Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра прикладной математики

ОТЧЕТ

Тема: ***Оценки законов распределений***

Направление: 01.03.02 Прикладная математика и информатика

Выполнил студент гр. 33631/4 Камалетдинова Ю.

Преподаватель Баженов А.

Санкт-Петербург
2019

Содержание

Постановка задачи

В данной лабораторной работе рассматриваются эмпирические функции распределения и ядерные оценки плотностей как статистические способы установления характера распределения. Требуется построить их для распределений из работы №1 на выборках размером $N = 20,\ 60,\ 100$ на отрезке $[-4,\ 4]$, а также сделать выводы о данных оценках законов распределений. Формулы распределений представлены ниже

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

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

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

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}}, & |x| \le \sqrt{3} \\ 0, & |x| > \sqrt{3} \end{cases} \text{ -- равномерное} \tag{4}$$

$$P(\lambda) = \frac{e^{-\lambda}}{k!} \lambda^k,\ \lambda = 2 \text{ -- Пуассон} \tag{5}$$

Описание метода

Пусть имеется некоторая выборка объемом $n:\ x_1, \ldots,\ x_n;\ x_i \in \mathbb{R}$. Эмпирической функцией распределения называют

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} u(x - x_i),\ \text{где} \tag{6}$$

$$u(z) = \begin{cases} 1, \ z \geq 0 \\ 0, \ z \leq 0 \end{cases} \quad \text{– фунция Хевисайда} \tag{7}$$

Ядерная оценка плотности определяется формулой

$$\hat{f}_{n,h_n}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right), \text{ где } K(u) \text{ – ядро, } h = h_n \text{ – параметр сглаживания} \tag{8}$$

Ядро $K(u)$ – это вещественнозначная функция со следующими свойствами

1. $K(u) \geq 0$
2. $K(-u) = K(u)$
3. $\int_{-\infty}^{+\infty} K(u)du = 1$

Ядерная оценка плотности сглаживает каждый элемент выборки до плавного участка, форма которого определяется функцией ядра $K(u)$. Затем функция суммирует все участки, чтобы получить оценку плотности. В данной работе будем использовать ядро Гаусса, заданное формулой

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{9}$$

Реализация

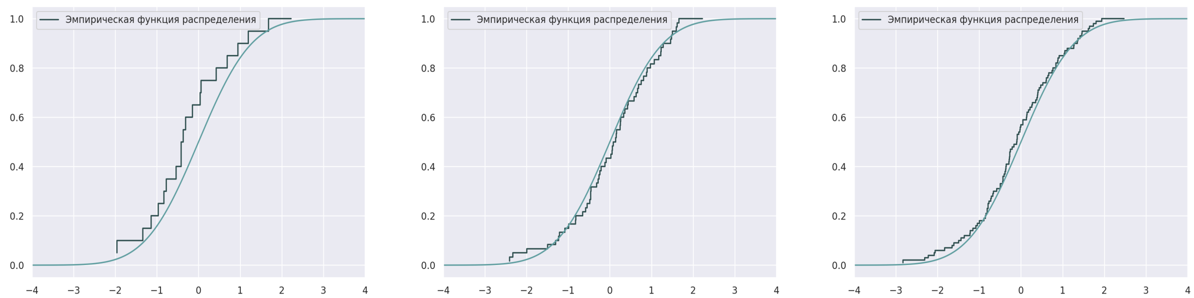
Для выполнения поставленной задачи будем пользоваться библиотеками для языка Python: *numpy, scipy* – расчеты, законы распределения вероятностей; *matplotlib, seaborn* – визуализация результатов. Ход работы:

- Задаем распределение с заданными параметрами

- Генерируем случайные выборки из распределений объемами $n = 20,\ 60,\ 100$
- Для отсортированных выборок из распределений задаем вектор значений $y = [\frac{1}{n}, \frac{2}{n}, \ldots, 1]$ и строим ступенчатый график - эмпирическую функцию распределения
- По формулам (8), (9) вычисляем ядерные оценки плотностей для параметров сглаживания h для всех выборок и строим графики

