

Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра «Прикладная математика»

**ОТЧЁТ ПО ЛАБОРАТОРНЫМ РАБОТАМ
ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА»**

Выполнил
студент группы 3630102/70201

Крупкина Дарья

Проверил
к. ф.-м. н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2020

Содержание

1	Постановка задачи	3
1.1	Задание	3
2	Теория	3
2.1	Распределения	3
2.2	Эмпирическая функция распределения	4
2.2.1	Статистический ряд	4
2.2.2	Определение	4
2.2.3	Описание	4
2.3	Оценки плотности вероятности	5
2.3.1	Определение	5
2.3.2	Ядерные оценки	5
3	Реализация	5
4	Результаты	6
4.1	Эмпирическая функция распределения	6
4.2	Ядерные оценки плотности распределения	8
5	Обсуждение	15
6	Приложения	16

Список таблиц

1	Статистический ряд	4
2	Статистический ряд	4

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

Для 5 распределений:

1. $N(x, 0, 1)$ – нормальное распределение
2. $C(x, 0, 1)$ – распределение Коши
3. $L(x, 0, \frac{1}{\sqrt{2}})$ – распределение Лапласа
4. $P(k, 10)$ – распределение Пуассона
5. $U(x, -\sqrt{3}, \sqrt{3})$ – равномерное распределение

1.1 Задание

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 Распределения

1. Нормальное распределение

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

2. Распределение Коши

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

3. Распределение Лапласа

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

4. Распределение Пуассона

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

5. Равномерное распределение

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

2.2 Эмпирическая функция распределения

2.2.1 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. Статистический ряд обычно записывается в виде таблицы

z	z_1	z_1	...	z_k
n	n_1	n_2	...	n_k

Таблица 1: Статистический ряд

2.2.2 Определение

Эмпирической (выборочной) функцией распределения (э. ф. р.) называется относительная частота события $X < x$, полученная по данной выборке:

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

2.2.3 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

$$F^*(x) = \frac{1}{n} \sum_{z_i < x} n_i \quad (7)$$

$F^*(x)$ — функция распределения дискретной случайной величины X_* , заданной таблицей распределения

X^*	z_1	z_1	...	z_k
P	$\frac{n_1}{n}$	$\frac{n_2}{n}$...	$\frac{n_k}{n}$

Таблица 2: Статистический ряд

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

$$F_n^*(x) \approx F_X(x) \quad (8)$$

2.3 Оценки плотности вероятности

2.3.1 Определение

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

$$f(x) \approx \hat{f}(x) \quad (9)$$

2.3.2 Ядерные оценки

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

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) \quad (10)$$

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

$$h_n \xrightarrow{n \rightarrow \infty} 0, \quad \frac{h_n}{n^{-1}} \xrightarrow{n \rightarrow \infty} \infty \quad (11)$$

Такие оценки называются непрерывными ядерными [1, с. 421-423].

Замечание. Свойство, означающее сближение оценки с оцениваемой величиной при $n \rightarrow \infty$ в каком-либо смысле, называется состоятельностью оценки.

Если плотность $f(x)$ кусочно-непрерывная, то ядерная оценка плотности является состоятельной при соблюдении условий, накладываемых на параметр сглаживания h_n , а также на ядро $K(u)$.

Гауссово (нормальное) ядро [2, с. 38]

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (12)$$

Правило Сильвермана [2, с. 44]

$$h_n = 1.06\hat{\sigma}n^{-\frac{1}{5}}, \quad (13)$$

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

3 Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm. Используются библиотеки `scipy` для генерации выборки и задания функции распределения и плотности вероятности, `matplotlib` для построения графиков. Исходный код лабораторной работы приведён в приложении в виде ссылки на репозиторий GitHub.

