Санкт-Петербургский Политехнический Университет им. Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Отчёт по лабораторной работе №4 по дисциплине “Математическая статистика”

**Эмпирические функции и ядерные оценки**

Выполнил студент:

Мишутин Д. В.

Группа:

3630102/70301

Проверил:

К.ф.-м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург

2020 г.

# Оглавление



# Список иллюстраций

# 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.

Распределения:

- Стандартное нормальное распределение:

$$N(x,0,1)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \quad (1.1)$$

- Стандартное распределение Коши:

$$C(x,0,1)=\frac{1}{\pi(1+x^2)} \quad (1.2)$$

- Распределение Лапласа:

$$L\left(x,0,\frac{1}{\sqrt{2}}\right)=\frac{1}{\sqrt{2}}e^{-\sqrt{2}|x|} \quad (1.3)$$

- Распределение Пуассона:

$$P(k,10)=\frac{10^k}{k!}e^{-10} \quad (1.4)$$

- Равномерное распределение:

$$U(x,-\sqrt{3},\sqrt{3})=\begin{cases}\frac{1}{2\sqrt{3}}, при|x|\le\sqrt{3}\\ 0, при|x|>\sqrt{3}\end{cases} \quad (1.5)$$

# 2 Теория

*Эмпирической функцией распределения (ЭФР)* $\dot{F}_n$ называется относительная частота события $X<x$, полученная по данной выборке суммированием частот $n_i$, для которых элементы $z_i$ статистического ряда меньше $x$:

$$\dot{F}_n(x)=\dot{F}(X<x)=\frac{1}{n}\sum_{z_i<x}n_i$$

ЭФР является оценкой, то есть приближённым значением, генеральной функции распределения.

$$\dot{F}_n(x) \approx F_X(x)$$

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

$$\hat{f}(x) \approx f(x)$$

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

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

Здесь $K(u)$ называется *ядерной функцией плотности (ЯФП)*, непрерывна и является плотностью вероятности, $\{h_n\}$ – любая последовательность положительных чисел, обладающая свойствами:

1. $\lim_{n \to \infty} h_n = 0$
2. $\lim_{n \to \infty} \frac{h_n}{n^{-1}} = \infty$

Гауссово ядро:

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

Правило Сильвермана:

$$h_n = 1.06 \hat{\sigma} n^{\frac{-1}{5}},$$

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

## 3 Реализация

Был использован язык *Python 3.8.2*: модуль *numpy* для генерации выборок с различными распределениями и математических расчётов, модуль *scipy* для функций и плотностей распределений, модуль *matplotlib* для построения и отображения графиков и гистограмм, и функция *gamma* из модуля *math* для вычисления вещественного факториала.