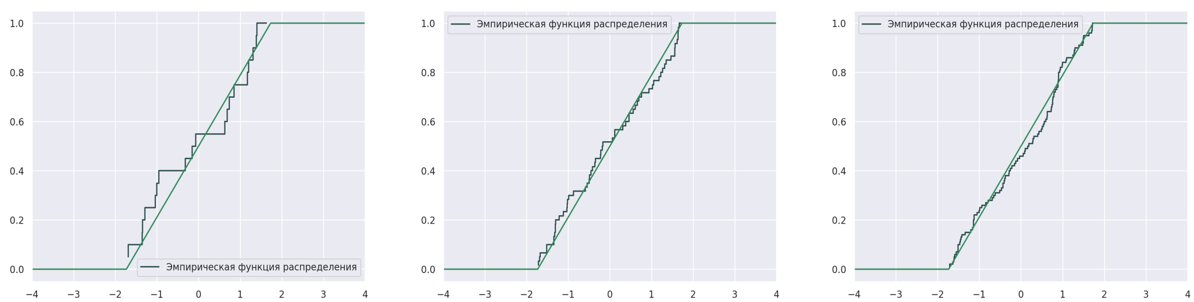
Результат

Эмпирические и теоретические функции распределения



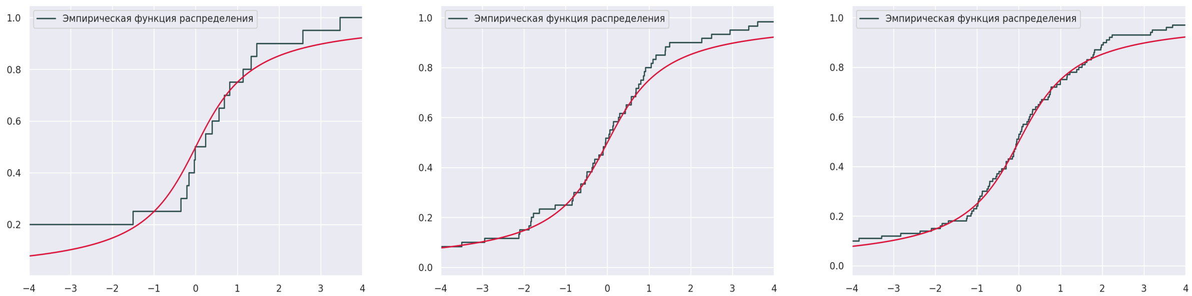
(a) (b) (c)

Рис. 1: Выборки из нормального распределения объемом: (a) 20; (b) 60; (c) 100



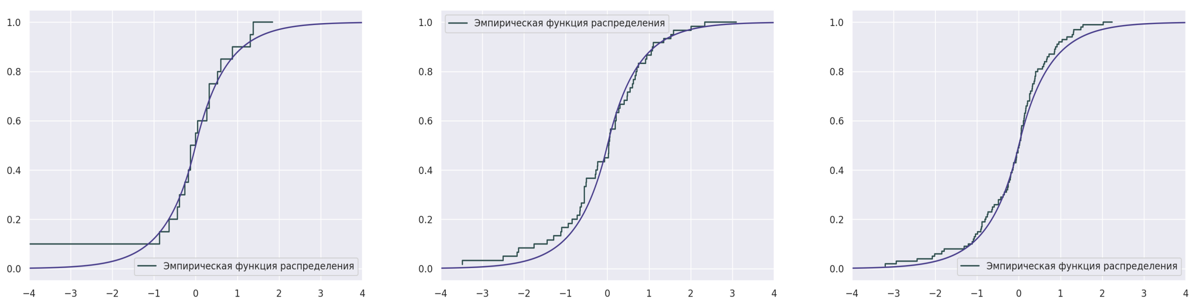
(a) (b) (c)