4 Результаты

4.1 Эмпирическая функция распределения

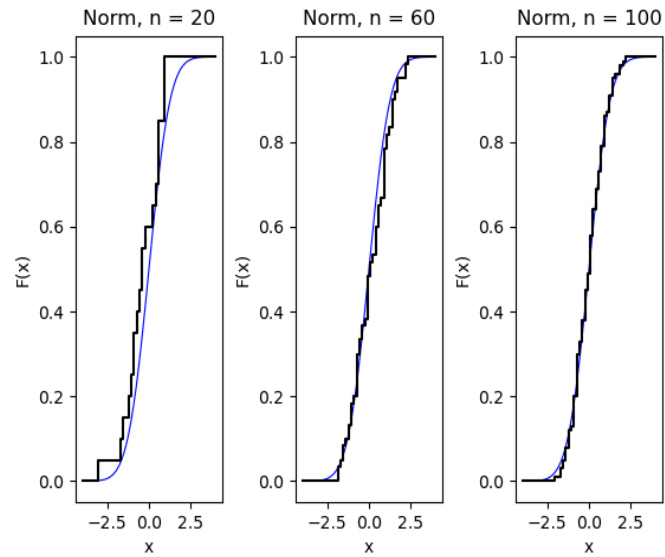


Рис. 1: Нормальное распределение

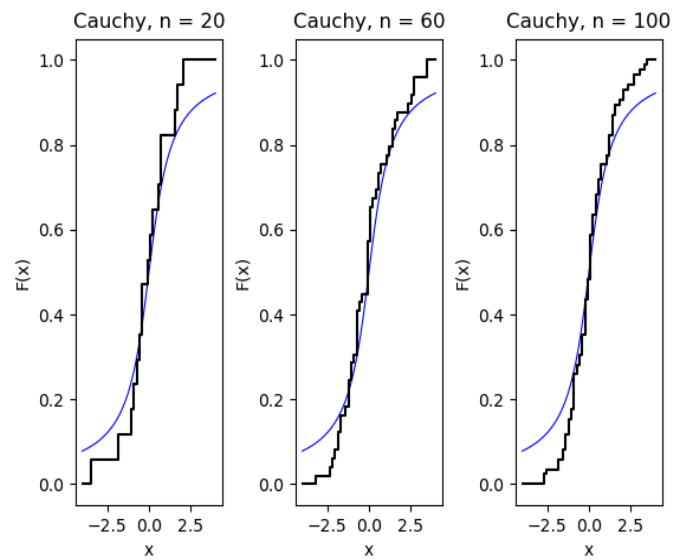


Рис. 2: Распределение Коши

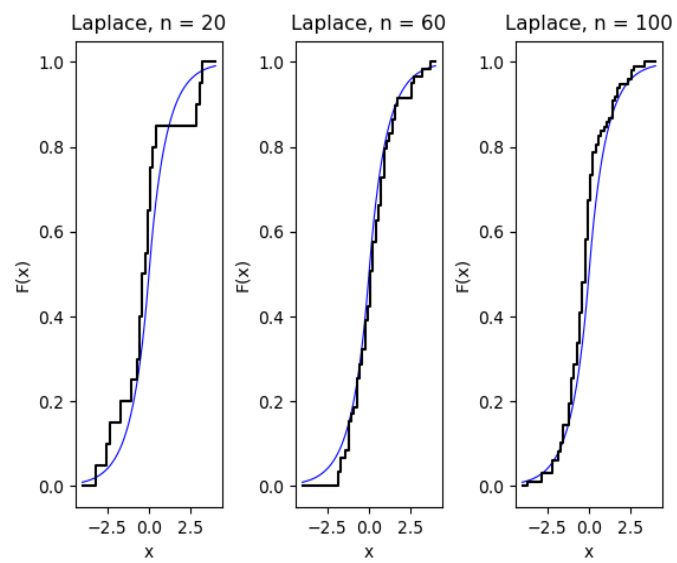


Рис. 3: Распределение Лапласа

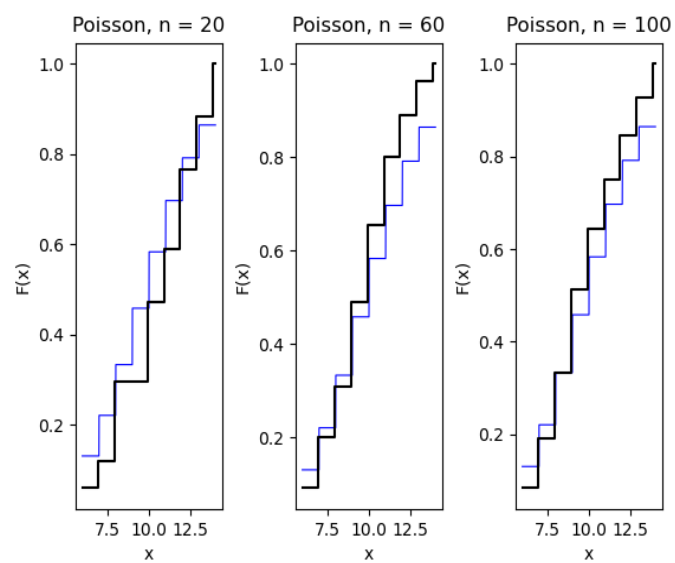


Рис. 4: Распределение Пуассона

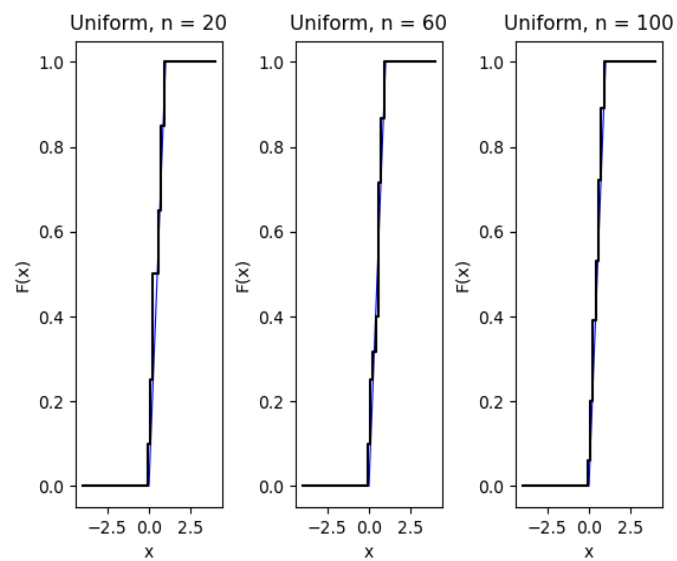


