Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Высшая школа прикладной математики

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №4

по дисциплине «Математическая статистика»

Выполнила
студентка гр.3630102/80101

А.А. Тимофеева

Руководитель доцент, к.ф.-м.н.

А.Н.Баженов

Санкт-Петербург 2021

СОДЕРЖАНИЕ

СПИСОК ИЛЛЮСТРАЦИЙ

СПИСОК ТАБЛИЦ

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 называется последовательность различных элементов выборки $z_1, z_2, \dots, z_k$, расположенных в возрастающем порядке с указанием частот $n_1, n_2, \dots, n_k$, с которыми эти элементы содержатся в выборке.

2.1.2 Эмпирическая функция распределения

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

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

$F^*(x) -$ 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 Оценки плотности вероятности

2.2.1 Определение

Оценкой плотности вероятности $f(x)$ называется функция $f(x)$, построенная на основе выборки, приближённо 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:

$$\widehat{f_n}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\,(\frac{x - x_i}{h_n}).$$

Здесь функция $K(u)$, называемая ядерной (ядром), непрерывна и является плотностью вероятности, $x_1, \dots, x_n$ — элементы выборки, $\{h_n\}_{\,n\in N}$ — любая последовательность положительных чисел, обладающая свойствами

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

Такие оценки называются непрерывными ядерными.

Гауссово (нормальное) ядро

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

Правило Сильвермана

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

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

3 РЕАЛИЗАЦИЯ

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm. Исходный код лабораторной работы приведён в приложении.