# 4 Результаты

## 4.1 Эмпирические функции распределения (ЭФР)

*Рис. 1 ЭФР для Стандартного нормального распределения*

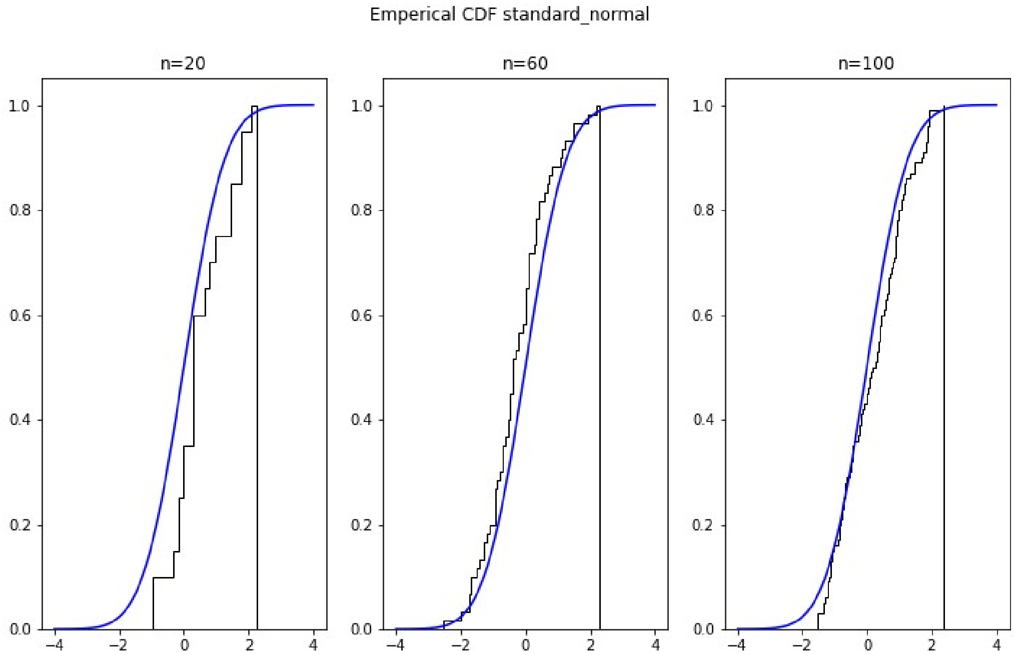



*Рис. 2 ЭФР для Стандартного распределения Коши*

