Санкт-Петербургский политехнический университет
Высшая школа прикладной математики и
вычислительной физики, ИПММ

Направление подготовки
01.03.02 «Прикладная математика и информатика»

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

Выполнил студент гр. 3630102/80201
Кирпиченко С. Р.
Руководитель
Баженов А. Н.

Санкт-Петербург
2021

Содержание

Список иллюстраций

Страница

Список таблиц

Страница

1 Постановка задачи

Для 5 распределений:

- 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 Эмпирическая функция распределения

2.1.1 Статистический ряд

Статистическим рядом назовем совокупность, состоящую из последовательности $\{z_i\}_{i=1}^k$ попарно различных элементов выборки, расположенных по возрастанию, и последовательности $\{n_i\}_{i=1}^k$ частот, с которыми эти элементы содержатся в выборке.

2.1.2 Эмпирическая функция распределения

Эмпирическая функция распределения (э. ф. р.) - относительная частота события $X < x$, полученная по данной выборке:

$$F_n^*(x) = P^*(X < x). \tag{1}$$

2.1.3 Нахождение э. ф. р.

$$F^*(x) = \frac{1}{n}\sum_{z_i<x} n_i. \quad (2)$$

$F^*(x)-$ 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 1: Таблица распределения

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

$$F_n^*(x) \approx F_X(x). \quad (3)$$

2.2 Оценки плотности вероятности

2.2.1 Определение

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

$$\widehat{f}(x) \approx f(x). \quad (4)$$

2.2.2 Ядерные оценки

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

$$\widehat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_n}\right). \quad (5)$$

$K(u)$ - ядро, т. е. непрерывная функция, являющаяся плотностью вероятности, $x_1, ..., x_n$ $-$ элементы выборки, а $\{h_n\}_{n\in\mathbb{N}}$ - последовательность элементов из $\mathbb{R}_+$ такая, что

$$h_n \xrightarrow[n\to\infty]{} 0; \quad nh_n \xrightarrow[n\to\infty]{} \infty. \quad (6)$$

Такие оценки называются непрерывными ядерными.

Гауссово ядро:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}. \tag{7}$$

Правило Сильвермана:

$$h_n = \left(\frac{4\hat{\sigma}^5}{3n} \right)^{1/5} \approx 1.06\hat{\sigma} n^{-1/5}, \tag{8}$$

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

3 Реализация

Лабораторная работа выполнена на языке Python 3.9 с использованием библиотек numpy, scipy, matplotlib, seaborn.

4 Результаты

Более насыщенными цветами выделены полученные результаты, более бледными - теоретические функции распределения и плотности вероятности.

