

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

1 Постановка задачи.....	4
2 Теория.....	4
3 Реализация.....	5
4 Результаты.....	6
4.1 Эмпирические функции распределения (ЭФР).....	6
4.2 Ядерные функции плотности (ЯФП).....	8
4.3 Таблицы. Средние модулей ошибок.....	15
5 Выводы.....	16
6 Литература.....	17
7 Приложения.....	17

Список иллюстраций и таблиц

ЭФР. Стандартное нормальное распределение.....	6
ЭФР. Стандартное распределение Коши.....	6
ЭФР. Распределение Лапласа.....	7
ЭФР. Равномерное распределение.....	7
ЭФР. Распределение Пуассона.....	8
ЯФП. Стандартное нормальное распределение при $n=20$	8
ЯФП. Стандартное нормальное распределение при $n=60$	9
ЯФП. Стандартное нормальное распределение при $n=100$	9
ЯФП. Стандартное распределение Коши при $n=20$	10
ЯФП. Стандартное распределение Коши при $n=60$	10
ЯФП. Стандартное распределение Коши при $n=100$	11
ЯФП. Распределение Лапласа при $n=20$	11
ЯФП. Распределение Лапласа при $n=60$	12
ЯФП. Распределение Лапласа при $n=100$	12
ЯФП. Равномерное распределение при $n=20$	13
ЯФП. Равномерное распределение при $n=60$	13
ЯФП. Равномерное распределение при $n=100$	14
ЯФП. Распределение Пуассона при $n=20$	14
ЯФП. Распределение Пуассона при $n=60$	15
ЯФП. Распределение Пуассона при $n=100$	15
Таблица. Стандартное нормальное распределение.....	15
Таблица. Стандартное распределение Коши.....	16
Таблица. Распределение Лапласа.....	16

Таблица. Равномерное распределение.....	16
Таблица. Распределение Пуассона.....	16

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:

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

2 Теория

Эмпирической функцией распределения (ЭФР) \hat{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 :

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

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

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

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:

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

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

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

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 вычисления описательных статистик, модуль *scipy* для функций и плотностей распределений, модуль *matplotlib* для построения и отображения графиков и гистограмм, модуль *math* для математических расчётов, модуль *pandas* для оптимального хранения статистических данных и функция *display* из модуля *IPython.display* для их корректного отображения в таблицах.

4 Результаты

4.1 Эмпирические функции распределения (ЭФР)

