

Санкт-Петербургский Политехнический Университет имени Петра Великого  
Институт Прикладной Математики и Механики  
Кафедра "Прикладная Математика"

**Отчет по лабораторной работе №4**  
**по дисциплине**  
**"Математическая Статистика"**

Выполнил студент:  
Тырыкин Я. А.  
группа 3630102/80401  
Проверил:  
к.ф.-м.н., доцент  
Баженов А. Н.

Санкт-Петербург  
2021

# Содержание

<b>1</b>	<b>Постановка задачи</b>	<b>4</b>
<b>2</b>	<b>Теория</b>	<b>4</b>
2.1	Распределения	4
2.2	Эмпирическая функция распределения	5
2.2.1	Статистический ряд	5
2.2.2	Определение	5
2.2.3	Описание	5
2.3	Оценки плотности вероятности	6
2.3.1	Определение	6
2.3.2	Ядерные оценки	6
2.3.3	Оценка качества ядерных приближений	6
<b>3</b>	<b>Модульная структура программы</b>	<b>7</b>
<b>4</b>	<b>Результаты</b>	<b>7</b>
4.1	Эмпирическая функция распределения	7
4.2	Ядерные оценки плотности	9
<b>5</b>	<b>Обсуждение</b>	<b>13</b>
5.1	Эмпирическая функция и ядерные оценки плотности распределения	13
<b>6</b>	<b>Ресурсы</b>	<b>14</b>

## Список иллюстраций

1	Нормальное распределение (1) . . . . .	7
2	Распределение Коши (2) . . . . .	7
3	Распределение Лапласа (3) . . . . .	8
4	Распределение Пуассона (4) . . . . .	8
5	Равномерное распределение (5) . . . . .	8
6	Нормальное распределение, $h_n = \frac{h_n}{2}$ (1) . . . . .	9
7	Нормальное распределение, $h_n = h_n$ (1) . . . . .	9
8	Нормальное распределение, $h_n = 2h_n$ (1) . . . . .	9
9	Распределение Коши $h_n = \frac{h_n}{2}$ (2) . . . . .	10
10	Распределение Коши $h_n = h_n$ (2) . . . . .	10
11	Распределение Коши $h_n = 2h_n$ (2) . . . . .	10
12	Распределение Лапласа $h_n = \frac{h_n}{2}$ (3) . . . . .	11
13	Распределение Лапласа $h_n = h_n$ (3) . . . . .	11
14	Распределение Лапласа $h_n = 2h_n$ (3) . . . . .	11
15	Распределение Пуассона $h_n = \frac{h_n}{2}$ (4) . . . . .	12
16	Распределение Пуассона $h_n = h_n$ (4) . . . . .	12
17	Распределение Пуассона $h_n = 2h_n$ (4) . . . . .	12
18	Равномерное распределение $h_n = \frac{h_n}{2}$ (5) . . . . .	13
19	Равномерное распределение $h_n = h_n$ (5) . . . . .	13
20	Равномерное распределение $h_n = 2h_n$ (5) . . . . .	13

# 1 Постановка задачи

Для 5 распределений:

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

Необходимо 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:

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

- Распределение Коши:

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

- Распределение Лапласа:

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

- Распределение Пуассона:

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

- Равномерное распределение:

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

## 2.2 Эмпирическая функция распределения

### 2.2.1 Статистический ряд

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

$z$	$z_1$	$z_2$	$\dots$	$z_k$
$n$	$n_1$	$n_2$	$\dots$	$n_k$

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:

$$F_n^*(x) = P^*(X < x) \quad (6)$$

### 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

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

$F^*(x)$  — функция распределения дискретной случайной величины  $X^*$ , заданной таблицей распределения:

$X^*$	$z_1$	$z_2$	$\dots$	$z_k$
$P$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\dots$	$\frac{n_k}{n}$

Таблица 2: Таблица распределения

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

$$F_n^*(x) \approx F_X(x) \quad (8)$$

## 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  $f(x)$

$$\hat{f}(x) \approx f(x) \quad (9)$$

### 2.3.2 Ядерные оценки

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

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) \quad (10)$$

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

$$h_n \xrightarrow{n \rightarrow \infty} 0; \quad nh_n \xrightarrow{n \rightarrow \infty} \infty \quad (11)$$

Такие оценки называются непрерывными ядерными.