Рис. 5: Равномерное распределение

4.2 Ядерные оценки плотности распределения

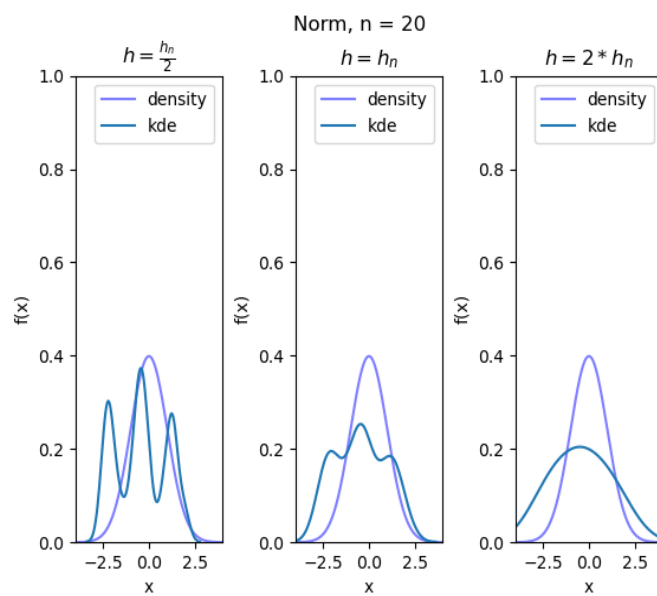


Рис. 6: Нормальное распределение, $n = 20$

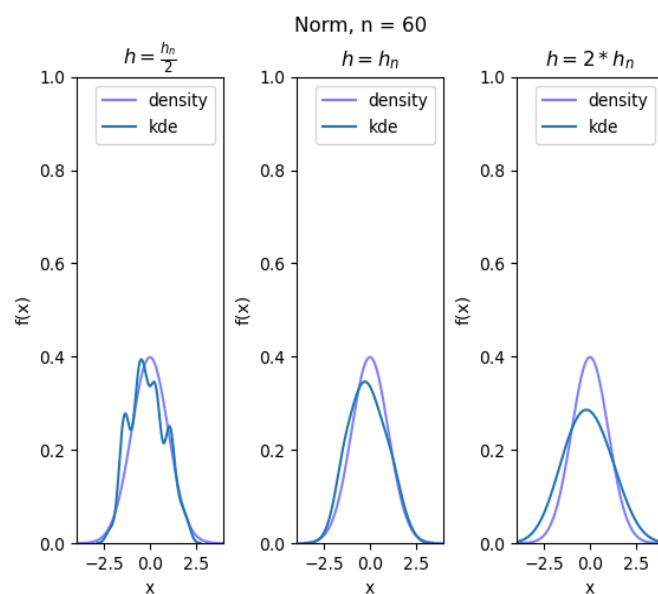


Рис. 7: Нормальное распределение, $n = 60$

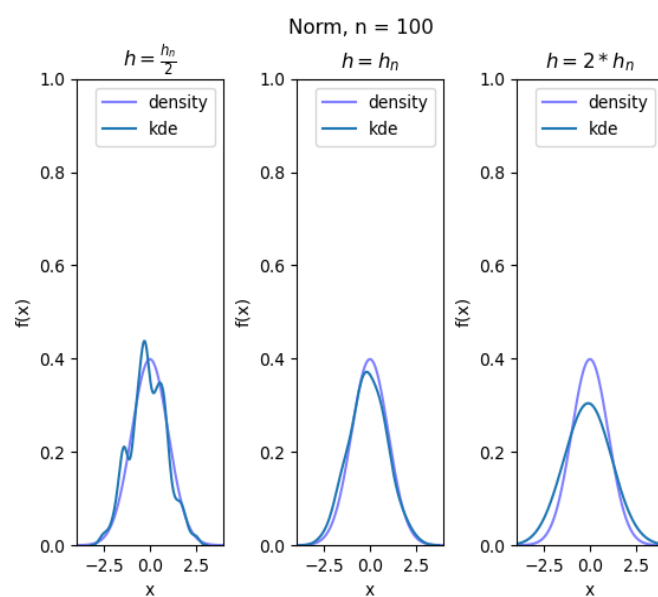


Рис. 8: Нормальное распределение, $n = 100$

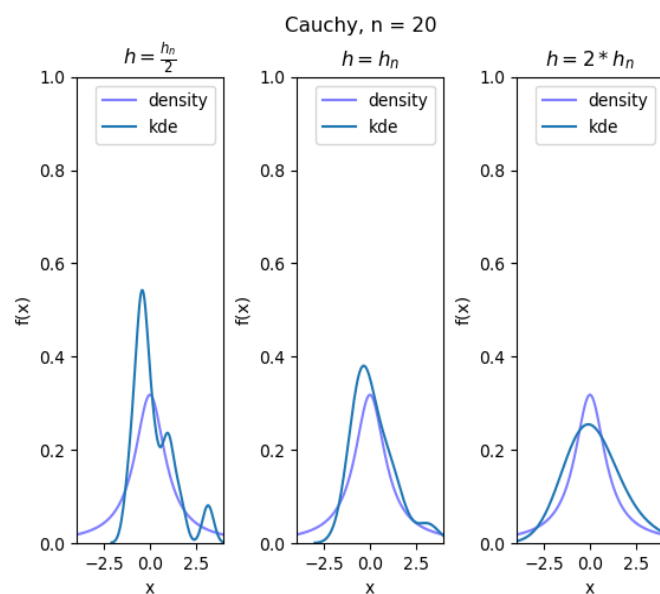


Рис. 9: Распределение Коши, $n = 20$

