

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра «Телематика (при ЦНИИ РТК)»

Отчет по лабораторной работе

Построение ядерных оценок плотности

По дисциплине «Теория вероятностей и математическая статистика»

Выполнил

Студент гр.3630201/80101

В.Н. Сеннов

Руководитель

доцент к.ф.-м.н.

А.Н. Баженов

«___» _____ 202__г.

Санкт-Петербург
2020

Содержание

1	Постановка задачи	4
2	Математическое описание	5
2.1	Ядерные оценки	5
2.2	Гауссово ядро	5
2.3	Выбор ширины полосы	5
3	Особенности реализации	6
4	Результаты работы программы	7
	Заключение	10
	Список литературы	11
A	Репозиторий с исходным кодом	12

Список таблиц

Список иллюстраций

- | | | |
|---|---|---|
| 1 | Ядерные оценки плотности для нормального распределения, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка. | 7 |
| 2 | Ядерные оценки плотности для распределения Коши, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка. . . . | 8 |
| 3 | Ядерные оценки плотности для распределения Лапласа, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка. . . . | 8 |
| 4 | Ядерные оценки плотности для распределения Пуассона, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка. | 9 |
| 5 | Ядерные оценки плотности для равномерного распределения, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка. | 9 |

1 Постановка задачи

Для заданных распределений нужно сгенерировать выборки размером 20, 60 и 100 элементов. Для каждой выборки нужно построить ядерные оценки плотности со следующими значениями ширины полосы:

1. ширина полосы по правилу Сильвермана;
2. половина ширины полосы по правилу Сильвермана;
3. ширина в два раза больше ширины полосы по правилу Сильвермана.

Для всех распределений необходимо построить на одном графике теоретическую плотность и ядерную оценку. Для непрерывных распределений график необходимо построить на промежутке $[-4; 4]$, для распределения Пуассона на $[6; 14]$.

Заданные распределения:

1. Нормальное (гауссово) распределение с параметрами $\mu = 0$, $\sigma = 1$;
2. Распределение Коши с параметрами $\mu = 0$, $\lambda = 1$;
3. Распределение Лапласа с параметрами $\mu = 0$, $\lambda = \frac{1}{\sqrt{2}}$;
4. Распределение Пуассона с параметром $\mu = 10$;
5. Равномерное распределение с параметрами $a = -\sqrt{3}$, $b = \sqrt{3}$.

2 Математическое описание

2.1 Ядерные оценки

Оценкой плотности вероятности $f(x)$ называется функция $\hat{f}(x)$, построенная на основе выборке и приближенно равная $f(x)$ [2].

Ядерной оценкой плотности $\hat{f}_n(x)$, построенной на основе выборки x_1, \dots, x_n называется функция:

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

где $h > 0$ — некий параметр, называемый шириной полосы, $K(x)$ — ядерная функция, для нее выполняется $\int_{-\infty}^{+\infty} K(t)dt = 1$ [1].

2.2 Гауссово ядро

В данной лабораторной в качестве ядерной функции было использовано Гауссово ядро — плотность вероятности стандартного нормального распределения:

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

2.3 Выбор ширины полосы

В рамках данной работы ширина полосы определялась по правилу Сильвермана:

$$h_n = 1.06 \hat{\sigma} n^{-1/5}, \quad (3)$$

где n — размер выборки, $\hat{\sigma}$ — выборочное стандартное отклонение.

3 Особенности реализации

Программа для выполнения лабораторной была написана на языке Python 3.8.2. Для генерации выборок использовался модуль **distributions**, написанный для лабораторной №1. Для построения графиков использовалась библиотека Matplotlib.

Для построения ядерных оценок плотности был написан модуль **kernel**. В нем отдельной функцией реализовано Гауссово ядро по формуле (2). Также отдельной функцией было реализовано вычисление значения ядерной оценки по формуле (1) с шириной полосы согласно правилу Сильвермана (3).

В приложении А приведена ссылка на репозиторий с исходным кодом.

4 Результаты работы программы

В этом разделе представлены результаты работы программы. Ядерные оценки плотности были вычислены по формулам (1) и (2).

На рис. 1 представлены ядерные оценки для нормального распределения.

