

Санкт-Петербургский политехнический университет Петра Великого  
Институт прикладной математики и механики  
**Высшая школа прикладной математики и вычислительной физики**

## **ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №4**

по дисциплине  
«Математическая статистика»

Выполнила студентка  
группы 3630102/80401

Мамаева Анастасия Сергеевна

Проверил  
Доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург  
2021

# СОДЕРЖАНИЕ

<b>СПИСОК ИЛЛЮСТРАЦИЙ . . . . .</b>	<b>3</b>
<b>1 Постановка задачи . . . . .</b>	<b>4</b>
<b>2 Теория . . . . .</b>	<b>4</b>
2.1 Эмпирическая функция распределения . . . . .	4
2.1.1 Статистический ряд . . . . .	4
2.1.2 Эмпирическая функция распределения . . . . .	4
2.1.3 Нахождение э. ф. р. . . . .	4
2.2 Оценки плотности вероятности . . . . .	5
2.2.1 Определение . . . . .	5
2.2.2 Ядерные оценки . . . . .	5
<b>3 Программная реализация . . . . .</b>	<b>5</b>
<b>4 Результаты . . . . .</b>	<b>6</b>
4.1 Эмпирическая функция распределения . . . . .	6
4.2 Ядерные оценки плотности распределения . . . . .	8
<b>5 Обсуждение . . . . .</b>	<b>15</b>
<b>6 Приложение . . . . .</b>	<b>16</b>

# СПИСОК ИЛЛЮСТРАЦИЙ

1	нормальное распределение . . . . .	6
2	распределение Коши . . . . .	6
3	распределение Лапласа . . . . .	7
4	распределение Пуассона . . . . .	7
5	равномерное распределение . . . . .	8
6	нормальное распределение, $n = 20$ . . . . .	8
7	нормальное распределение, $n = 60$ . . . . .	9
8	нормальное распределение, $n = 100$ . . . . .	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 Эмпирическая функция распределения

#### 2.1.1 Статистический ряд

Статистическим рядом назовем совокупность, состоящую из последовательности  $\{z_i\}_{i=1}^k$  попарно различных элементов выборки, расположенных по возрастанию, и последовательности  $\{n_i\}_{i=1}^k$  частот, с которыми эти элементы содержатся в выборке.

#### 2.1.2 Эмпирическая функция распределения

Эмпирическая функция распределения (э. ф. р.) - относительная частота события  $X < x$ , полученная по данной выборке:

$$F_n^*(x) = P^*(X < x). \quad (1)$$

#### 2.1.3 Нахождение э. ф. р.

$$F_n^*(x) = \frac{1}{n} \sum_{z_i < x} n_i. \quad (2)$$

$F_n^*(x)$  — функция распределения дискретной случайной величины  $X^*$ , заданной таблицей распределения

$X^*$	$z_1$	$z_2$	...	$z_k$
$P$	$n_1/n$	$n_2/n$	...	$n_k/n$

Таблица 1: Таблица распределения

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

$$F_n^*(x) \approx F_X(x). \quad (3)$$

## 2.2 Оценки плотности вероятности

### 2.2.1 Определение

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

$$\hat{f}(x) \approx f(x). \quad (4)$$

### 2.2.2 Ядерные оценки

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

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right). \quad (5)$$

$K(u)$  - ядро, т. е. непрерывная функция, являющаяся плотностью вероятности,  $x_1, \dots, x_n$  - элементы выборки, а  $\{h_n\}_{n \in \mathbb{N}}$  - последовательность элементов из  $\mathbb{R}_+$  такая, что

$$h_n \xrightarrow{n \rightarrow \infty} 0; \quad nh_n \xrightarrow{n \rightarrow \infty} \infty. \quad (6)$$

Такие оценки называются непрерывными ядерными.

Гауссово ядро:

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

Правило Сильвермана:

$$h_n = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5} \approx 1.06\hat{\sigma}n^{-1/5}, \quad (8)$$

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

## 3 Программная реализация

Лабораторная работа выполнена на языке Python версии 3.7 в среде разработки JupyterLab. Использовались дополнительные библиотеки:

1. `scipy` (генерация выборок)
2. `statsmodels` (построение эмпирических функций распределения)
3. `matplotlib`, `seaborn` (визуализация)
4. `numpy` (вычисление ряда числовых характеристик)

В приложении находится ссылка на GitHub репозиторий с исходным кодом.