Рис. 2: Выборки из равномерного распределения объемом: (a) 20; (b) 60; (c) 100



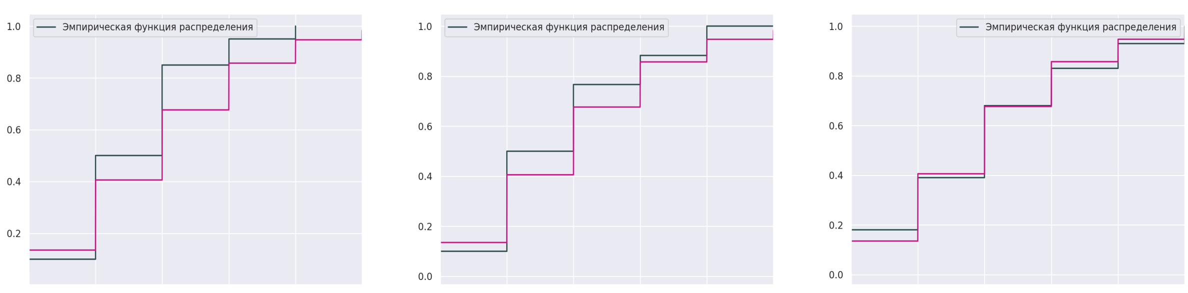
(a) (b) (c)

Рис. 3: Выборки из распределения Коши объемом: (a) 20; (b) 60; (c) 100



(a) (b) (c)

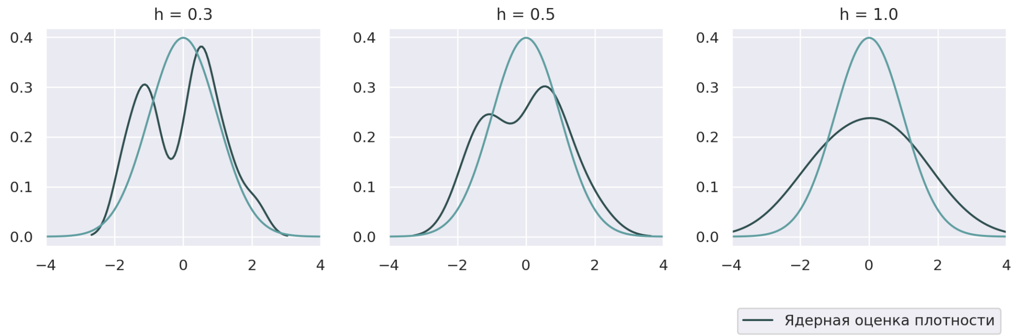
Рис. 4: Выборки из распределения Лапласа объемом: (a) 20; (b) 60; (c) 100



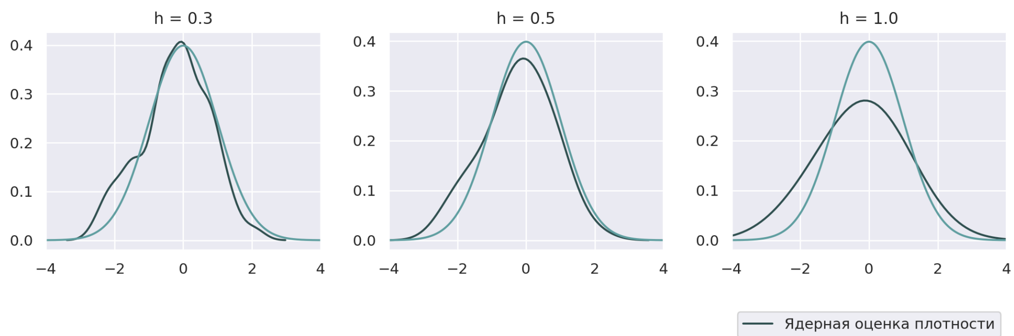
(a) (b) (c)