Рис. 1 ЭФР для Стандартного нормального распределения

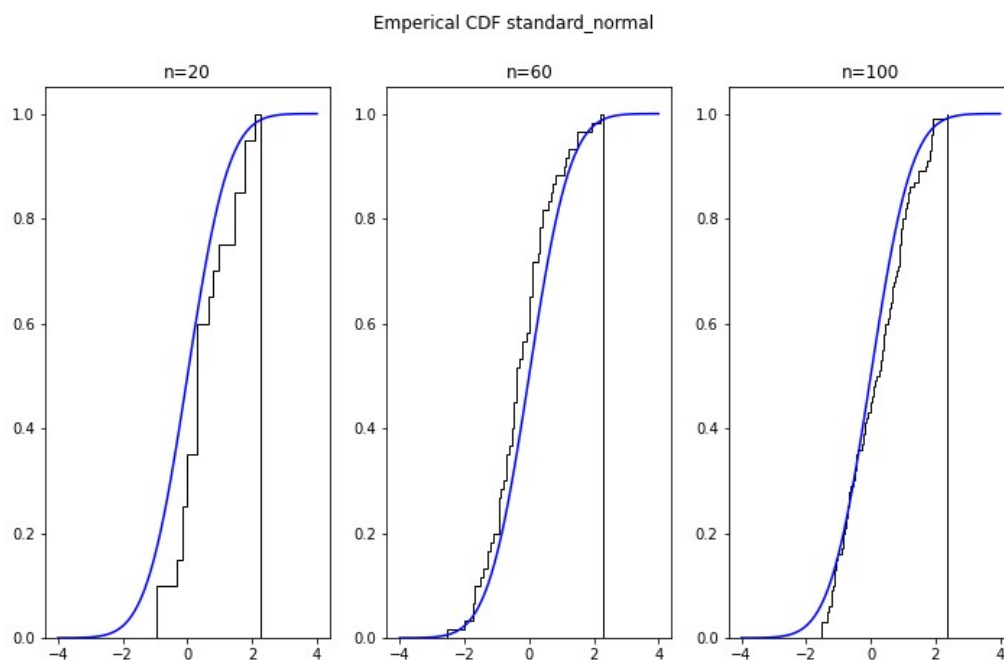


Рис. 2 ЭФР для Стандартного распределения Коши

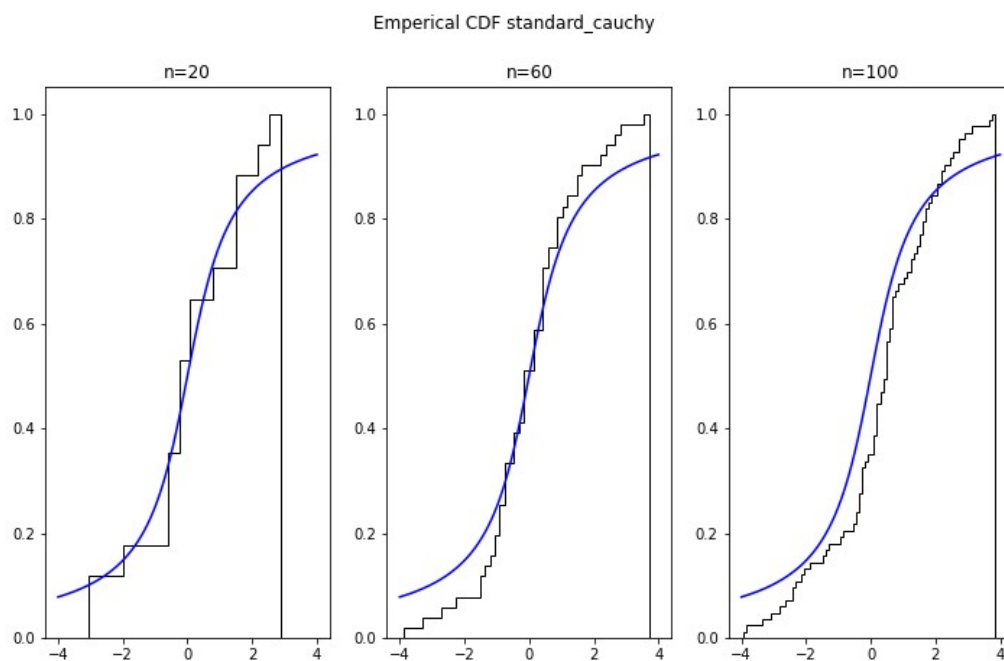


Рис. 3 ЭФР для распределения Лапласа

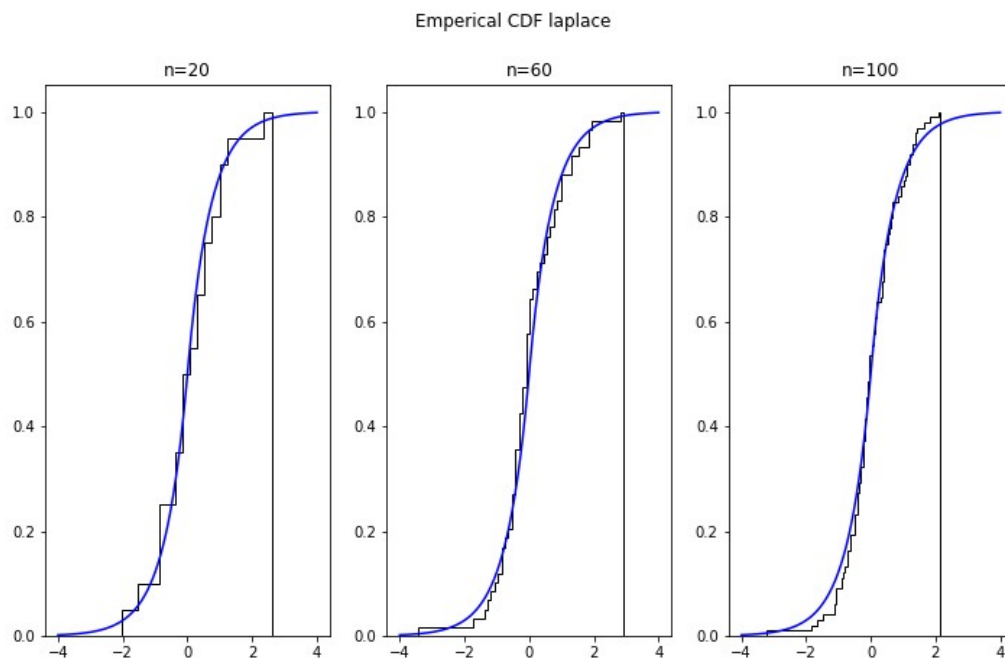


Рис. 4 ЭФР для Равномерного распределения

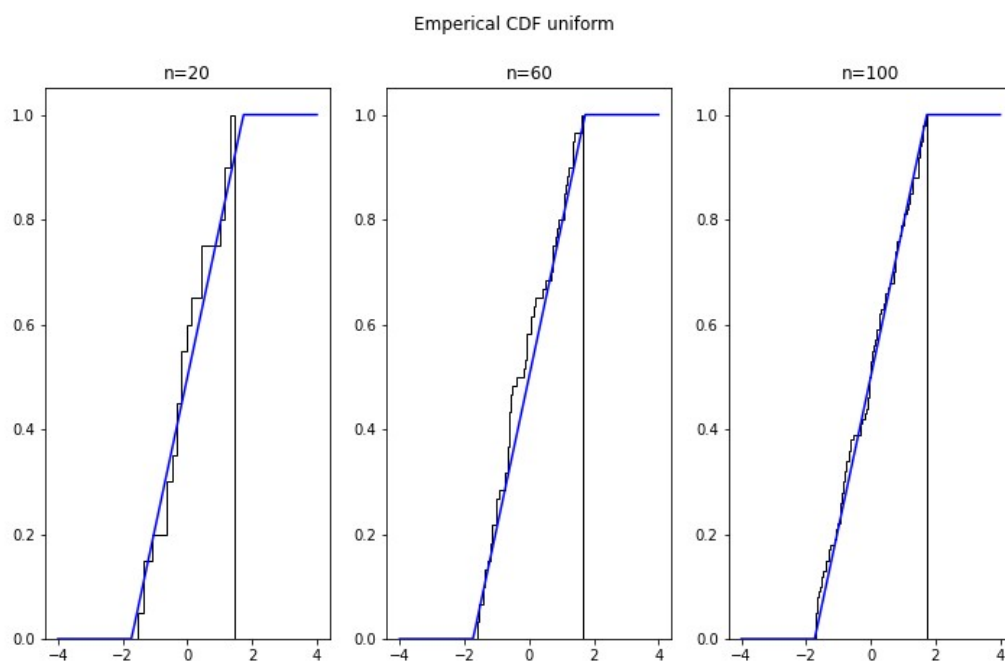
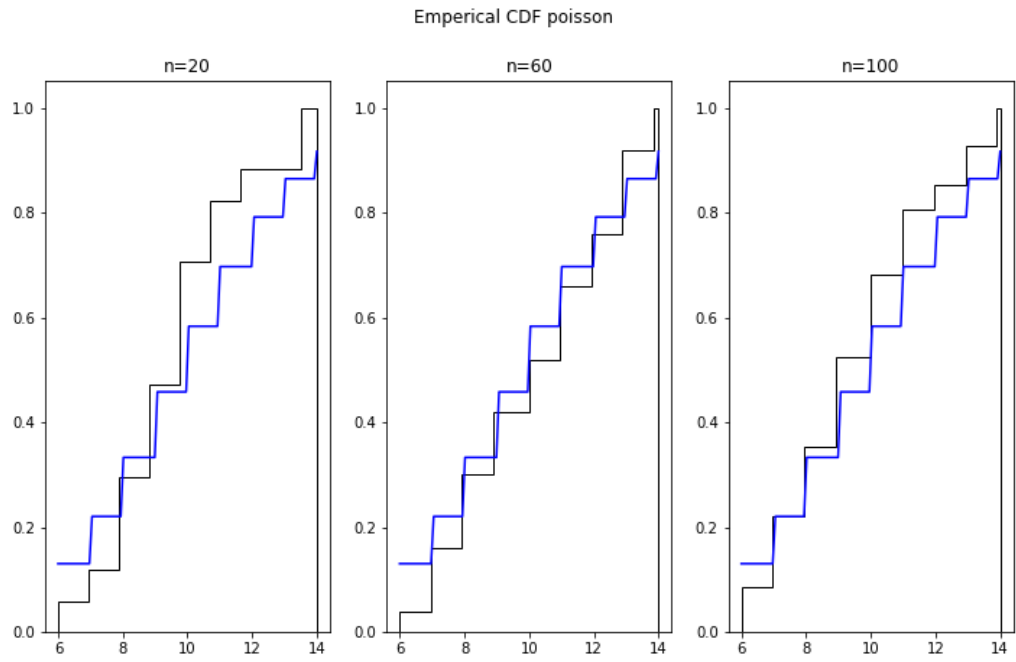