4.1 Эмпирическая функция распределения

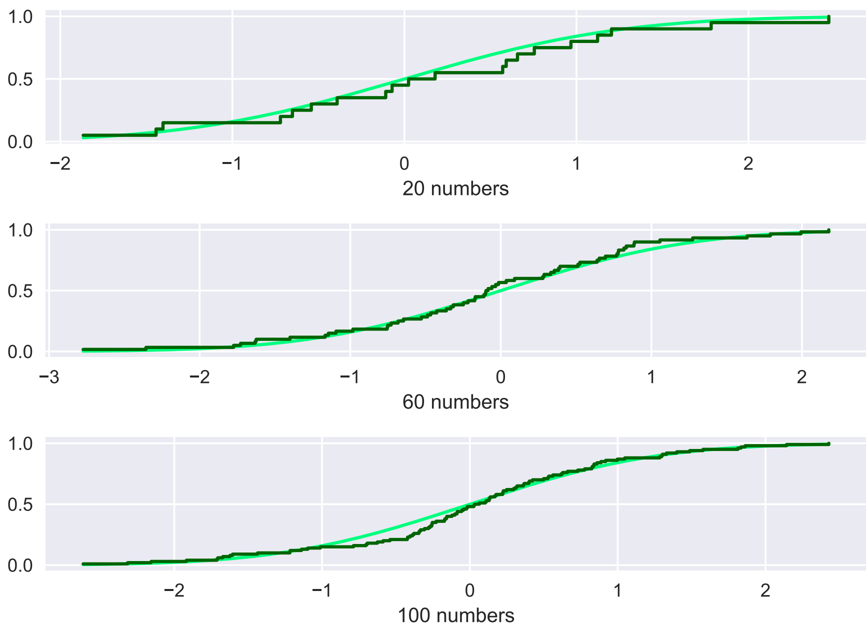

Рис. 1: Нормальное распределение

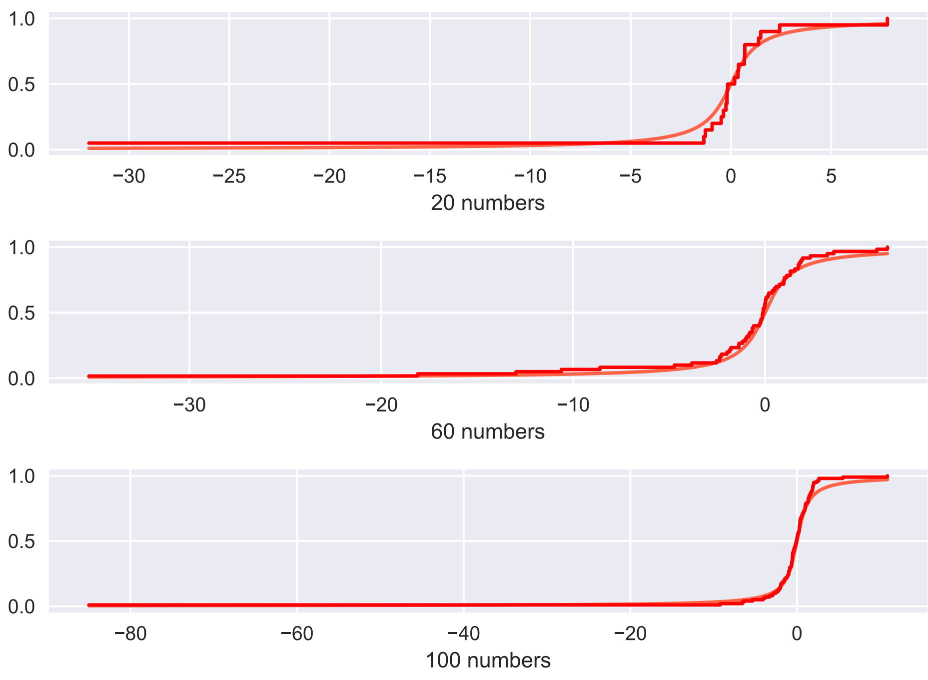

Рис. 2: Распределение Коши

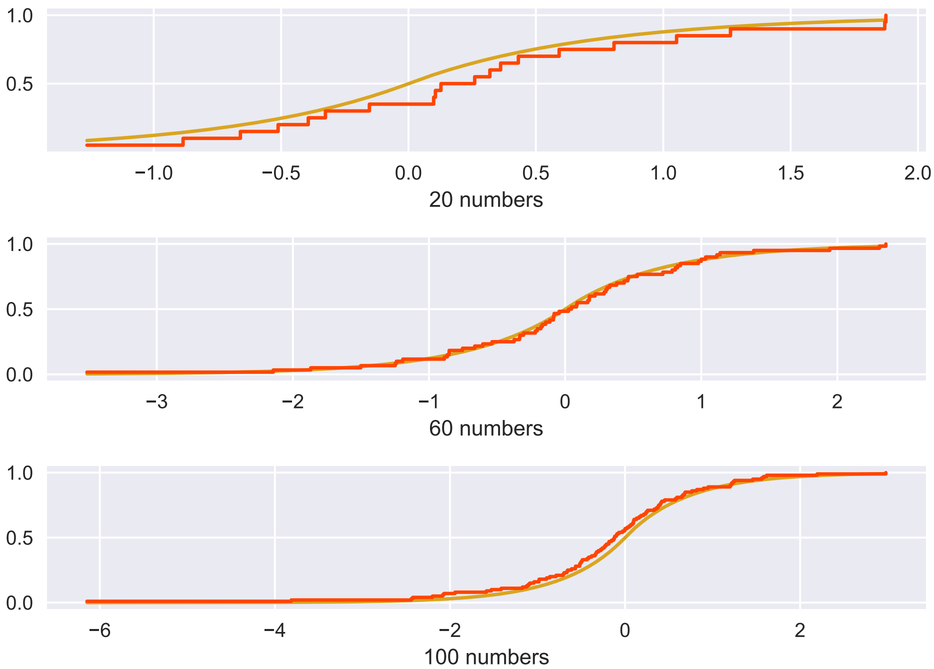

Рис. 3: Распределение Лапласа

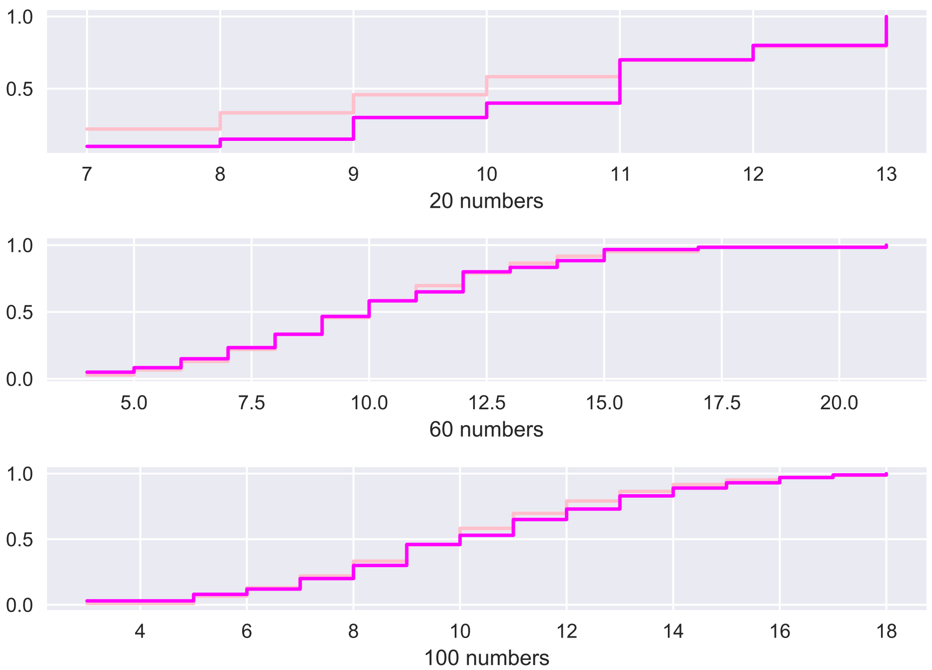

Рис. 4: Распределение Пуассона

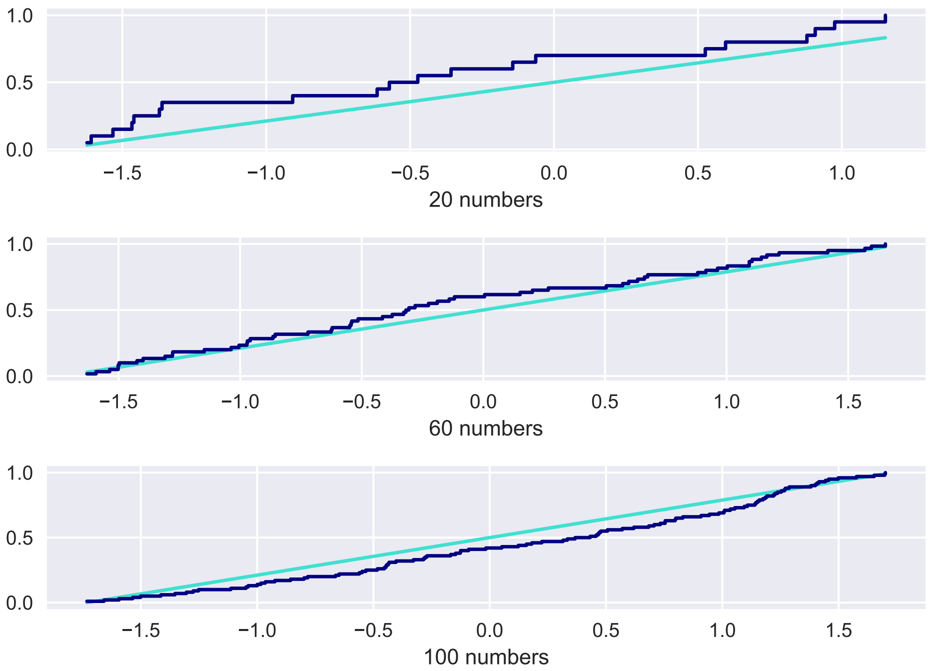

Рис. 5: Равномерное распределение