Рис. 5: Выборки из распределения Пуассона объемом: (a) 20; (b) 60; (c) 100

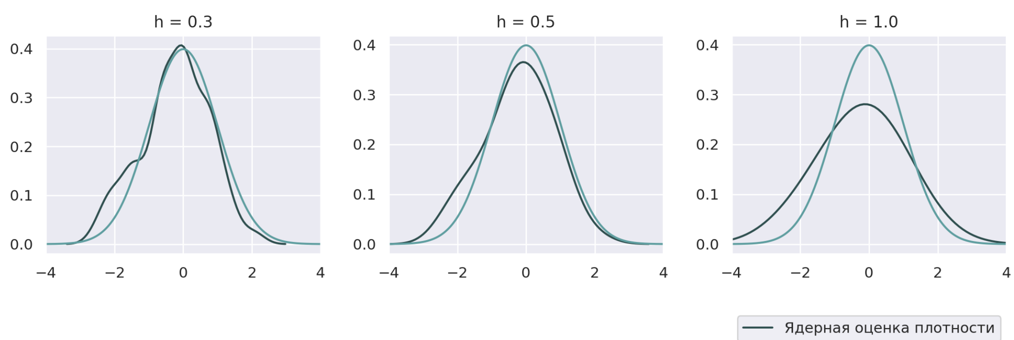
Ядерные оценки плотностей и функции плотностей распределения



(a)

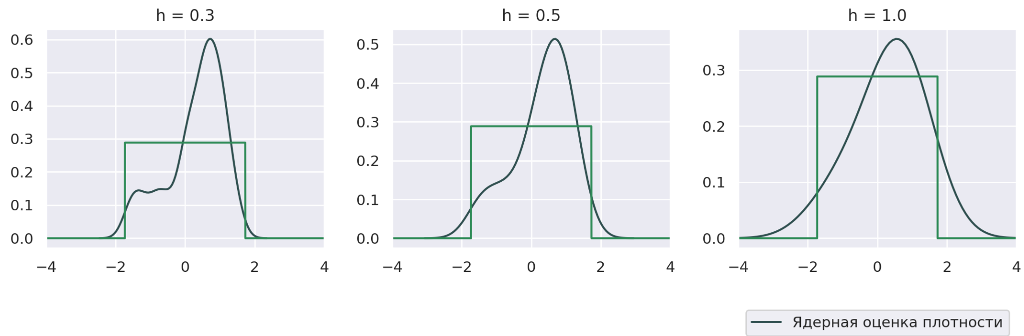


(b)

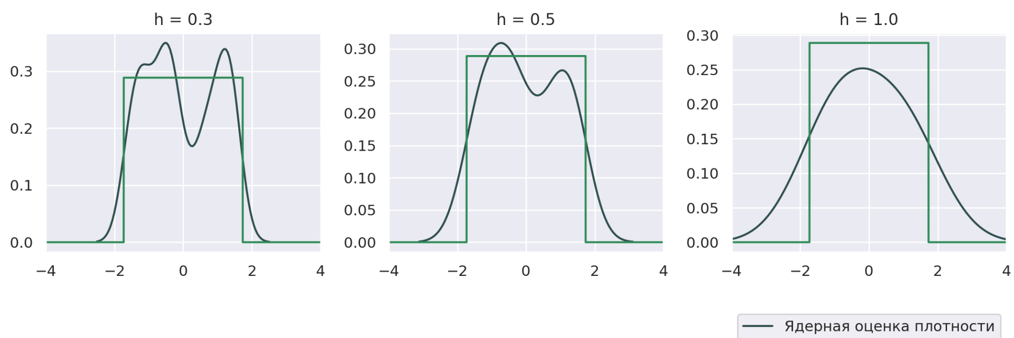


(c)