Рис. 5 ЭФР для распределения Пуассона



4.2 Ядерные функции плотности (ЯФП)

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

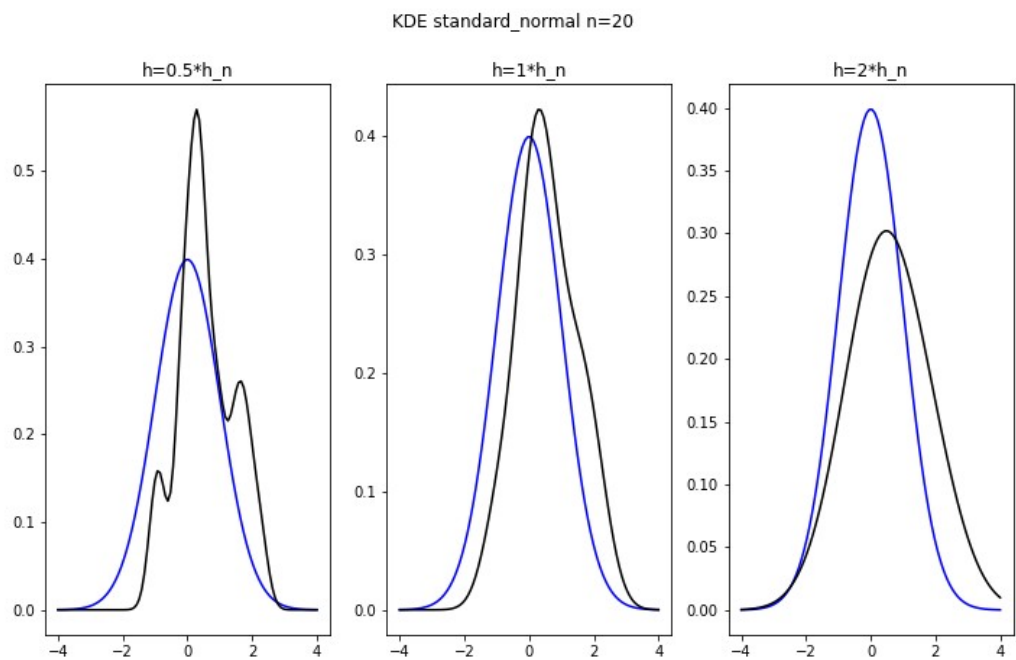


Рис. 7 ЯФП для Стандартного нормального распределения при $n=60$

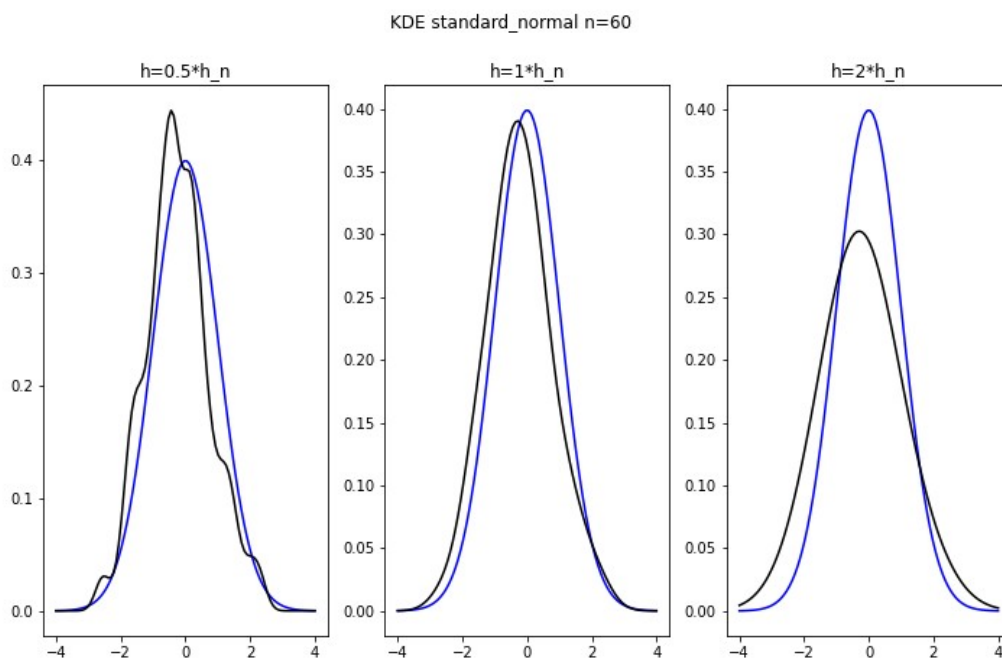


Рис. 8 ЯФП для Стандартного нормального распределения при $n=100$

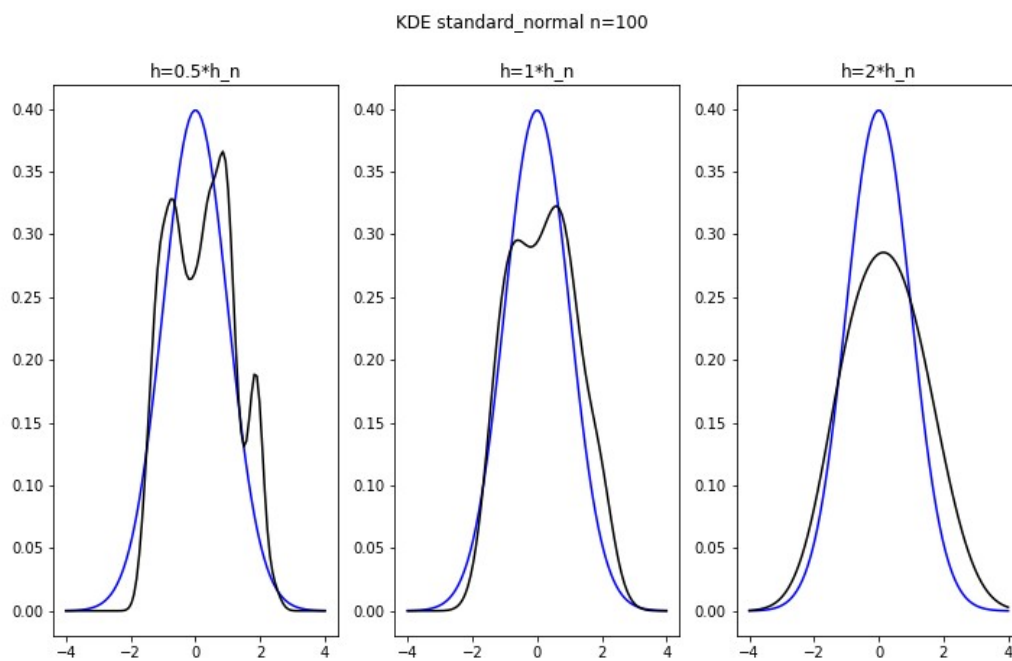


Рис. 9 ЯФП для Стандартного распределения Коши при $n=20$

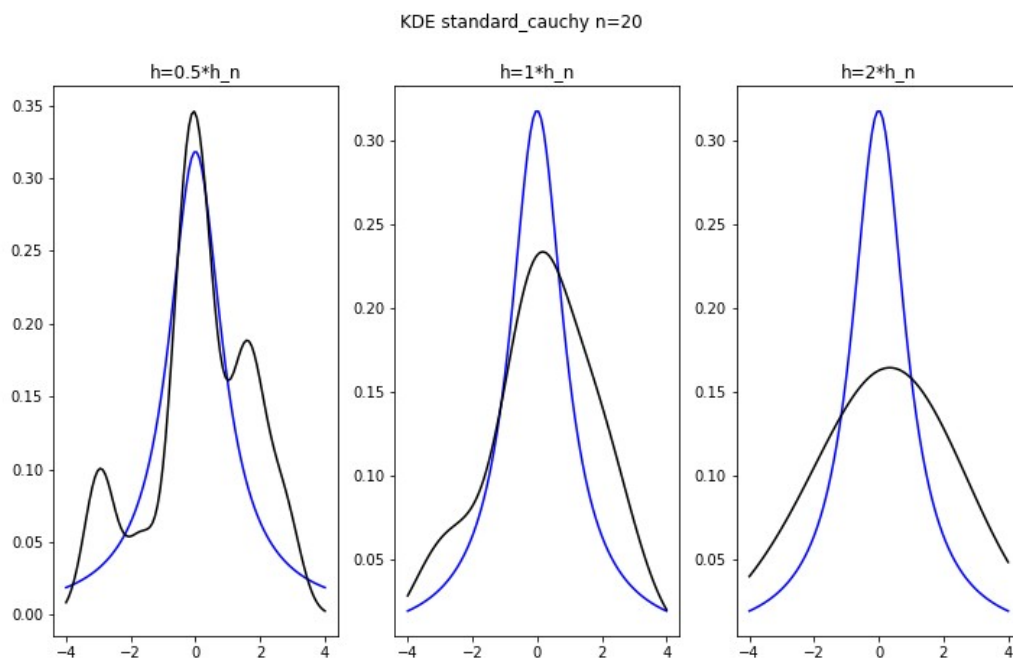


