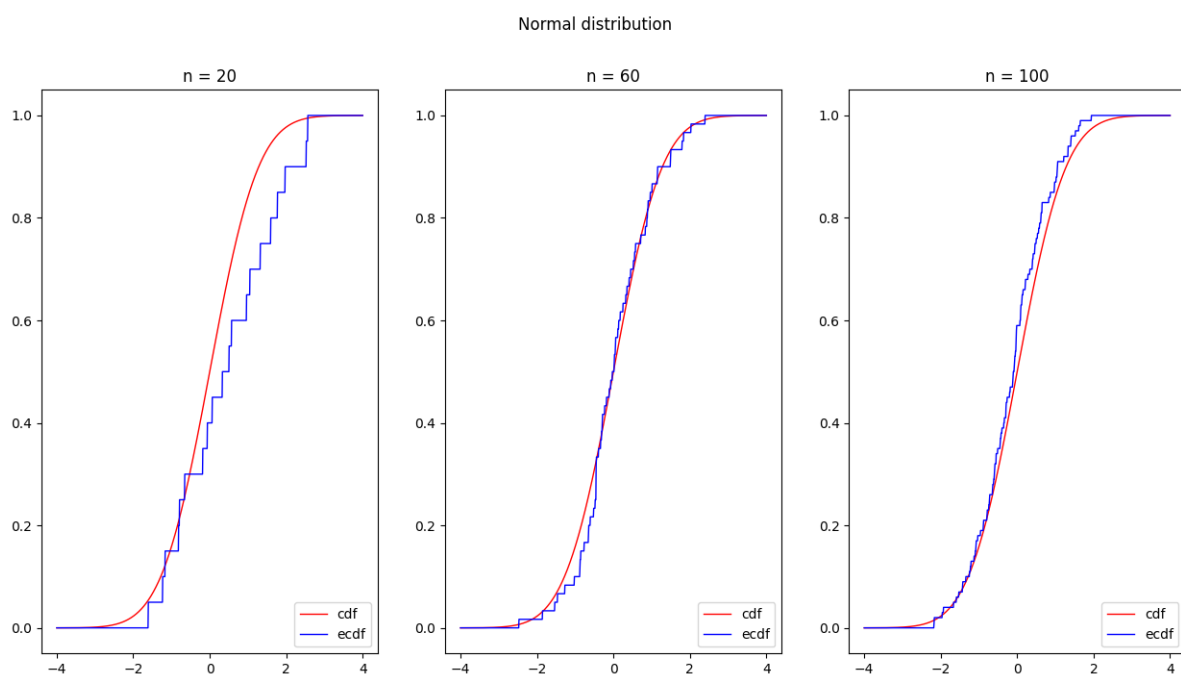


Рис. 1. Нормальное распределение

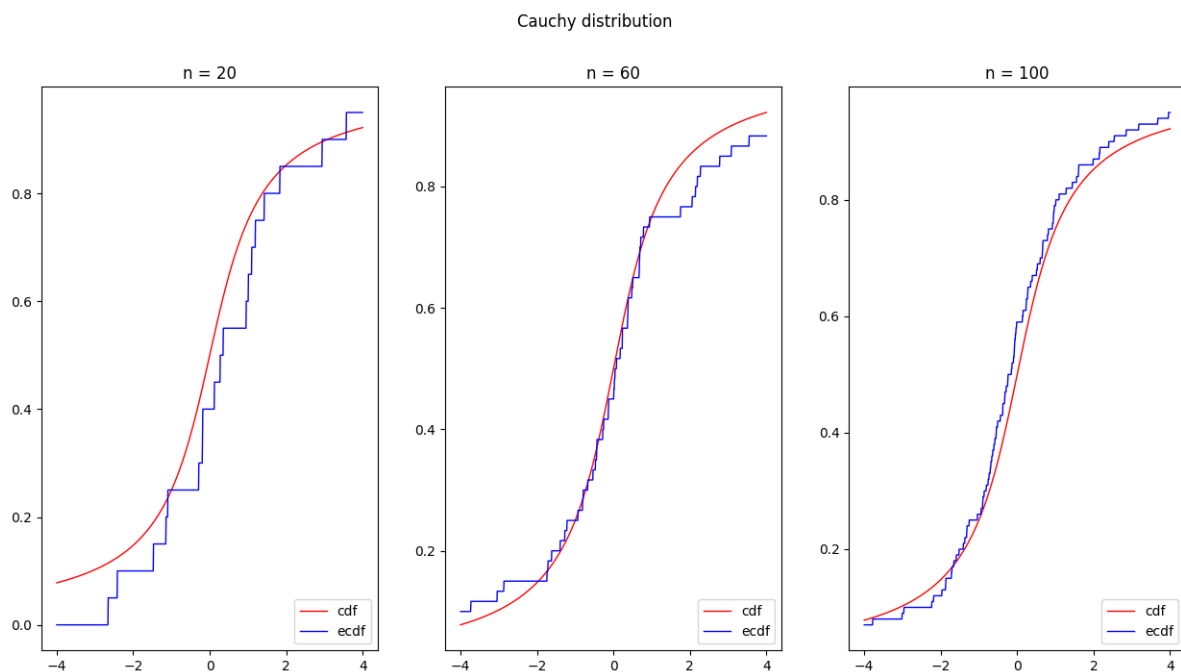


Рис. 2. Нормальное распределение

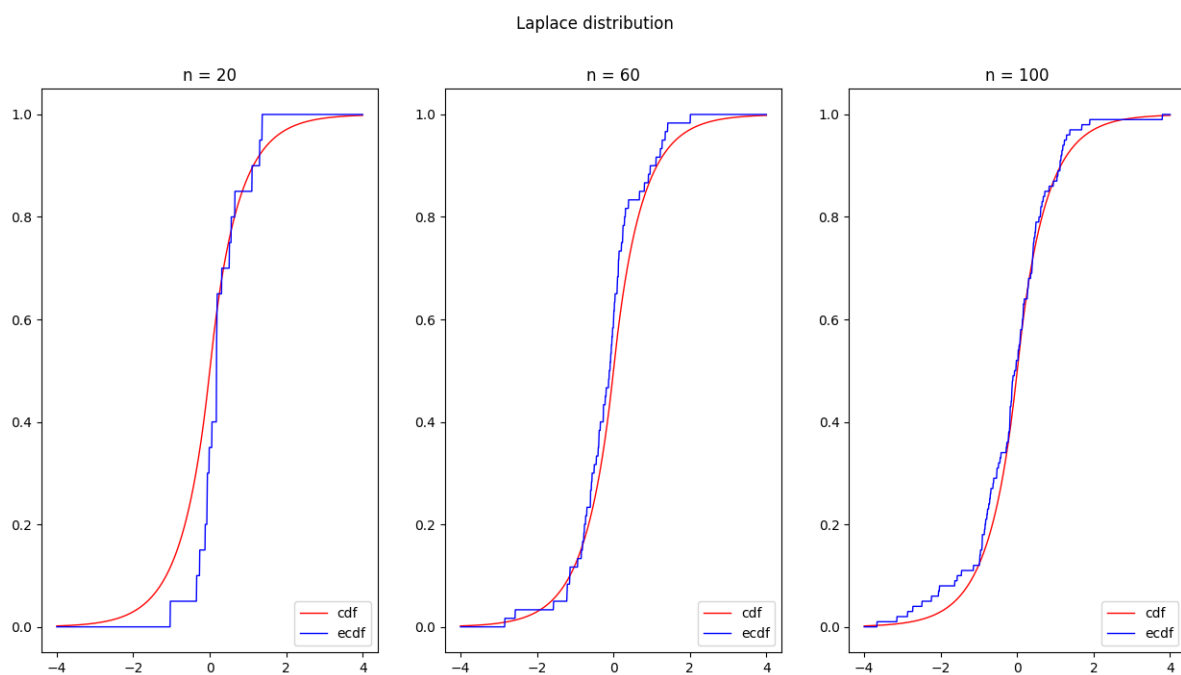


Рис. 3. Нормальное распределение

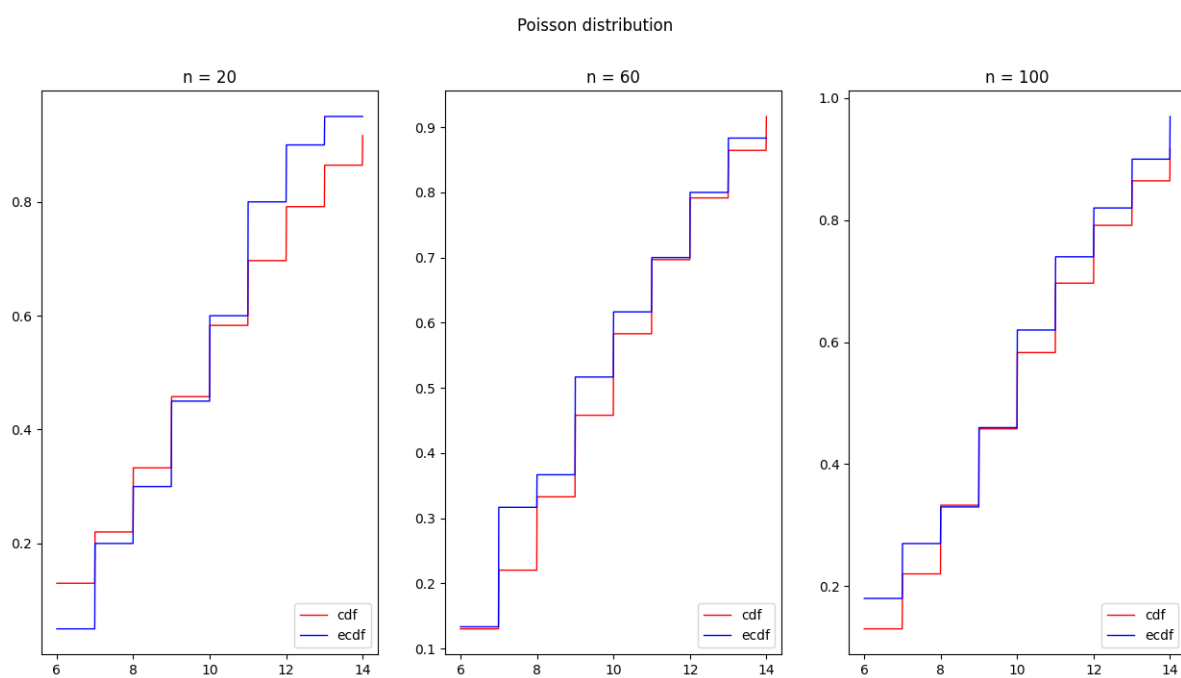


Рис. 4. Нормальное распределение

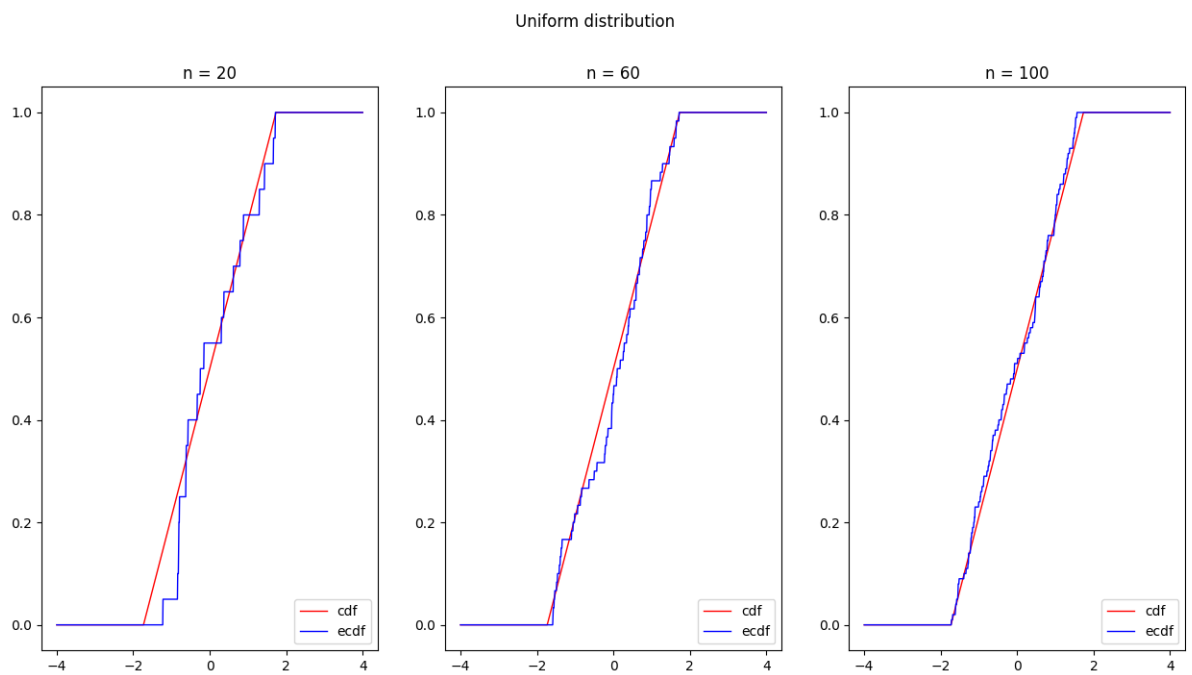