## 4 Результаты

### 4.1 Эмпирическая функция распределения

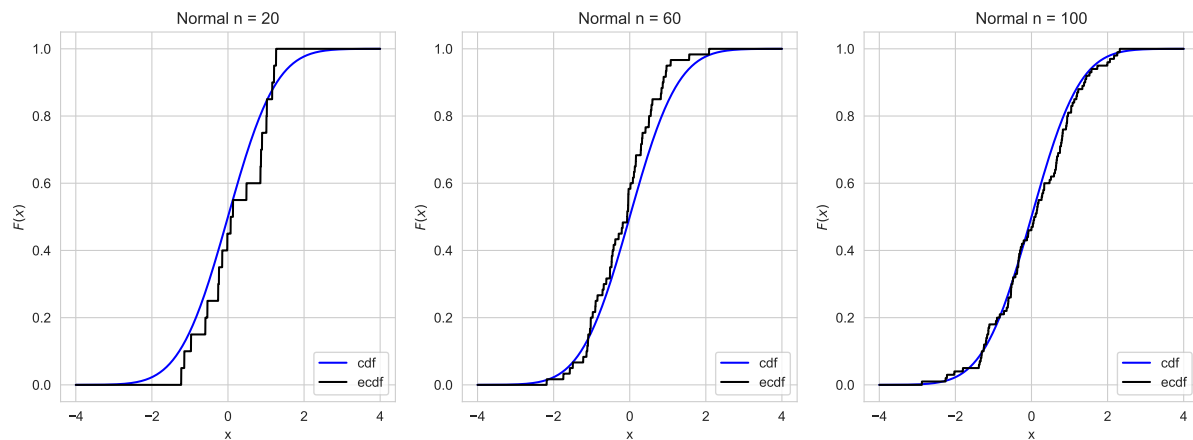


Рис. 1: нормальное распределение

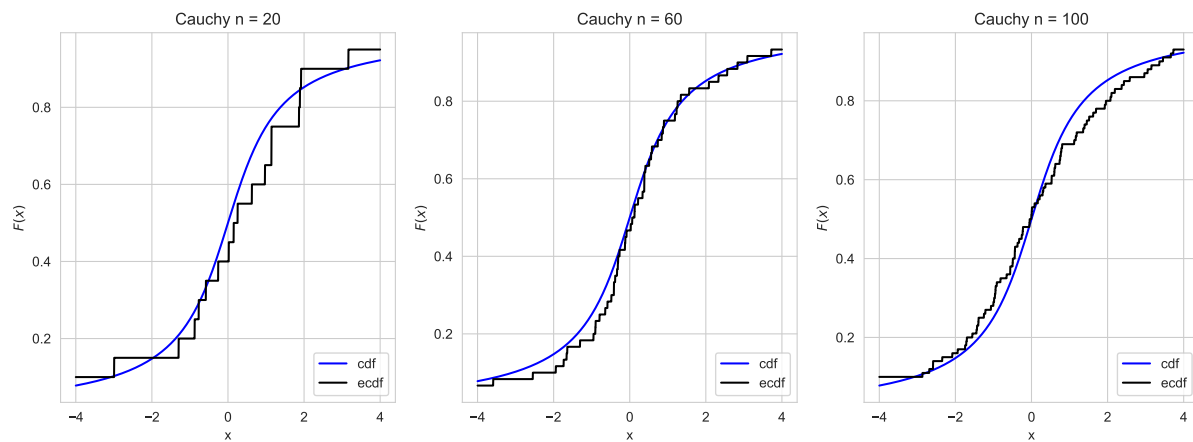


Рис. 2: распределение Коши

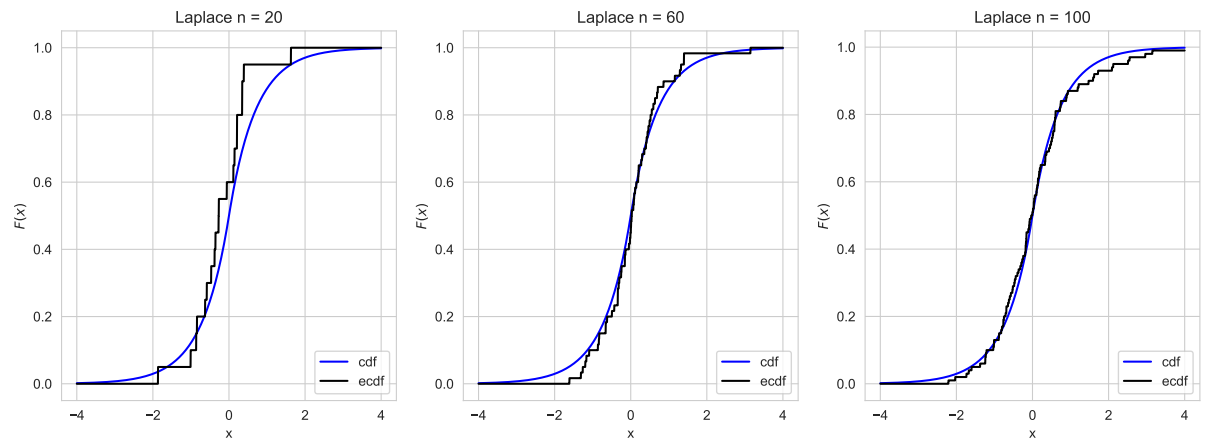