Рис. 10 ЯФП для Стандартного распределения Коши при $n=60$

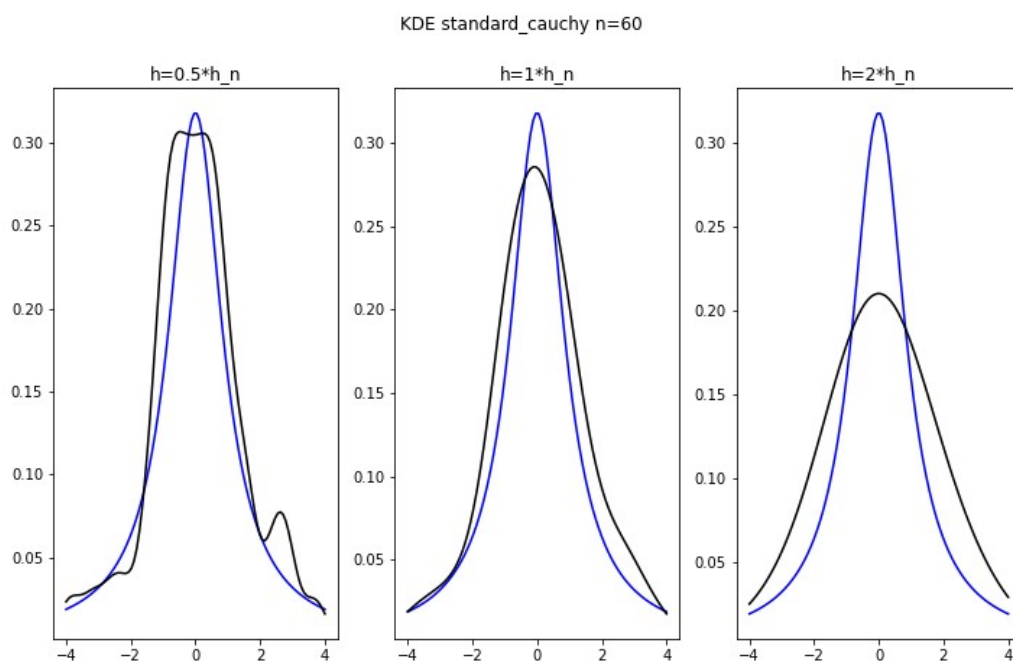


Рис. 11 ЯФП для Стандартного распределения Коши при $n=100$

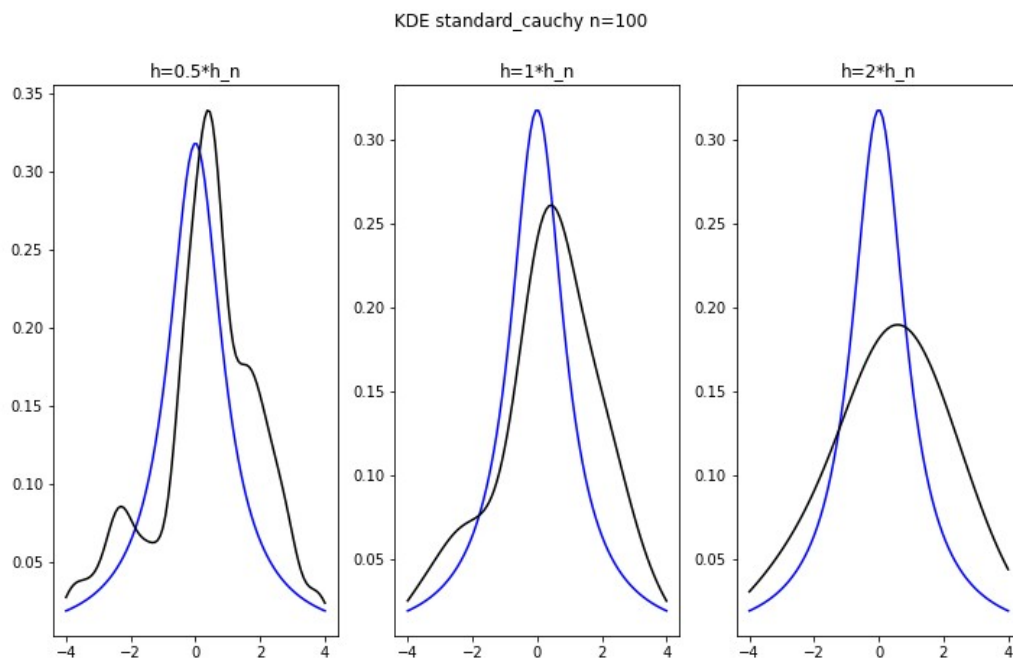


Рис. 12 ЯФП для распределения Лапласа при $n=20$

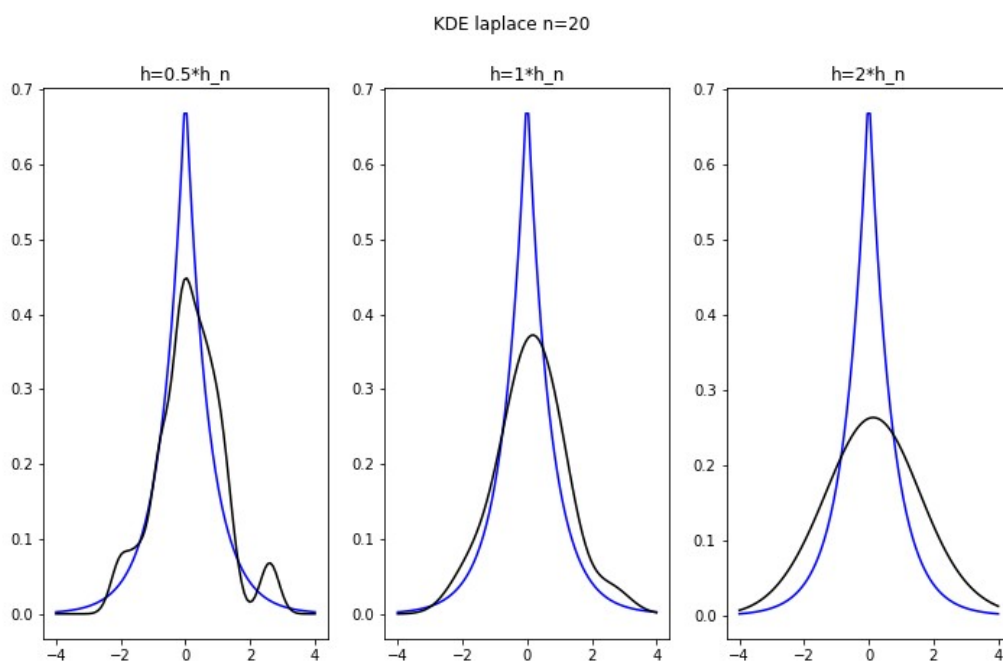


Рис. 13 ЯФП для распределения Лапласа при $n=60$

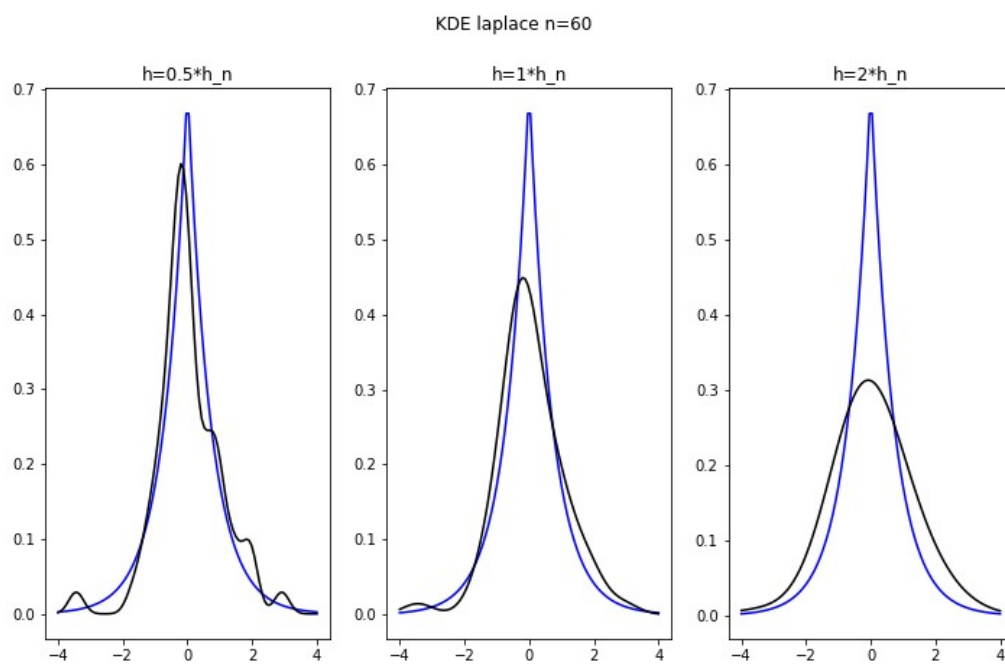


Рис. 14 ЯФП для распределения Лапласа при $n=100$

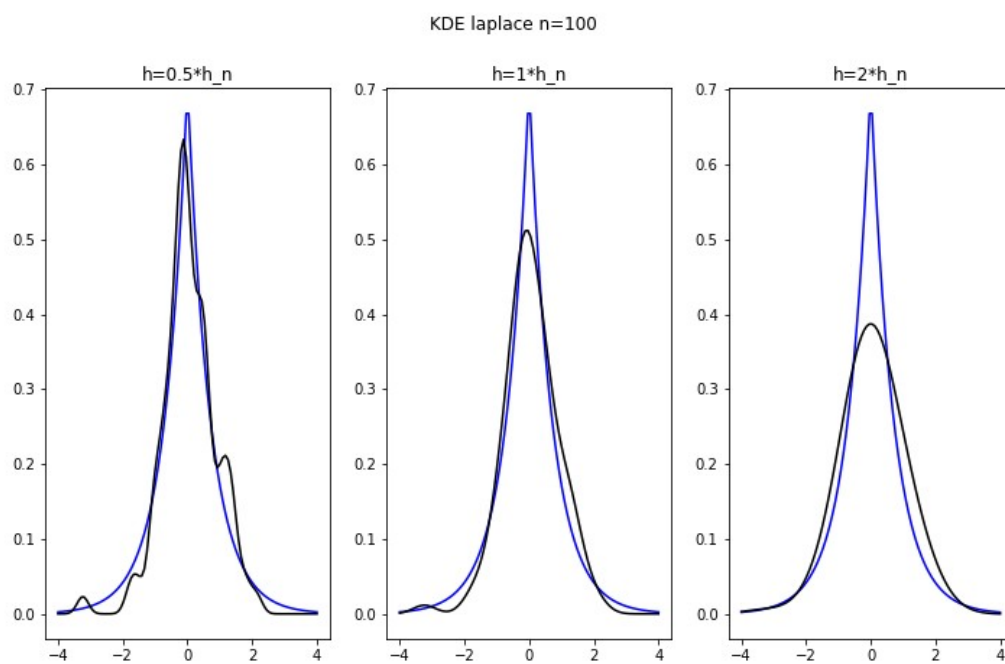


Рис. 15 ЯФП для Равномерного распределения при $n=20$