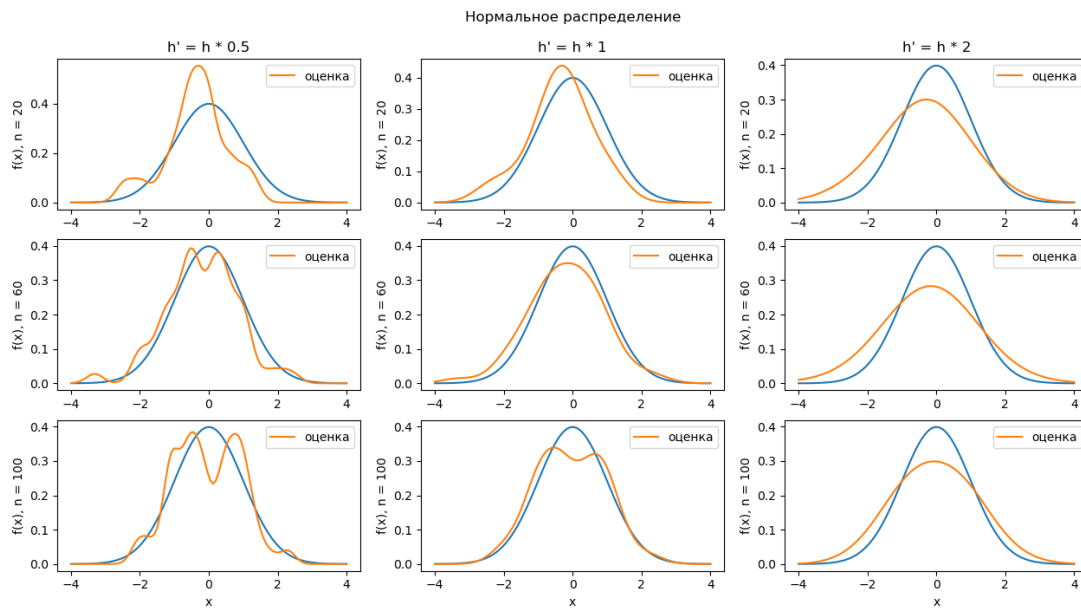


Рис. 1: Ядерные оценки плотности для нормального распределения, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка.

На рис. 2 представлены ядерные оценки для распределения Коши.

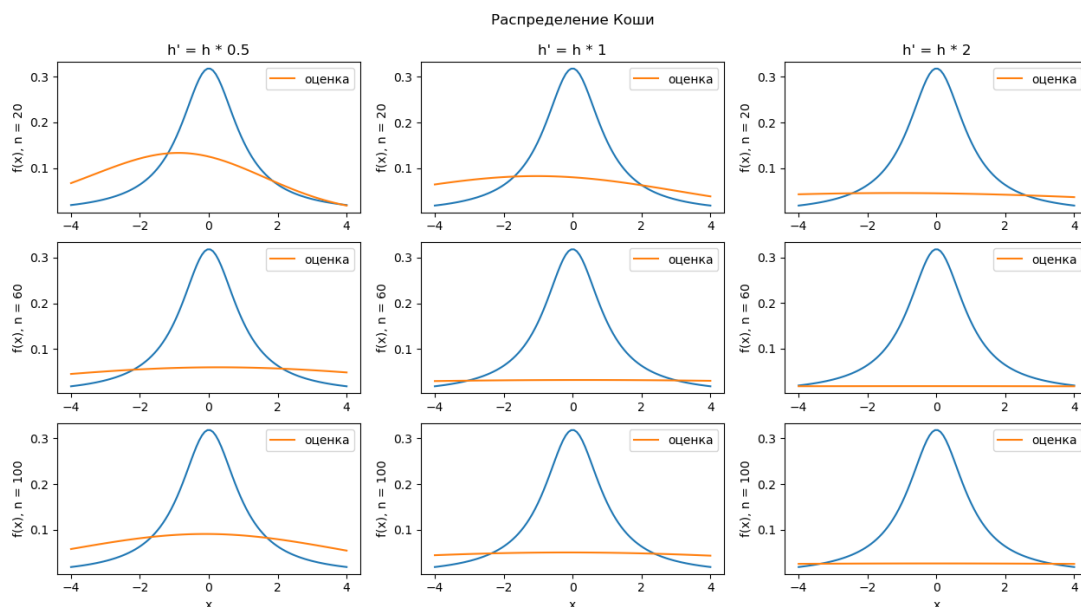


Рис. 2: Ядерные оценки плотности для распределения Коши, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка.