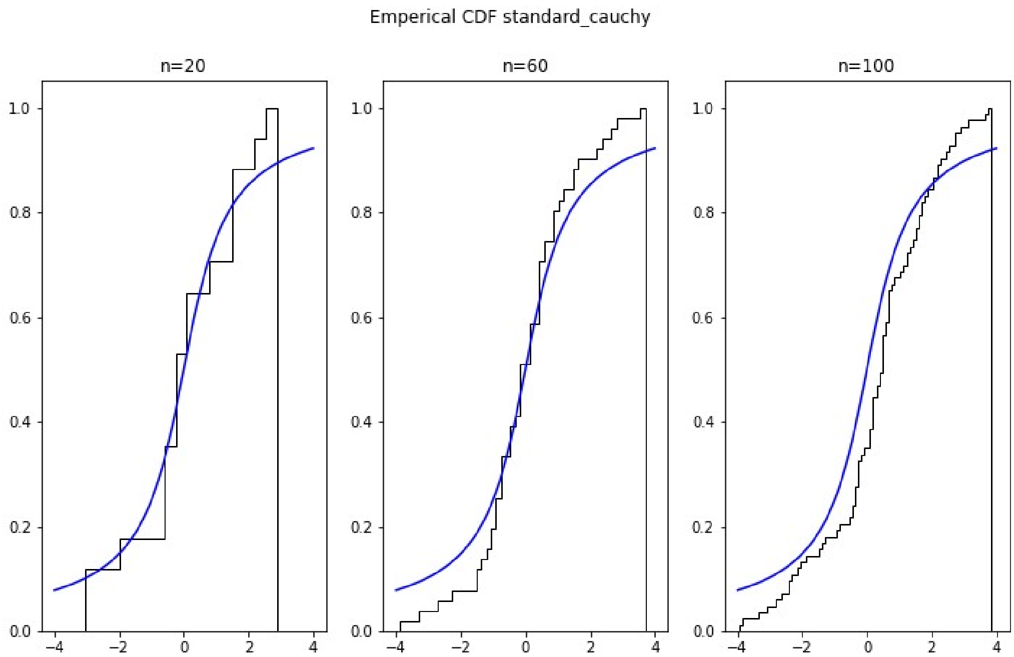


*Рис. 3 ЭФР для распределения Лапласа*

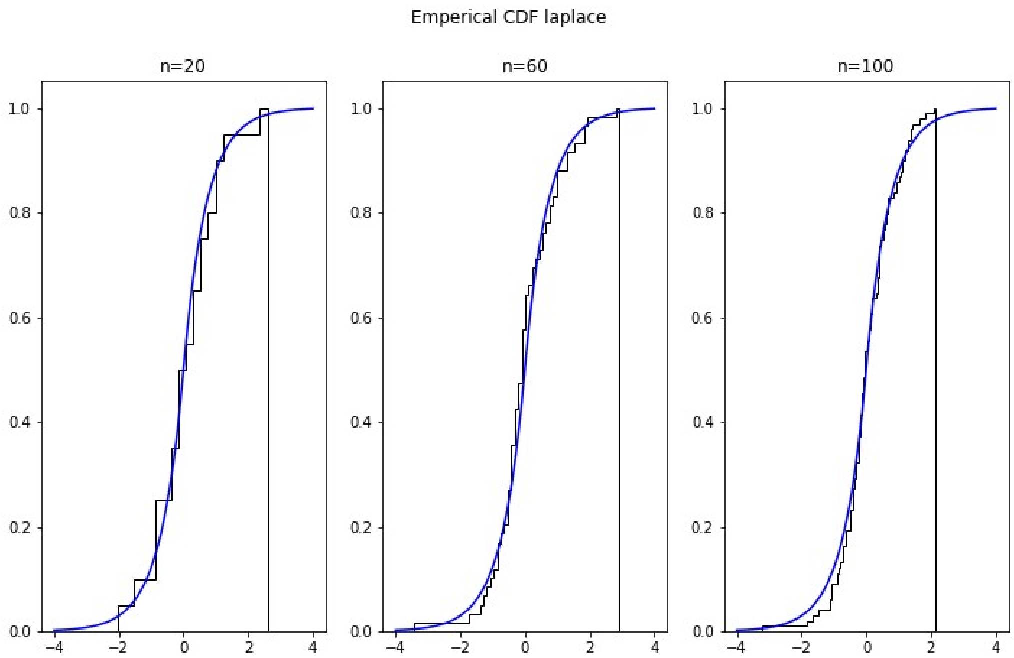



*Рис. 4 ЭФР для Равномерного распределения*

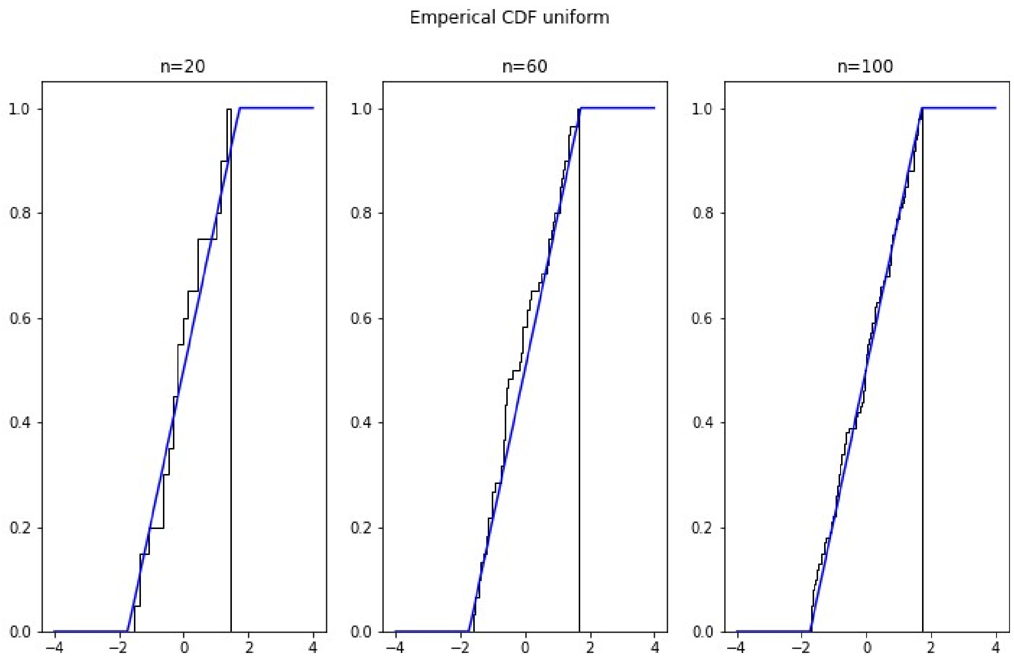
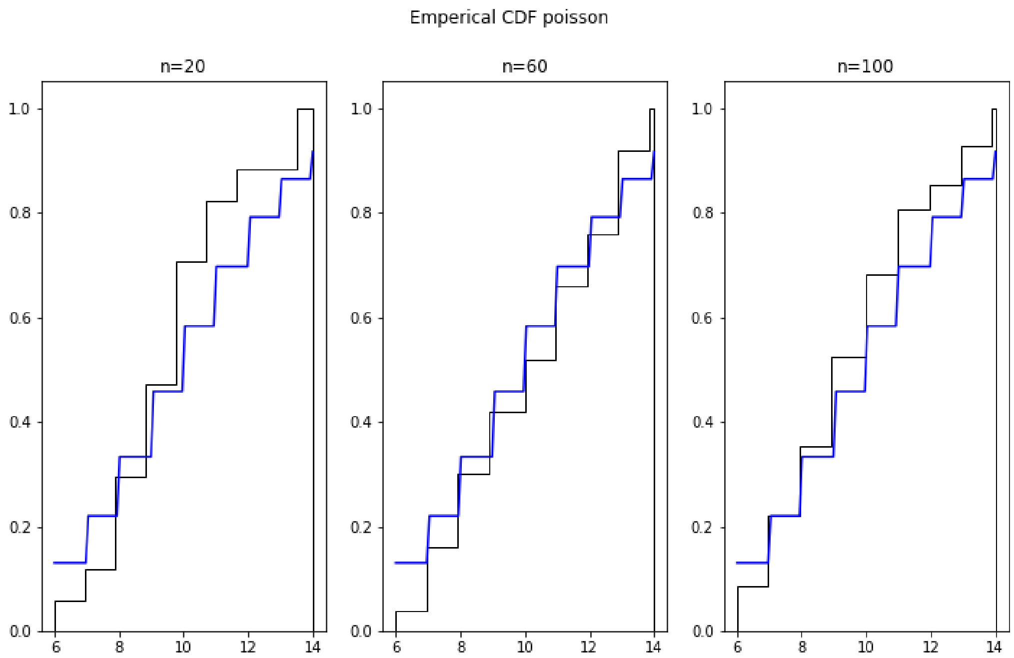