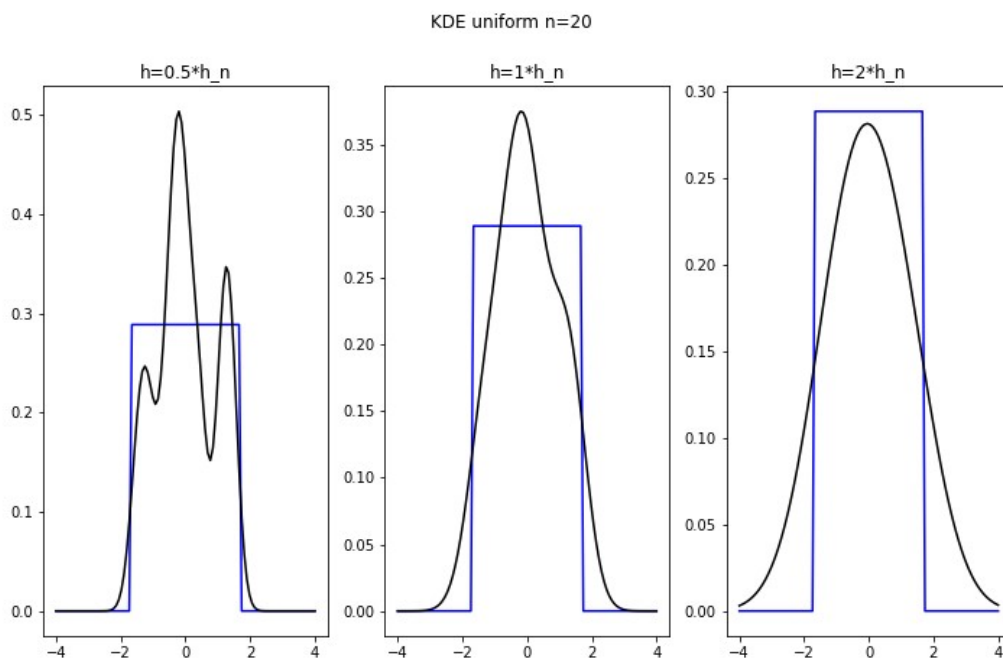


Рис. 16 ЯФП для Равномерного распределения при $n=60$

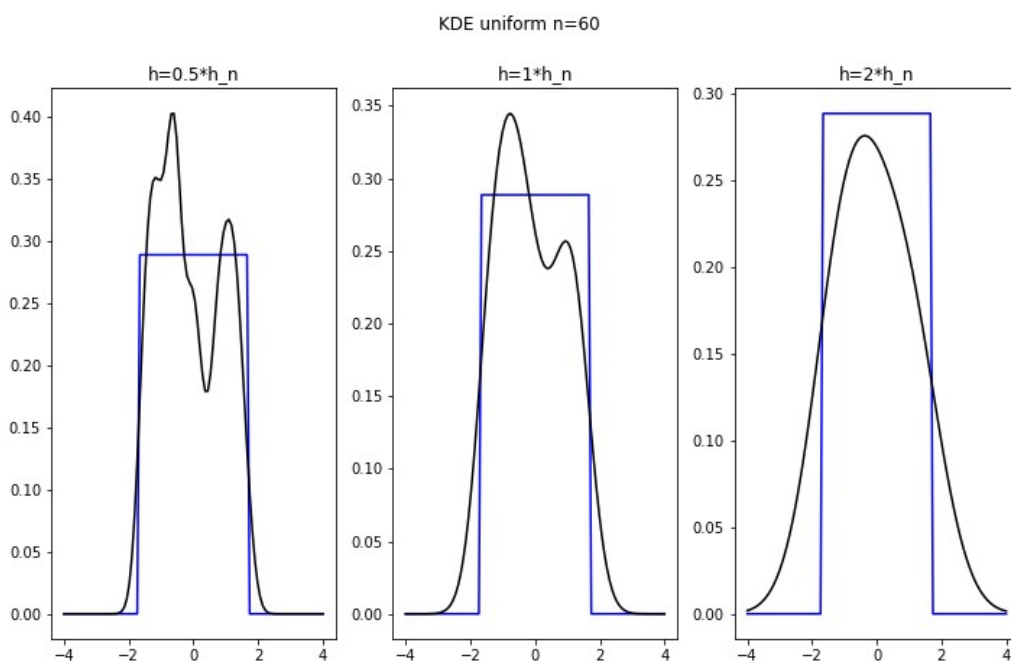


Рис. 17 ЯФП для Равномерного распределения при $n=100$

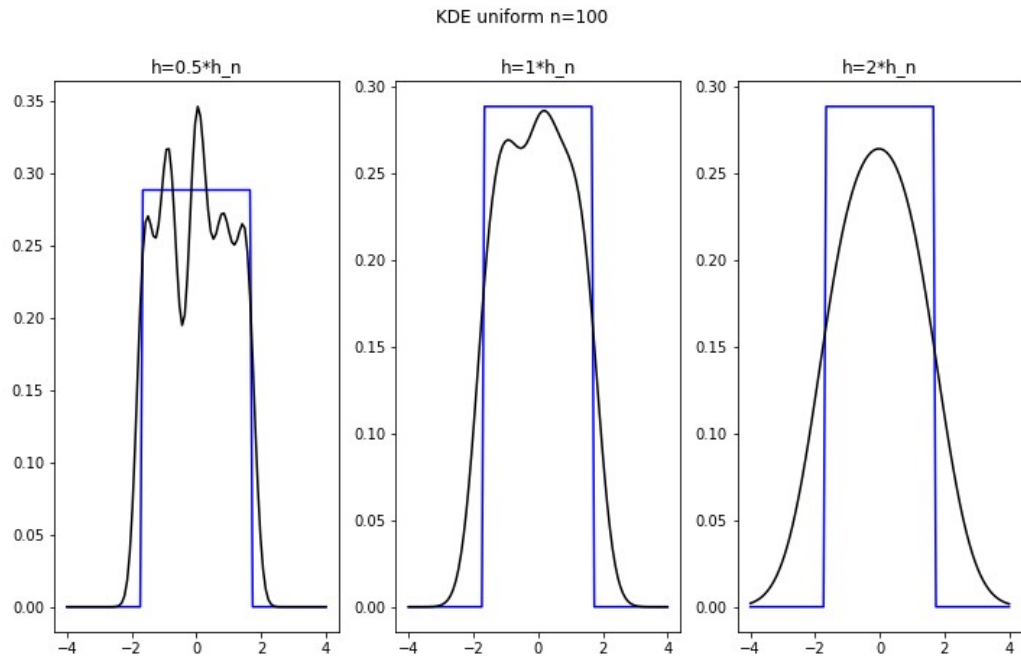


Рис. 18 ЯФП для распределения Пуассона при $n=20$

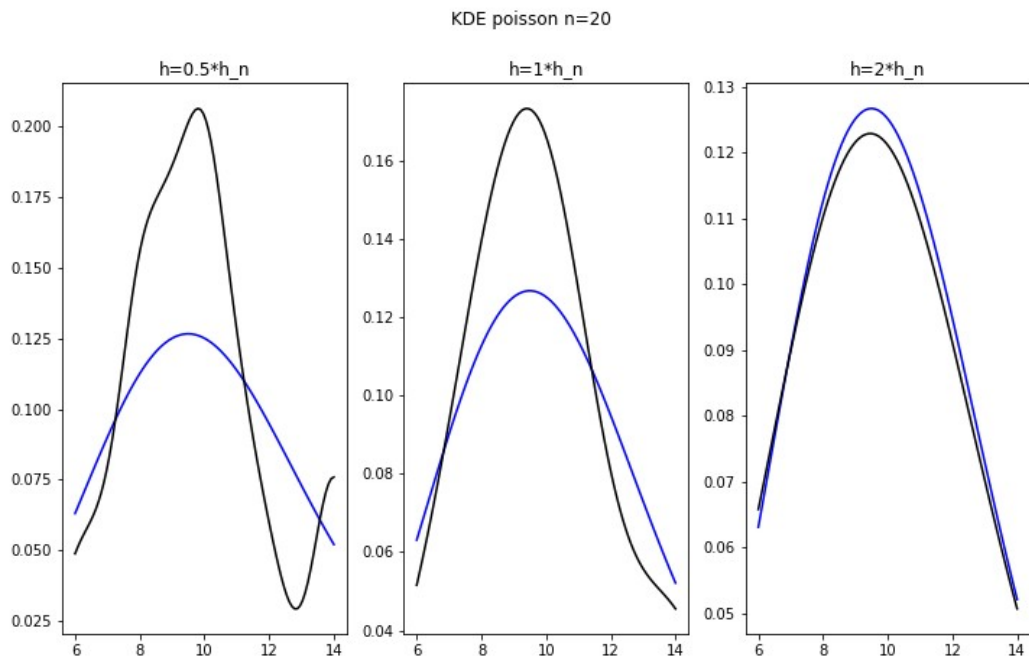


Рис. 19 ЯФП для распределения Пуассона при $n=60$

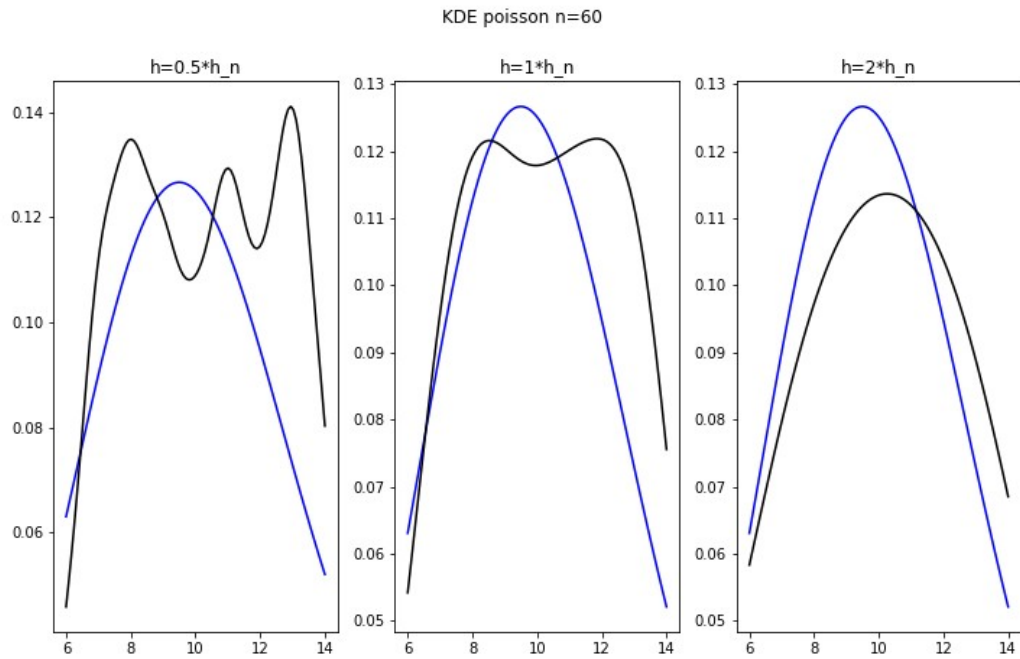
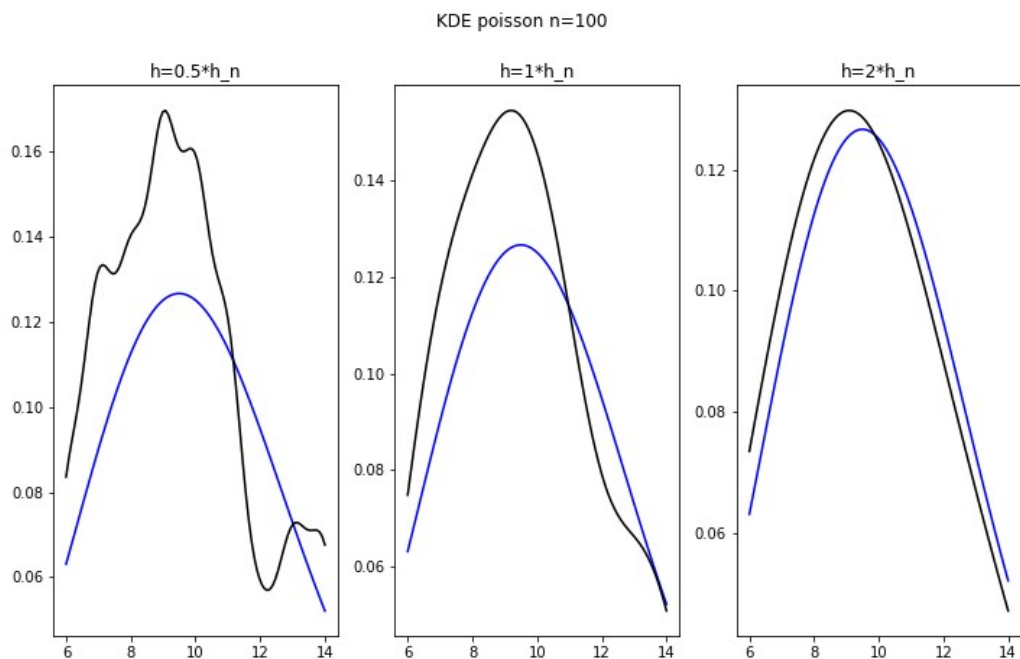


Рис. 20 ЯФП для распределения Пуассона при $n=100$