Рис. 3: распределение Лапласа

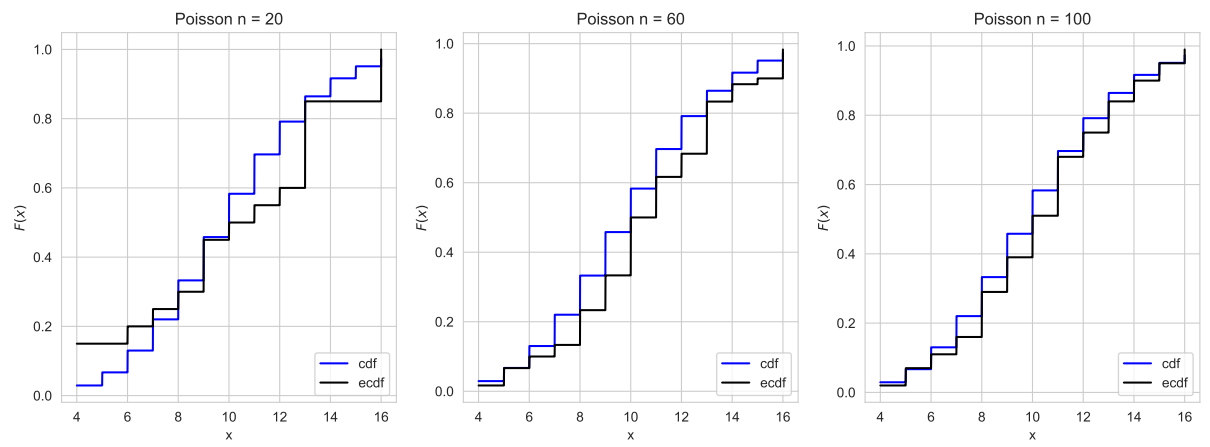


Рис. 4: распределение Пуассона

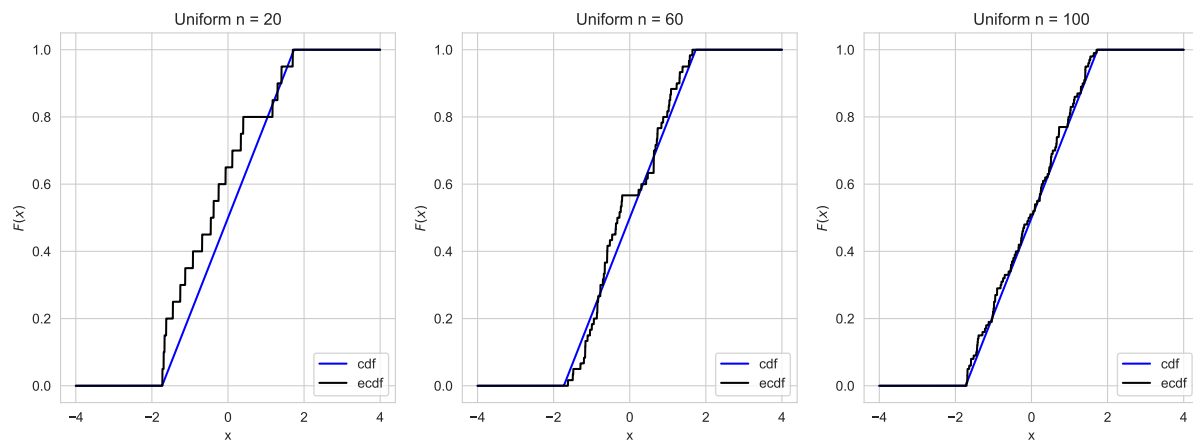


Рис. 5: равномерное распределение

## 4.2 Ядерные оценки плотности распределения

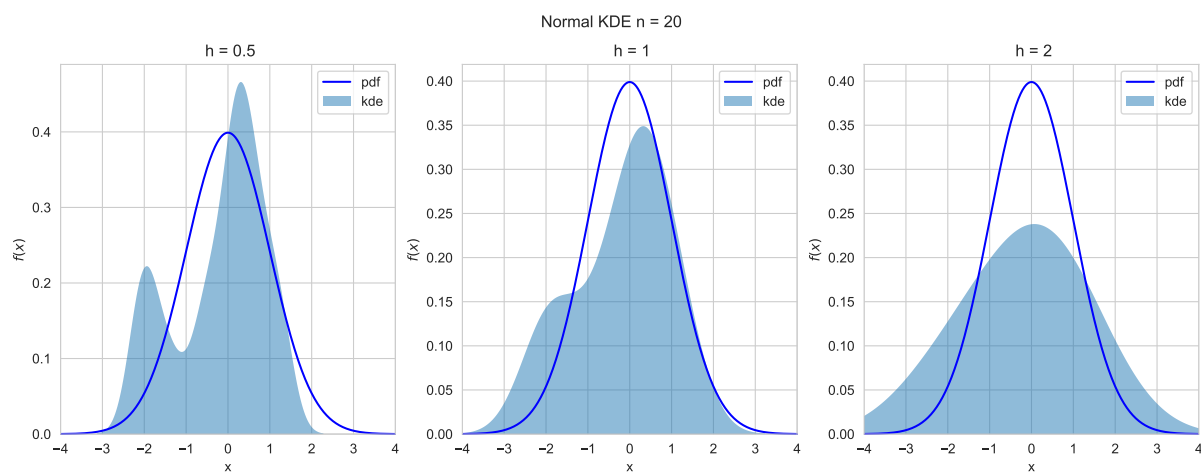