*Рис. 5 ЭФР для распределения Пуассона*




## 4.1 Ядерные функции плотности (ЯФП)

*Рис. 6 ЯФП для Стандартного нормального распределения при n=20*

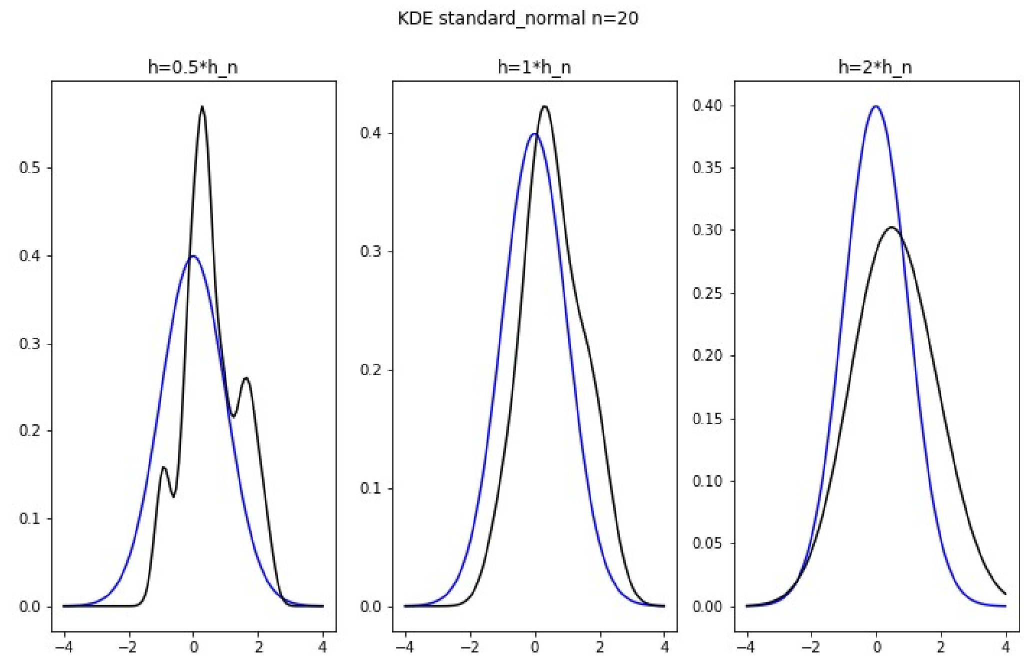


*Рис. 7 ЯФП для Стандартного нормального распределения при n=60*

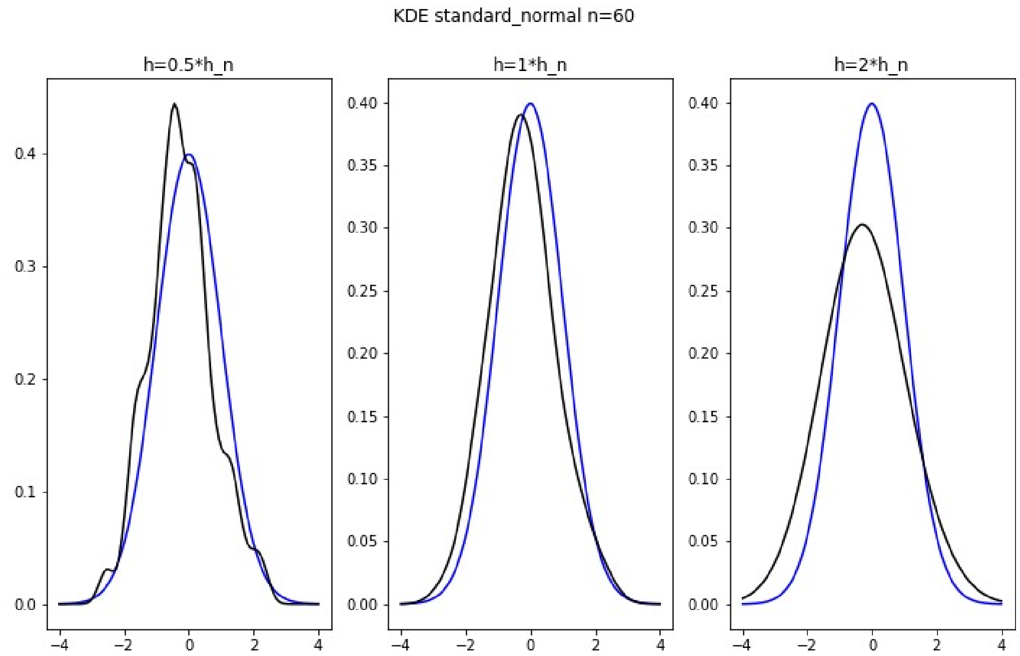



*Рис. 8 ЯФП для Стандартного нормального распределения при n=100*

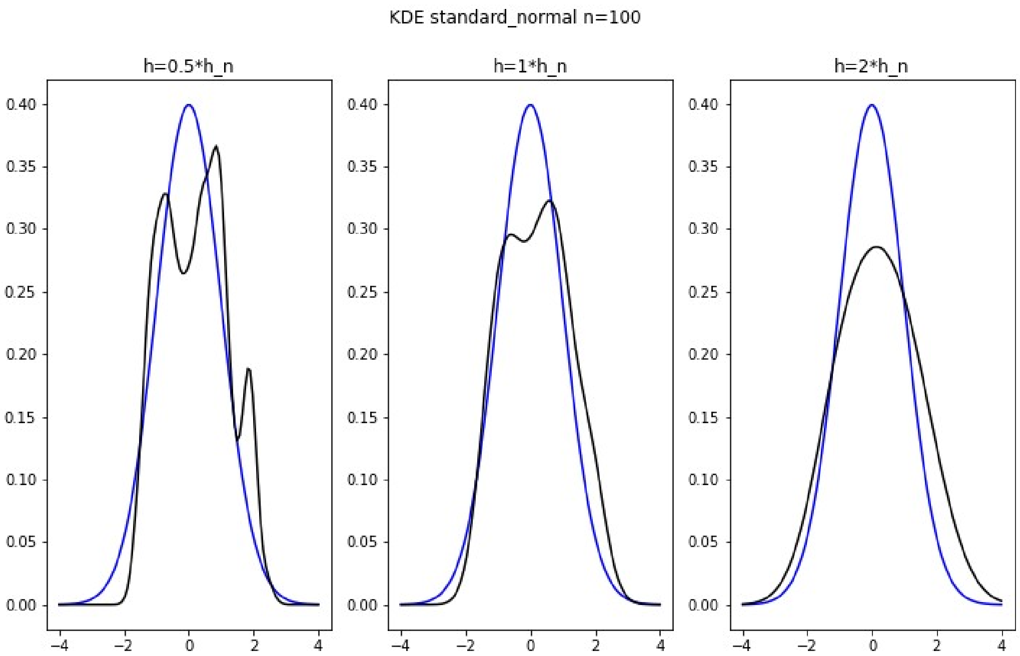


