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:
Самутичев Евгений Романович
группа: 3630102/70201

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

Санкт-Петербург
2020 г.

Содержание

Список иллюстраций

1 Постановка задачи

Для каждого из 5 распределений:

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

сгенерировать выборки размером 20, 60 и 100 элементов. Построить на них эмпирические функции распределения и ядерные оценки плотности/функции распределения на отрезке $[-4, 4]$ для непрерывных распределений и на отрезке $[6, 14]$ для распределения Пуассона.

2 Теория

2.1 Эмпирическая функция распределения

Эмпирической функцией распределения, построенной по выборке $(x_1, ..., x_n)$ объема n называется случайная функция $F_n^* : \mathbb{R} \times \Omega \to [0, 1]$, которая имеет вид

$$F_n^*(y) = \frac{1}{n} \sum_{i=1}^{n} I(x_i < y) \tag{1}$$

где I - индикатор события $x_i < y$ [1]

2.2 Ядерная оценка плотности распределения

Пусть $(x_1, ..., x_n)$ - выборка полученная по распределению с некоторой плотностью f, требуется оценить функцию f. *Ядерным оценщиком плотности* называется [2]

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{2}$$

где K - т.н. *ядро* (некоторая неотрицательная функция), $h > 0$ - сглаживающий параметр, именуемый *шириной полосы.*

Как правило используется нормальное (или гауссово) ядро, в силу его удобных математических свойств:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{3}$$

В случае если используется гауссово ядро и оцениваемая плотность является гауссовой, оптимальный выбор для h определяется т.н. *правилом Сильвермана* [2]:

$$h_n = \left(\frac{4s_n^5}{3n}\right)^{\frac{1}{5}} \approx 1.06 s_n n^{-\frac{1}{5}} \tag{4}$$

где s_n - выборочное среднеквадратичное отклонение (корень из выборочной дисперсии)

3 Реализация

Работа выполнена с использованием языка **Python** в интегрированной среде разработки **PyCharm**, были задействованы библиотеки:

- **NumPy** - векторизация вычислений, работа с массивами данных
- **SciPy** - модуль **stats** для генерации данных по распределениям, вычисления ядерной оценки плотности
- **Matplotlib** - построение графиков

Исходный код работы приведен в приложении.