4.2 Ядерные оценки плотности распределения

4.2.1 Нормальное распределение

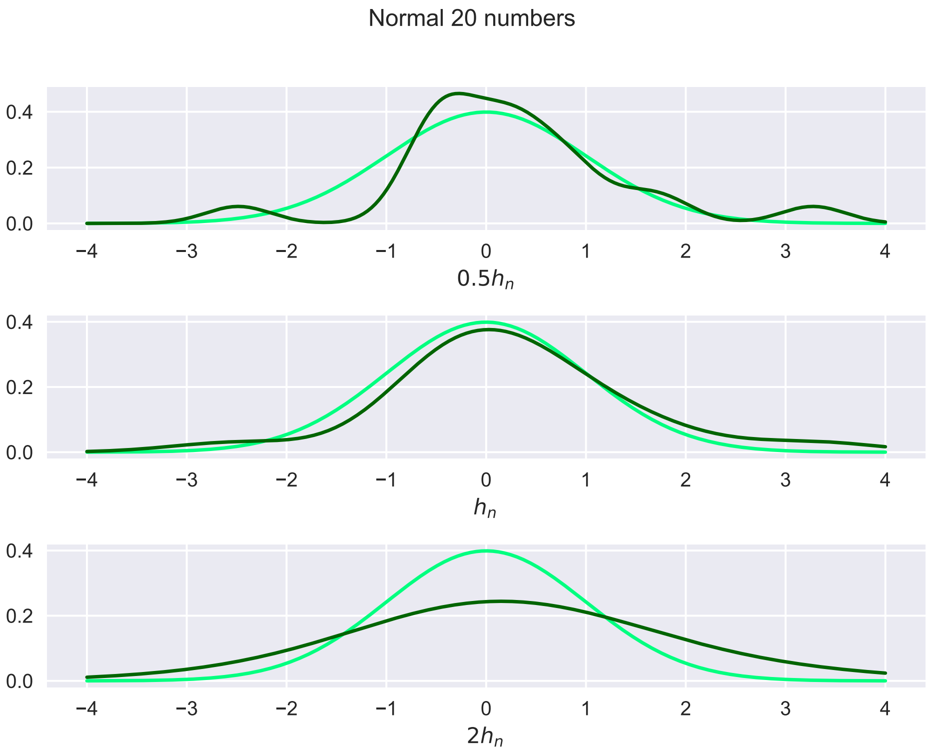

Рис. 6: Нормальное распределение, 20 чисел

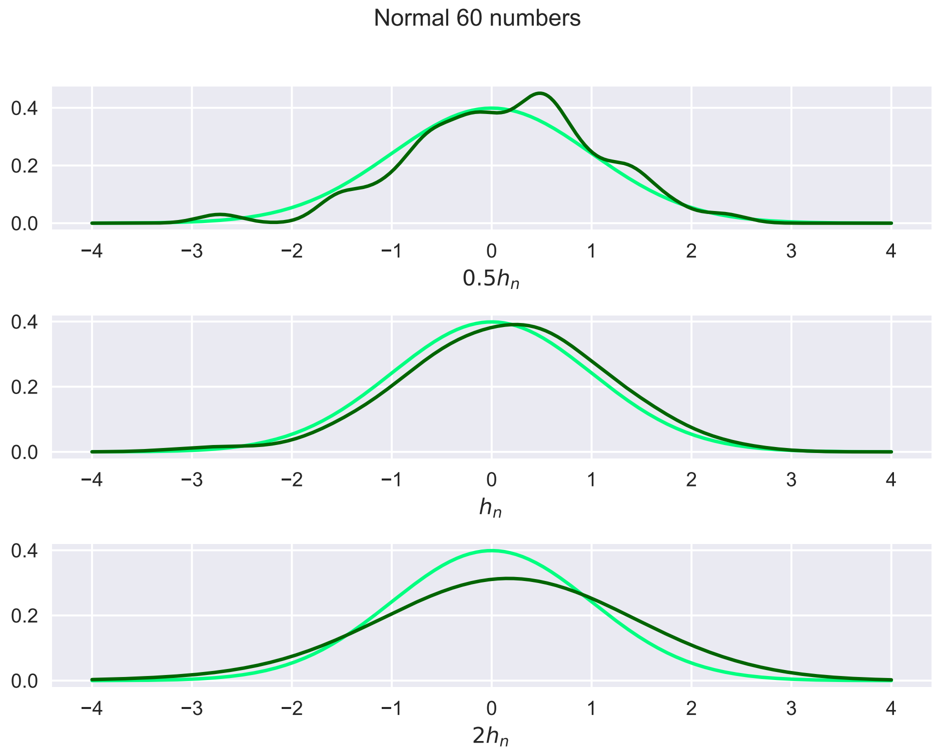

Рис. 7: Нормальное распределение, 60 чисел

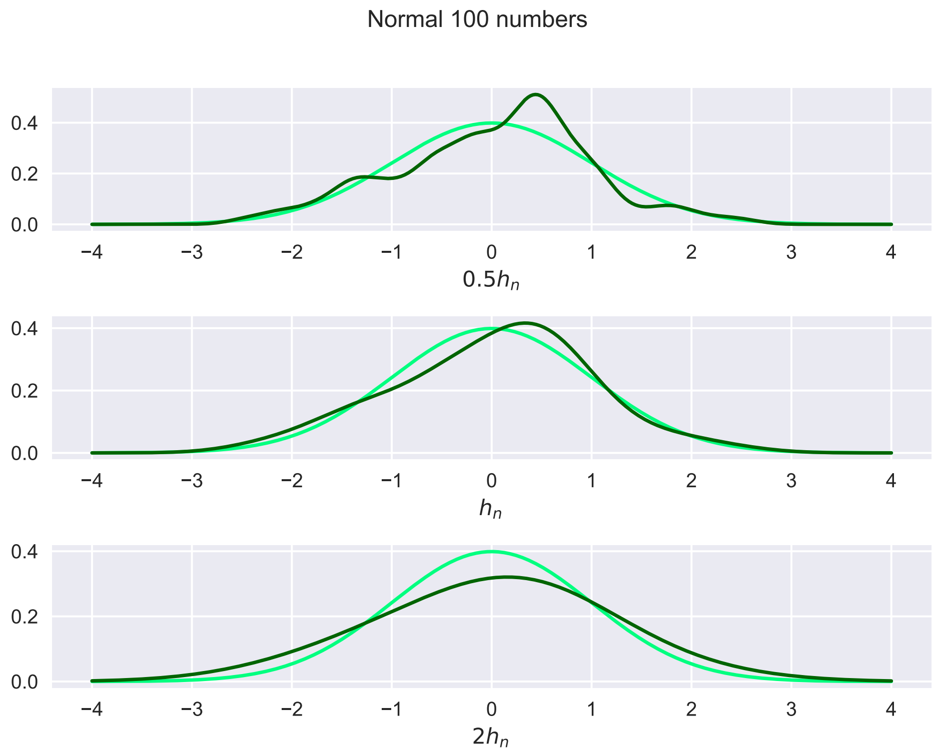

Рис. 8: Нормальное распределение, 100 чисел

4.2.2 Распределение Коши

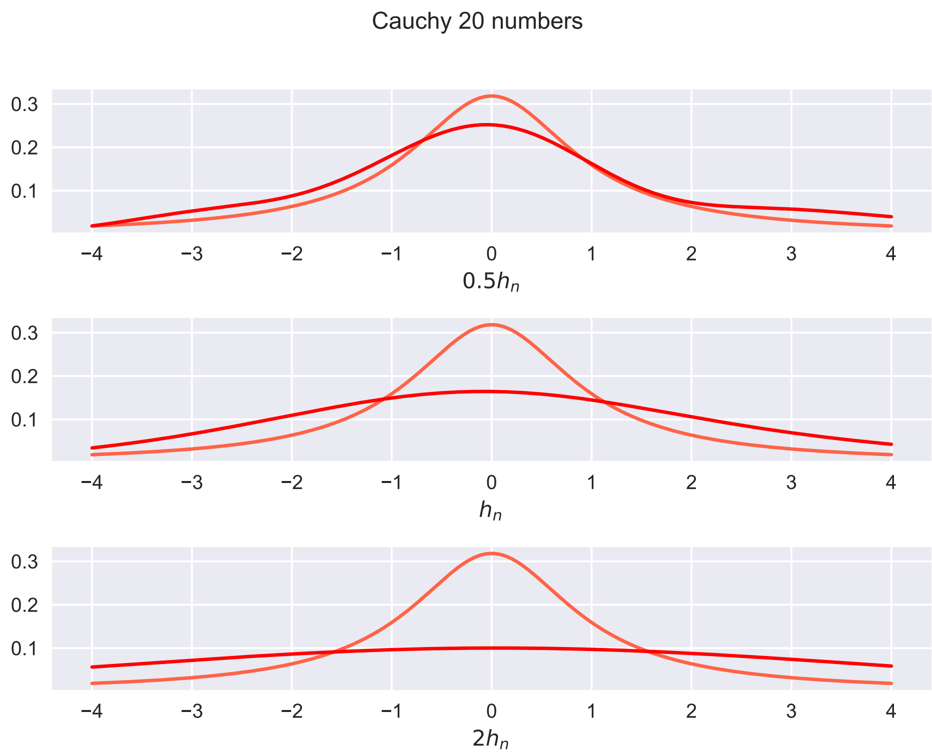

Рис. 9: Распределение Коши, 20 чисел