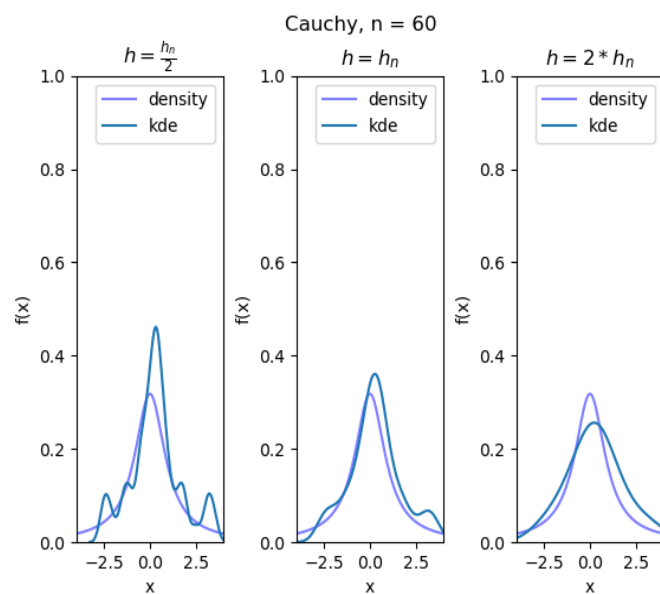


Рис. 10: Распределение Коши, $n = 60$

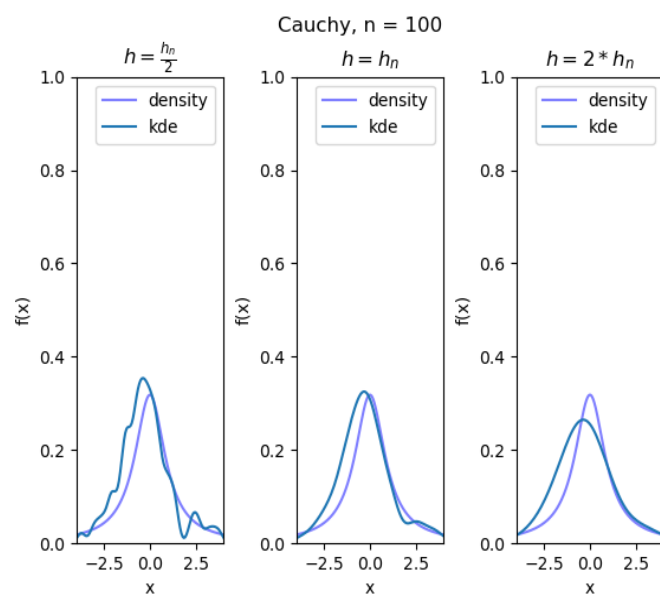


Рис. 11: Распределение Коши, $n = 100$

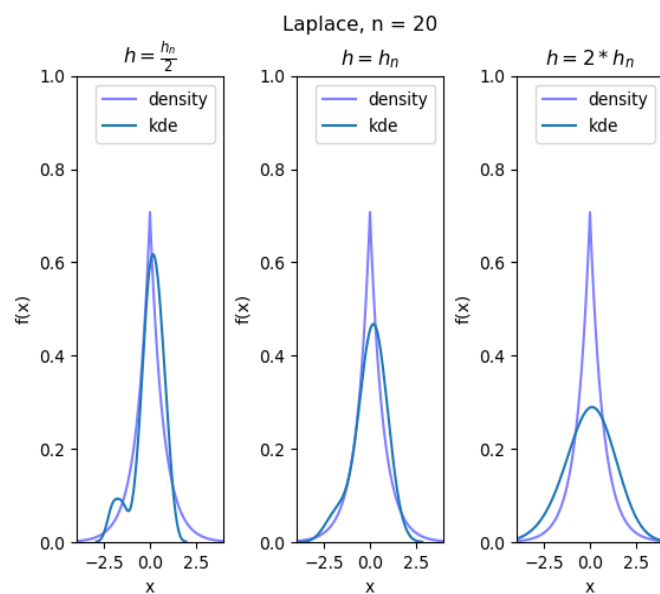


Рис. 12: Распределение Лапласа, $n = 20$

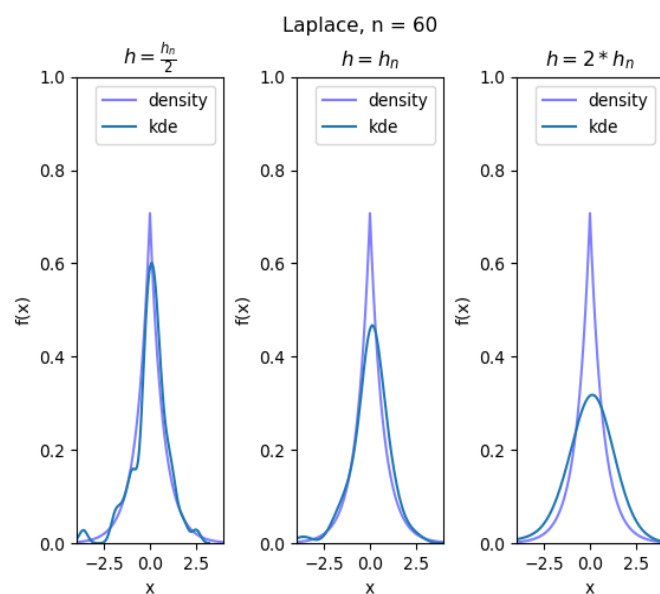


Рис. 13: Распределение Лапласа, $n = 60$

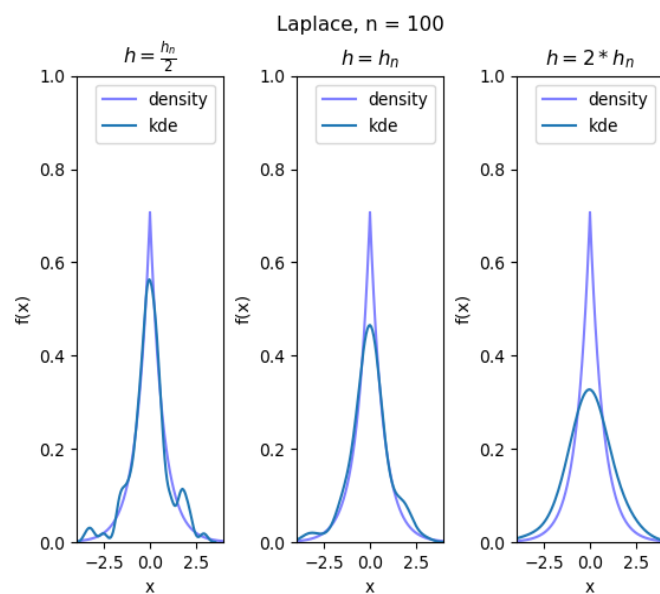


Рис. 14: Распределение Лапласа, $n = 100$