Рис. 5. Нормальное распределение

4.2. Ядерные оценки плотности распределения

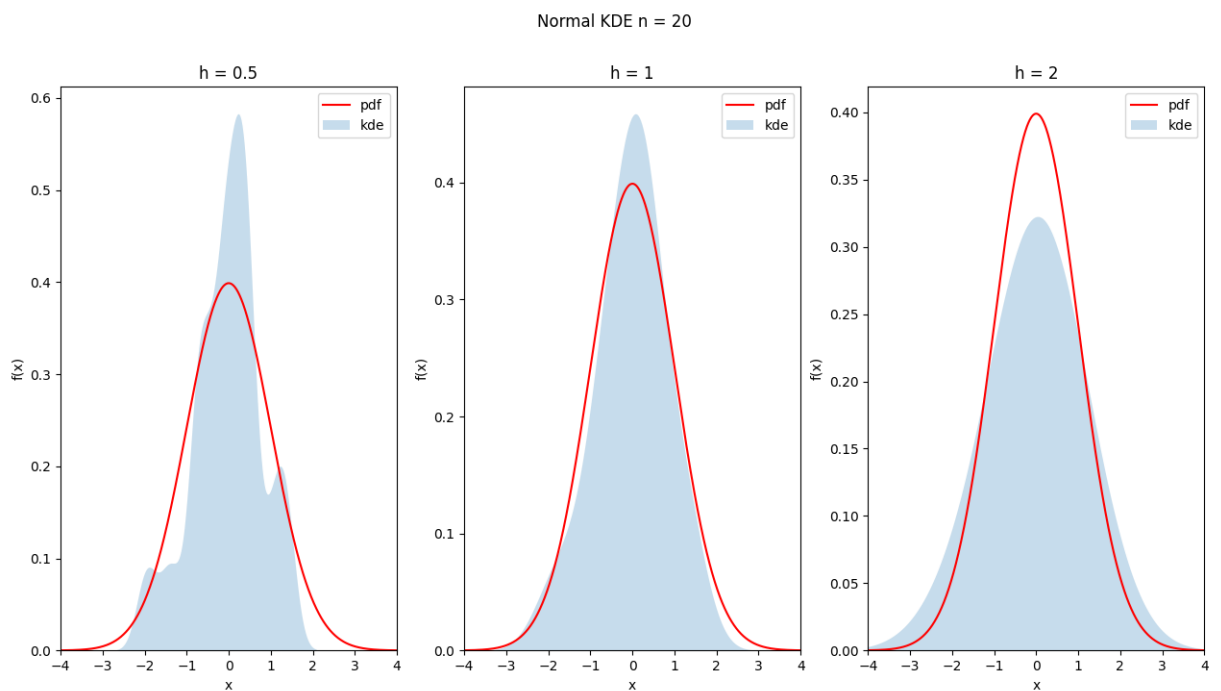


Рис. 6. Нормальное распределение $n = 20$

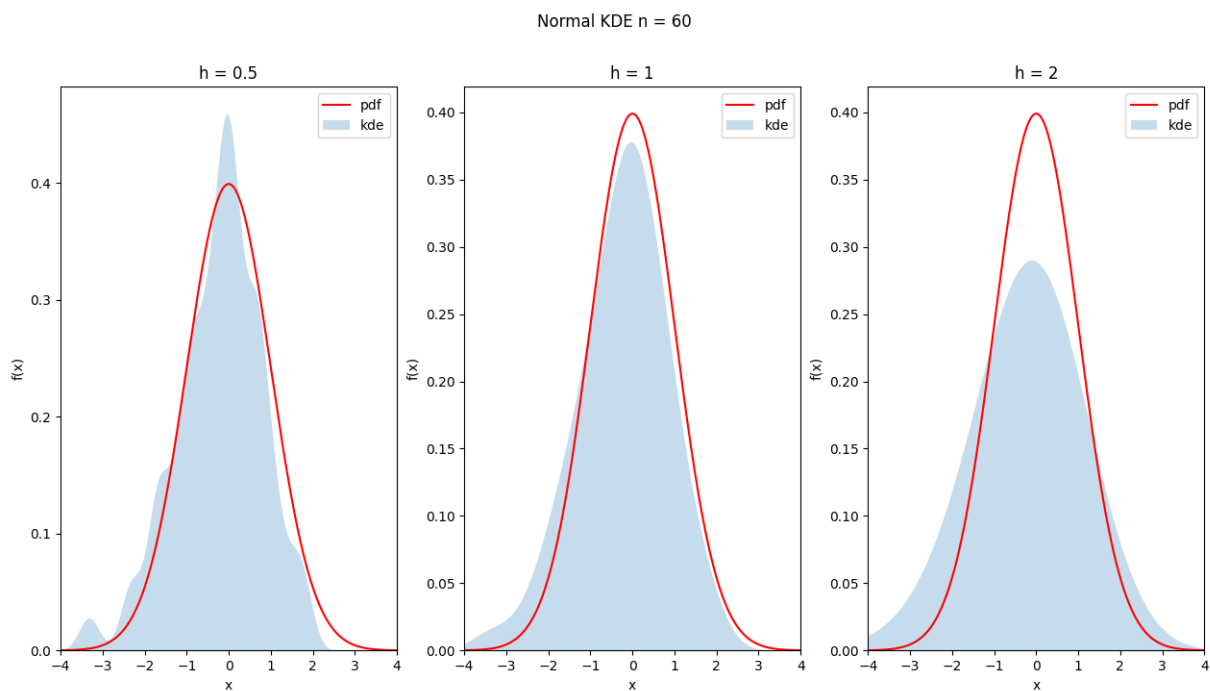


Рис. 7. Нормальное распределение $n = 60$

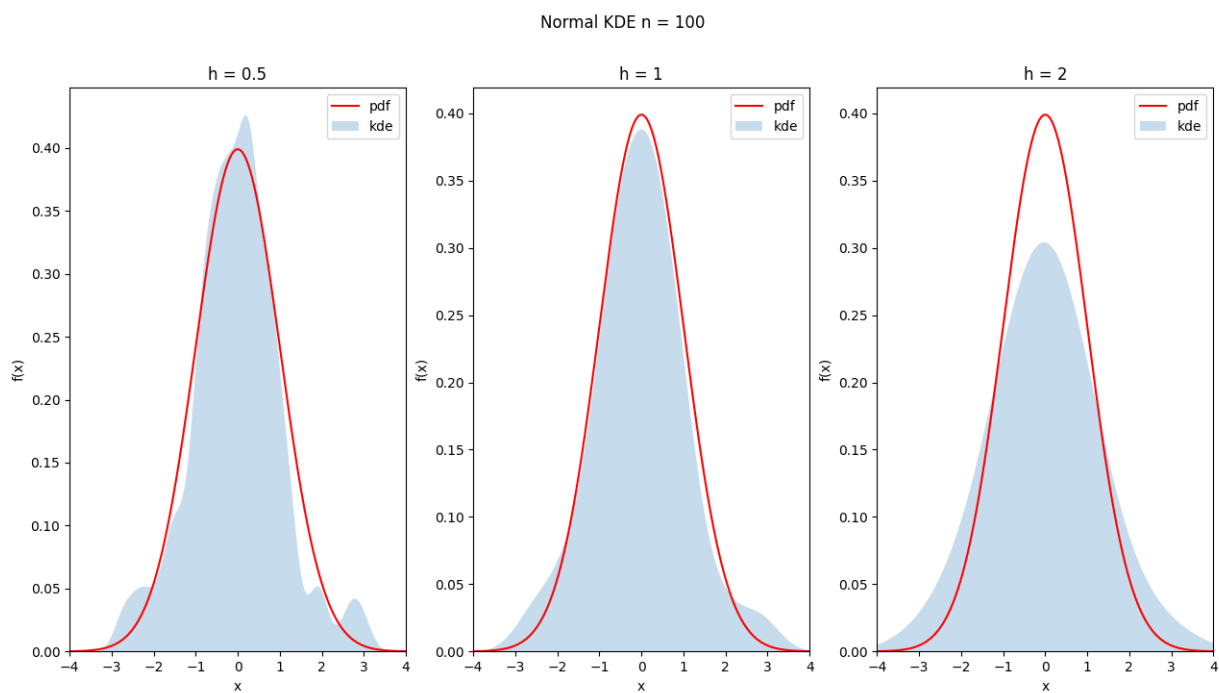


Рис. 8. Нормальное распределение $n = 60$