4 Результаты

4.1 Эмпирические функции распределения

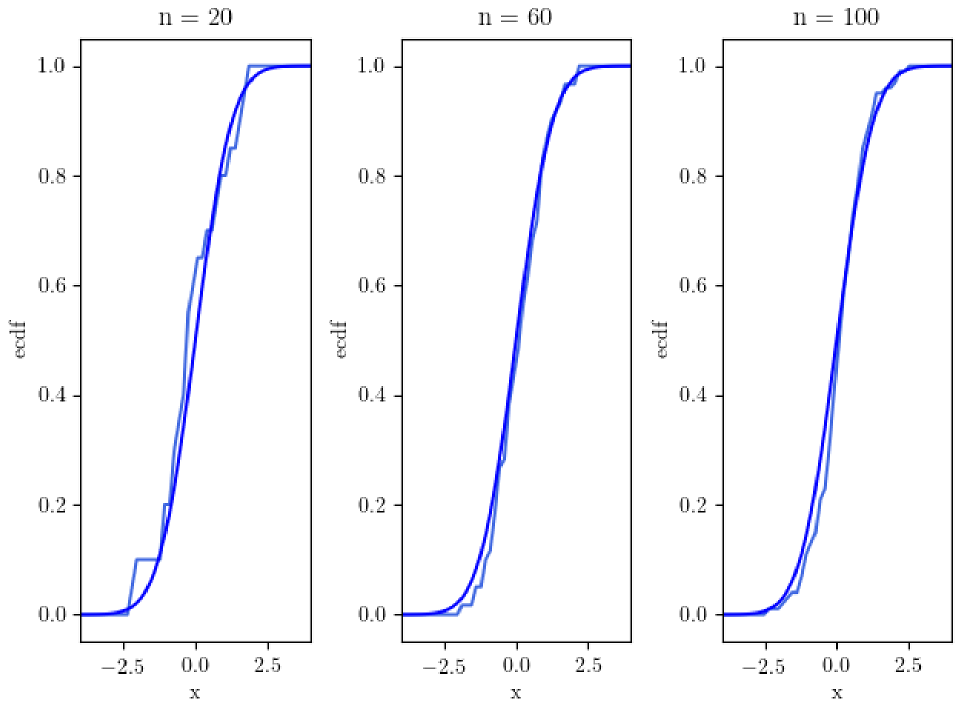

Рис. 1: Нормальное распределение

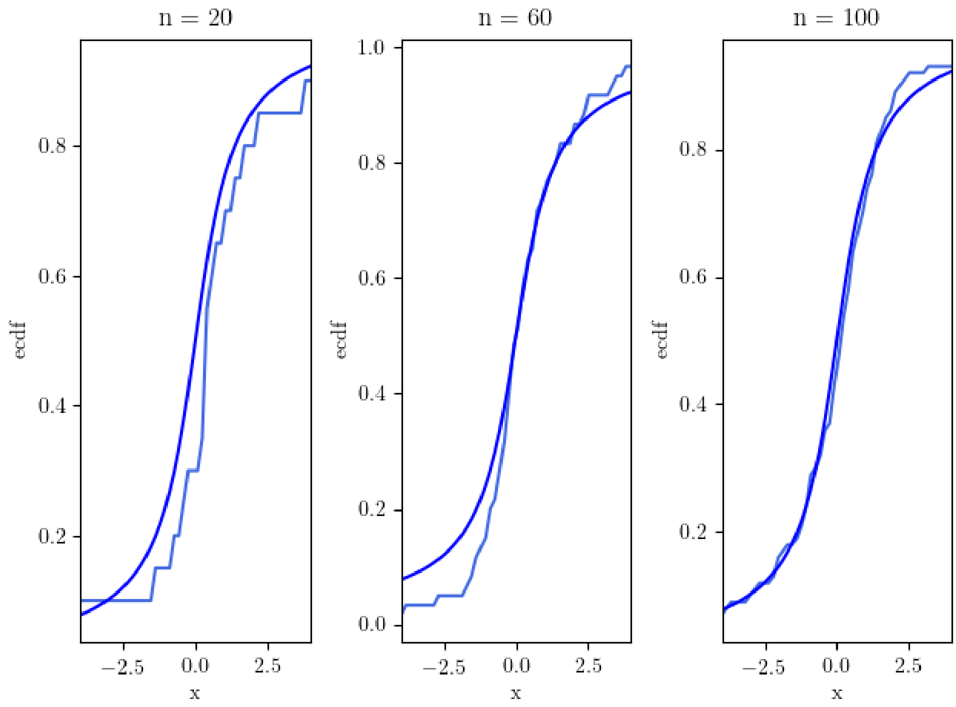

Рис. 2: Распределение Коши

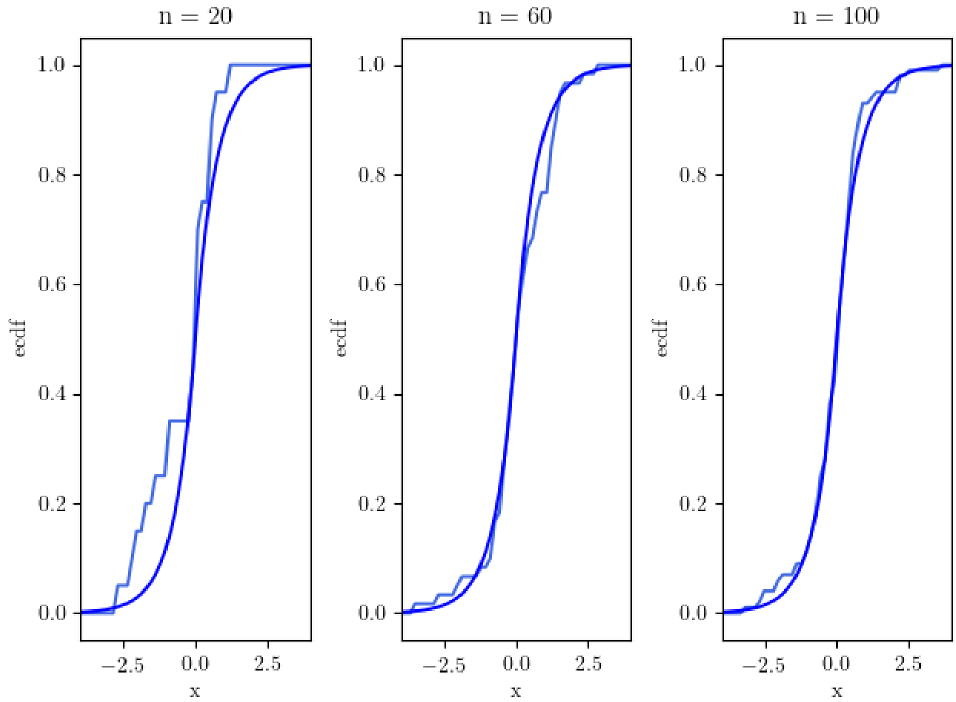

Рис. 3: Распределение Лапласа

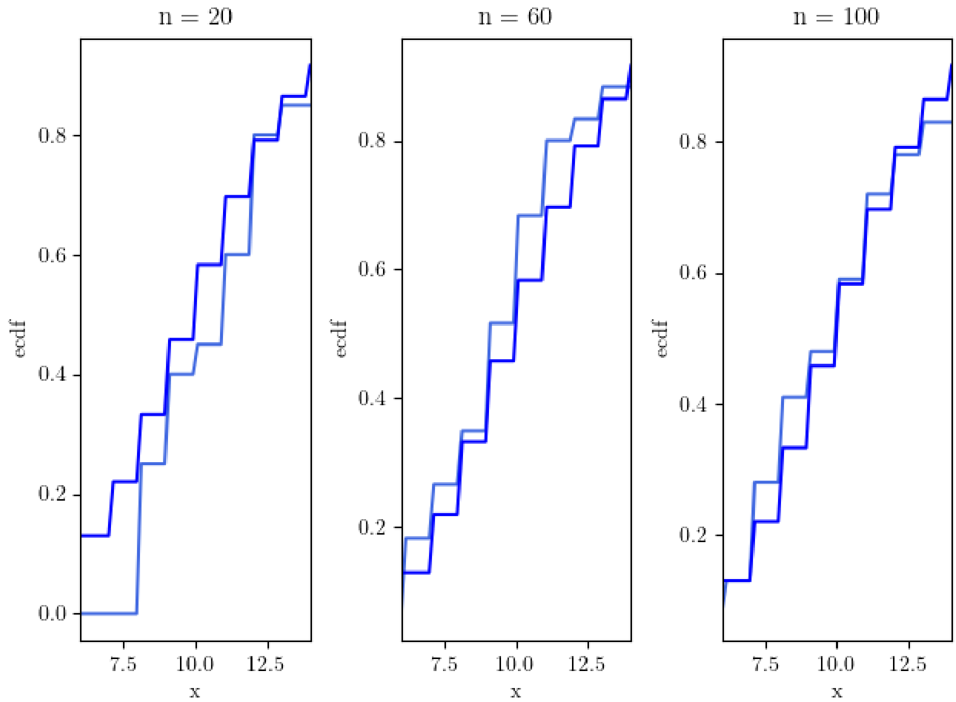

