Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторной работе №4
по дисциплине
«Математическая статистика»

Выполнил студент:

Кондратьев Д. А.
группа: 3630102/70301

Проверил:

к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

Список иллюстраций

1. Постановка задачи

Для 5-ти рапределений:

- Нормальное распределение $N(x, 0, 1)$;
- Распределение Коши $C(x, 0, 1)$;
- Распределение Лапласа $L(x, 0, \frac{1}{\sqrt{2}})$;
- Распределение Пуассона $P(k, 10)$;
- Равномерное Распределение $U(x, -\sqrt{3}, \sqrt{3})$;

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

- Нормальное распределение

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

- Распределение Коши

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

- Распределение Лапласа

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

- Распределение Пуассона

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

- Равномерное Распределение

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

2.2. Эмпирическая функция распределения

2.2.1. Статистический ряд

Статистическим рядом называется последовательность различных элементов выборки $z_1, z_2, ..., z_k$, расположенных в возрастающем порядке с указанием частот $n_1, n_2, ..., n_k$, с которыми эти элементы содержатся в выборке. Статистический ряд обычно записывается в виде таблицы:

z	z_1	z_2	...	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:

$$F_n^*(x) = P^*(X < x) \tag{6}$$

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

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

$F^*(x)$ — 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 2. Таблица распределения

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

$$F_n^*(x) \approx F_X(x). \tag{8}$$

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 $f(x)$:

$$\hat{f}(x) \approx f(x). \tag{9}$$

2.3.2. Ядерные оценки

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

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K(\frac{x - x_i}{h_n}). \tag{10}$$

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

$$h_n \xrightarrow[n \to \infty]{} 0; \quad \frac{h_n}{n^{-1}} \xrightarrow[n \to \infty]{} \infty. \tag{11}$$

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

Замечание. Свойство, означающее сближение оценки с оцениваемой величиной при $n \to \infty$ 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]:

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

Правило Сильвермана [2, с. 44]:

$$h_n = 1.06\hat{\sigma}n^{-1/5}, \tag{13}$$

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

3. Реализация

Лабораторная работа выполнена на программном языке *Python 3.8* в среде разработки *Jupyter Notebook 6.0.3*. В работе использовались следующие пакеты языка *Python*:

- *numpy* — для генерации выборки и работы с массивами;
- *matplotlib.pyplot* и *seaborn* — для построения графиков;
- *scipy.stats* — содержит все необходимые распределения.

Ссылка на исходный код лабораторной работы приведена в приложении.