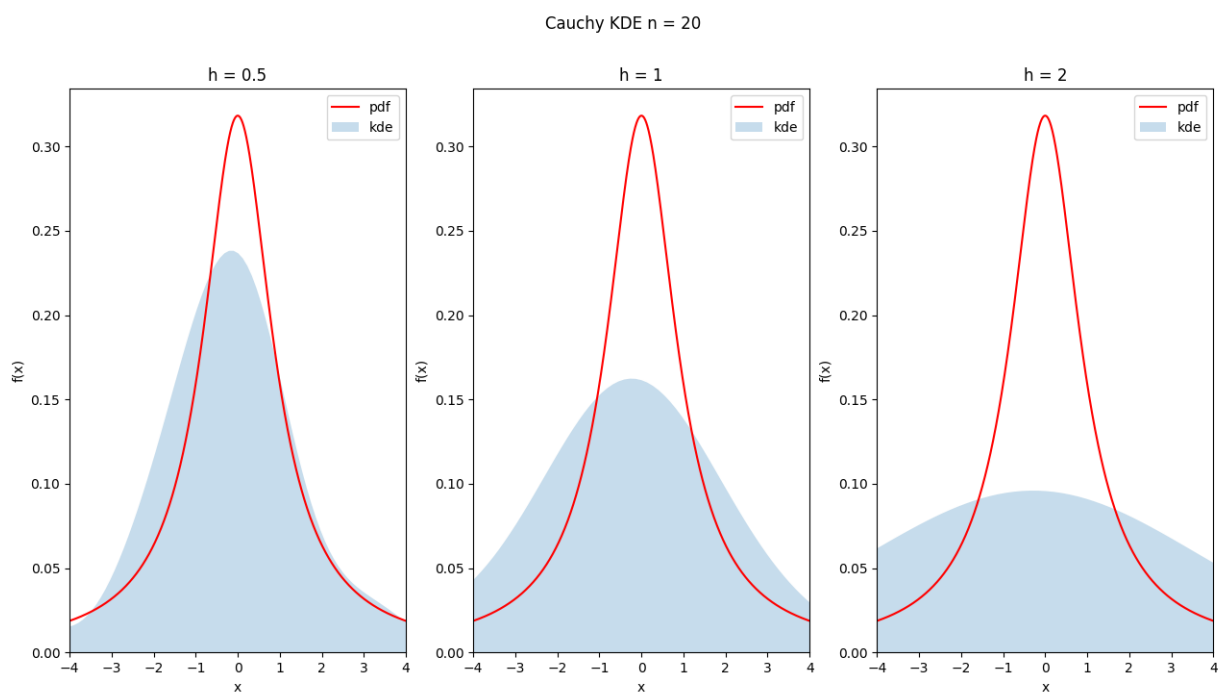


Рис. 9. Распределение Коши $n = 20$

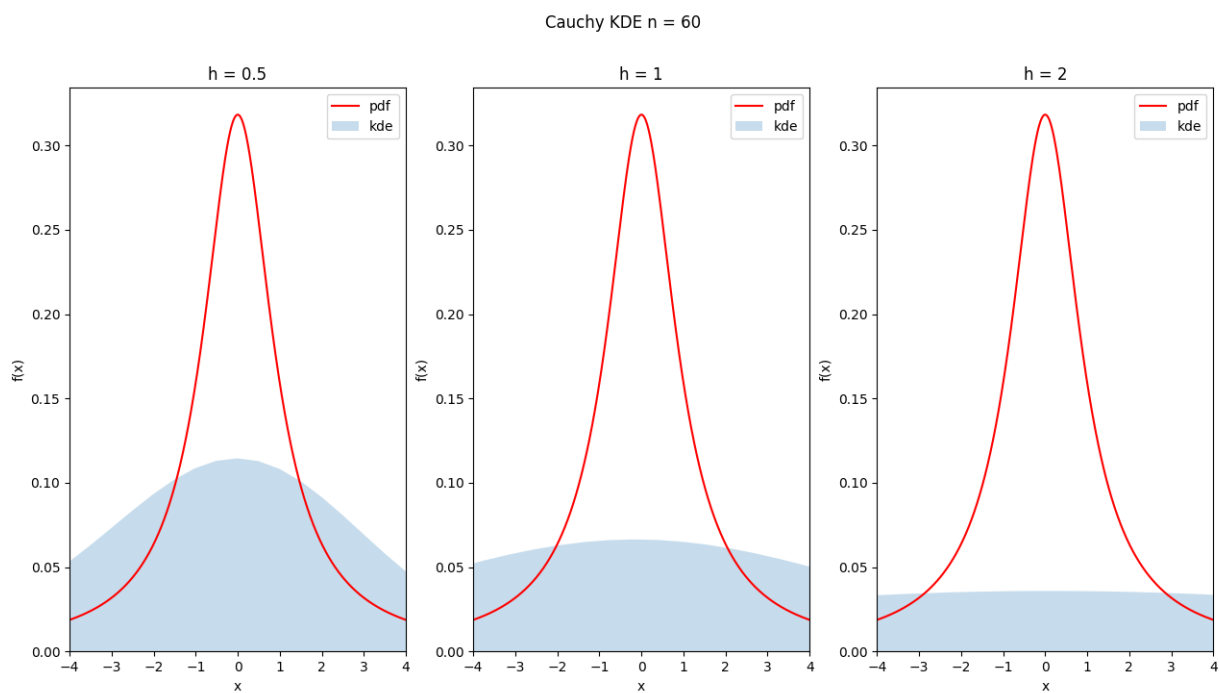


Рис. 10. Распределение Коши $n = 60$

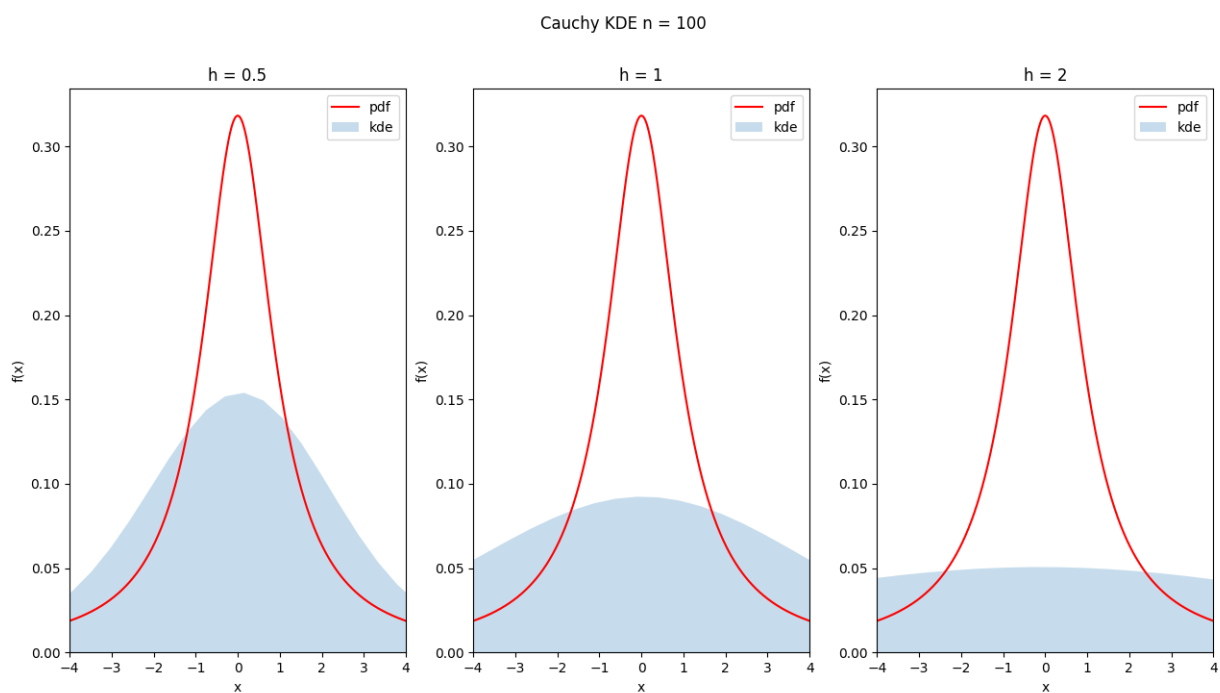


Рис. 11. Распределение Коши $n = 100$

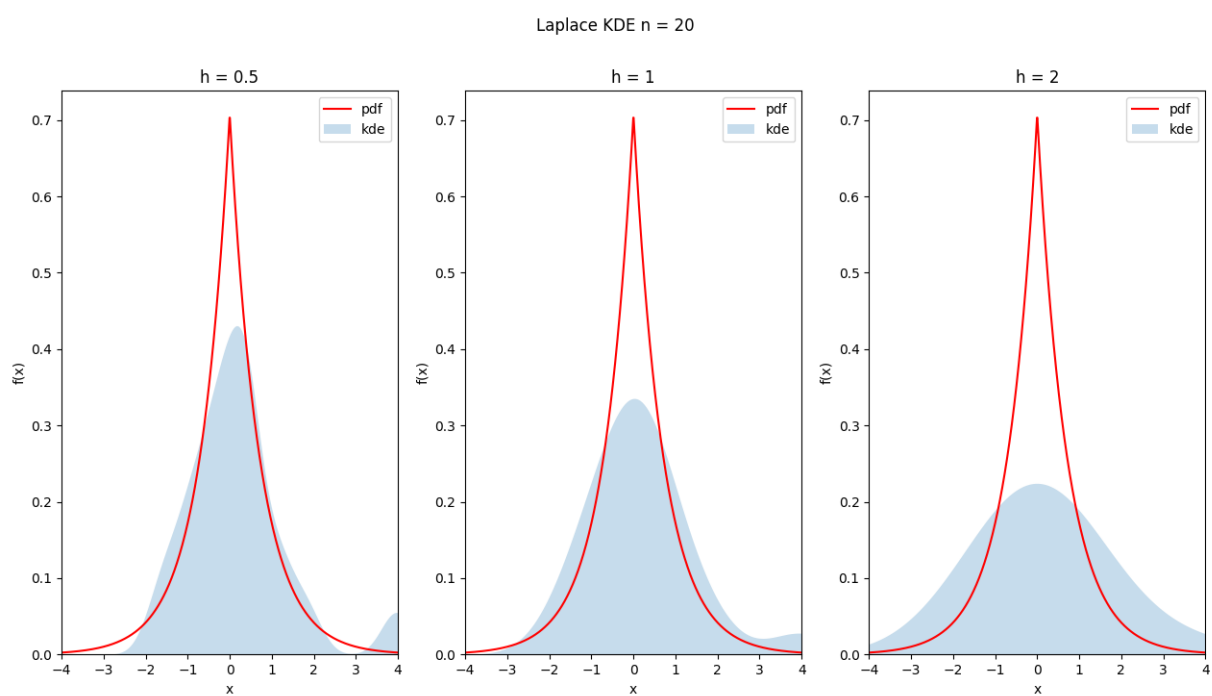


Рис. 12. Распределение Лапласа $n = 20$

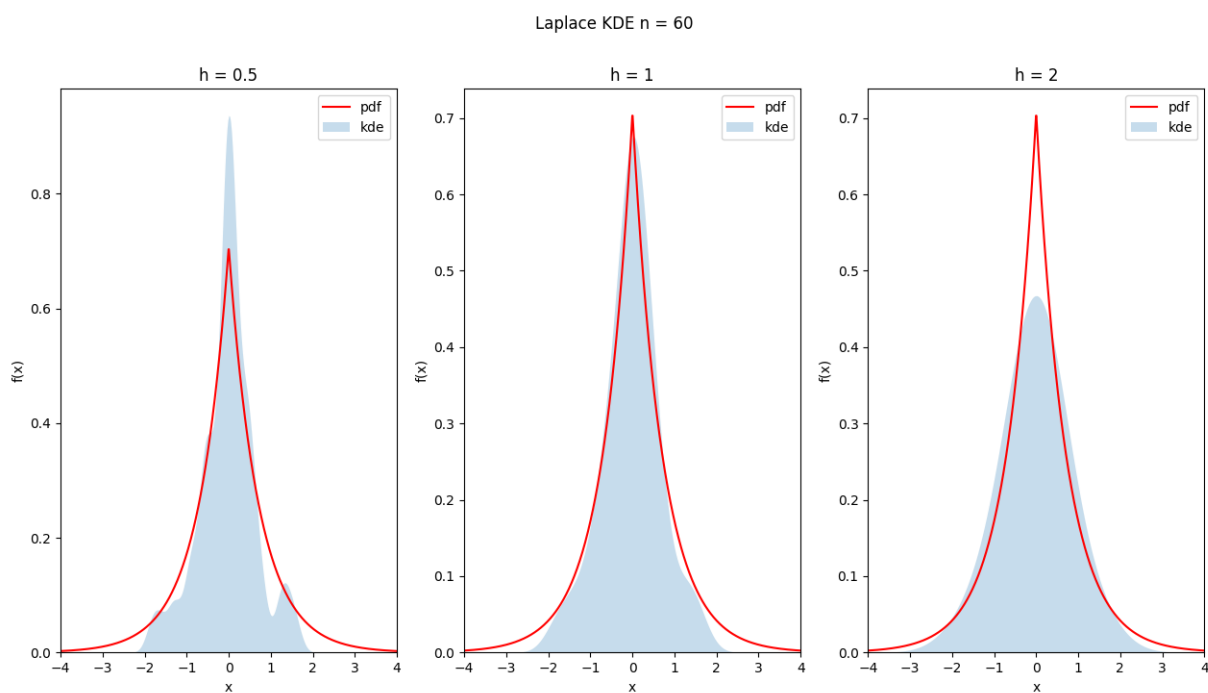