Рис. 6: нормальное распределение,  $n = 20$



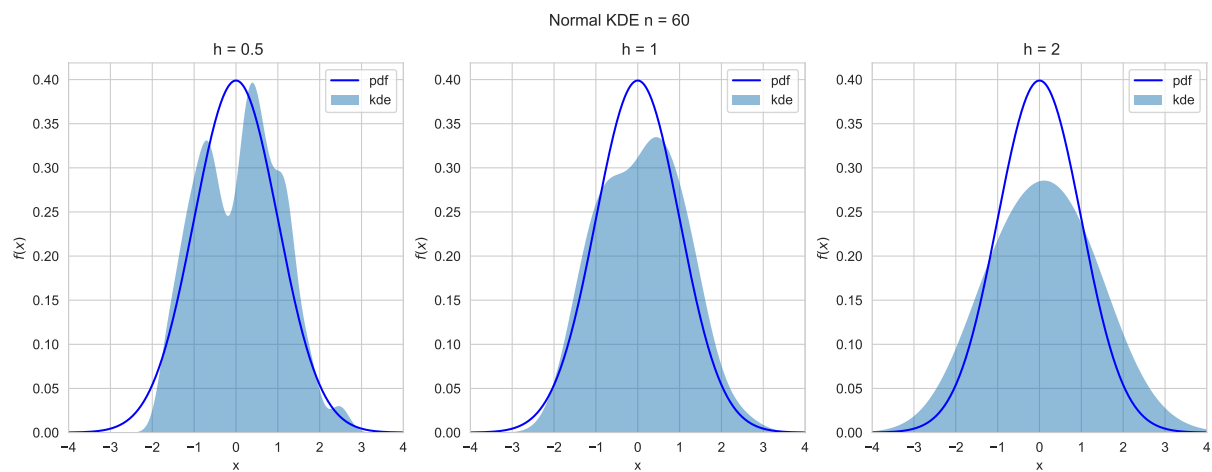


Рис. 7: нормальное распределение,  $n = 60$

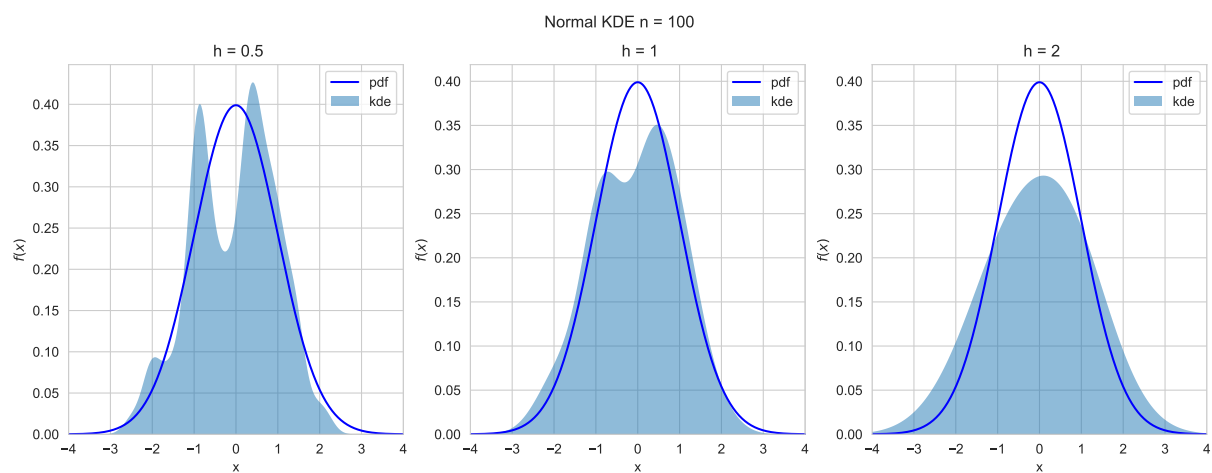
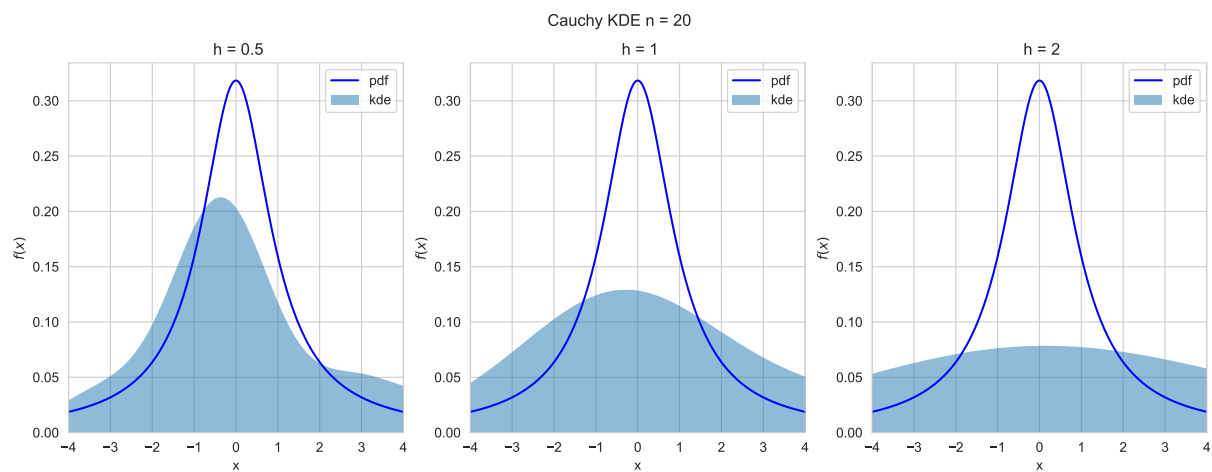
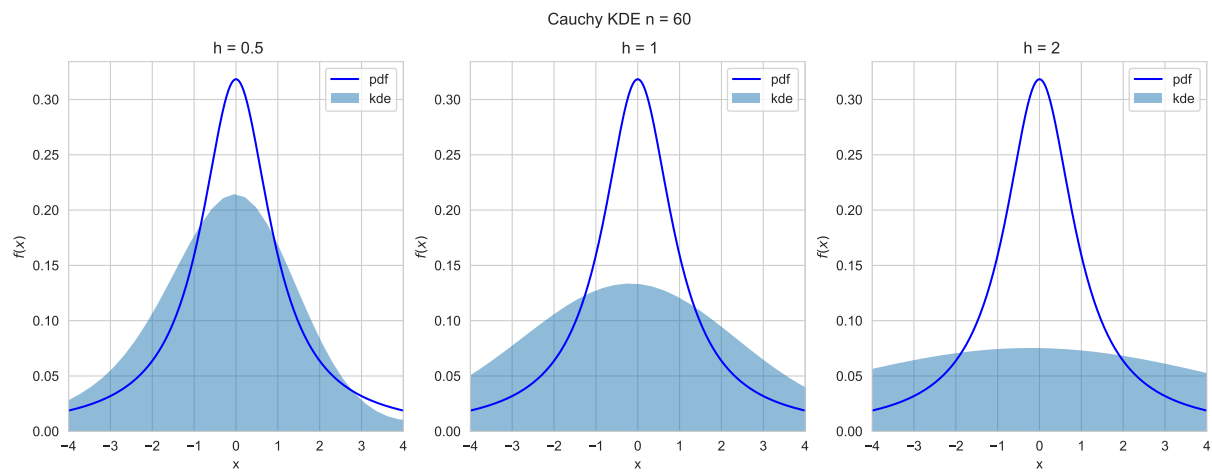