4. Результаты

4.1. Эмпирическая функция распределения

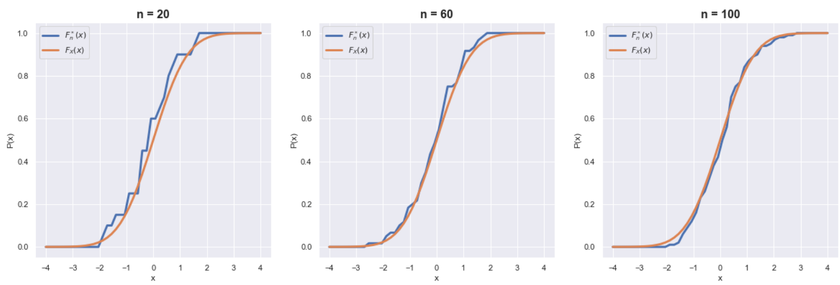

Рис. 1. Нормальное распределение

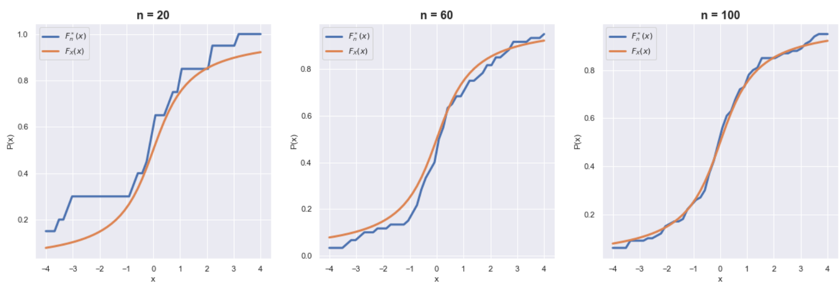

Рис. 2. Распределение Коши

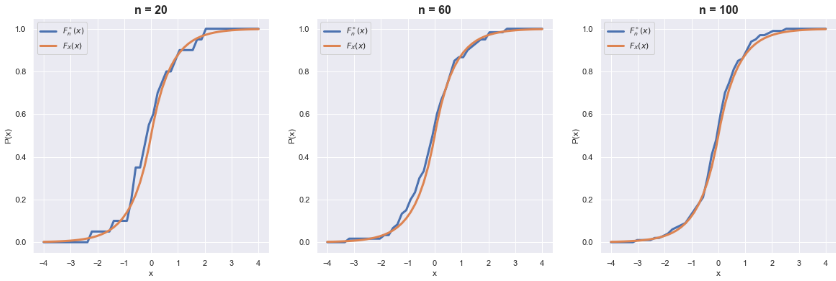

Рис. 3. Распределение Лапласа

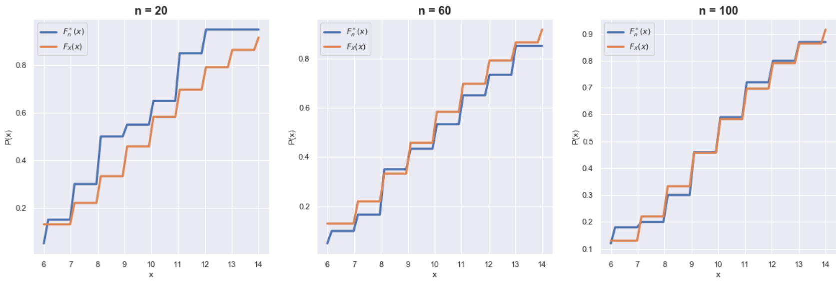

Рис. 4. Распределение Пуассона

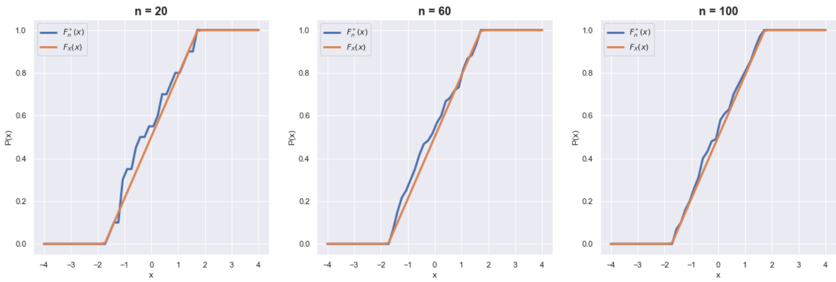

Рис. 5. Равномерное распределение

4.2. Ядерные оценки плотности распределения

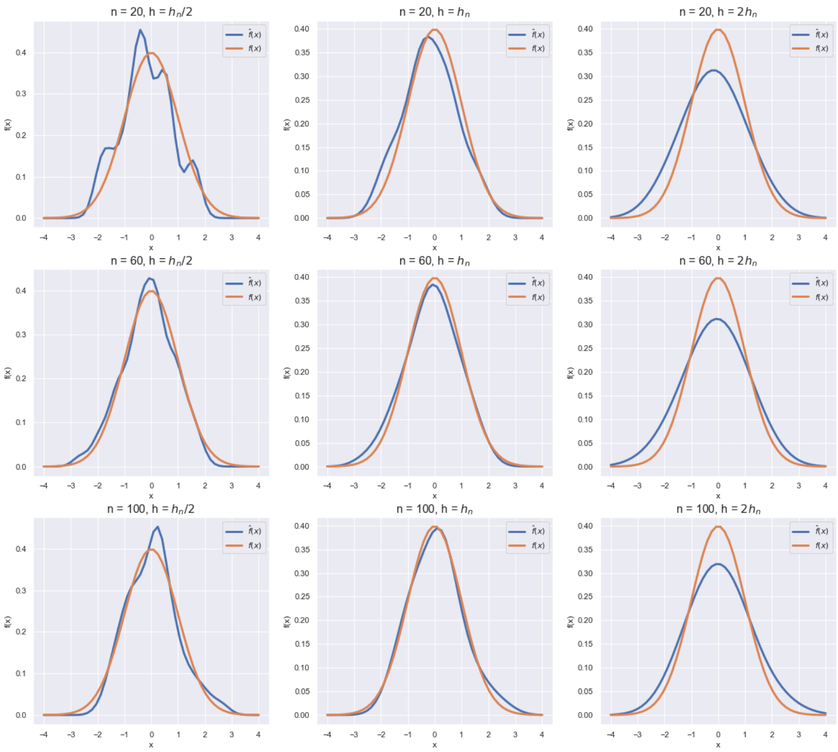

Рис. 6. Нормальное распределение

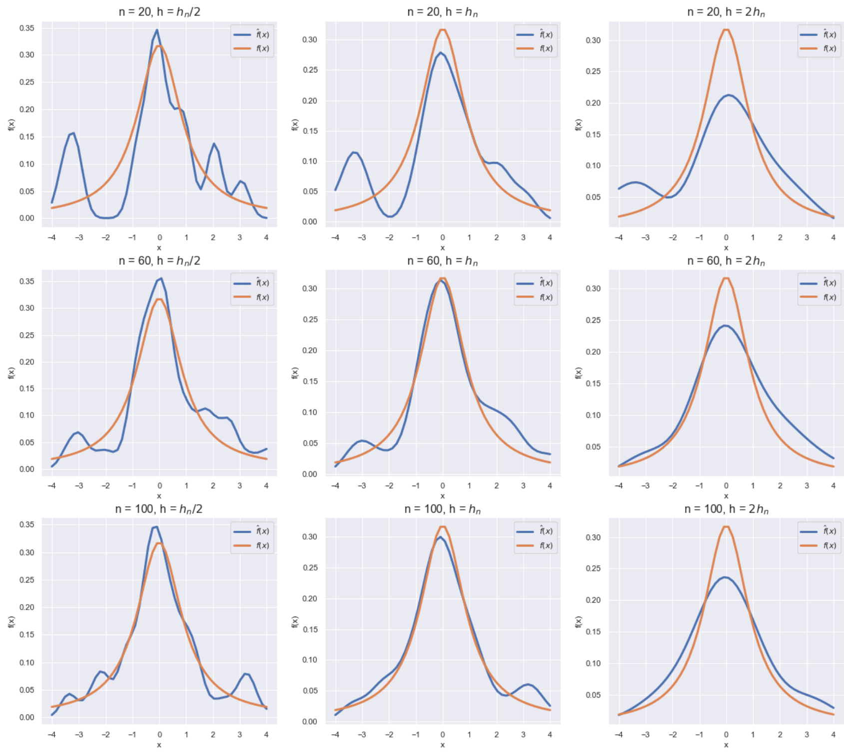

Рис. 7. Распределение Коши