Рис. 13. Распределение Лапласа $n = 60$

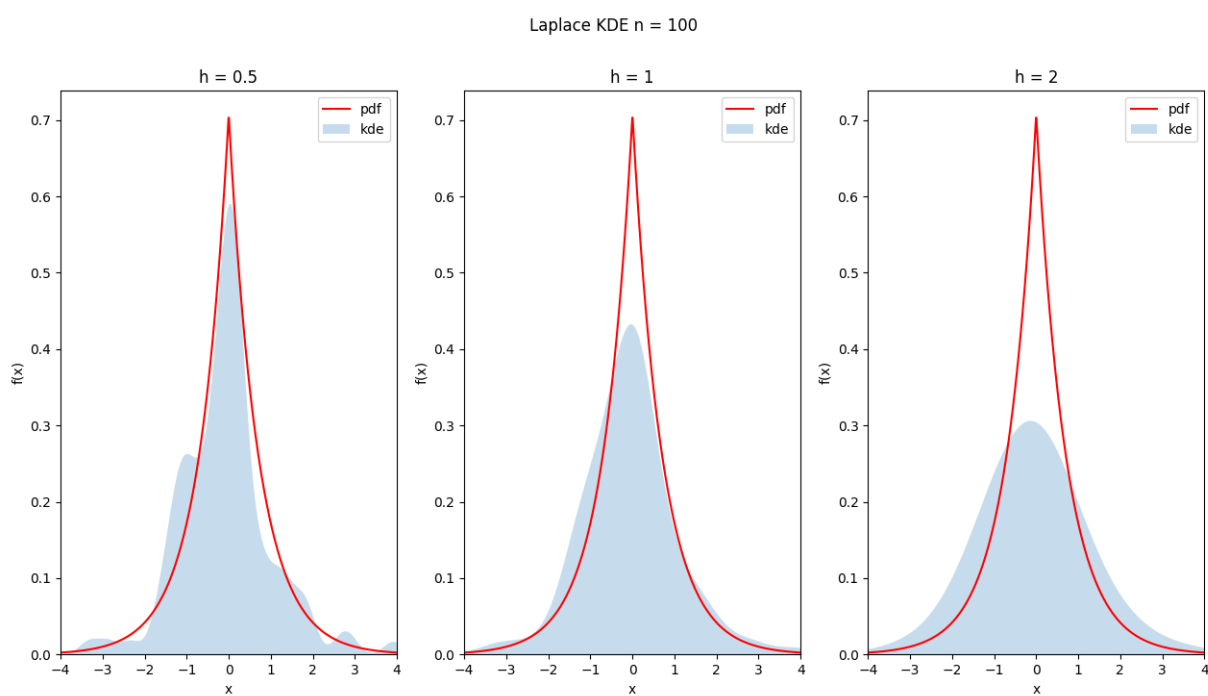


Рис. 14. Распределение Лапласа $n = 100$

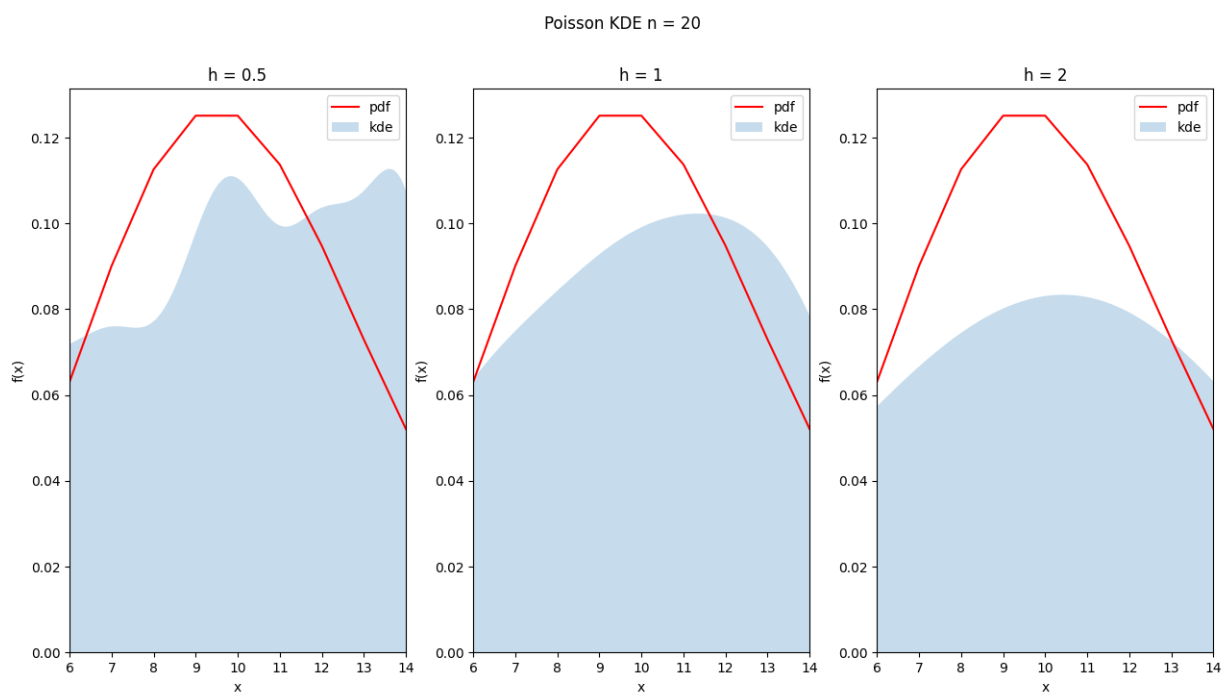


Рис. 15. Распределение Пуассона $n = 20$

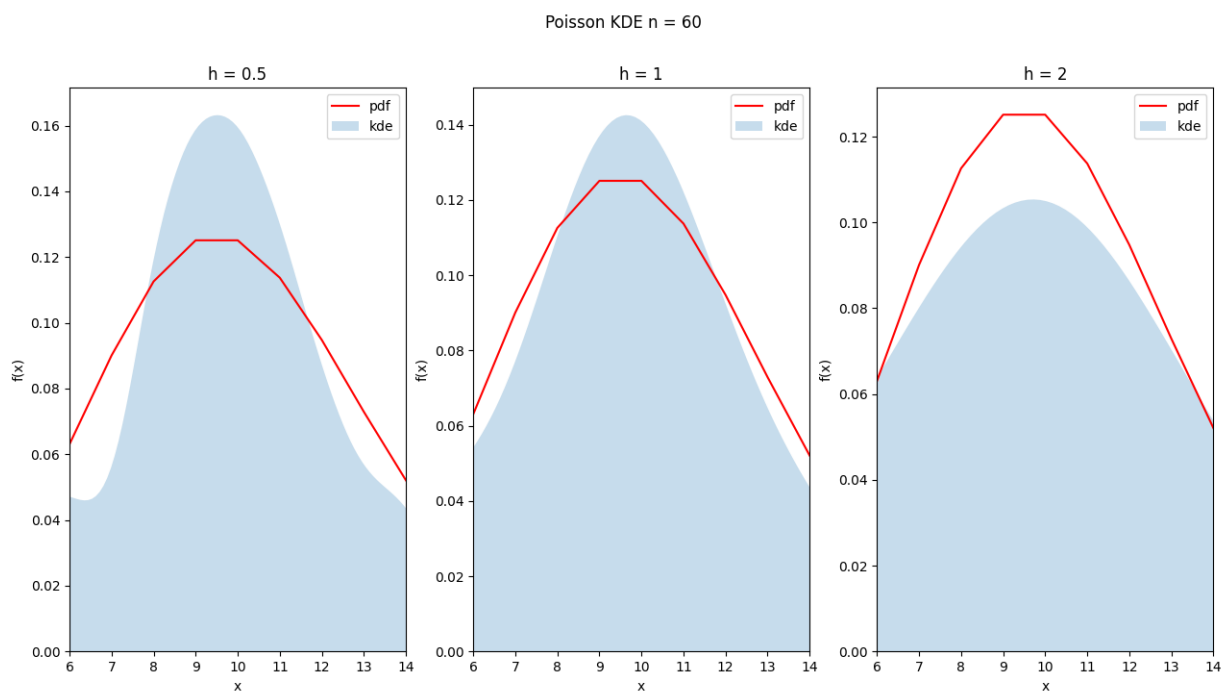


Рис. 16. Распределение Пуассона $n = 60$

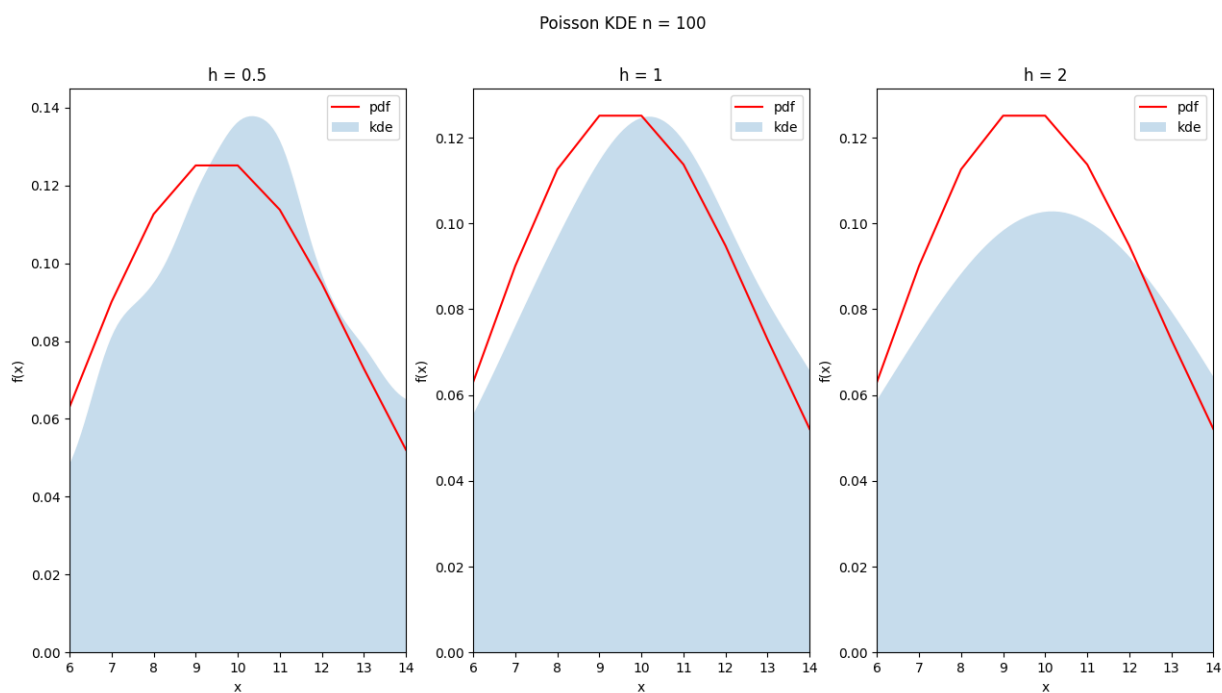