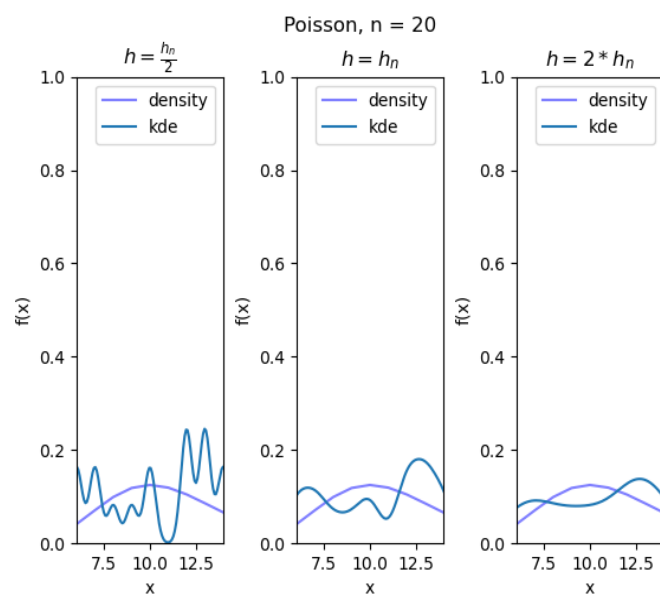


Рис. 15: Распределение Пуассона, $n = 20$

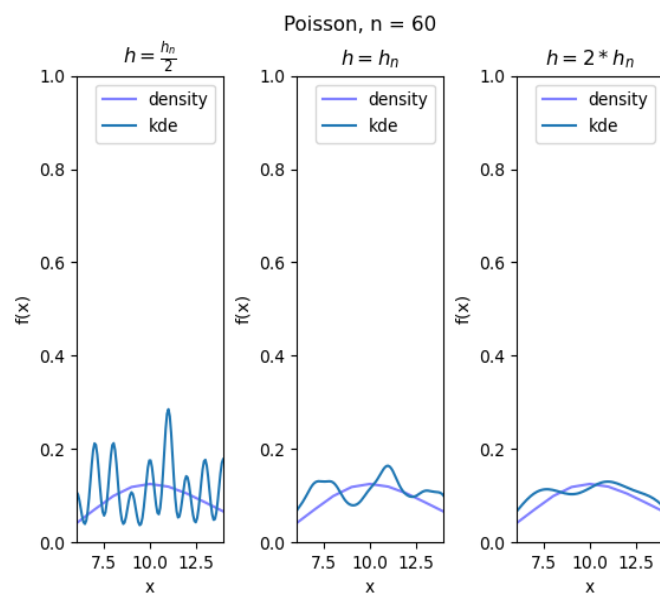


Рис. 16: Распределение Пуассона, $n = 60$

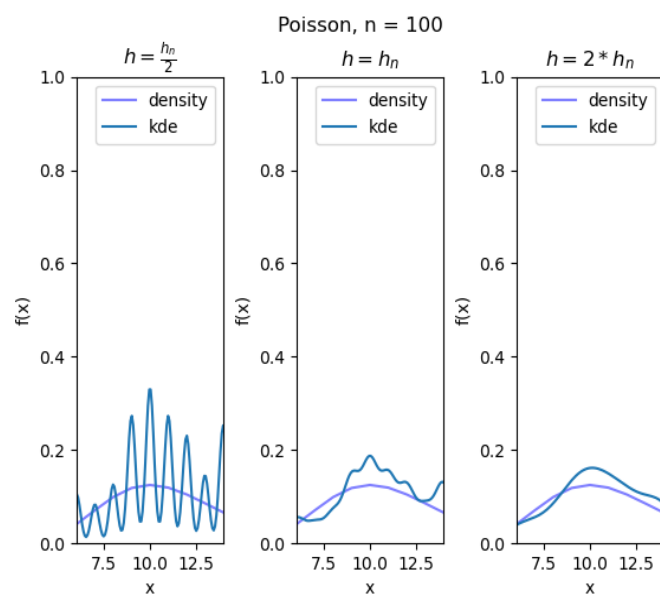


Рис. 17: Распределение Пуассона, $n = 100$

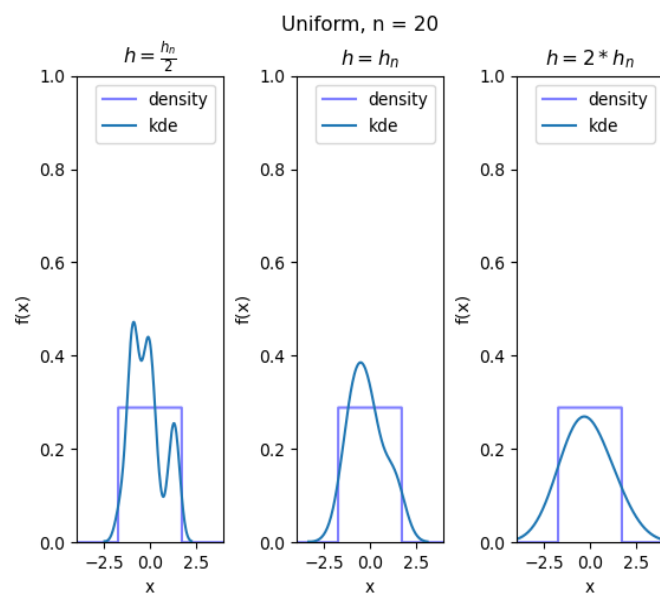


Рис. 18: Равномерное распределение, $n = 20$

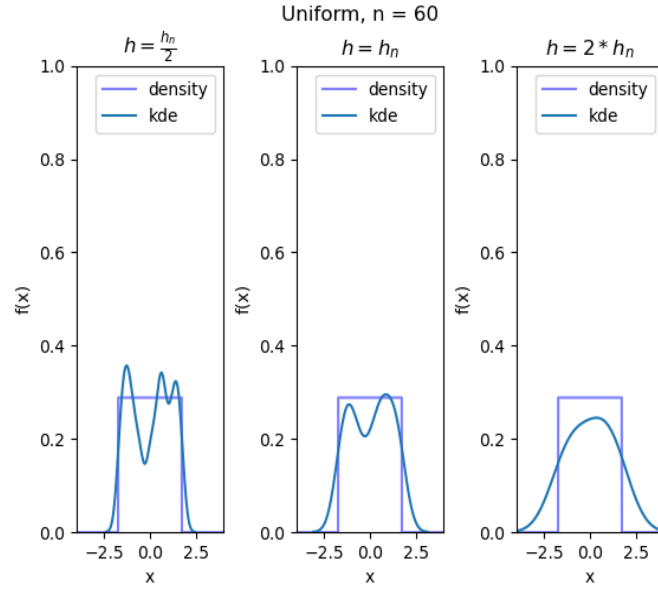