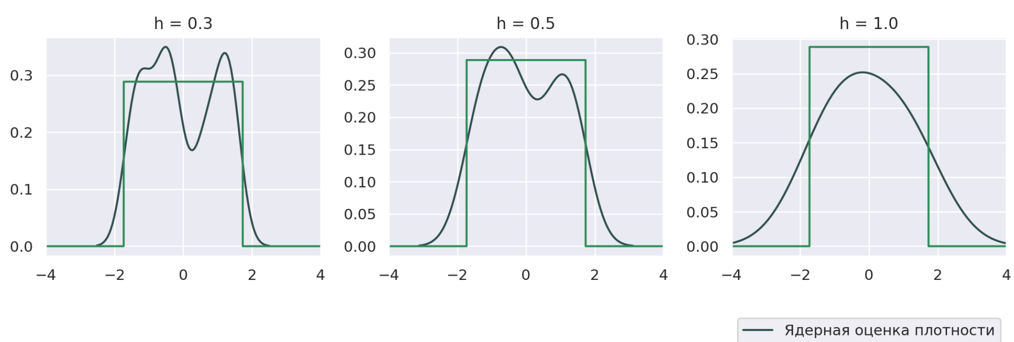
Рис. 6: Для выборок из нормального распределения объемом: (a) 20; (b) 60; (c) 100



(a)

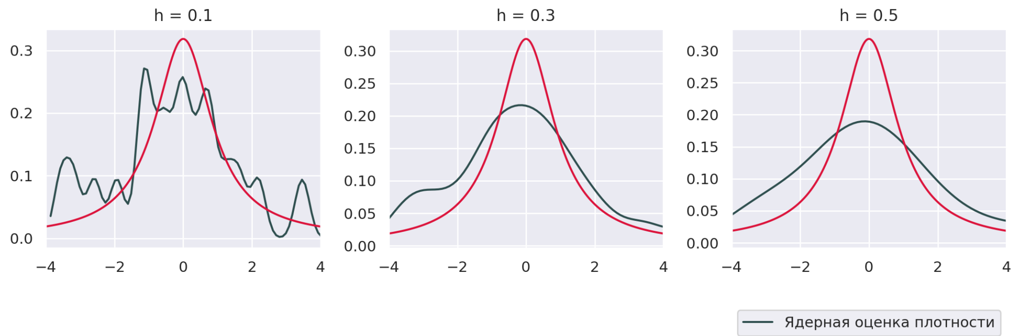


(b)



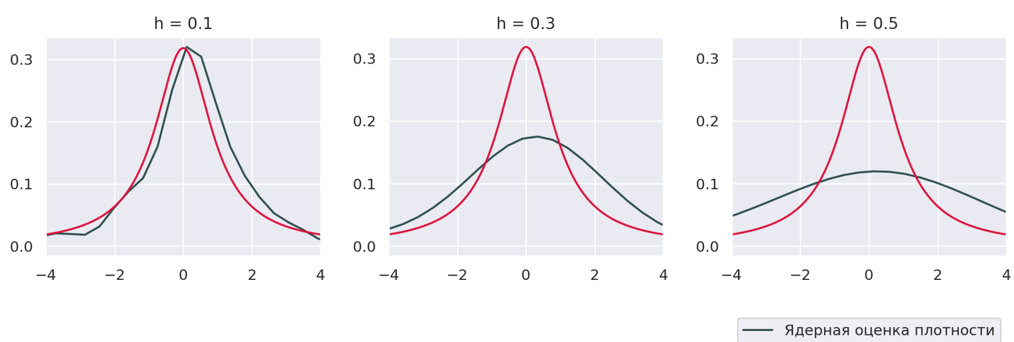
(c)

Рис. 7: Для выборок из равномерного распределения объемом: (a) 20; (b) 60; (c) 100



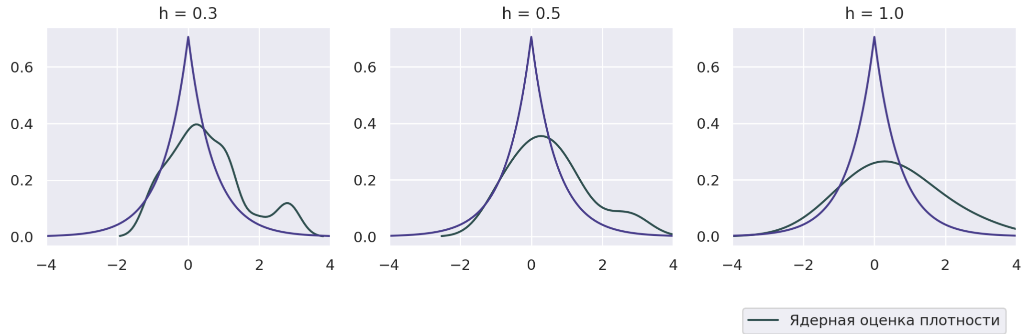
(a)