Рис. 4: Распределение Пуассона

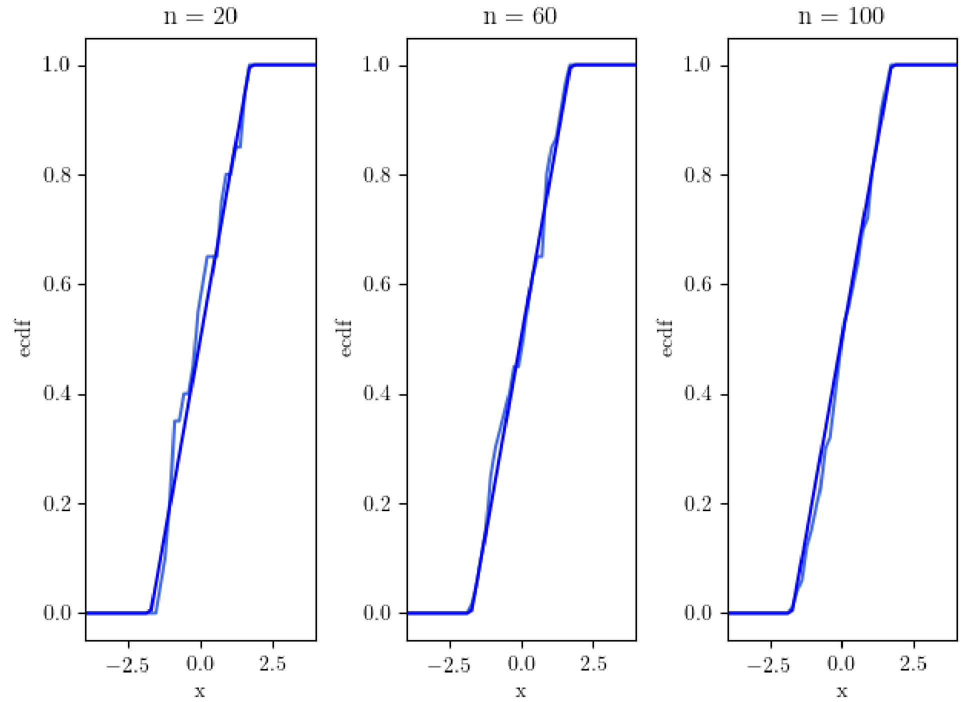

Рис. 5: Равномерное распределение

4.2 Ядерные оценки плотности распределения

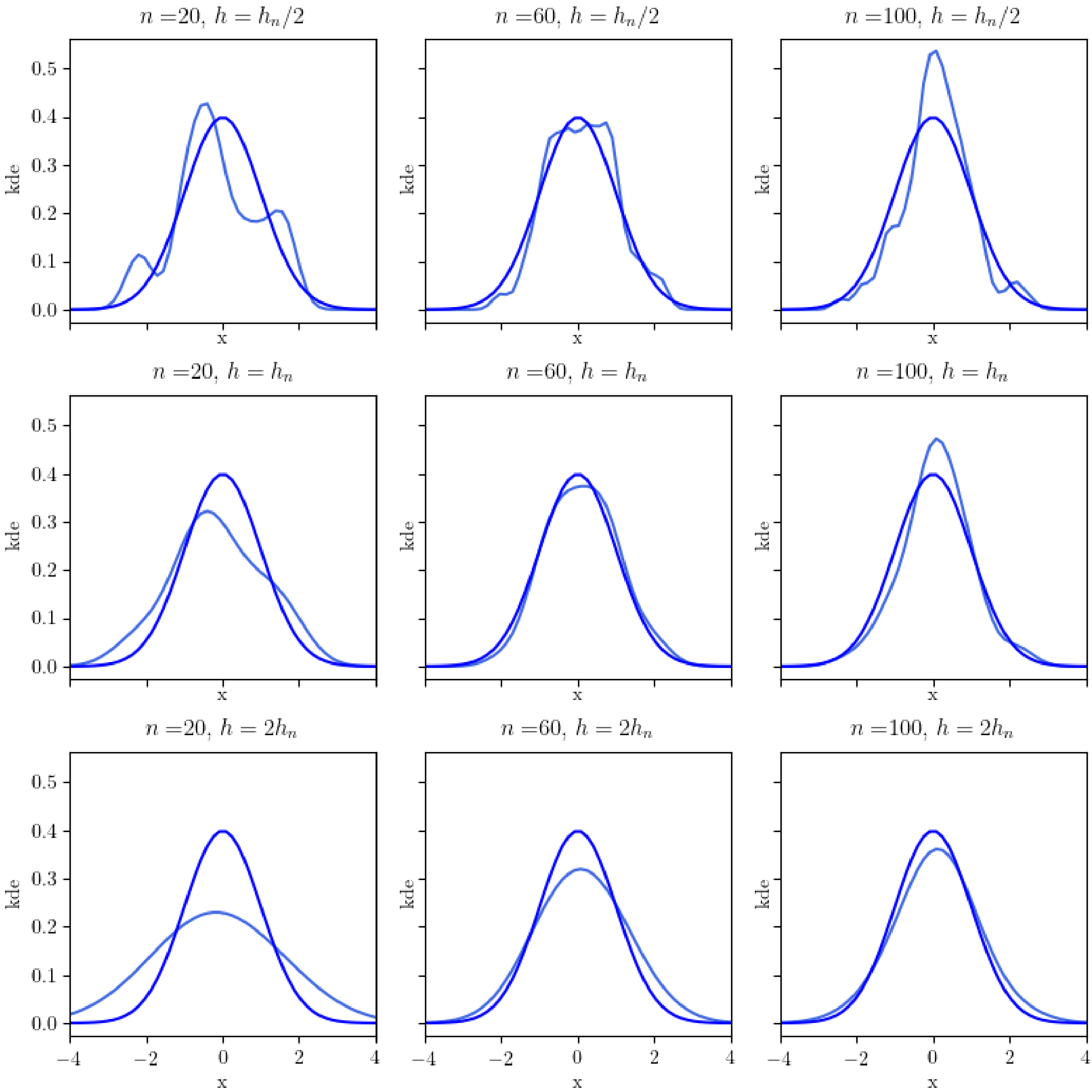

Рис. 6: Нормальное распределение

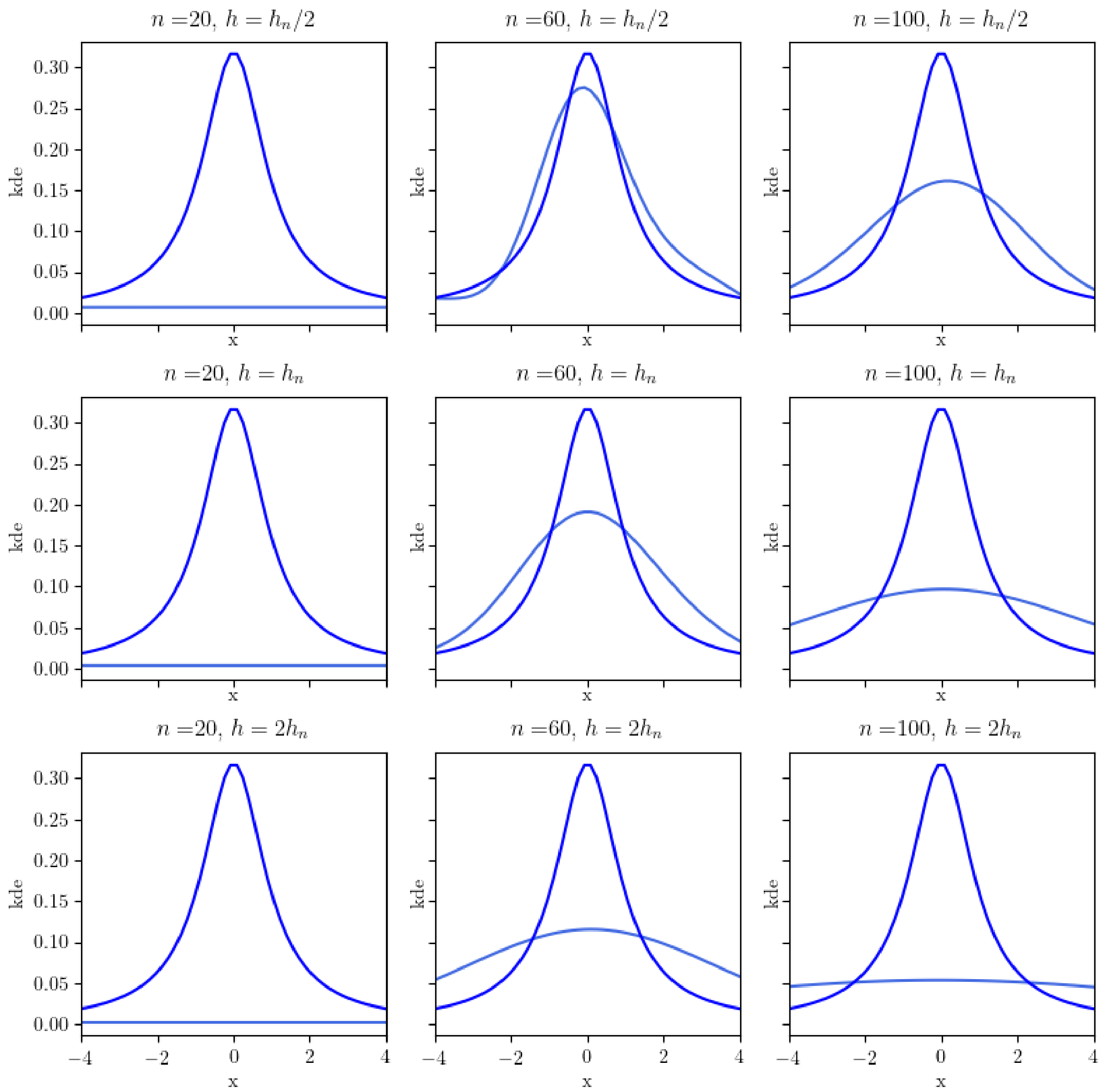

Рис. 7: Распределение Коши