*Рис. 9 ЯФП для Стандартного распределения Коши при n=20*

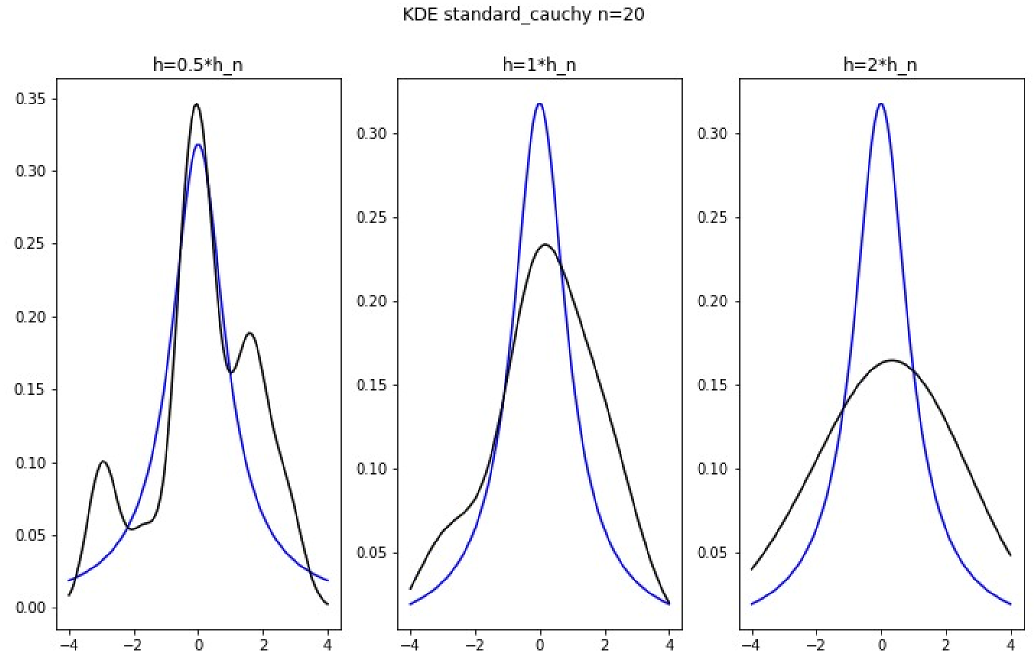



*Рис. 10 ЯФП для Стандартного распределения Коши при n=60*

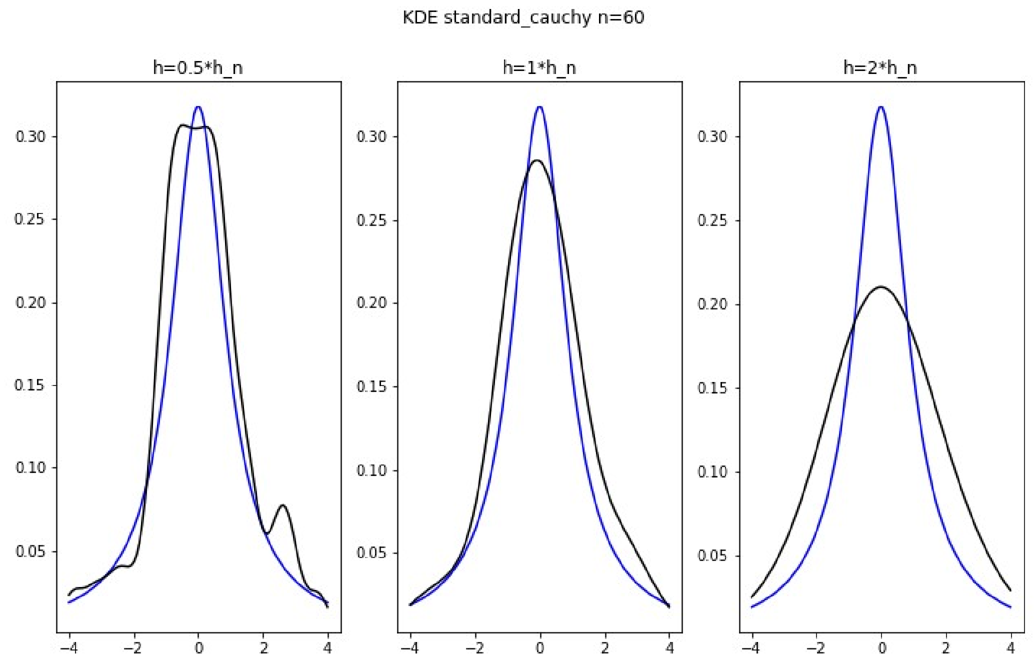


*Рис. 11 ЯФП для Стандартного распределения Коши при n=100*

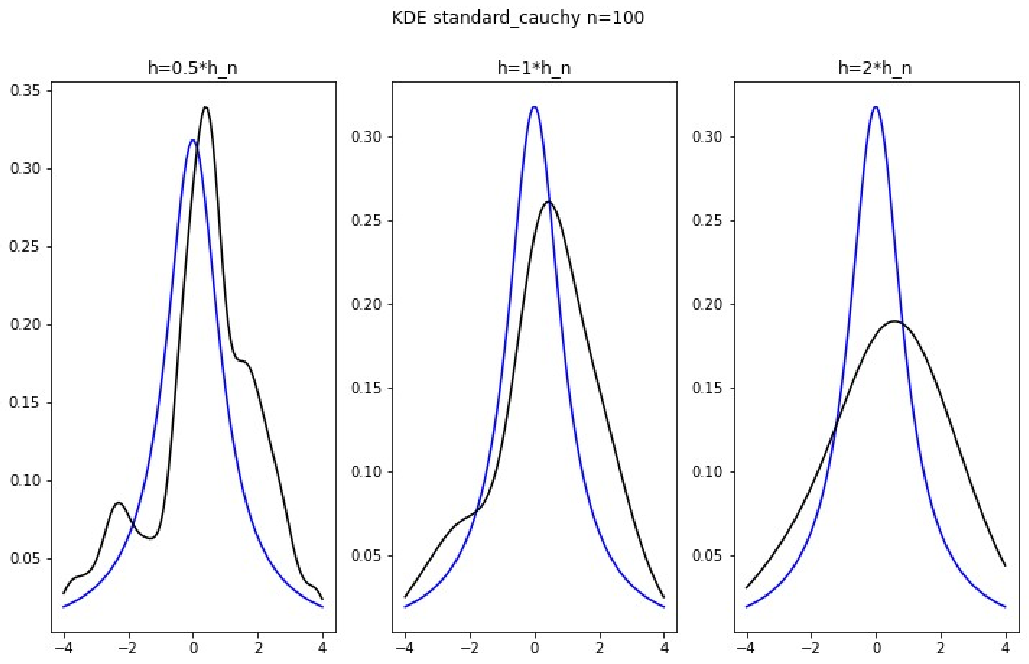



*Рис. 12 ЯФП для распределения Лапласа при n=20*

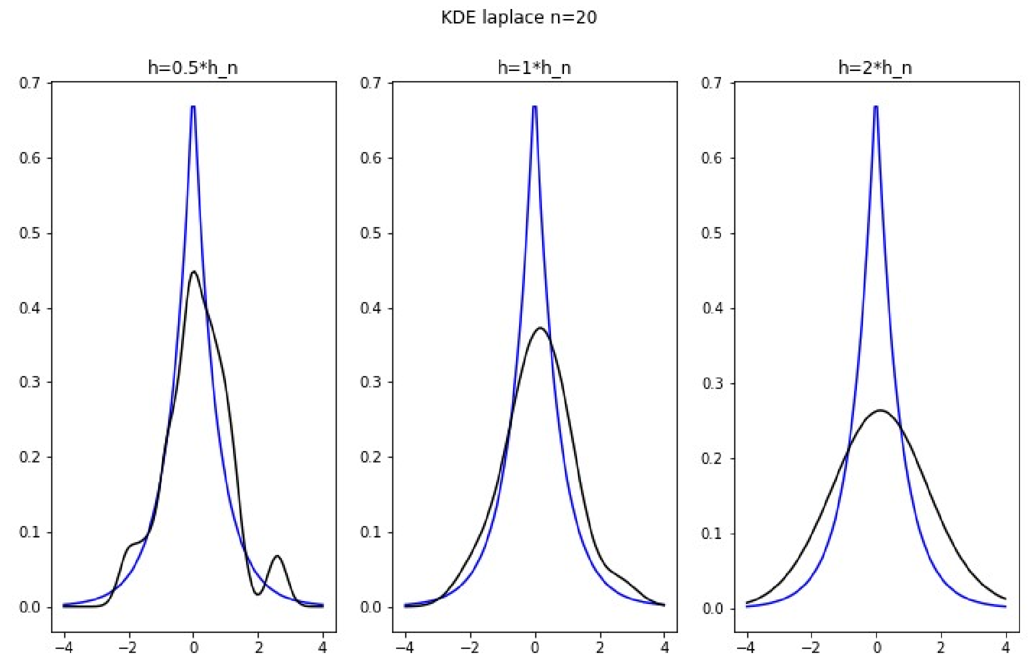