4 РЕЗУЛЬТАТЫ

4.1 Эмпирическая функция распределения

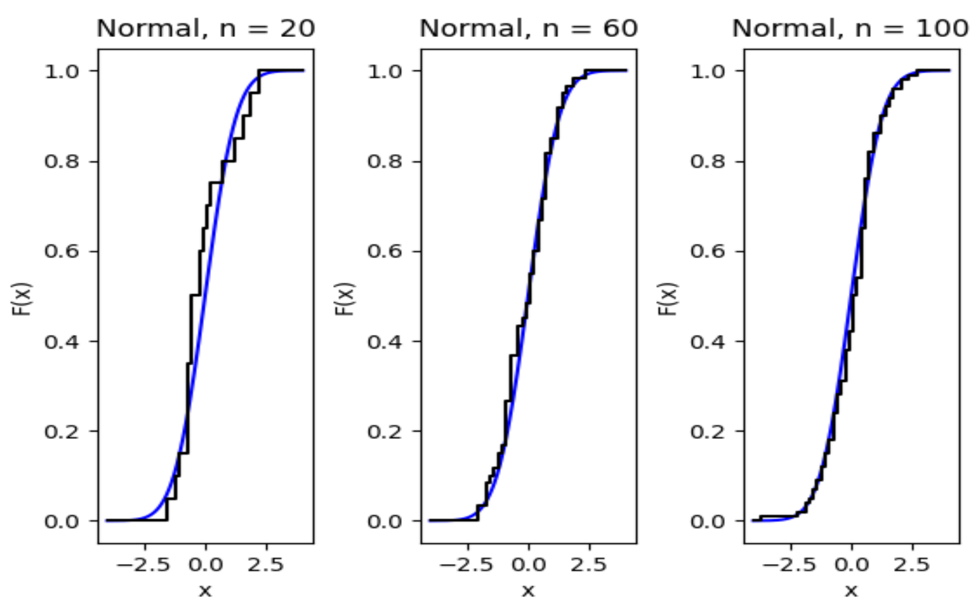

Рисунок 1: Нормальное распределение

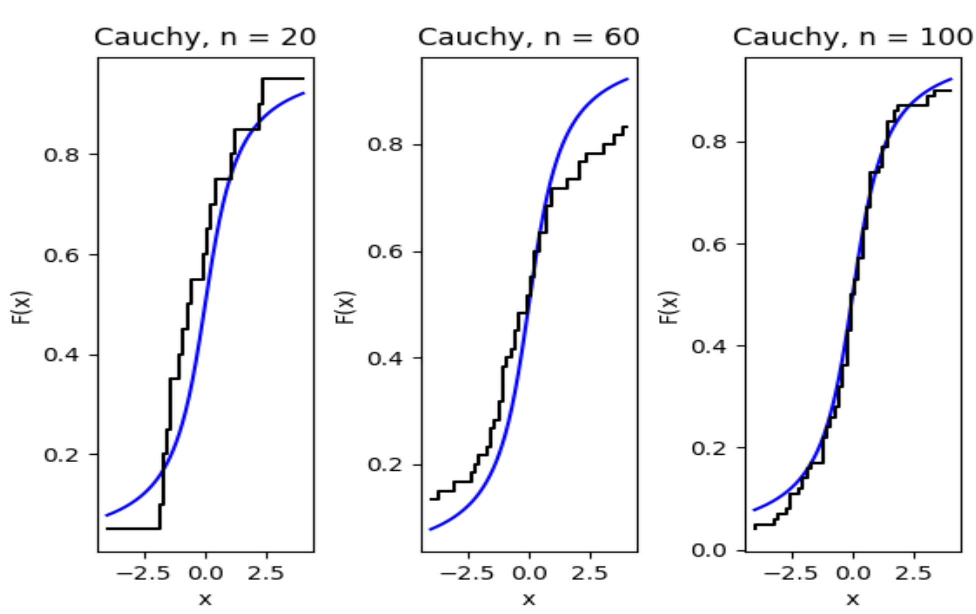

Рисунок 2: Распределение Коши

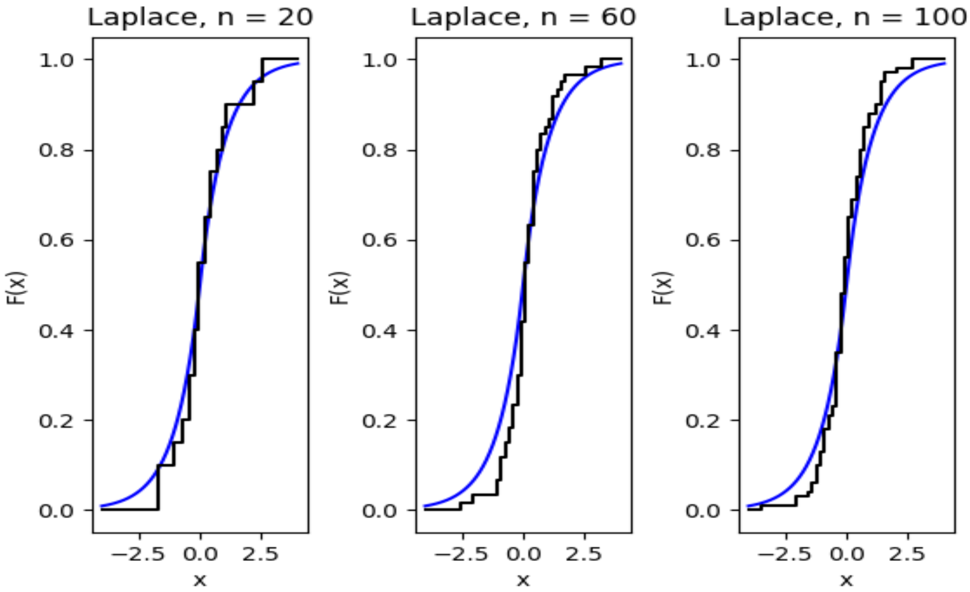

Рисунок 3: Распределение Лапласа

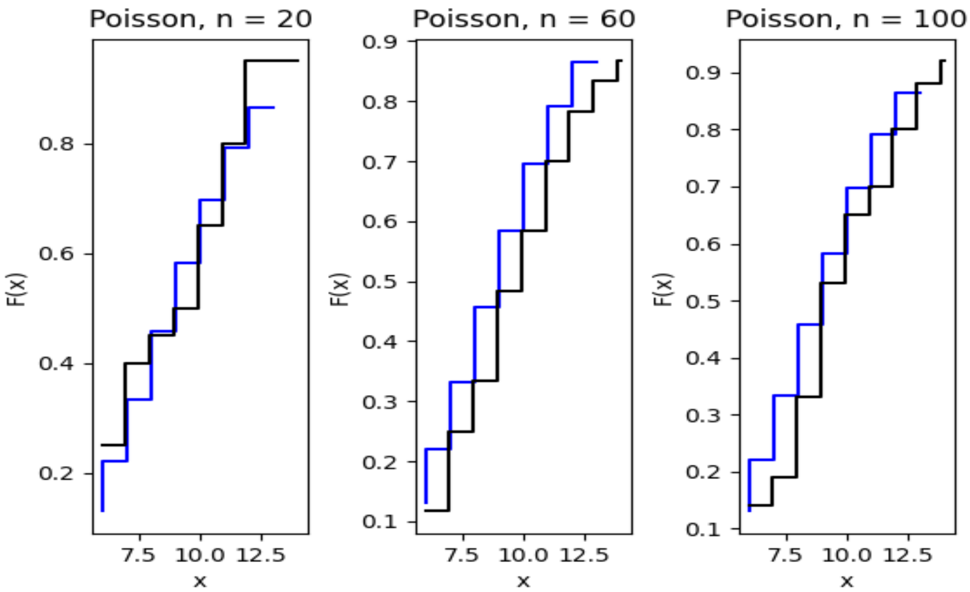