Гауссово (нормальное) ядро:

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

Правило Сильвермана:

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

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

### 2.3.3 Оценка качества ядерных приближений

После получения результатов возникает необходимость оценить качество ядерных приближений. Приведем в пример один из приемов количественных описаний сходства кривых - метод Фреше.

Расстояние Фреше — это мера сходства кривых, принимающая во внимание число и порядок точек вдоль кривых. Расстояние названо по имени французского математика Мориса Фреше. Метрика Фреше принимает во внимание течение двух кривых, поскольку пары точек, расстояние между которыми определяет расстояние Фреше, «пробегают» вдоль кривых. Расстояние Фреше между двумя кривыми — это не длина

самого короткого поводка, с которым можно пройти все пути, а самый короткий, при котором можно пройти этот путь.

Определим кривую как непрерывное отображение  $f : [a, b] \rightarrow V$ , где  $a, b \in R$  и  $a \leq b$  и  $(V, d)$  — метрическое пространство. Даны две кривые  $f : [a, b] \rightarrow V$  и  $g : [a', b'] \rightarrow V$ , их расстояние Фреше определено в виде:

### 3 Модульная структура программы

Лабораторная работа выполнена с применением средств языка Python версии 3.7 в среде разработки PyCharm IDE (в частности, с применением встроенных методов библиотеки SciPy и Matplotlib). Исходный код лабораторной работы находится по ссылке в приложении к отчёту.

## 4 Результаты

### 4.1 Эмпирическая функция распределения

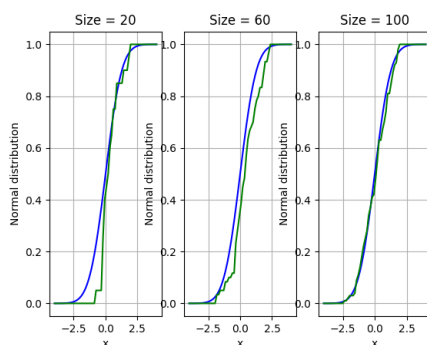


Рис. 1: Нормальное распределение (1)

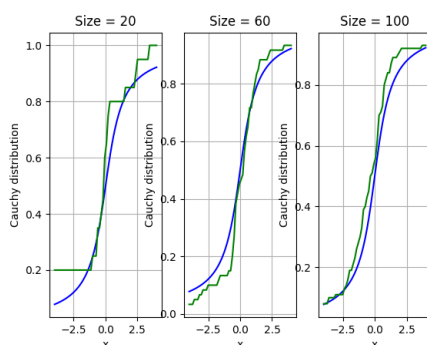


Рис. 2: Распределение Коши (2)

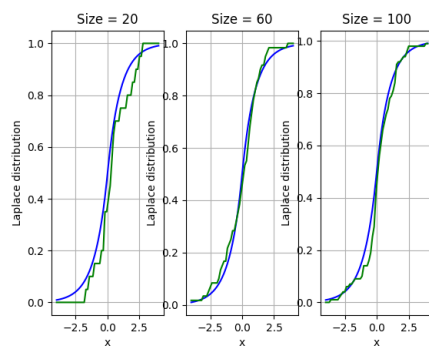


Рис. 3: Распределение Лапласа (3)

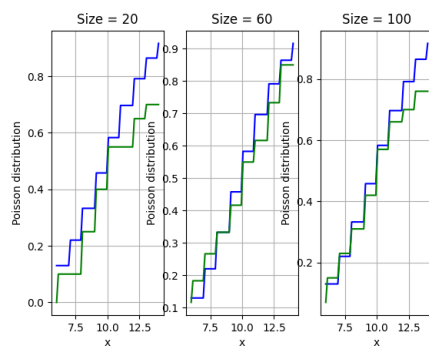


Рис. 4: Распределение Пуассона (4)