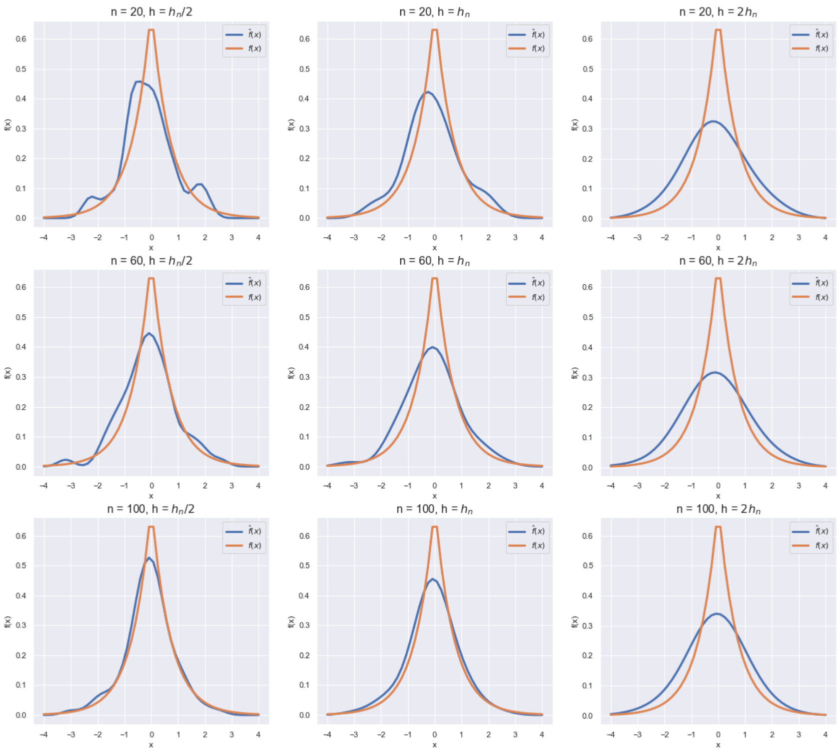

Рис. 8. Распределение Лапласа

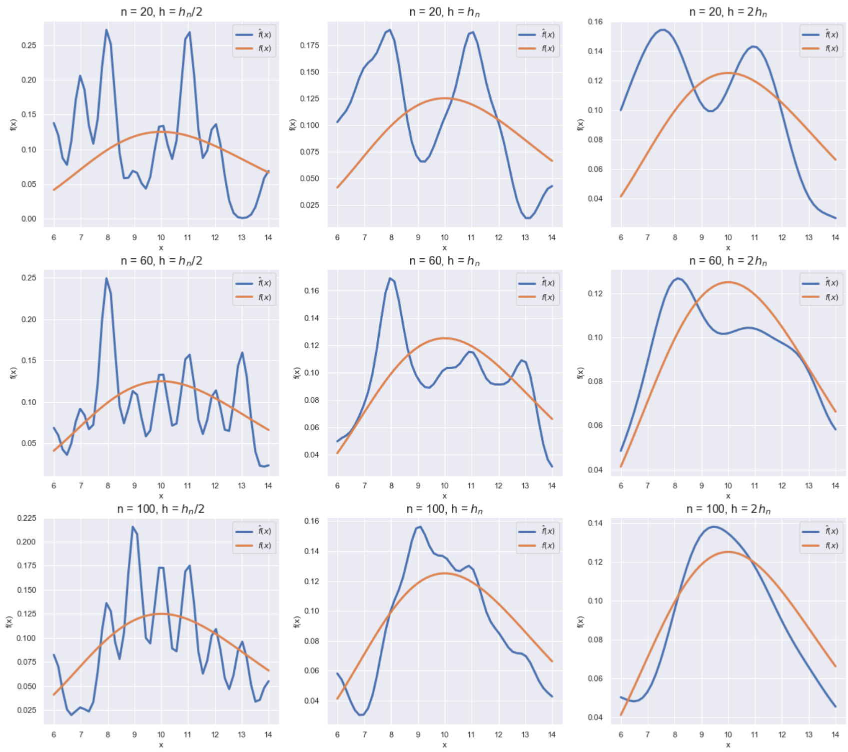

Рис. 9. Распределение Пуассона

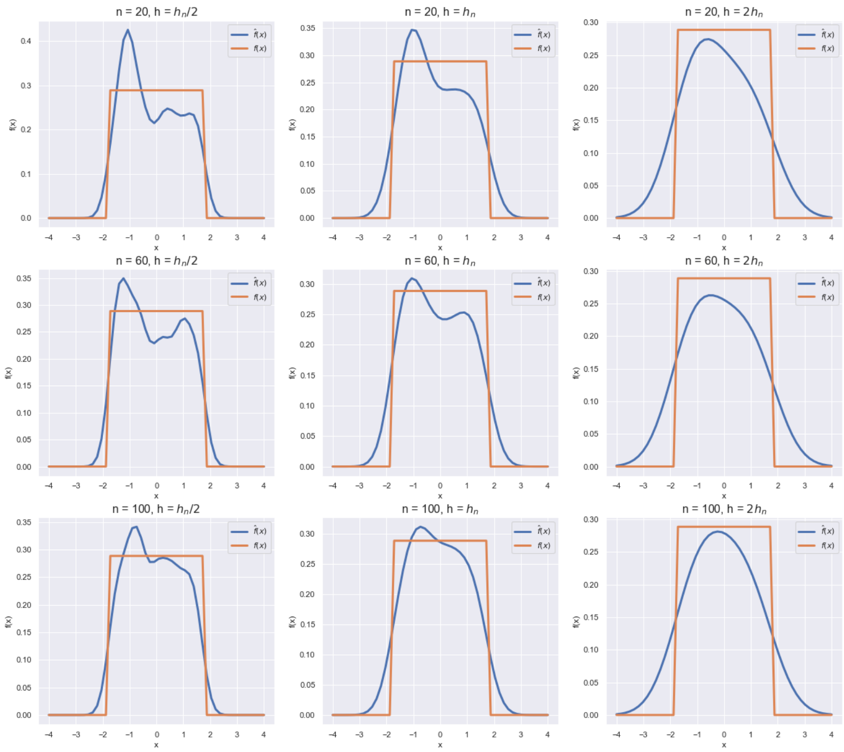

Рис. 10. Равномерное распределение

5. Обсуждение

Исходя из полученных результатов можно сделать следующие выводы:

- При увеличении размера выборки качество оценки эмпирической функцией эталонной функции распределения возрастает.
- При увеличении размера выборки качество ядерной оценки плотности распределения возрастает.
- Наилучшее качество ядерной оценки плотности распределения достигается при следующих параметрах:
  1) для нормального распределения — $h = h_n$;
  2) для распределения Коши — $h = h_n$;
  3) для распределения Лапласа — $h = \frac{h_n}{2}$ или $h = h_n$;
  4) для распределения Пуассона — $h = 2h_n$;
  5) для равномерного распределения — $h = \frac{h_n}{2}$ или $h = h_n$.

6. Литература

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.

3) Kernel density estimation. URL: `https://en.wikipedia.org/wiki/Kernel_density_estimation`

7. Приложение

1) Код лабораторной. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_4/Lab_4.ipynb`

2) Код отчёта. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_4/Lab_report_4.tex`