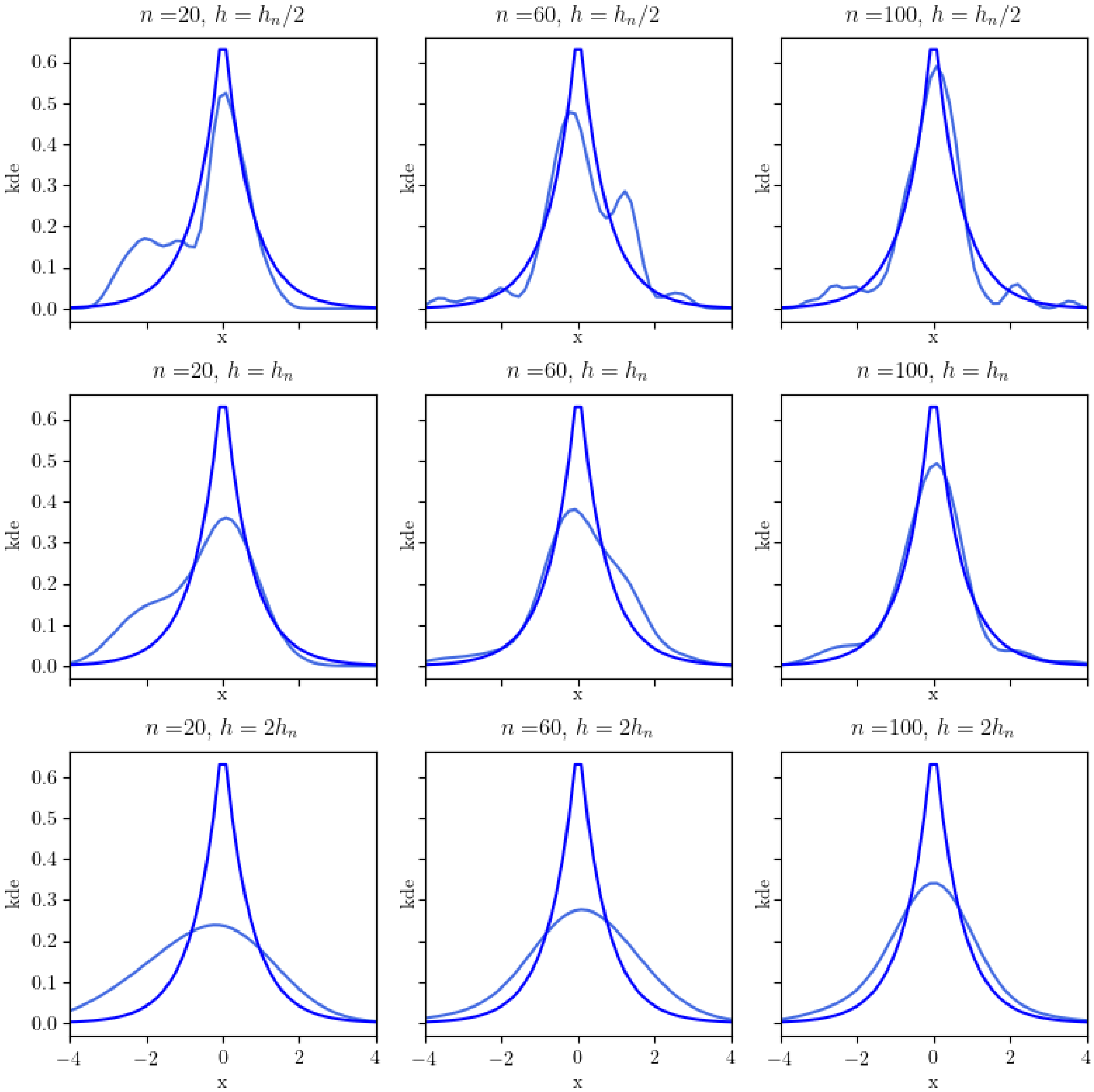

Рис. 8: Распределение Лапласа

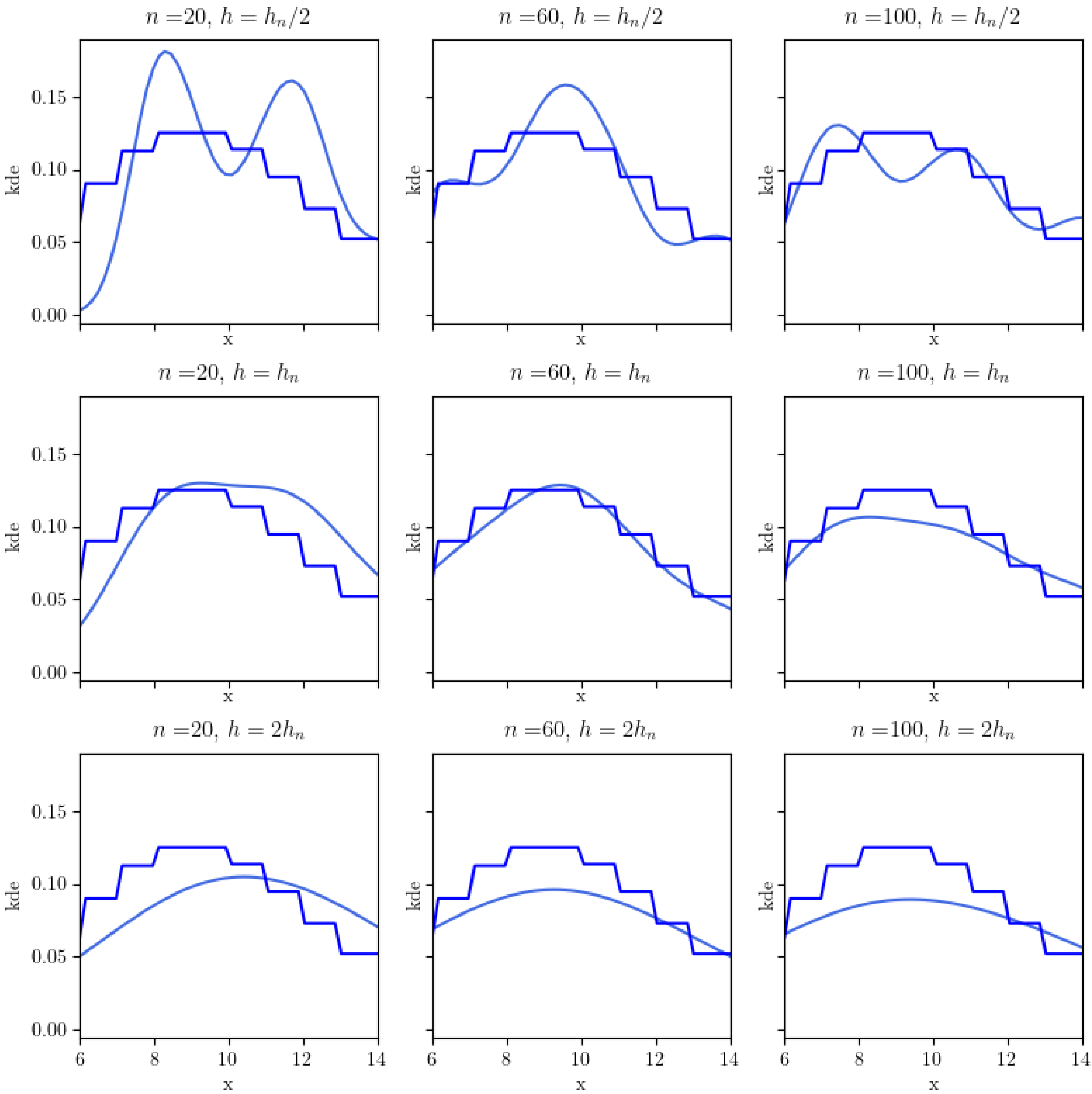

Рис. 9: Распределение Пуассона

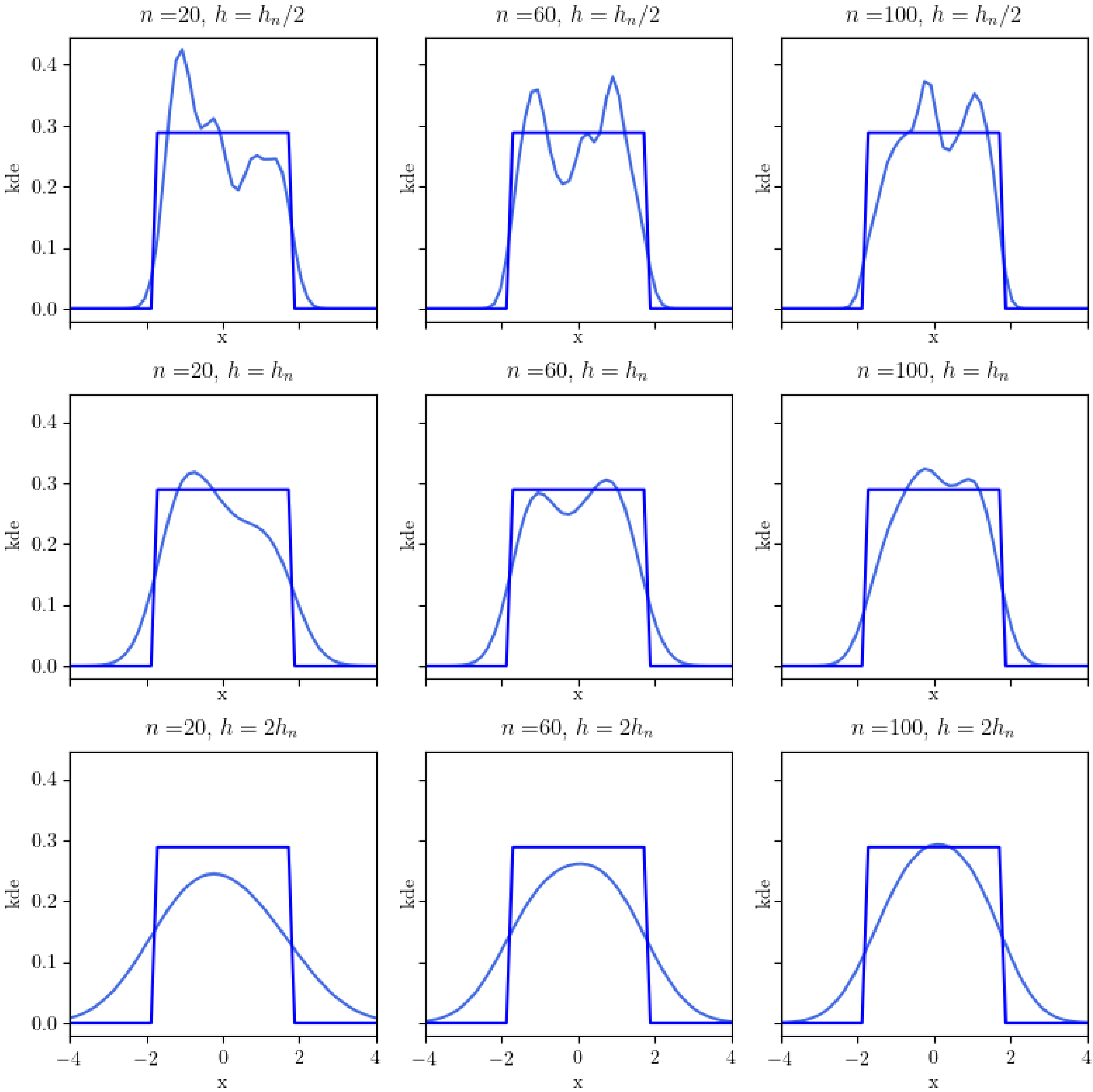

Рис. 10: Равномерное распределение

5 Обсуждение

5.1 Эмпирическая функция распределения