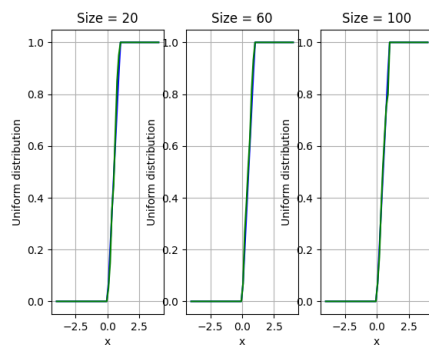


Рис. 5: Равномерное распределение (5)



## 4.2 Ядерные оценки плотности

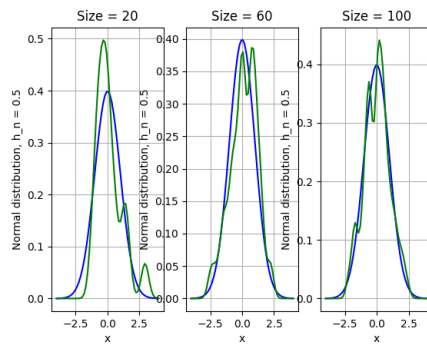


Рис. 6: Нормальное распределение,  $h_n = \frac{h_n}{2}$  (1)

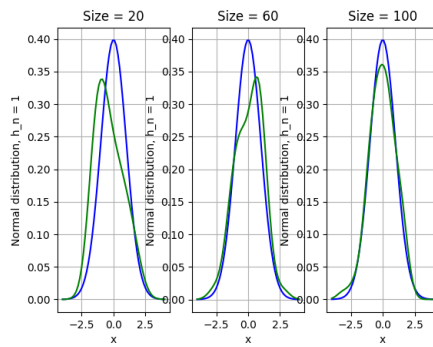


Рис. 7: Нормальное распределение,  $h_n = h_n$  (1)

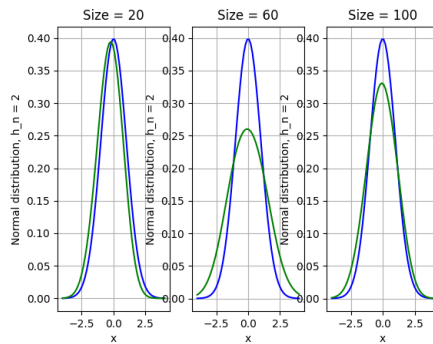


Рис. 8: Нормальное распределение,  $h_n = 2h_n$  (1)

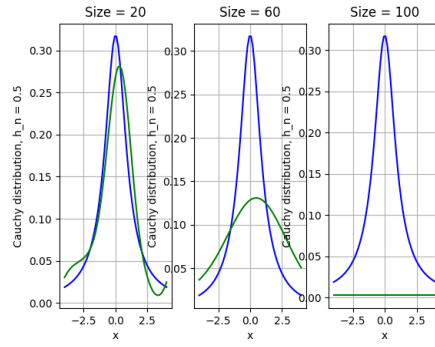


Рис. 9: Распределение Коши  $h_n = \frac{h_n}{2}$  (2)

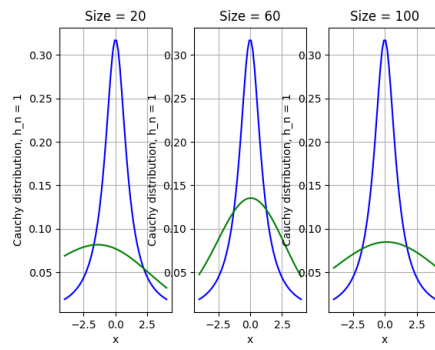


Рис. 10: Распределение Коши  $h_n = h_n$  (2)