Рисунок 4: Распределение Пуассона

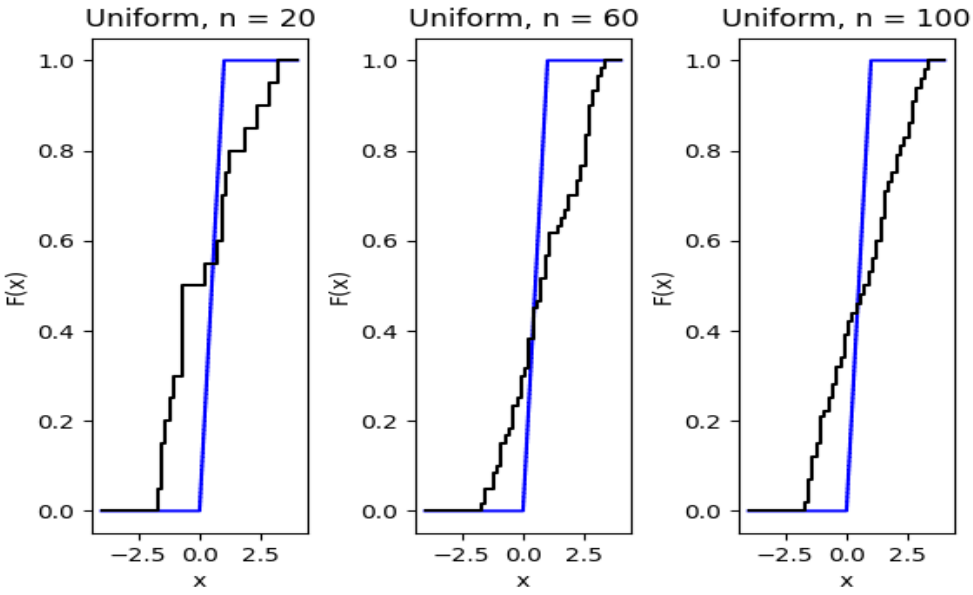

Рисунок 5: Равномерное распределение

4.2 Ядерные оценки плотности распределения

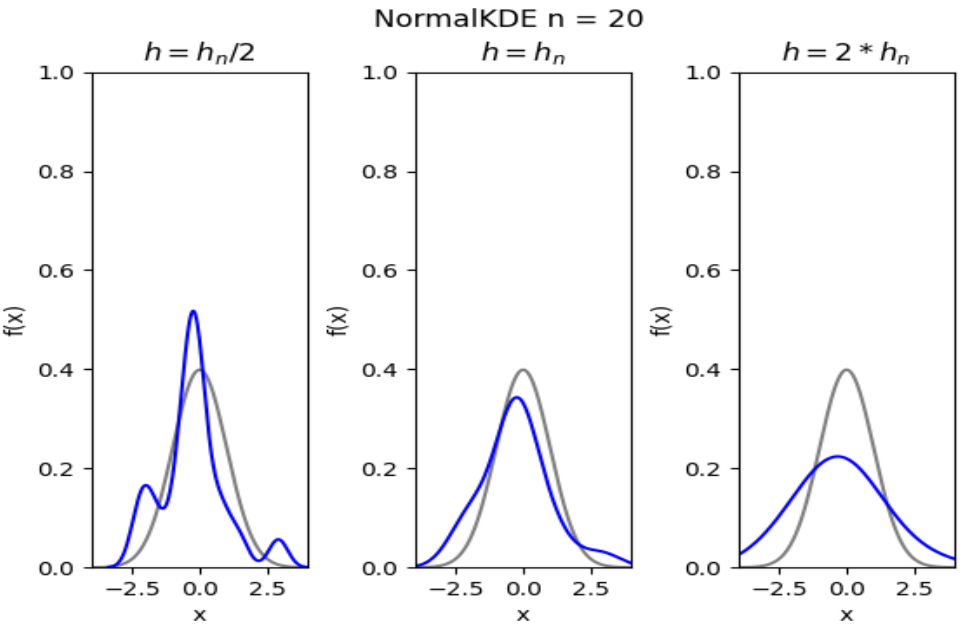

Рисунок 6: Нормальное распределение n = 20

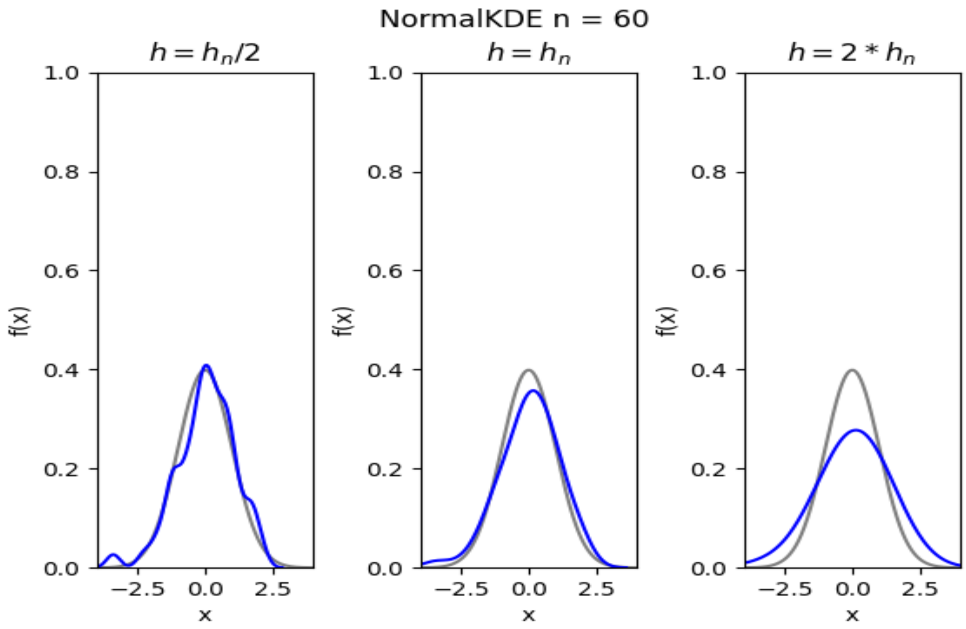

Рисунок 7: Нормальное распределение n = 60

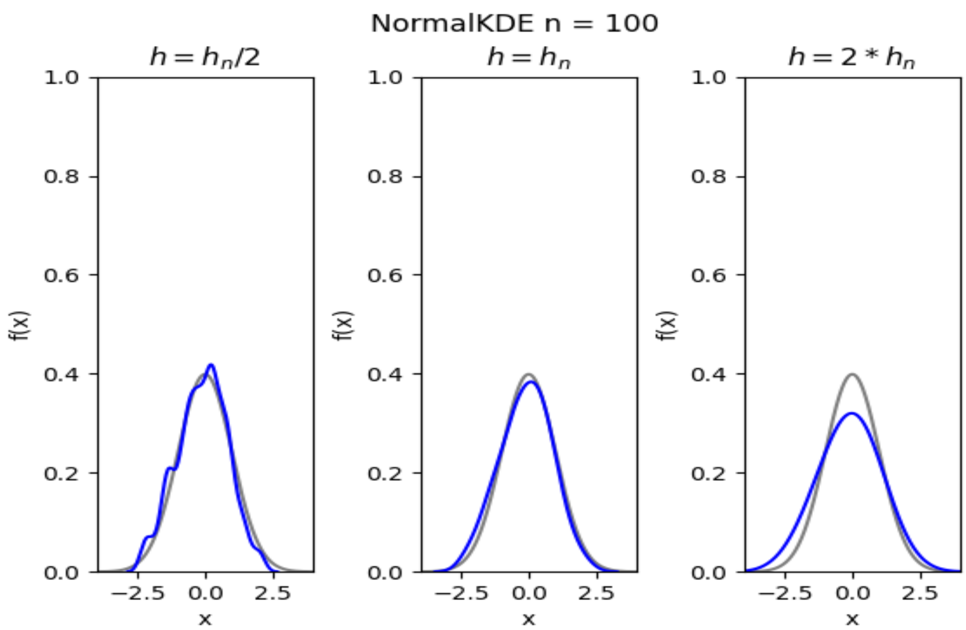