(b)

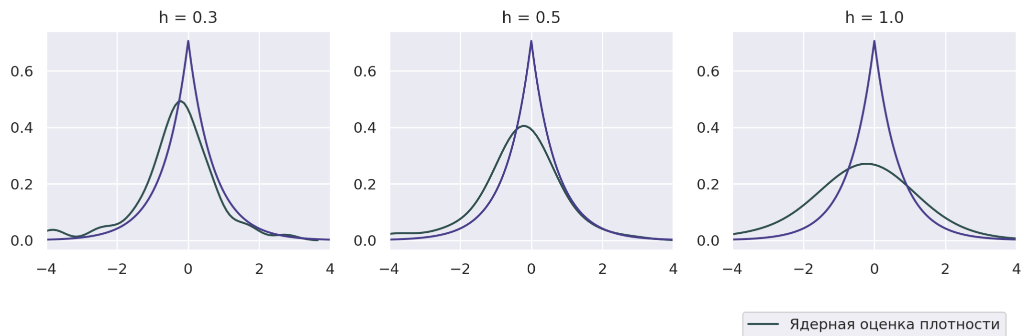


(c)

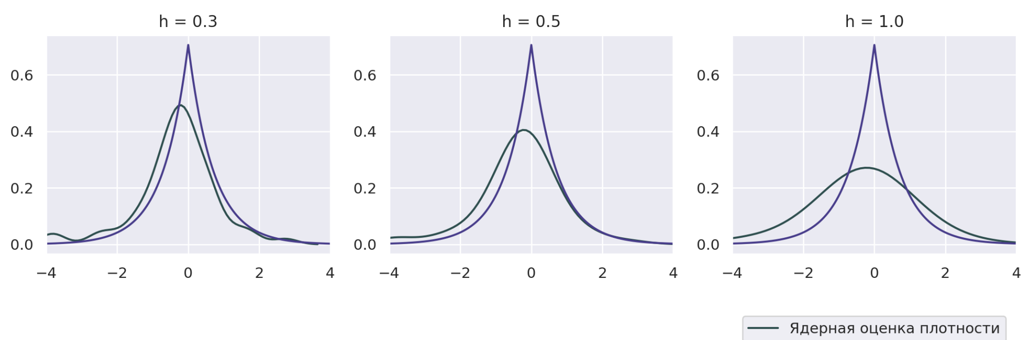
Рис. 8: Для выборок из распределения Коши объемом: (a) 20; (b) 60; (c) 100



(a)

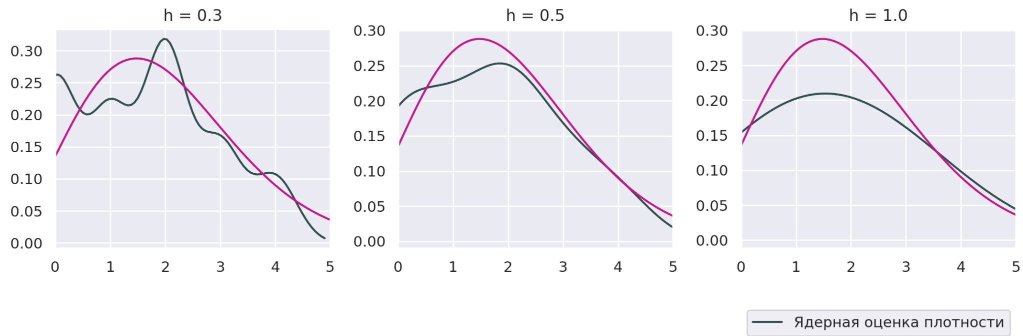


(b)