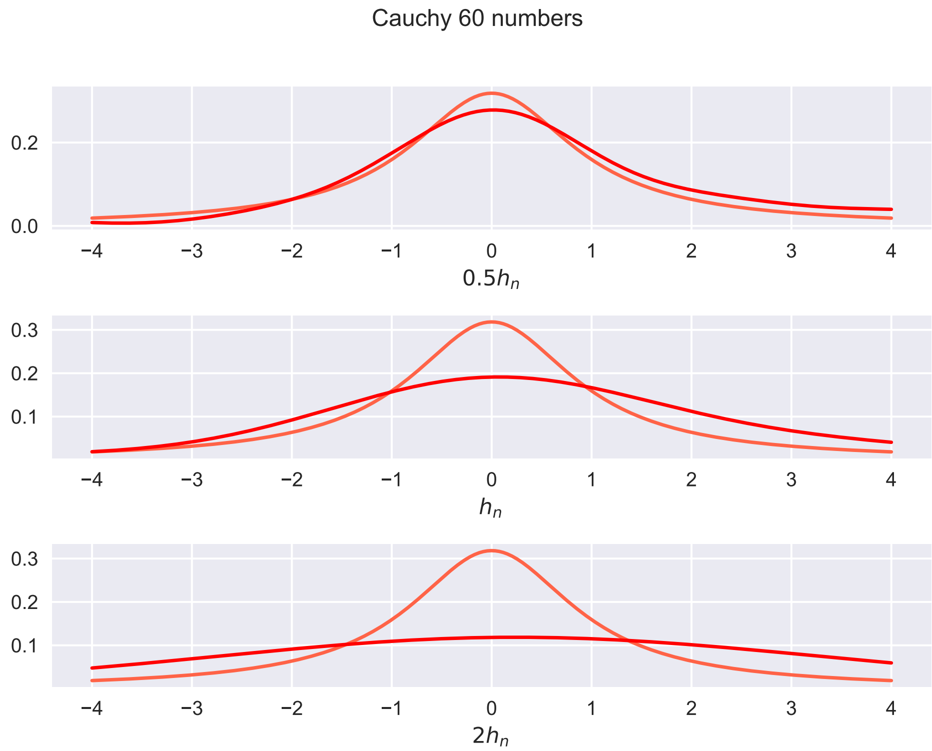

Рис. 10: Распределение Коши, 60 чисел

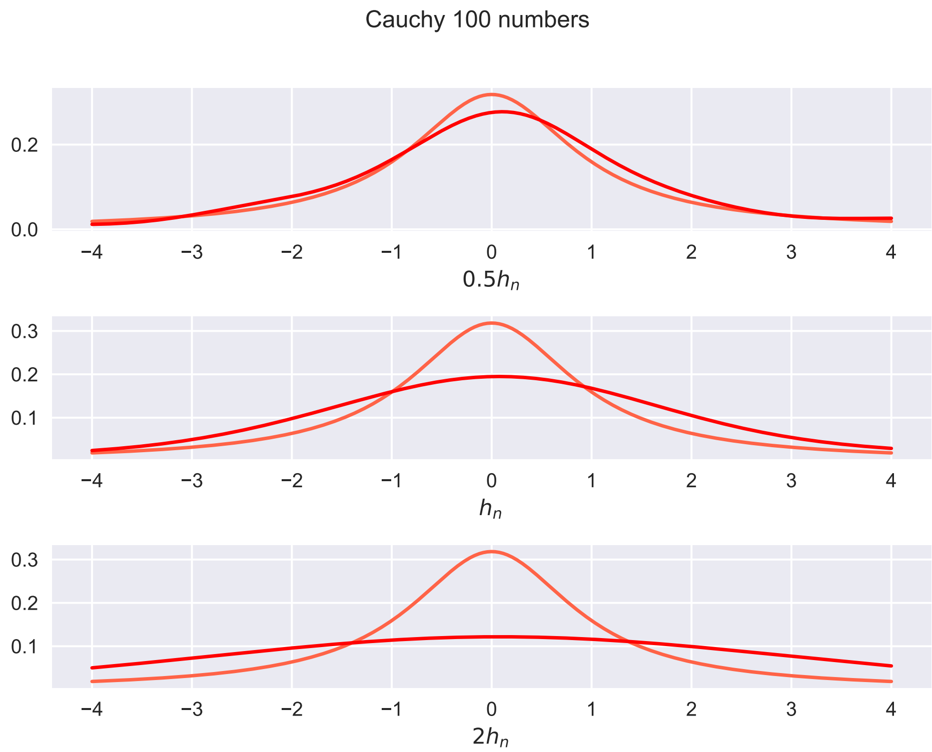

Рис. 11: Распределение Коши, 100 чисел

4.2.3 Распределение Лапласа

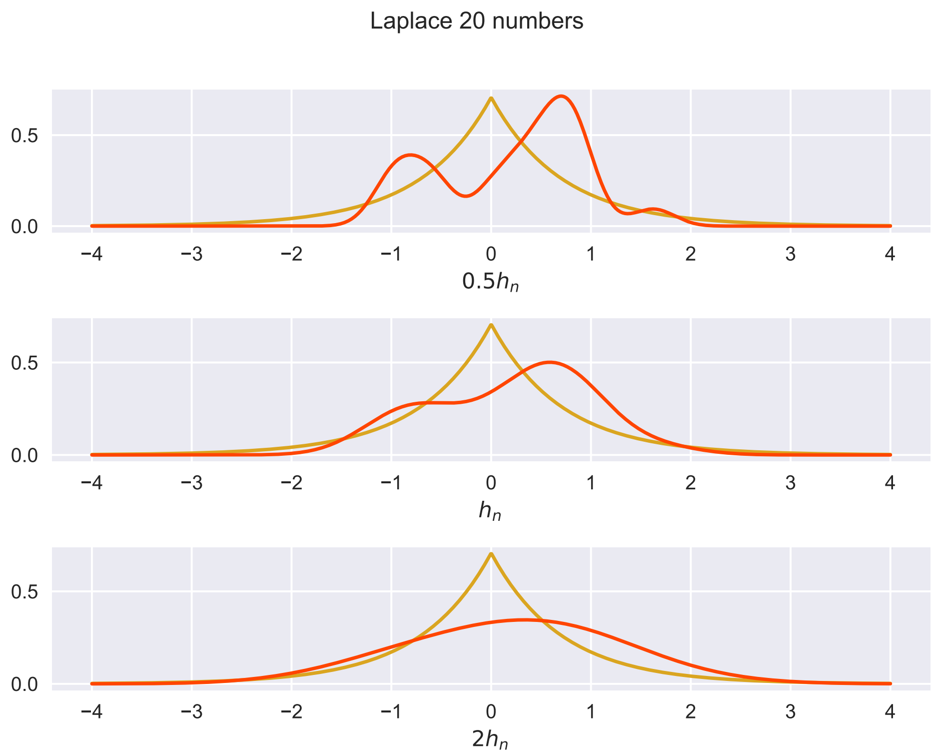

Рис. 12: Распределение Лапласа, 20 чисел

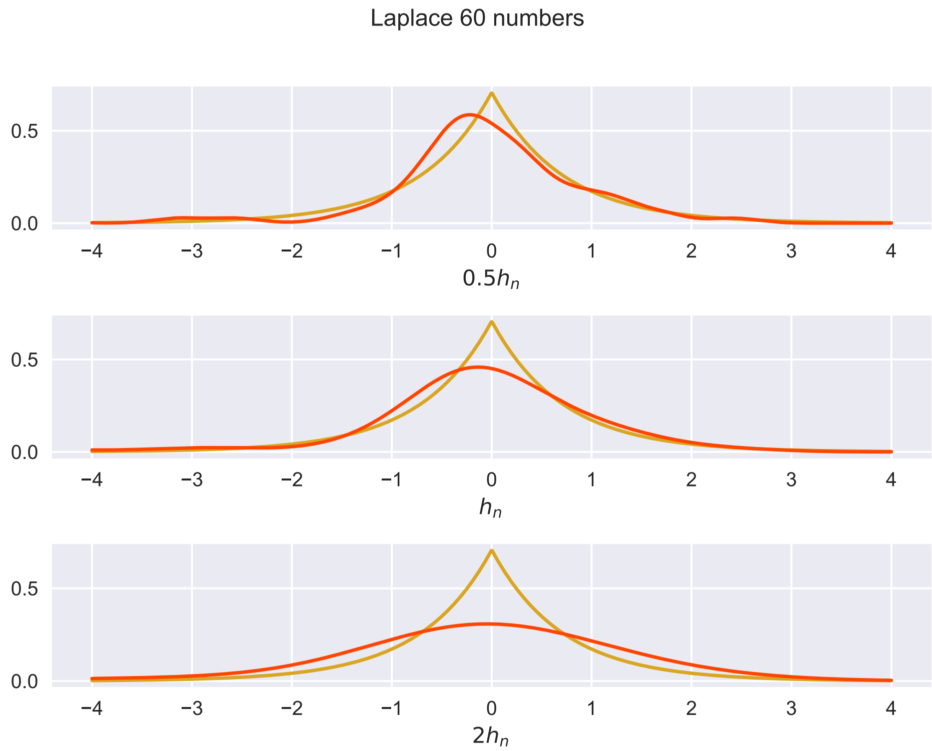

Рис. 13: Распределение Лапласа, 60 чисел

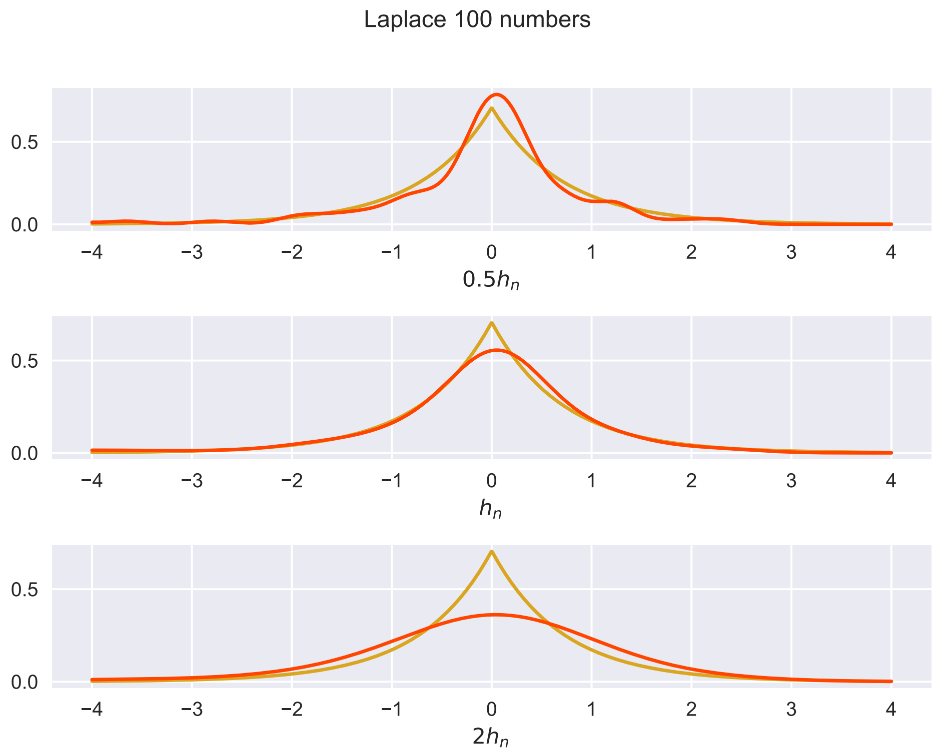

Рис. 14: Распределение Лапласа, 100 чисел