Рисунок 8: Нормальное распределение n = 100

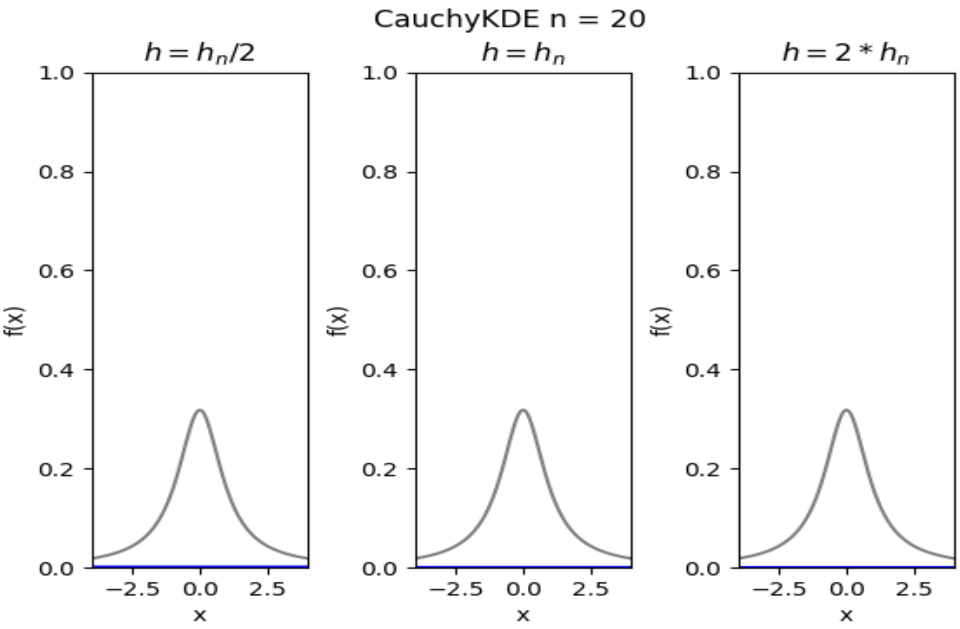

Рисунок 9: Распределение Коши n = 20

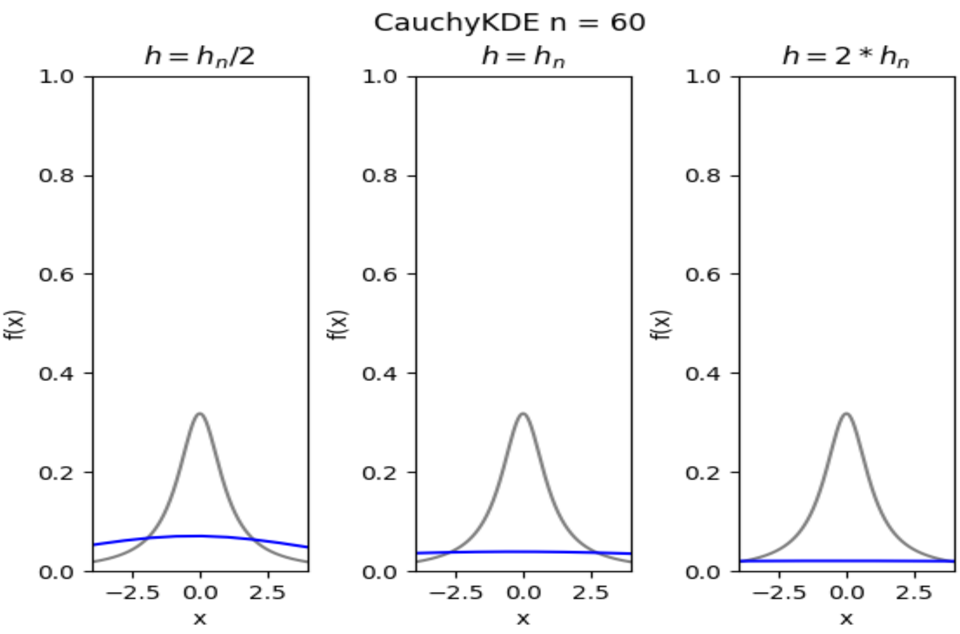

Рисунок 10: Распределение Коши n = 60

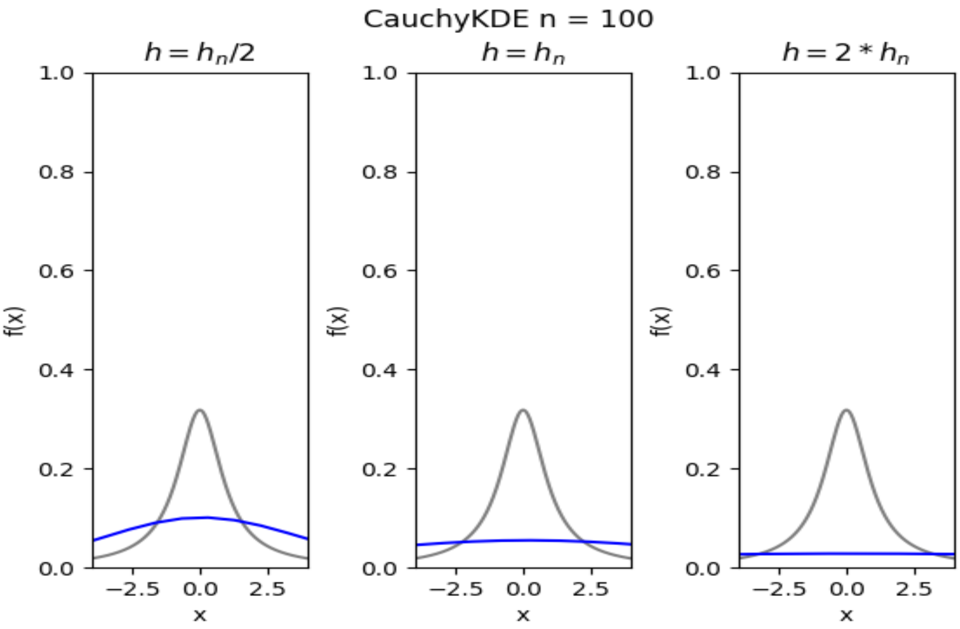

Рисунок 11: Распределение Коши n = 100

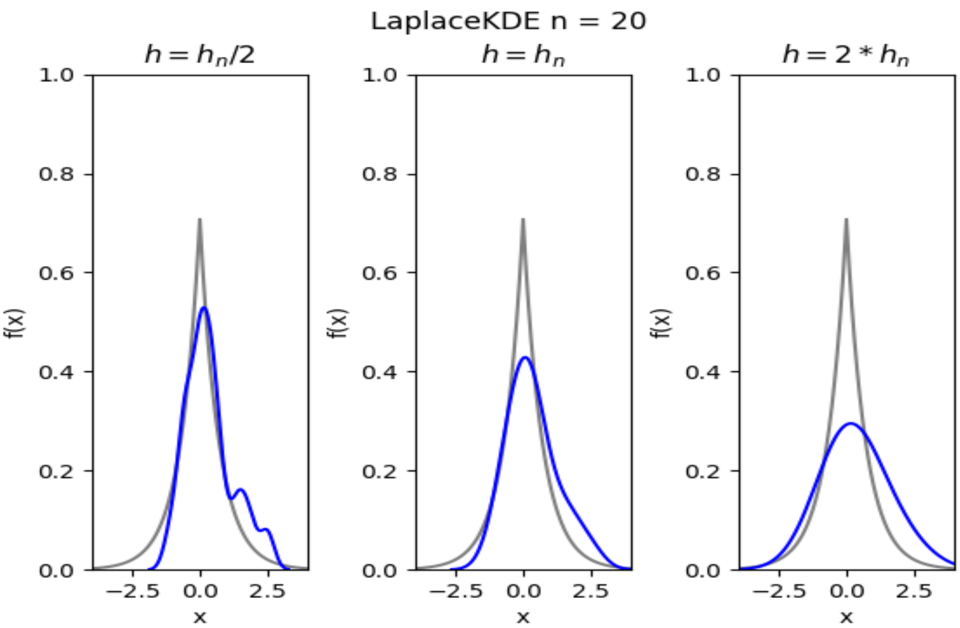