Рис. 8: нормальное распределение,  $n = 100$

Рис. 9: распределение Коши,  $n = 20$ Рис. 10: распределение Коши,  $n = 60$

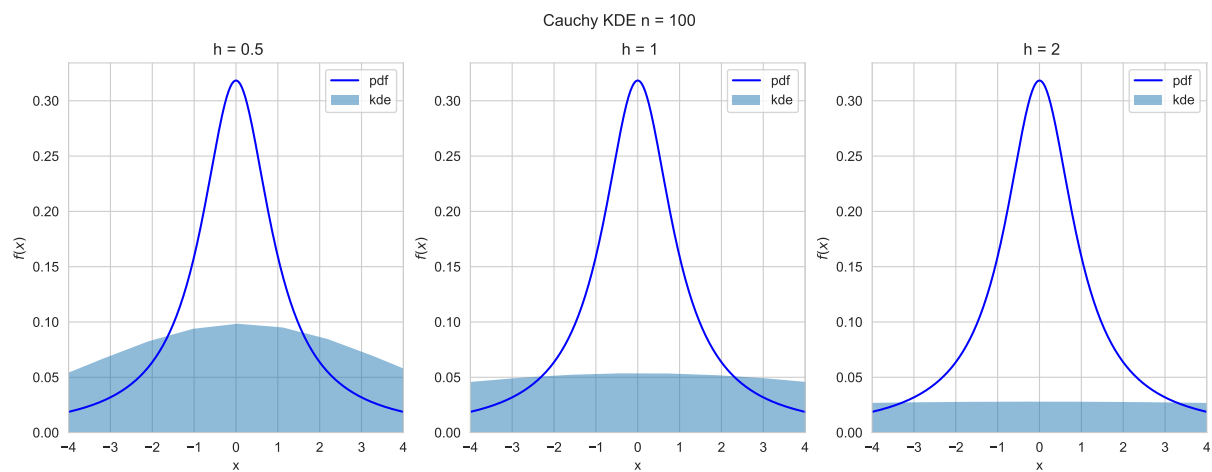


Рис. 11: распределение Коши,  $n = 100$

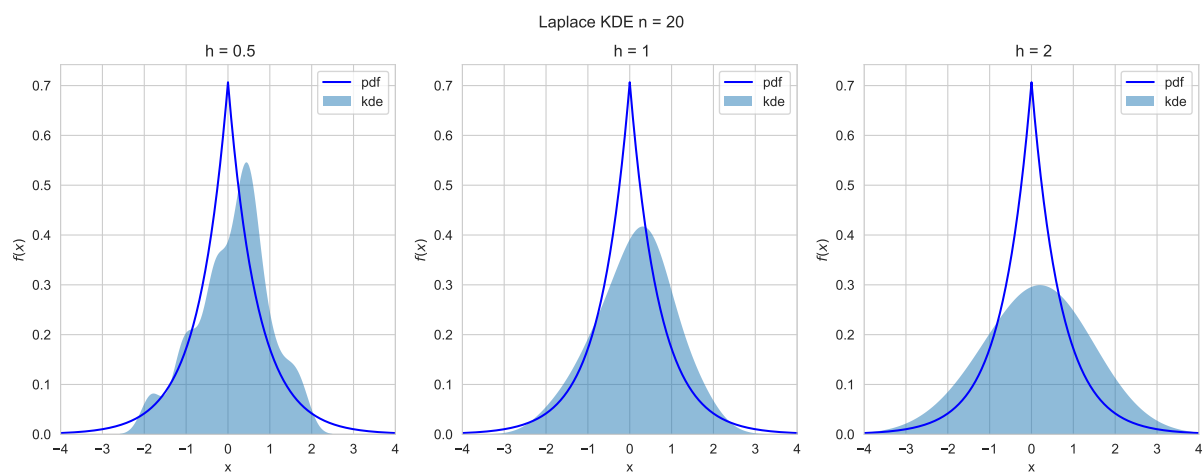