4.2.4 Распределение Пуассона

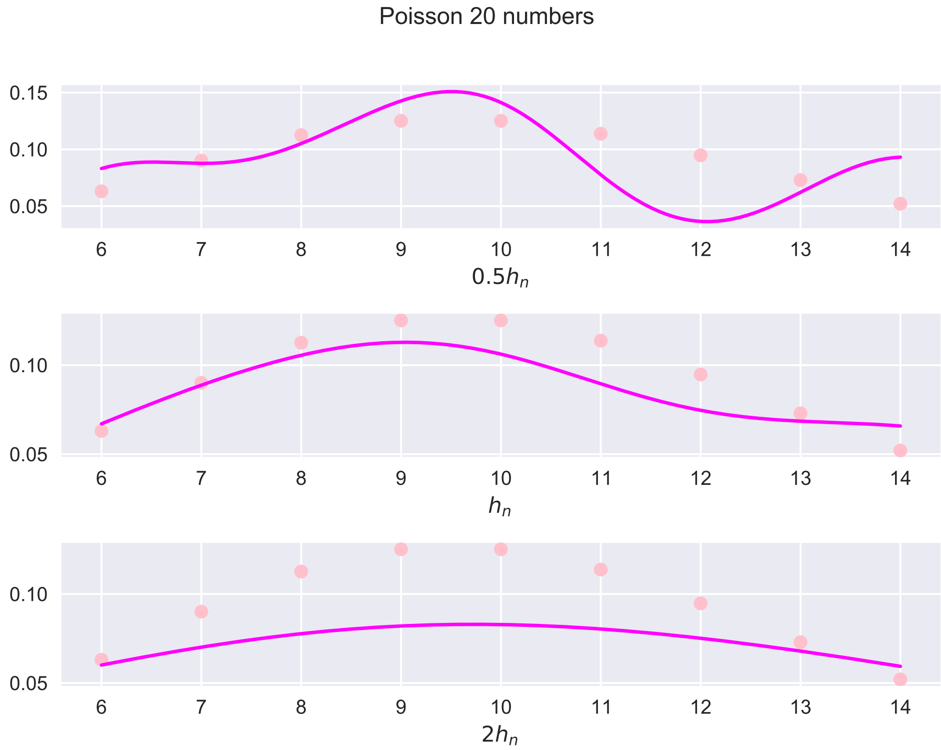

Рис. 15: Распределение Пуассона, 20 чисел

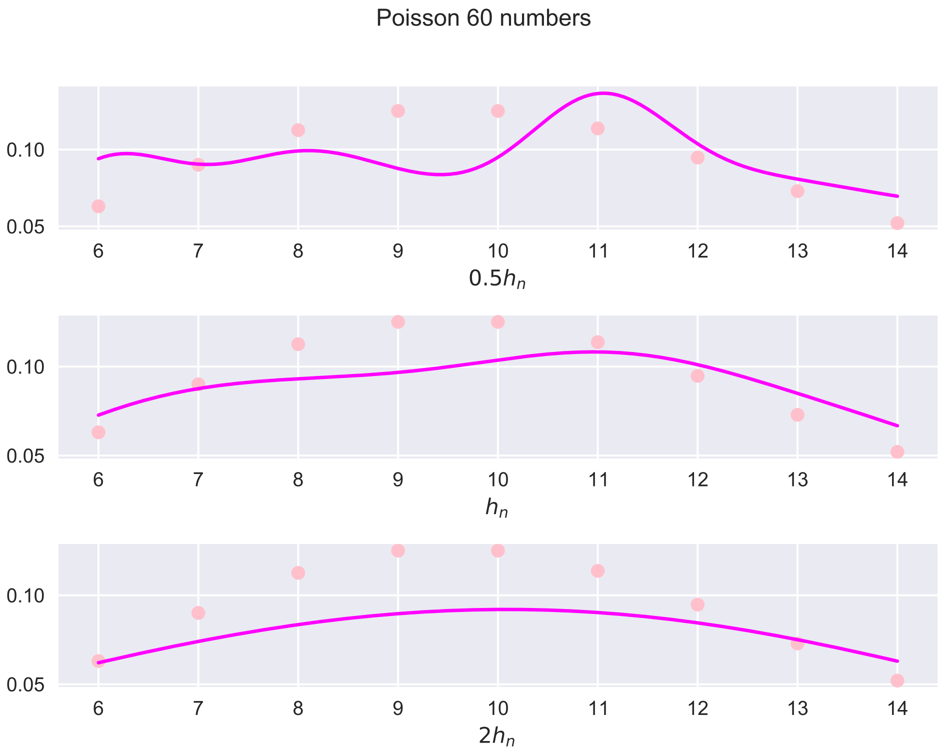

Рис. 16: Распределение Пуассона, 60 чисел

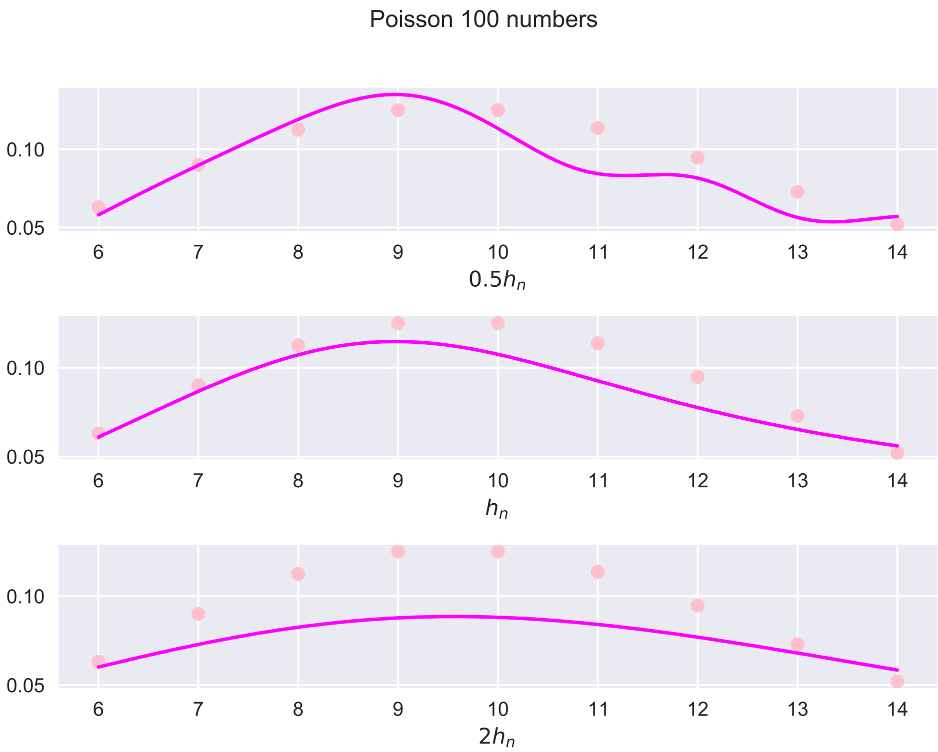

Рис. 17: Распределение Пуассона, 100 чисел

4.2.5 Равномерное распределение

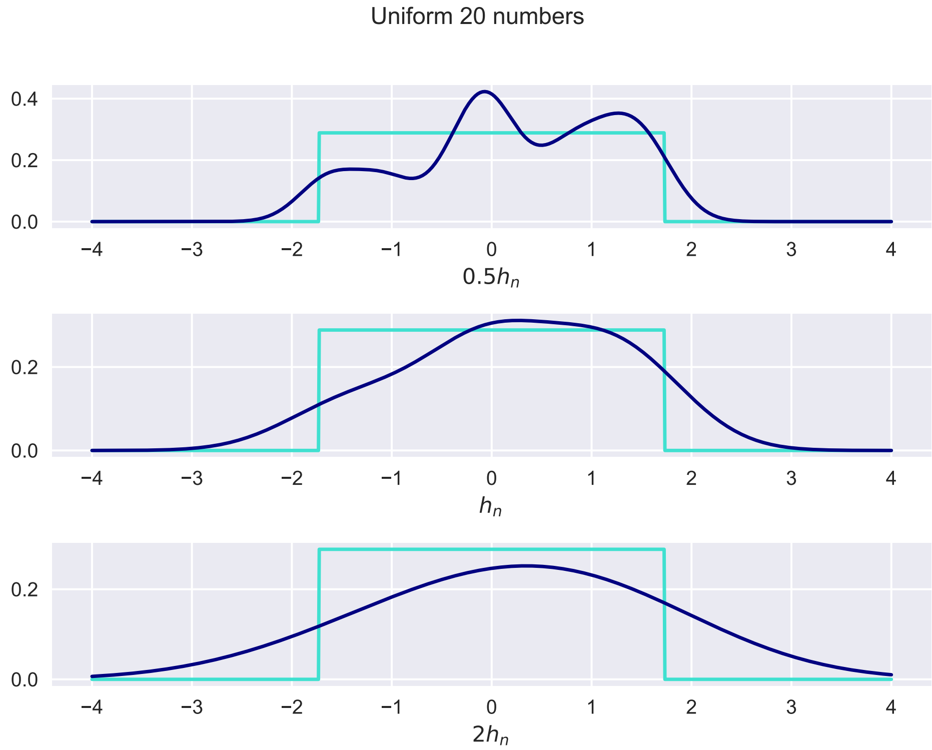

Рис. 18: Равномерное распределение, 20 чисел