На рис. 3 представлены ядерные оценки для распределения Лапласа.

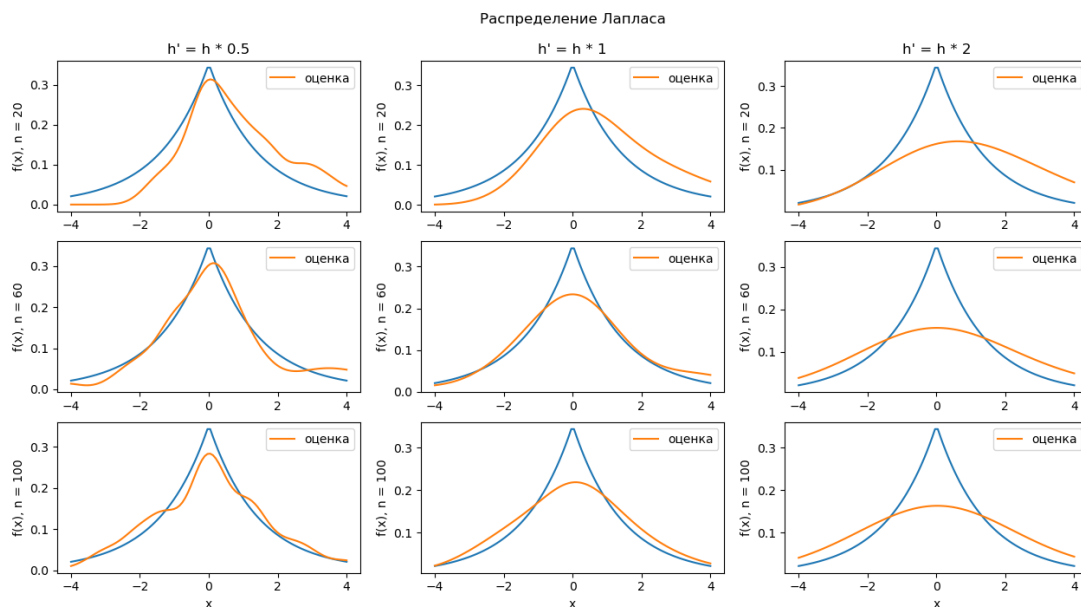


Рис. 3: Ядерные оценки плотности для распределения Лапласа, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка.

На рис. 4 представлены ядерные оценки для распределения Пуассона.

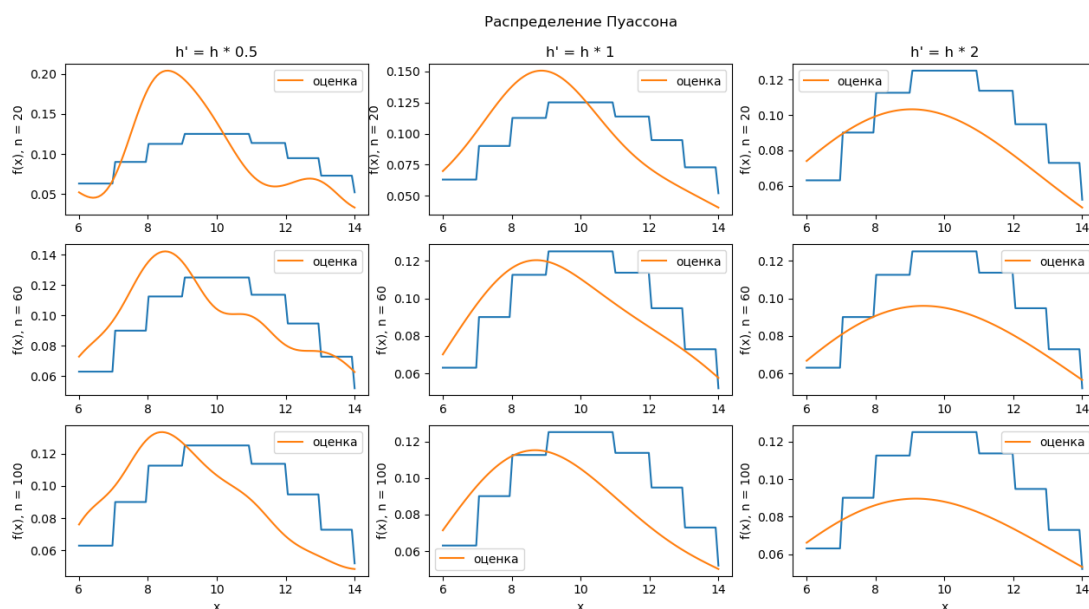


Рис. 4: Ядерные оценки плотности для распределения Пуассона, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка.