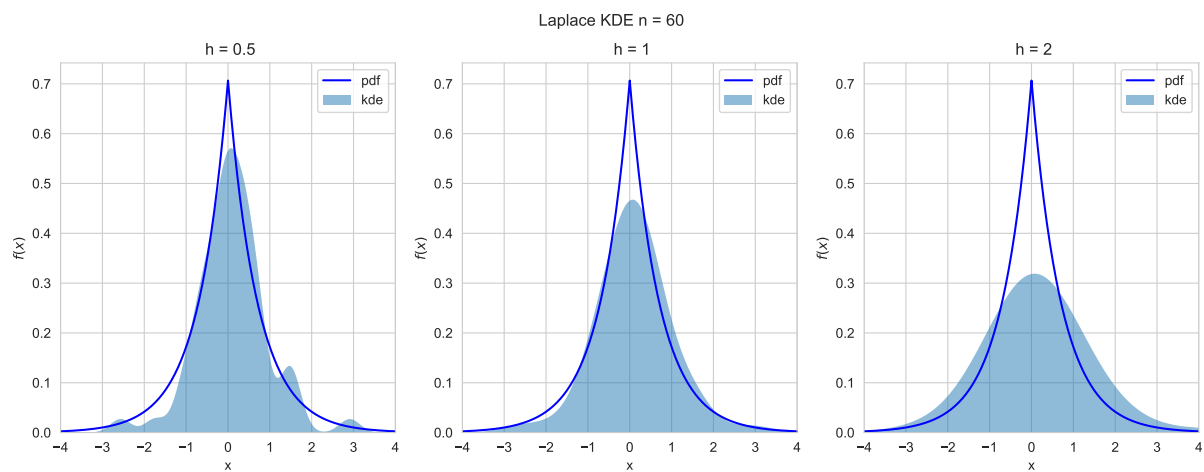
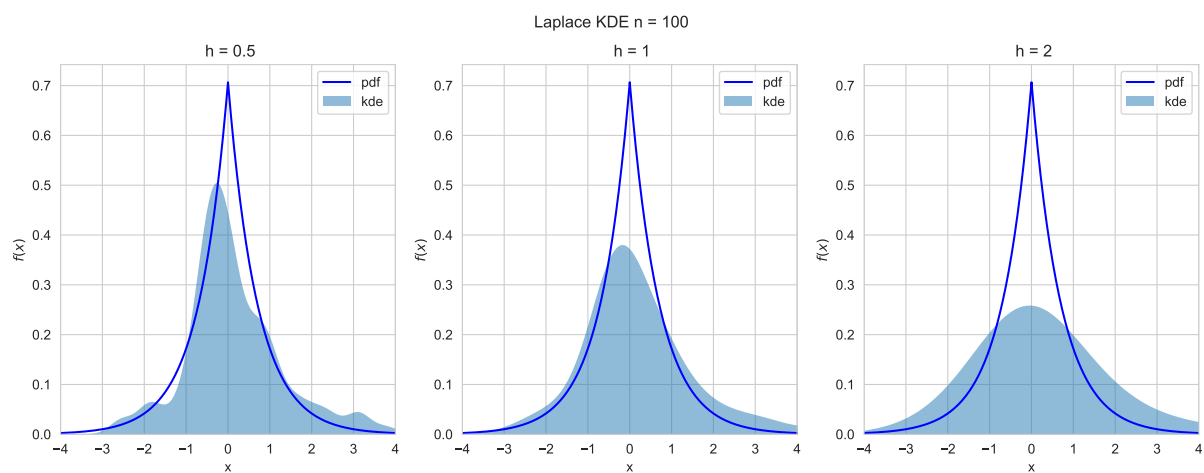


Рис. 12: распределение Лапласа,  $n = 20$

Рис. 13: распределение Лапласа,  $n = 60$ Рис. 14: распределение Лапласа,  $n = 100$