Рис. 17. Распределение Пуассона $n = 100$

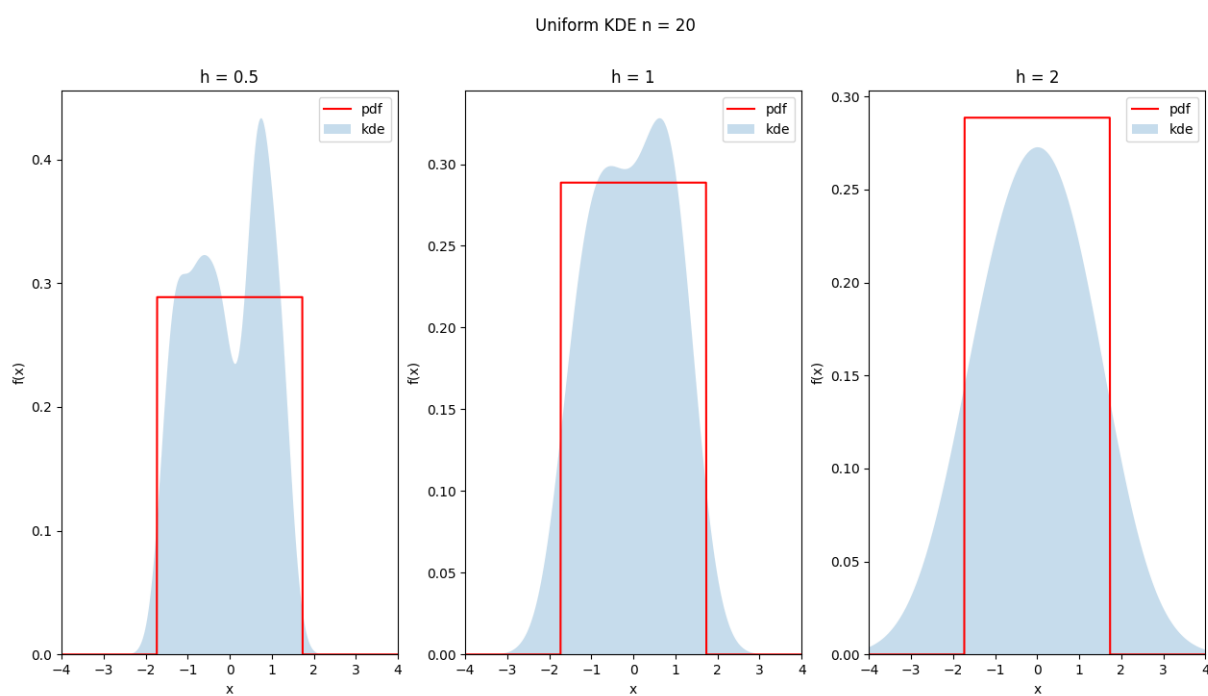


Рис. 18. Равномерное распределение $n = 20$

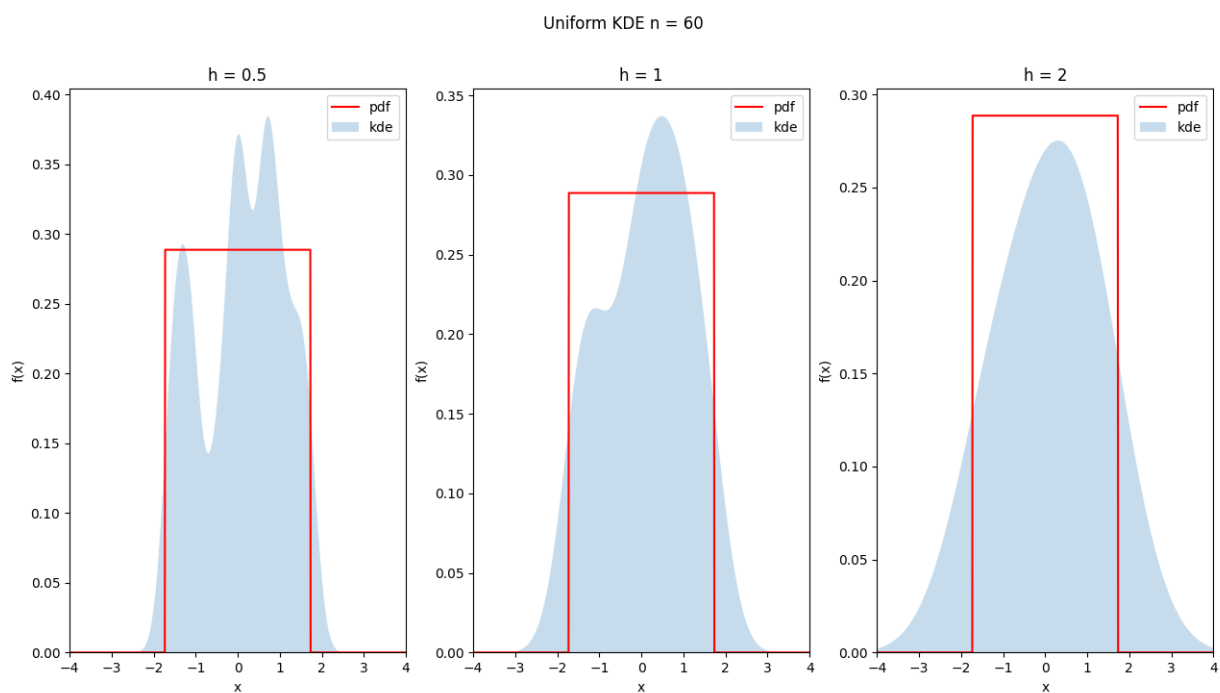


Рис. 19. Равномерное распределение $n = 60$

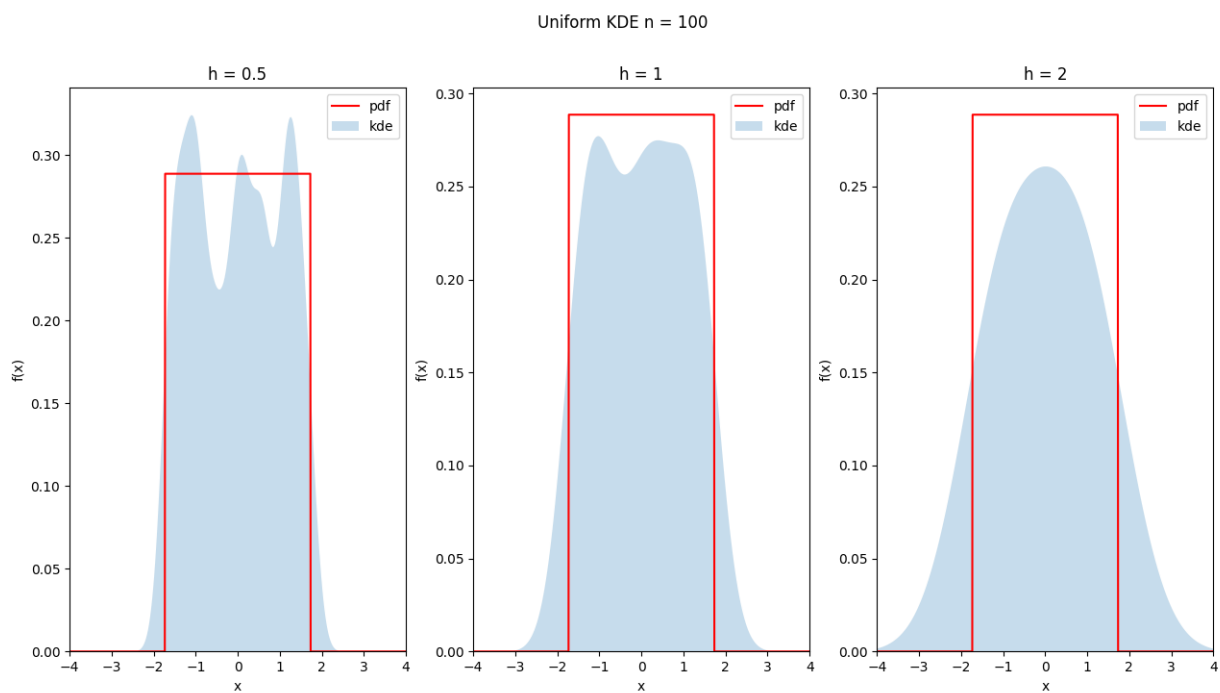


Рис. 20. Равномерное распределение $n = 100$

5. Обсуждение

Из иллюстрации с эмпирическими функциями вытекает, что чем больше выборка, ступенчатая эмпирическая функция распределения тем лучше приближает функцию распределения

реальной выборки. Так и согласится с теоремой Гливенко — Кантелли, что $\lim_{n \rightarrow \infty} \sup_{x \in R} |F_n^*(x) - F(x)| = 0$ почти наверное.

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.

Каждое распределение имеет свой самый подходящий параметр h для описания ядерной оценки: для нормального, равномерного и пуассоновского распределений оптимальным значением параметра является $h = h_n$, а для распределений Коши и Лапласа - $h = h_n/2$.

Также можно увидеть, что чем больше коэффициент при параметре сглаживания h_n тем меньше изменений знака производной у аппроксимирующей функции, вплоть до того, что при $h = 2h_n$ функция становится унимодальной на рассматриваемом промежутке. Также видно, что при $h = 2h_n$ по полученным приближениям становится сложно сказать плотность вероятности какого распределения они должны повторять, так как они очень похожи между собой.