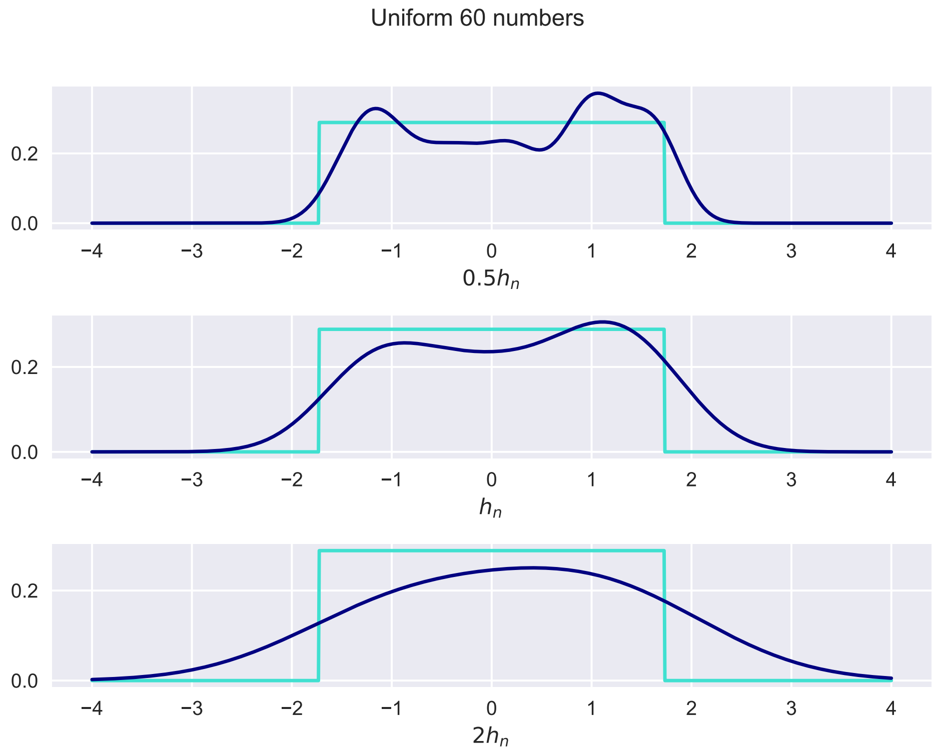

Рис. 19: Равномерное распределение, 60 чисел

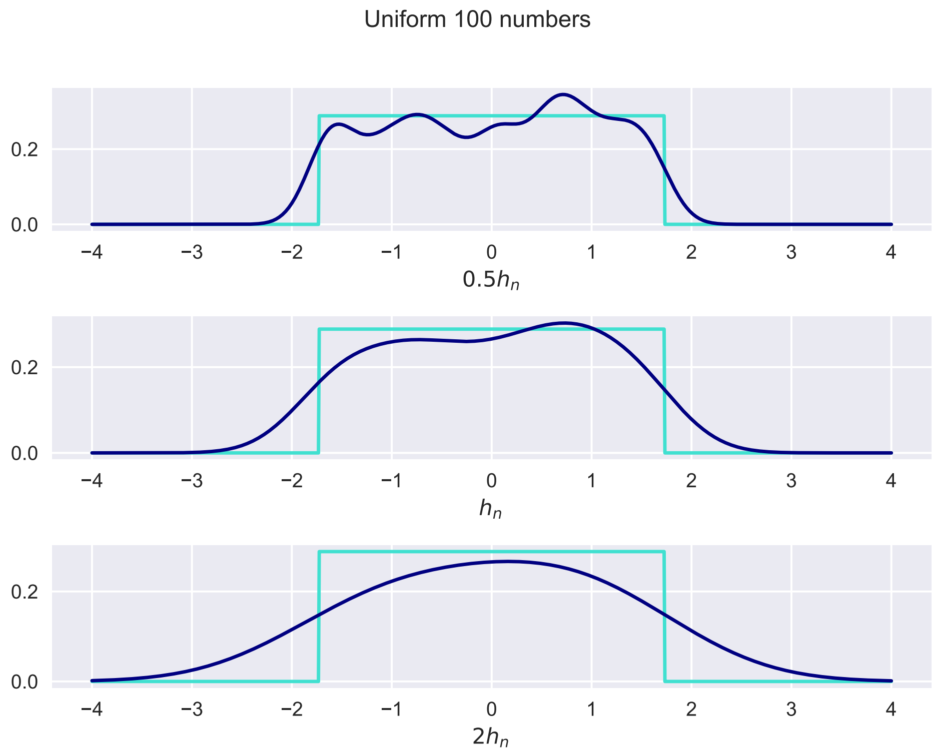

Рис. 20: Равномерное распределение, 100 чисел

5 Обсуждение

5.1 Эмпирическая функция и ядерные оценки плотности распределения

По графикам эмпирической функции распределения видно, что с увеличением мощности выборки растет точность приближения оценки к теоретической функции распределения вероятности. Хуже всего эмпирическая функция распределения приближает теоретическую на равномерном распределении.

По графикам ядерных оценок видно, что при любом выборе параметра h_n увеличение мощности выборки положительно сказывается на точности оценки.

Для различных распределений больше подходят различные параметры h_n: для нормального, пуассоновского и равномерного распределений большую эффективность показывает параметр h_n. Для распределений Коши и Лапласа - $\frac{h_n}{2}$.

Также можно сделать вывод, что увеличении коэффициента при параметре h_n ведет к сглаживанию ядерной оценки: количество перегибов у функции уменьшается, при значении параметра $2h_n$ ядерные оценки становятся сложно различимыми у разных распределений.

Примечание

С исходным кодом работы и данного отчета можно ознакомиться в репозитории `https://github.com/Stasychbr/MatStat`