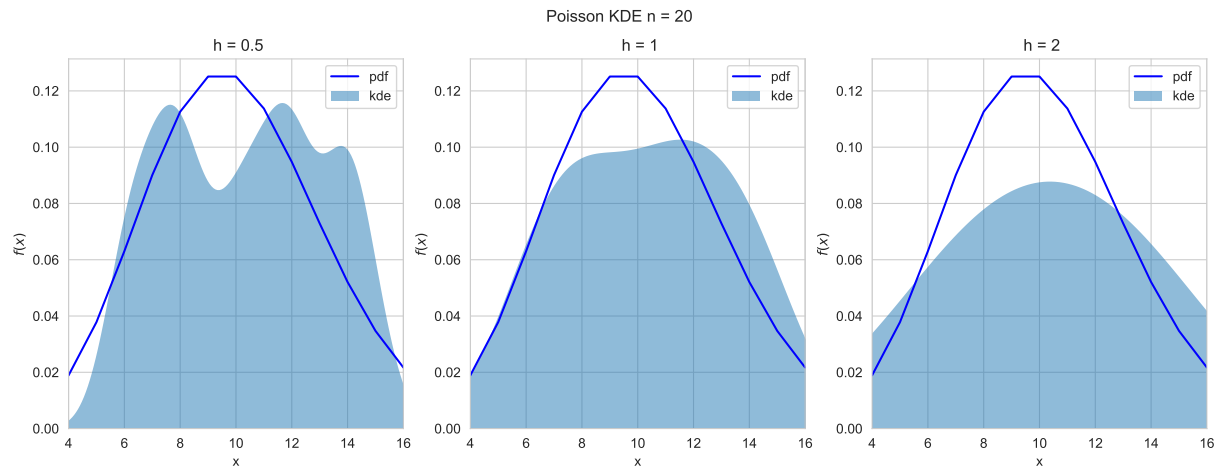


Рис. 15: распределение Пуассона,  $n = 20$

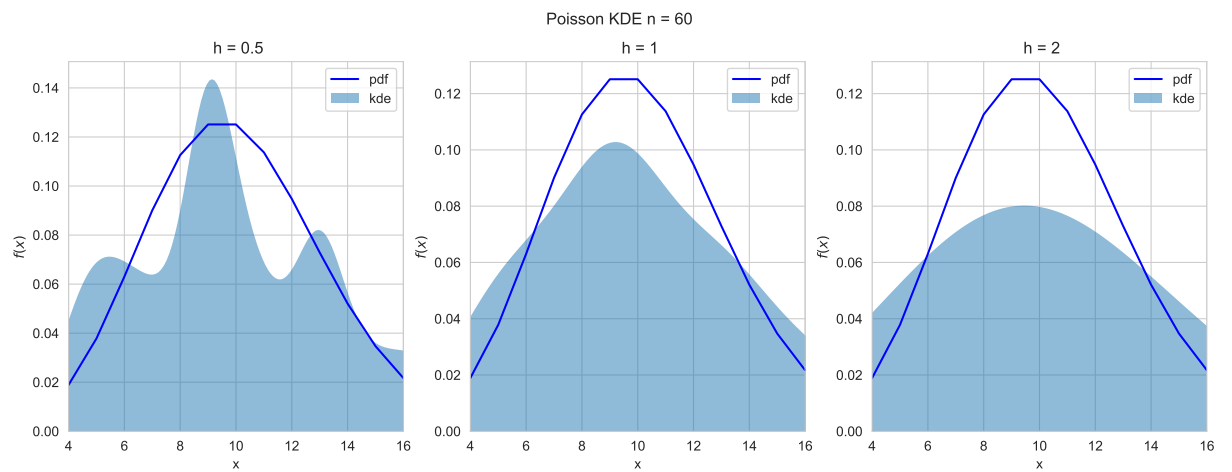


Рис. 16: распределение Пуассона,  $n = 60$

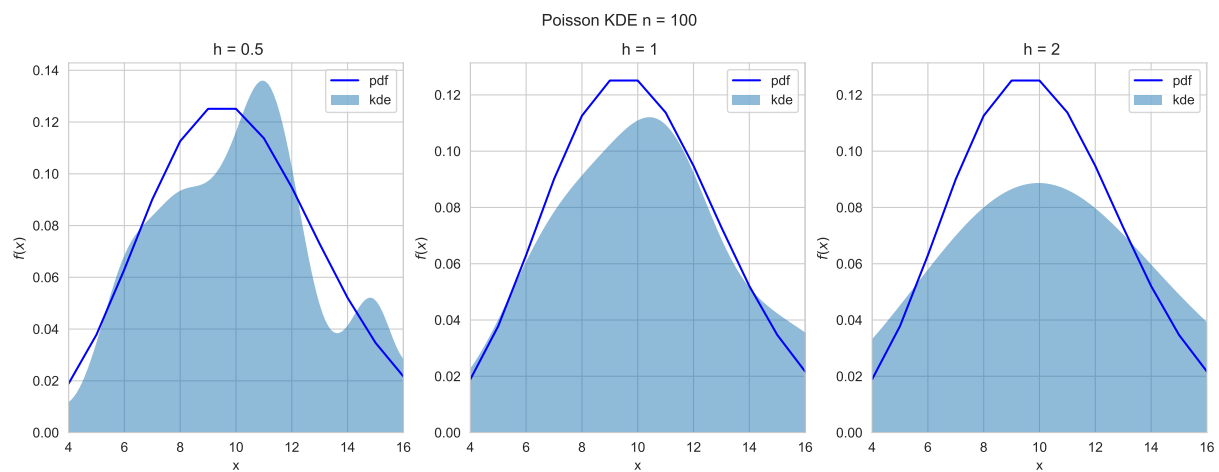


Рис. 17: распределение Пуассона,  $n = 100$

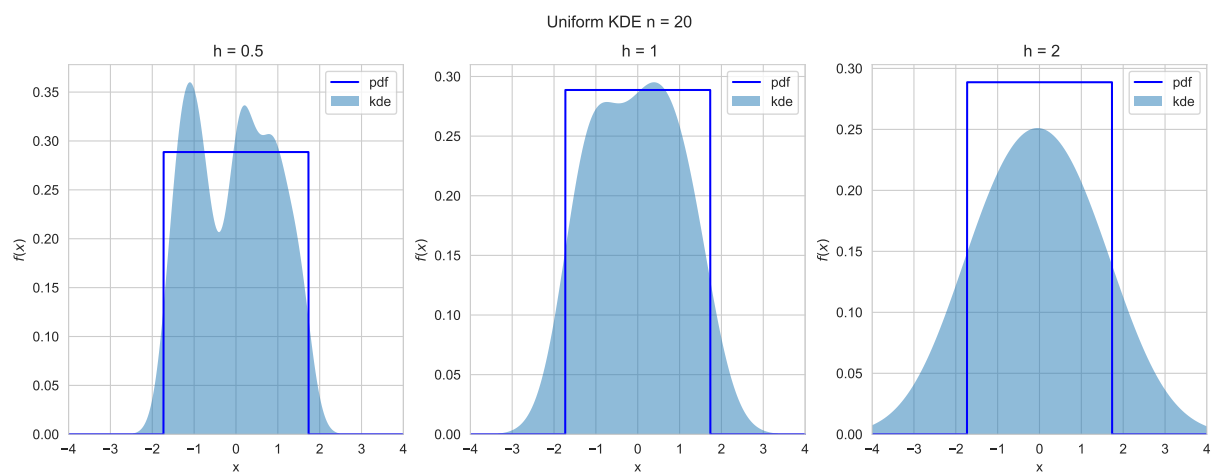


Рис. 18: равномерное распределение,  $n = 20$

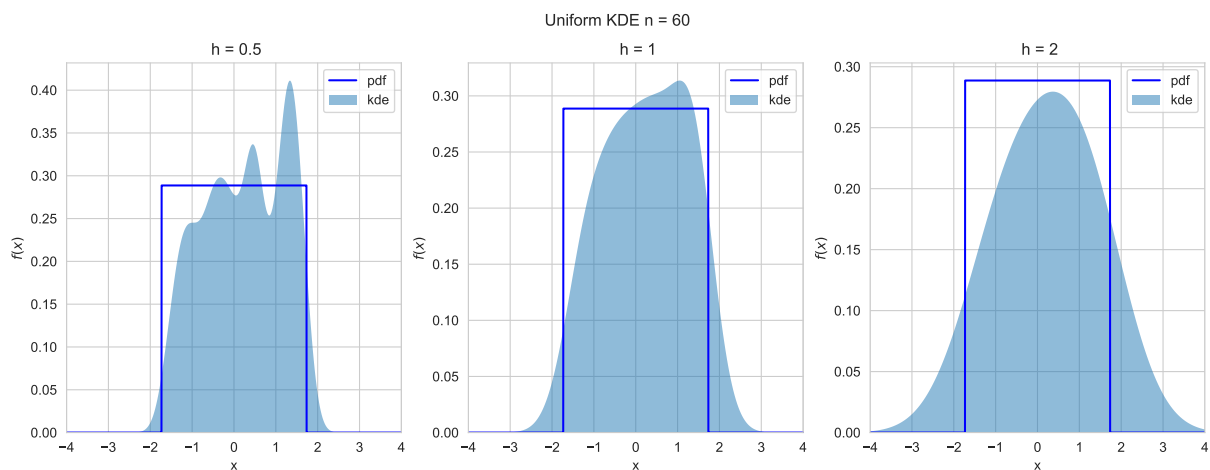


Рис. 19: равномерное распределение,  $n = 60$

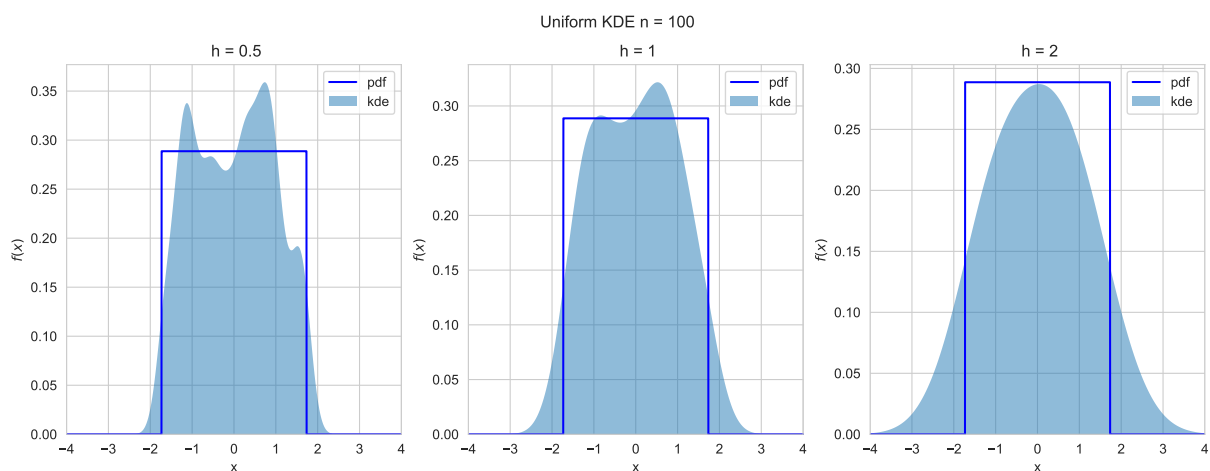


Рис. 20: равномерное распределение,  $n = 100$

## 5 Обсуждение

Можем наблюдать на иллюстрациях с эмпирическими функциями, что ступенчатая эмпирическая функция распределения тем лучше приближает функцию распределения реальной выборки, чем мощнее эта выборка. Заметим так же, что для распределения Пуассона и равномерного распределения отклонение функций друг от друга наибольшее.

Рисунки, посвященные ядерным оценкам, иллюстрируют сближение ядерной оценки и функции плотности вероятности для всех  $h$  с ростом размера выборки. Для распределения Пуассона

наиболее ярко видно, как сглаживает отклонения увеличение параметра сглаживания  $h$ .

В зависимости от особенностей распределений для их описания лучше подходят разные параметры  $h$  в ядерной оценке: 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  $\hat{h}_n$ , тем меньше изменений знака производной у аппроксимирующей функции, вплоть до того, что при  $h = h_n/2$  функция становится унимодальной на рассматриваемом промежутке. Также видно, что при  $h = h_n/2$  по полученным приближениям становится сложно сказать плотность вероятности какого распределения они должны повторять, так как они очень похожи между собой.

## 6 Приложение

Код программы GitHub URL:

<https://github.com/Brightest-Sunshine/Math-Statistic-2021/blob/main/Lab4/Lab4.ipynb>