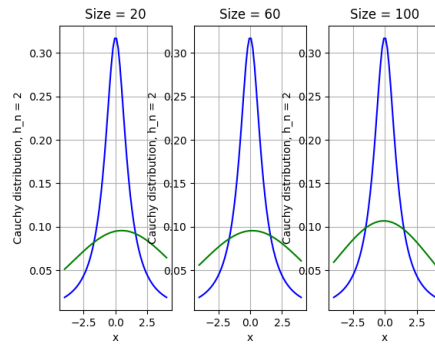


Рис. 11: Распределение Коши  $h_n = 2h_n$  (2)

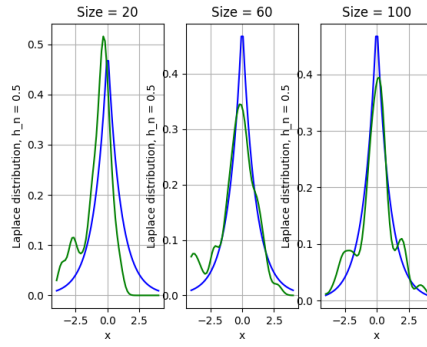


Рис. 12: Распределение Лапласа  $h_n = \frac{h_n}{2}$  (3)

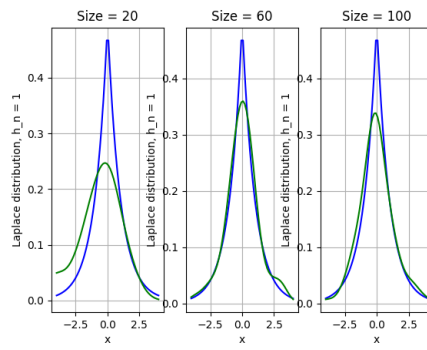


Рис. 13: Распределение Лапласа  $h_n = h_n$  (3)

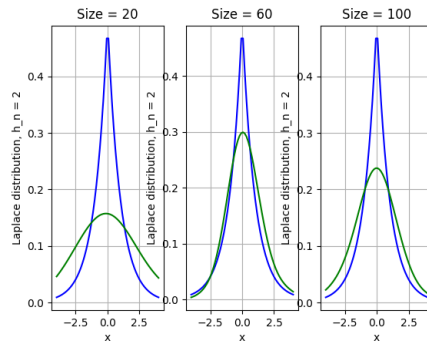


Рис. 14: Распределение Лапласа  $h_n = 2h_n$  (3)

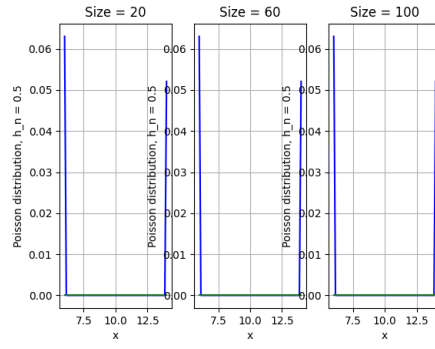


Рис. 15: Распределение Пуассона  $h_n = \frac{h_n}{2}$  (4)

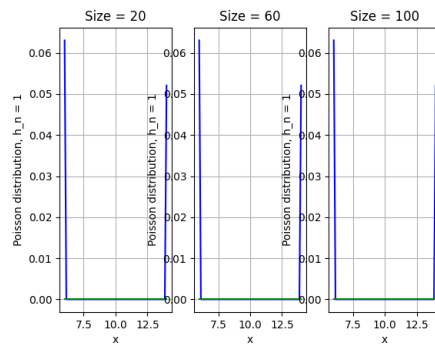


Рис. 16: Распределение Пуассона  $h_n = h_n$  (4)