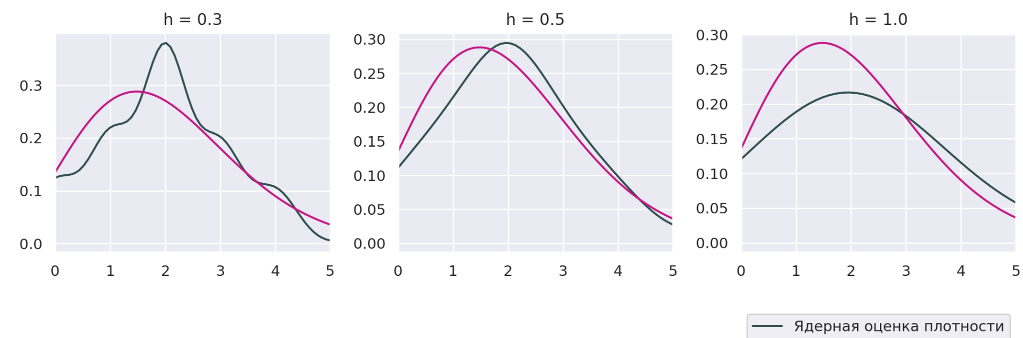


(c)

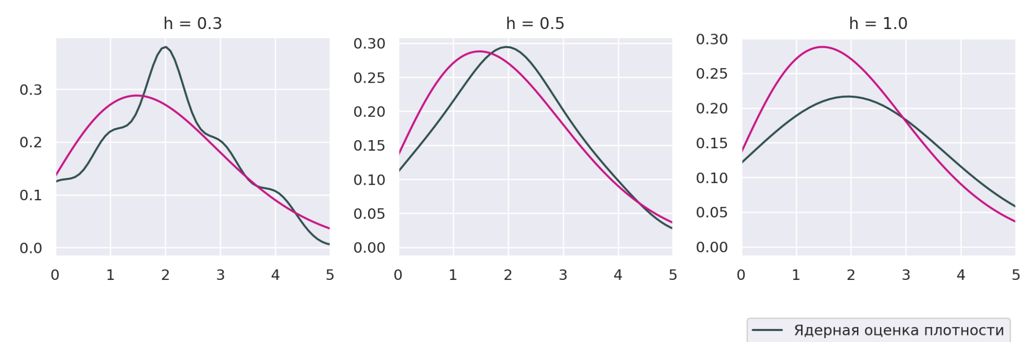
Рис. 9: Для выборок из распределения Лапласа объемом: (a) 20; (b) 60; (c) 100



(a)



(b)



(c)

Рис. 10: Для выборок из распределения Пуассона объемом: (a) 20; (b) 60; (c) 100

Описывая полученные результаты, можно заключить, что чем больше выборка, тем точнее эмпирическая функция распределения оценивает теоретическую.

Точность ядерной оценки плотности сильно варьируется в зависимости от значения сглаживающего параметра h. Так, при $h \to 0$ оценка плотности точна на выборочных данных, но только на них, и такая функция не способна описать характер распределения. Выбрав $(n+1)$-ое значение из распределения, мы столкнемся с тем, что, вероятнее всего, статистическая функция плохо оценит значение теоретической функции плотности вероятности для данного выборочного элемента. В таком случае можно говорить о плохой способности функции к обобщению.

Напротив, при увеличении параметра h ядерная оценка плотности показывает себя плохо даже на выборочных данных и вообще не позволяет понять характера распределения.

Выбор параметра сглаживания следует производить исходя из того, насколько плотно распределение объектов выборки; большей плотности соответствует выбор меньшего параметра и наоборот.

Список литературы

[1] *Conlen, M.* Kernel Density Estimation (2019). URL: https://mathisonian.github.io/kde/