Рис. 19: Равномерное распределение, $n = 60$

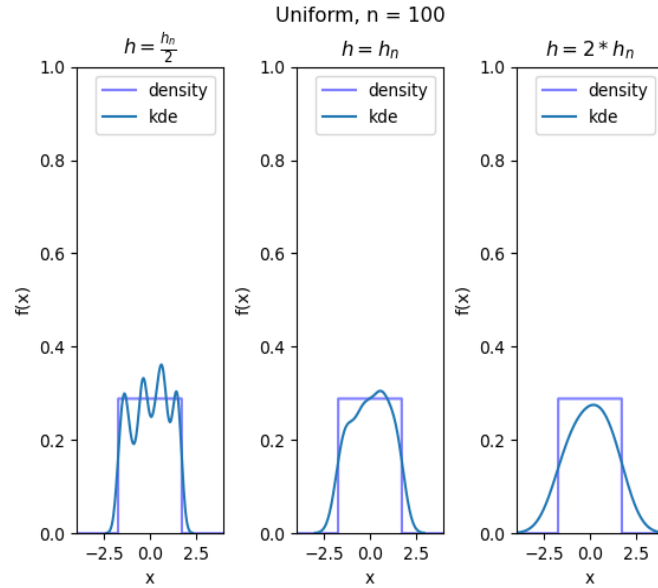


Рис. 20: Равномерное распределение, $n = 100$

5 Обсуждение

По полученным результатам можно сделать вывод, с увеличением размера выборки эмпирическая и теоретическая функции распределения сближаются. При этом для распределения Пуассона эти функции имеют наибольшее отклонение друг от друга. В качестве вывода для ядерных оценок можно сказать, что с ростом размера выборки для всех h функция плотности распределения и ядерная оценка сближаются. Для

части распределений (нормального, Коши) точнее оказывается результат при $h = h_n$, в то время как для равномерного распределения, распределения Пуассона при $h = 2h_n$; для распределения Лапласа - при $h = \frac{h_n}{2}$. Распределение Пуассона имеет результат с наибольшим отклонением ядерной оценки от плотности распределения.

6 Приложения

Код программы на GitHub, URL: <https://github.com/DariaKrup/Statistics>

Список литературы

- [1] Вероятностные разделы математики. Учебник для бакалавров технических направлений. //Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.
- [2] Анатольев, Станислав (2009) «Непараметрическая регрессия», Квантиль, №7, стр. 37-52.