*Рис. 13 ЯФП для распределения Лапласа при n=60*

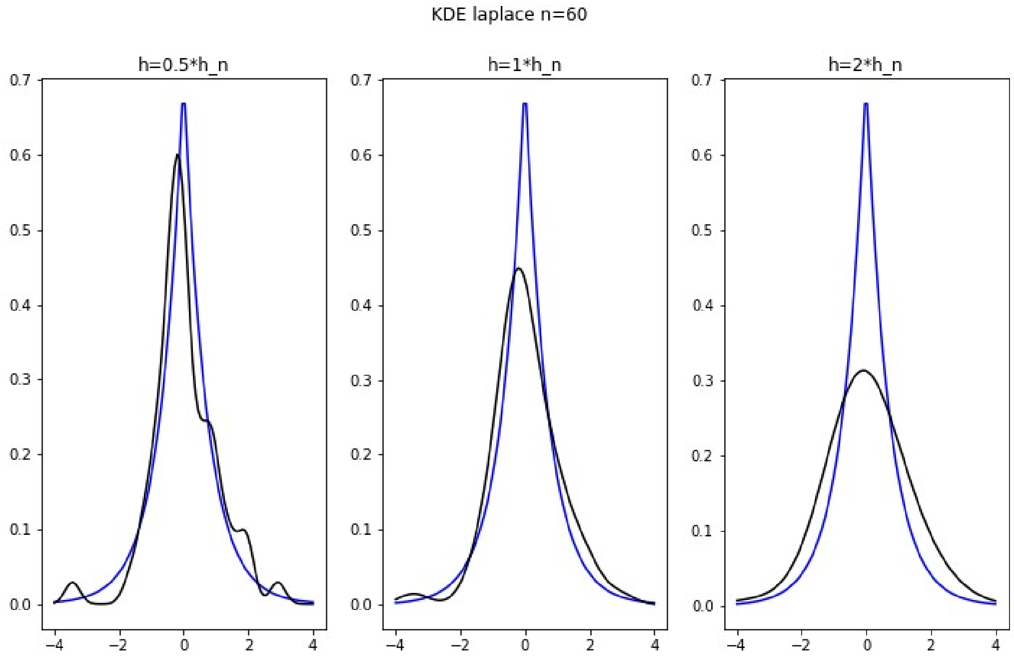



*Рис. 14 ЯФП для распределения Лапласа при n=100*

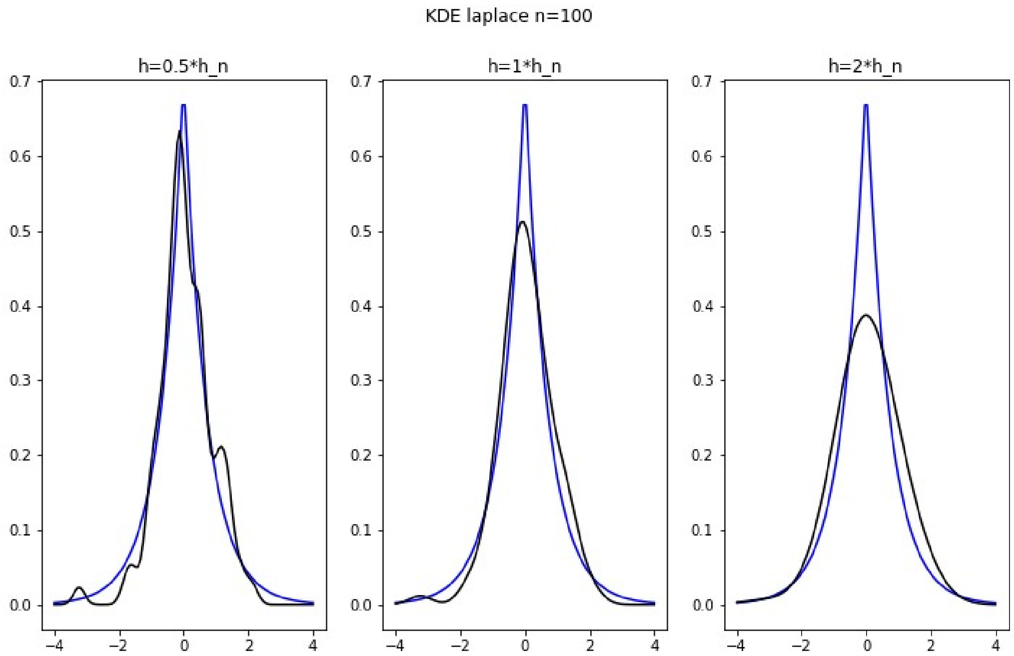


*Рис. 15 ЯФП для Равномерного распределения при n=20*

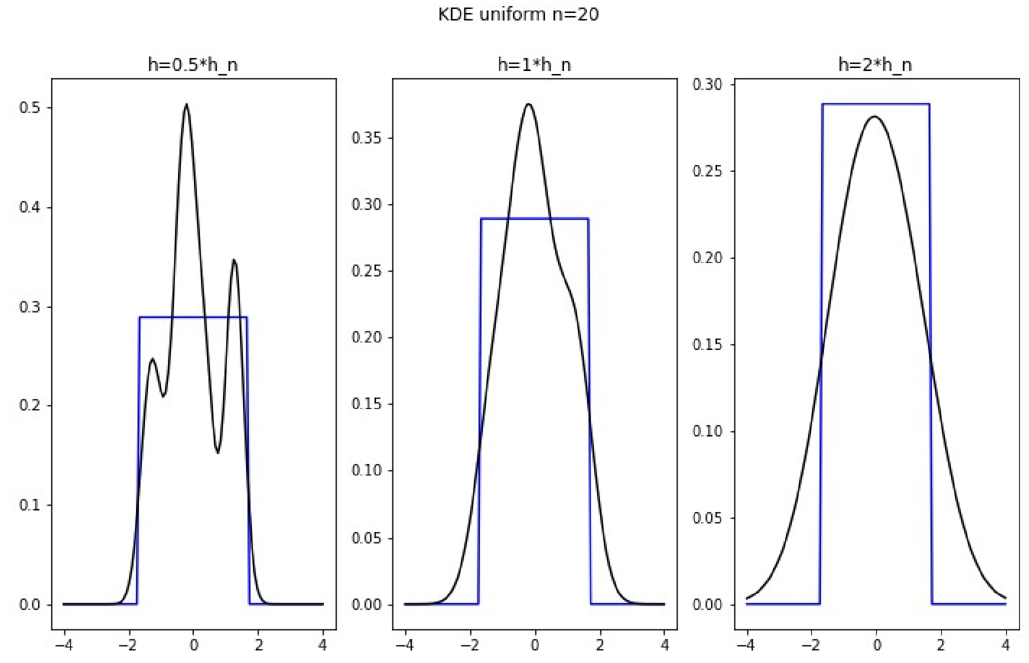



*Рис. 16 ЯФП для Равномерного распределения при n=60*

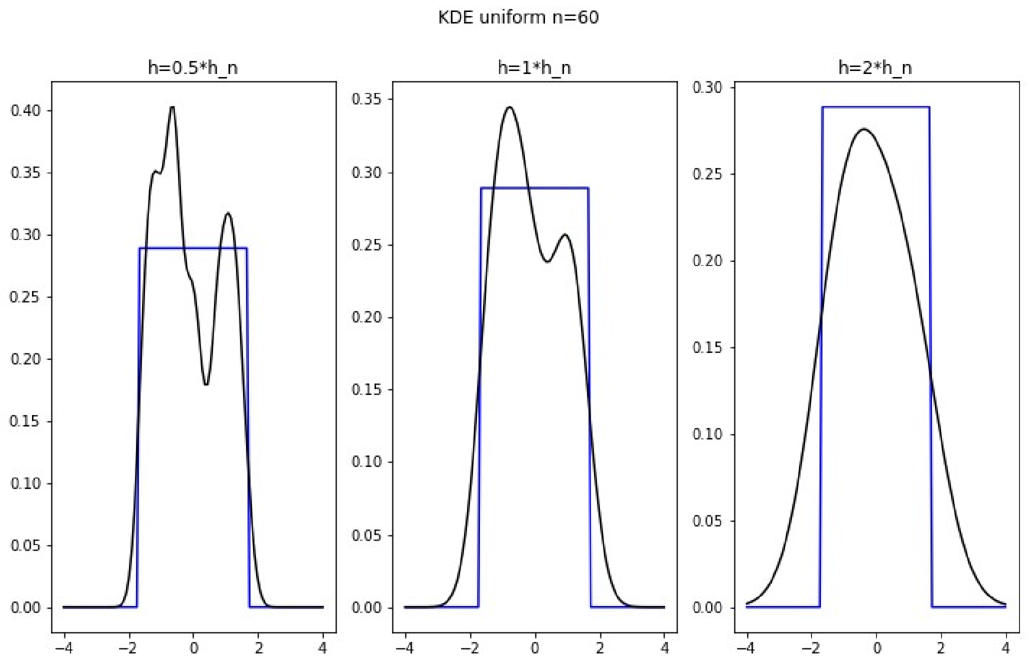