На рис. 5 представлены ядерные оценки для равномерного распределения.

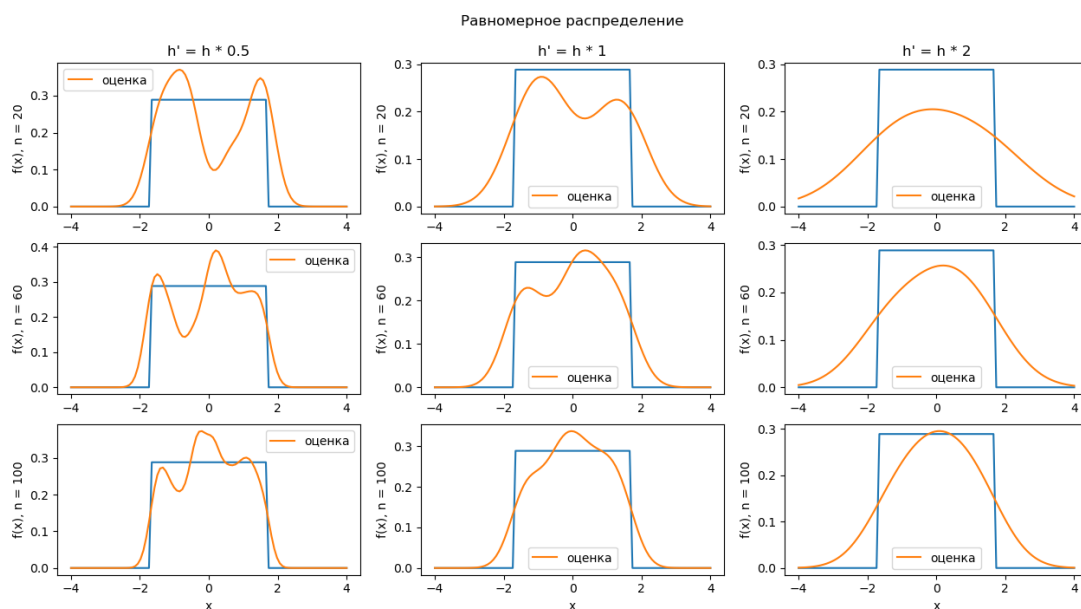


Рис. 5: Ядерные оценки плотности для равномерного распределения, где h — ширина полосы по формуле (3), h' — ширина полосы, с которой строилась ядерная оценка.

Заключение

В рамках лабораторной работы были построены ядерные оценки плотности для заданных распределений с различными значениями ширины полосы. Для полученных оценок были построены графики.

По построенным графикам видно, что чем меньше значение ширины полосы, тем чувствительнее ядерная оценка к отдельным значениям выборки, тем менее она гладкая. Также можно заметить, что в среднем оценка близка к теоретической плотности, если выбирать ширину полосы согласно правилу Сильвермана.

Также стоит заметить, что для распределения Коши оценка получилась слишком сглаженная из-за того, что выборочное стандартное отклонение чересчур велико. Такая оценка не может быть названа состоятельной.

Также можно заметить, что при увеличении размера выборки точность ядерной оценки не сильно возрастает.

Программа для лабораторной была написана языке Python 3.8.2, для построения графиков использовалась библиотека Matplotlib.

Список литературы

- [1] Kernel density estimation. // Wikipedia, the free encyclopedia. — URL: https://en.wikipedia.org/wiki/Kernel_density_estimation. — (дата обращения: 18.11.2020)
- [2] Теоритическое приложение к лабораторным работам №1-4 по дисциплине «Математическая статистика». — Спб.: Сантк-Петербургский политехнический университет, 2020. — 12 с.

А Репозиторий с исходным кодом

Исходный код программы для данной лабораторной размещен на сервисе GitHub.

Ссылка на репозиторий: <https://github.com/Vovan-S/TV-Lab1>.