Существует **теорема** [1]: *Пусть* $(x_1, ..., x_n)$ *- выборка из распределения с некоторой функцией распределения* F *и пусть* F_n^* *- эмпирическая функция распределения построенная по этой выборке. Тогда* $F_n^*(y) \xrightarrow{p} F(y), \forall y \in \mathbb{R}$ Полученные графики подтверждают данный теоретический факт, с ростом n эмпирическая функцяи распределения все ближе к истинной.

5.2 Ядерная оценка плотности распределения

Для нормального распределения наилучшие результаты показал выбор h по правилу Сильвермана, что обосновано теоретически т.к. он оптимален в некотором смысле (см. Теория), как и для распределения Пуассона. Для распределения Лапласа хорошие результаты в приближении плотности распределения имеем как при h_n, так и при $0.5h_n$. Плотность равномерного распределения аппроксимируется неудачно т.к. оно далеко от гауссова, как и распределение Коши.

6 Приложения

1. Исходный код лабораторной `https://github.com/zhenyatos/statlabs/tree/master/Lab4`

Список литературы

[1] Н. И. Чернова, *Математическая статистика: Учеб. пособие*. Новосиб. гос. ун-т. Новосибирск, 2007. 148 стр.

[2] Ядерная оценка плотности // Википедия. [2020—2020]. Дата обновления: 05.01.2020. URL: https://ru.wikipedia.org/?oldid=104368872 (дата обращения: 05.01.2020).