4.3 Таблицы. Средние модули ошибок

Строки – мощности выборок, столбцы – коэффициенты при h_n , маркером помечены минимальные значения.

Таблица 1 Стандартное нормальное распределение

$E(z)$	0.5	1.0	2.0
$n=20$	0.061887	0.051959	0.052813

n=60	0.028175	0.026624	0.039276
n=100	0.035016	0.028130	0.037715

Таблица 2 Стандартное распределение Коши

E(z)	0.5	1.0	2.0
n=20	0.036591	0.037148	0.052669
n=60	0.022031	0.021232	0.039087
n=100	0.046387	0.042639	0.048923

Таблица 3 Распределение Лапласа

E(z)	0.5	1.0	2.0
n=20	0.035431	0.040874	0.067724
n=60	0.032295	0.030991	0.052448
n=100	0.024103	0.023139	0.034833

Таблица 4 Равномерное распределение

E(z)	0.5	1.0	2.0
n=20	0.041468	0.039627	0.052521
n=60	0.030078	0.033001	0.051592
n=100	0.025169	0.030812	0.054162

Таблица 5 Распределение Пуассона

E(z)	0.5	1.0	2.0
n=20	0.036118	0.020863	0.002852
n=60	0.022694	0.013810	0.011530
n=100	0.025138	0.017411	0.006402

5 Выводы

Чем больше выборка, тем лучше ЭФР приближает эталонную функцию распределения.

Исходя из данных в таблицах, для распределений

1. Стандартного нормального: лучшие оценки при $h=h_n$
2. Стандартного Коши: для малых выборок лучшая оценка будет при $h=\frac{1}{2}h_n$, при больших – $h=h_n$
3. Лапласа: для малых будет $h=\frac{1}{2}h_n$, для больших – $h=h_n$
4. Равномерного: для малых – $h=h_n$, для больших – $h=\frac{1}{2}h_n$
5. Пуассона: лучшие оценки при $h=2h_n$

6 Литература

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

[Документация по *scipy*](#)

[Pandas обзор](#)

7 Приложения

[Код лабораторной](#)