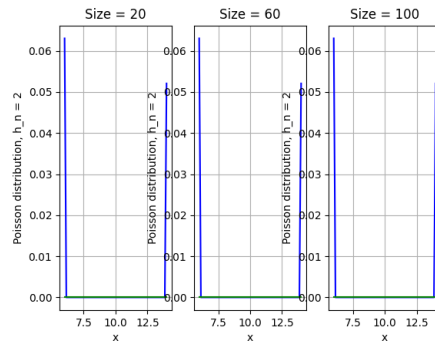


Рис. 17: Распределение Пуассона  $h_n = 2h_n$  (4)

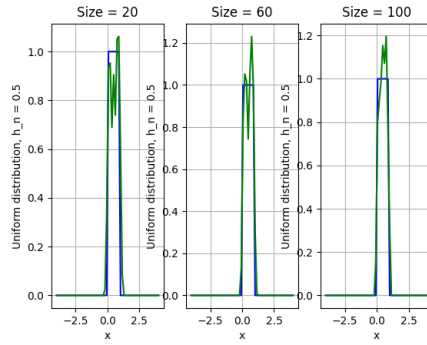


Рис. 18: Равномерное распределение  $h_n = \frac{h_n}{2}$  (5)

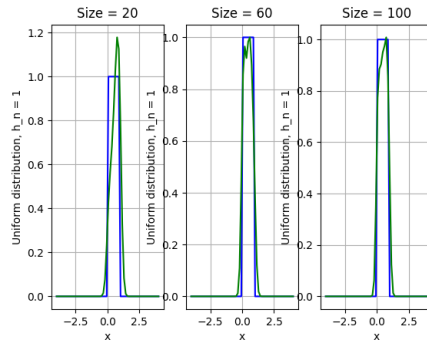


Рис. 19: Равномерное распределение  $h_n = h_n$  (5)

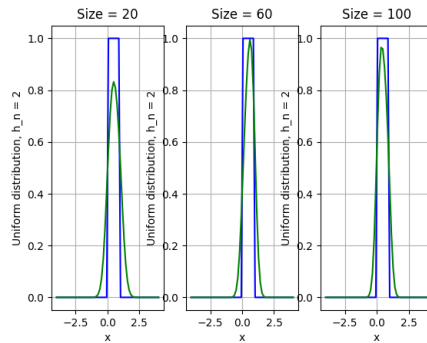


Рис. 20: Равномерное распределение  $h_n = 2h_n$  (5)

## 5 Обсуждение

### 5.1 Эмпирическая функция и ядерные оценки плотности распределения

Можем наблюдать на иллюстрациях с э. ф. р., что ступенчатая эмпирическая функция распределения тем лучше приближает функцию распределения реальной

выборки, чем мощнее эта выборка. Заметим так же, что для распределения Пуассона и равномерного распределения отклонение функций друг от друга наибольшее. //

Рисунки, посвященные ядерным оценкам, иллюстрируют сближение ядерной оценки и функции плотности вероятности для всех  $h$  с ростом размера выборки. Для распределения Пуассона наиболее ярко видно, как сглаживает отклонения увеличение параметра сглаживания  $h$ .

В зависимости от особенностей распределений для их описания лучше подходят разные параметры  $h$ . В зависимости от особенностей распределений для их описания лучше подходят разные параметры  $h = 2h_n$ , для распределения Лапласа -  $h = h_n/2$ , а для нормального распределения и распределения Коши -  $h = h_n$ . Такие значения дают вид ядерной оценки наиболее близкий к плотности, характерной данным распределениям.

Также можно увидеть, что чем больше коэффициент при параметре сглаживания  $\widehat{h}_n$  тем меньше изменений знака производной у аппроксимирующей функции, вплоть до того, что при  $h = 2h_n$  функция становится унимодальной на рассматриваемом промежутке. Также видно, что при  $h = 2h_n$  по полученным приближениям становится сложно сказать плотность вероятности какого распределения они должны повторять, так как они очень похожи между собой.

## 6 Ресурсы

Код программы, реализующей отрисовку обозначенных распределений:

<https://github.com/YaroslavAggressive/Mathematical-statistics-lab-works>