*Рис. 17 ЯФП для Равномерного распределения при n=100*

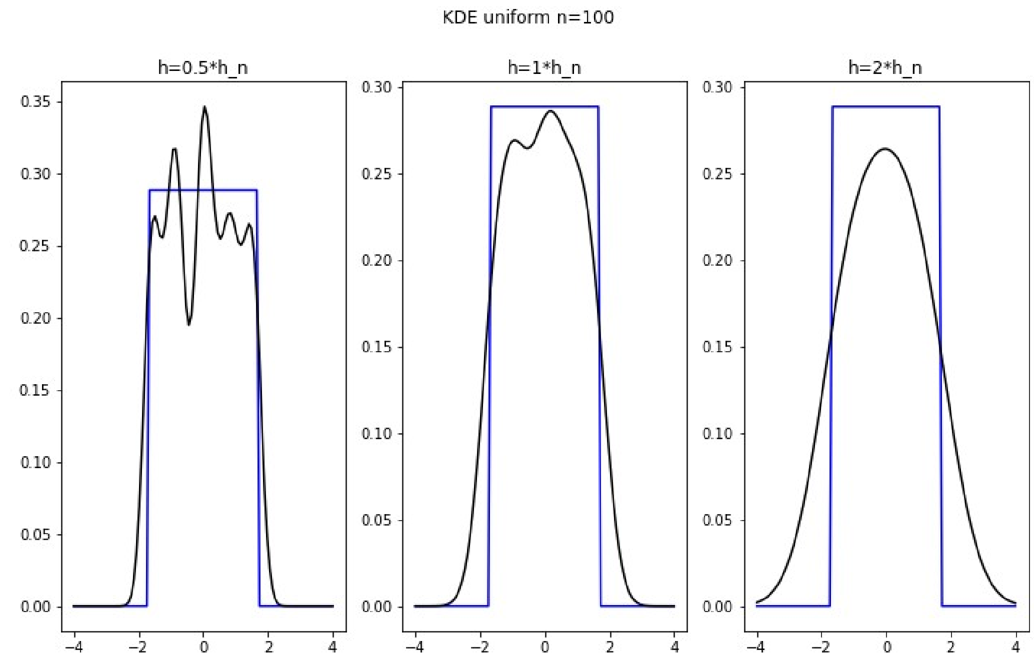



*Рис. 18 ЯФП для распределения Пуассона при n=20*

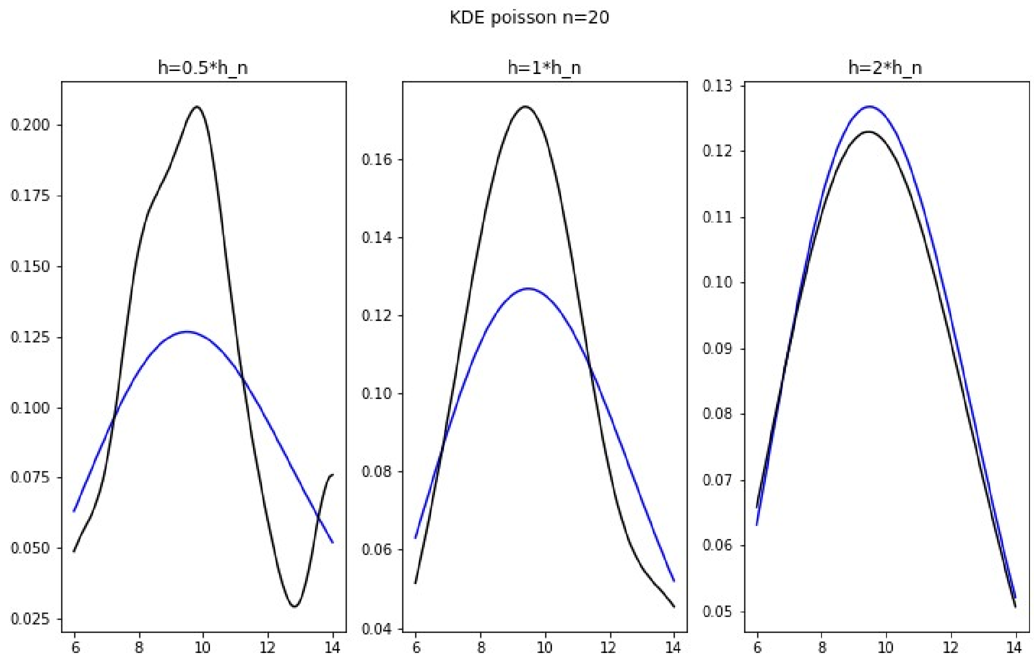


*Рис. 19 ЯФП для распределения Пуассона при n=60*

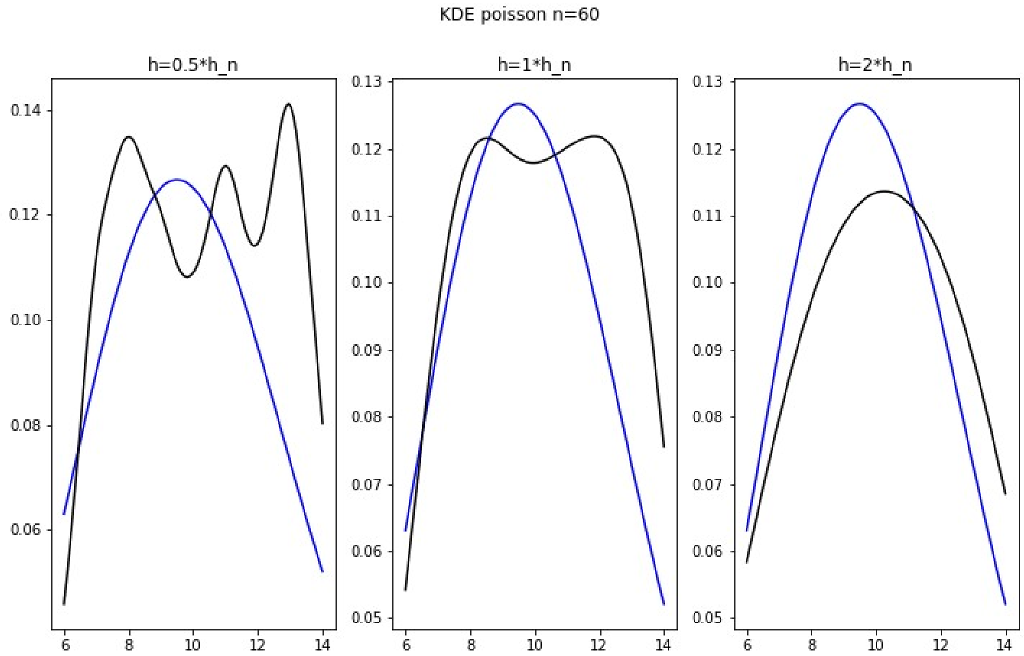
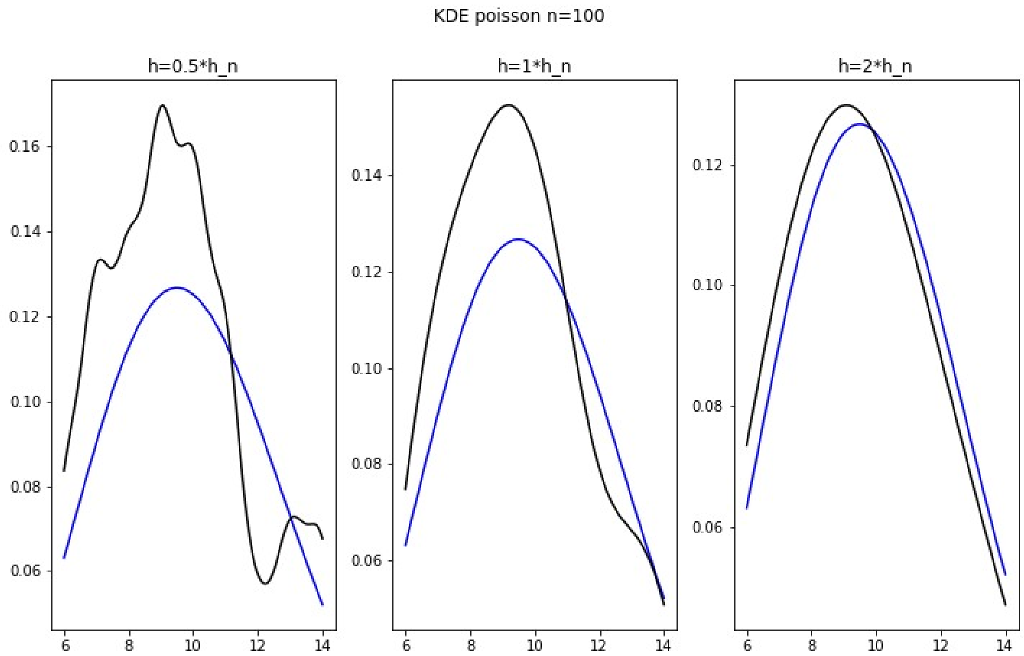



*Рис. 20 ЯФП для распределения Пуассона при n=100*




# 5 Выводы

ЭФР лучше приближает эталонную функцию на больших выборках.

Для Стандартного нормального и Стандартного распределения Коши при всех исследуемых $n$ были достигнуты наилучшие оценки при $h=h_n$. Для распределения Пуассона оценки получились лучше при $h=2h_n$.

## 6 Литература

[Основы работы с *numpy* (отдельная глава курса)]

[Документация по *scipy*]

## 7 Приложения

[Код лабораторной]