Рисунок 12: Распределение Лапласа n = 20

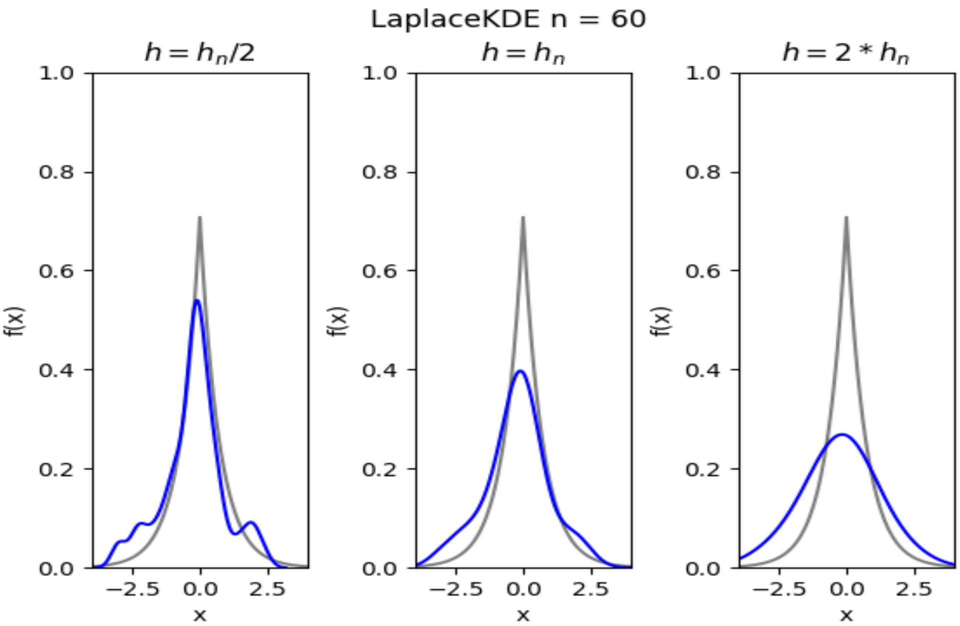

Рисунок 13: Распределение Лапласа n = 60

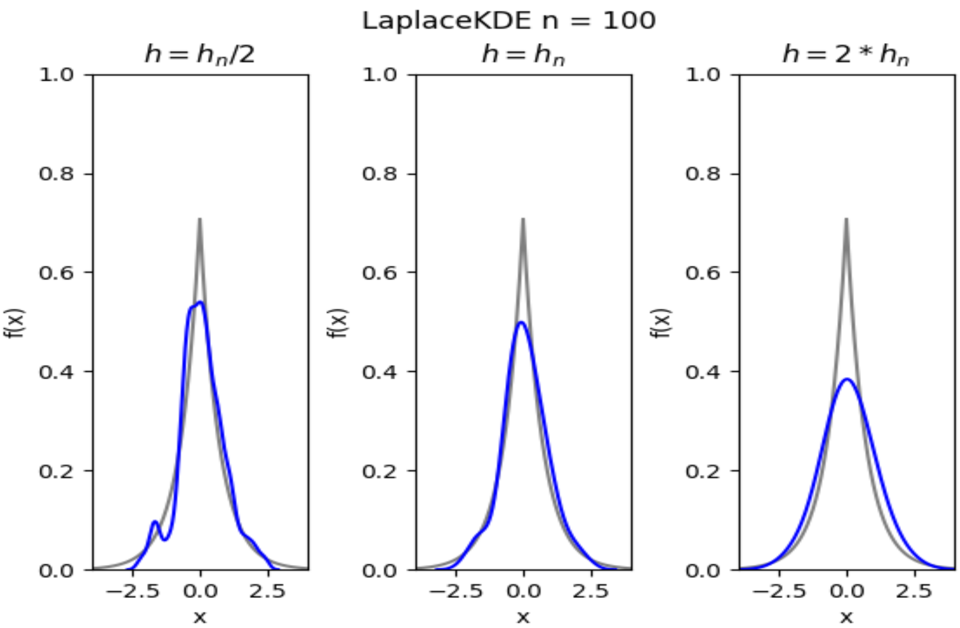

Рисунок 14: Распределение Лапласа n = 100

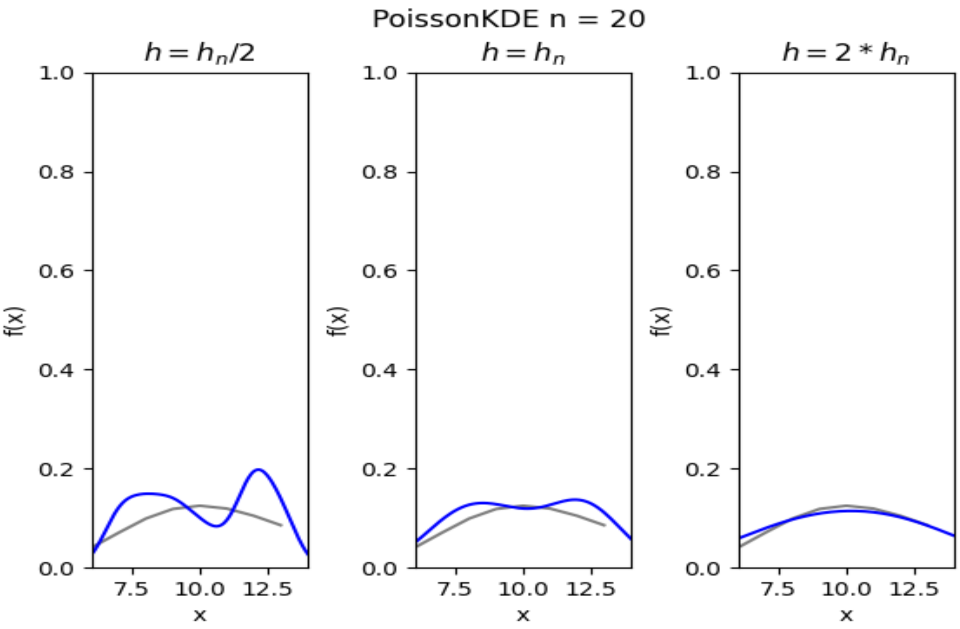

Рисунок 15: Распределение Пуассона n = 20

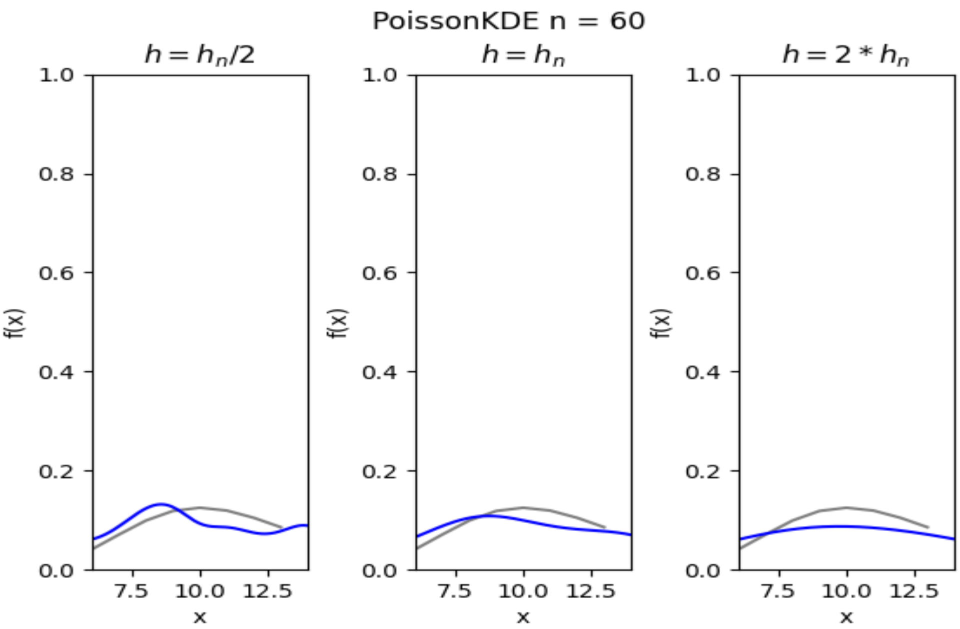

Рисунок 16: Распределение Пуассона n = 60

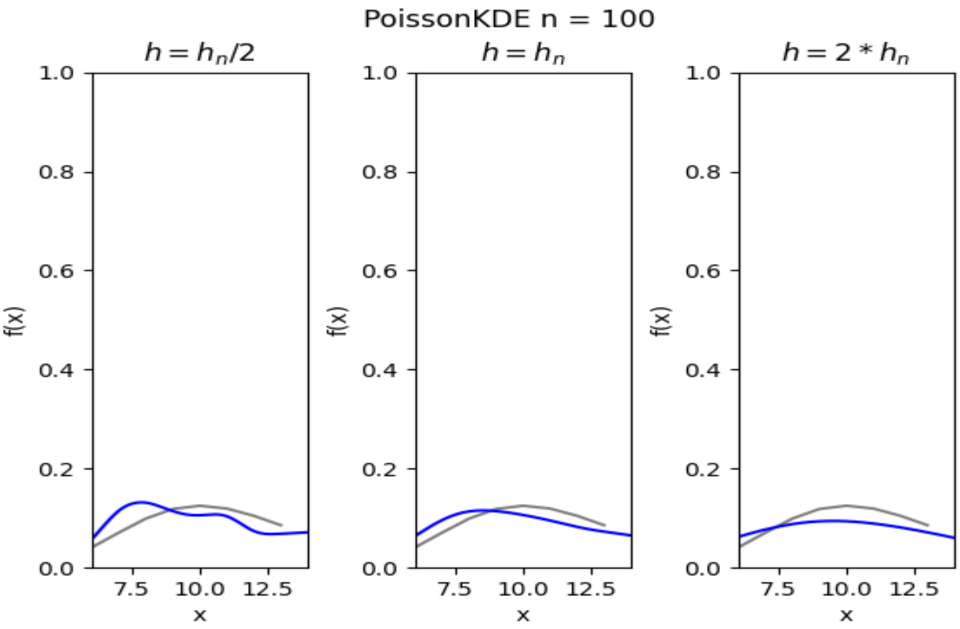

Рисунок 17: Распределение Пуассона n = 100

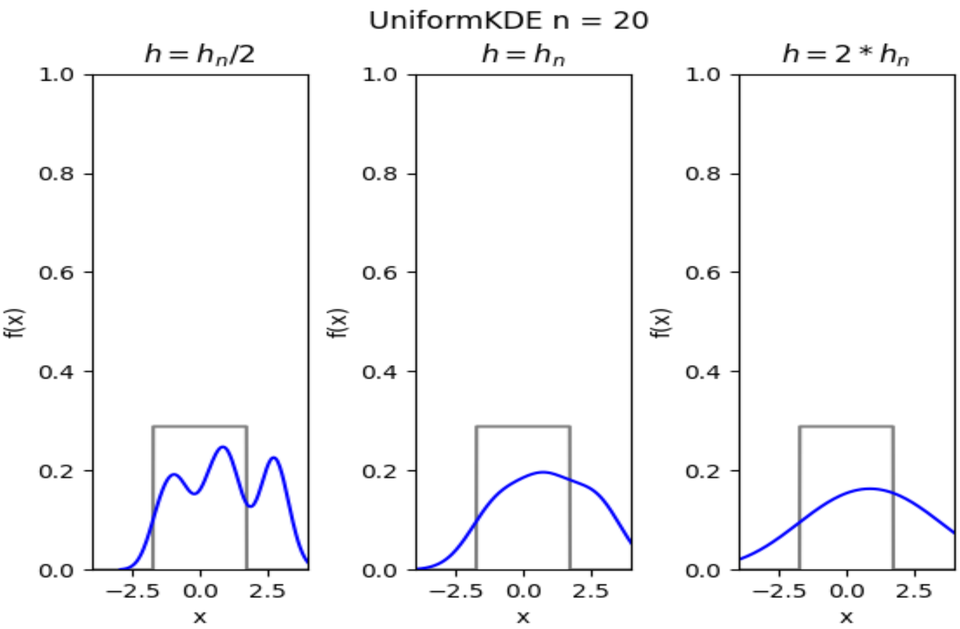

Рисунок 18: Равномерное распределение n = 20

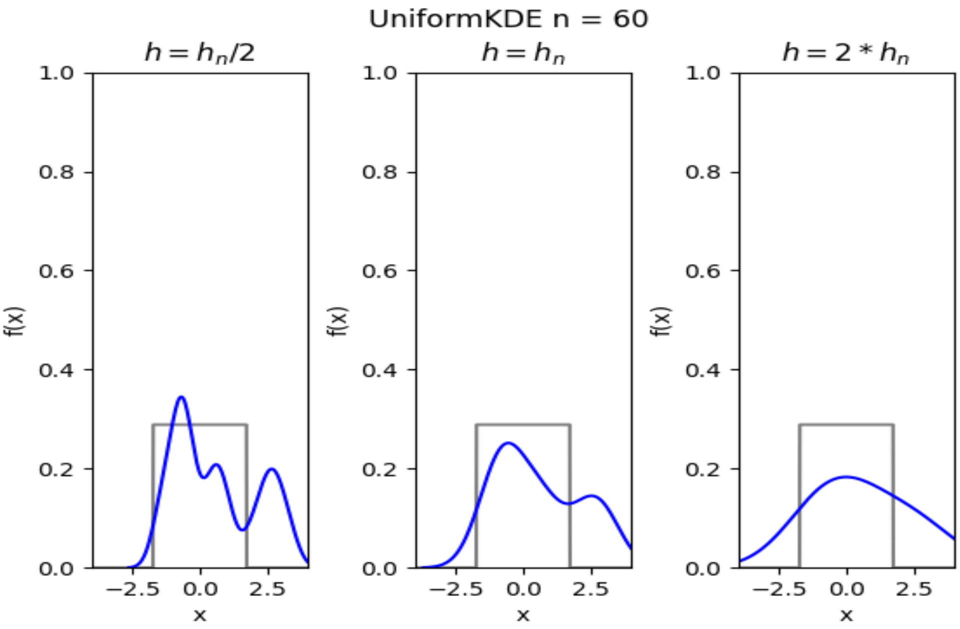

Рисунок 19: Равномерное распределение n = 60

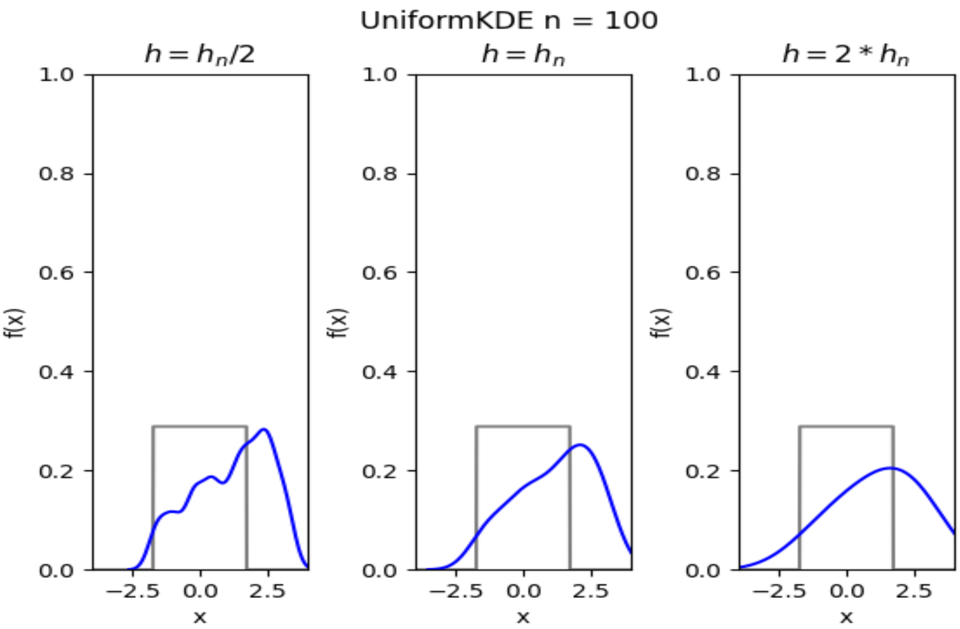

Рисунок 20: Равномерное распределение n = 100

5 ОБСУЖДЕНИЕ

Можем наблюдать на иллюстрациях с э. ф. р., 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 равномерного распределения и распределения Пуассона лучше подойдет параметр $h = 2h_n$, для распределения Лапласа − $h = h_n/2$, а для нормального и Коши − $h = h_n$. Такие значения дают вид ядерной оценки наиболее близкий к плотности, характерной данным распределениям.

Также можно увидеть, что чем больше коэффициент при параметре сглаживания $\widehat{h_n}$, 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.

6 ПРИЛОЖЕНИЕ

Код программы URL:https://github.com/tmffv/MathStat/blob/master/lab4/lab4.py