

**Санкт-Петербургский политехнический университет Петра Великого**

**Институт прикладной математики и механики**

**Кафедра «Телематика (при ЦНИИ РТК)»**

**Отчет по лабораторной работе**

**Построение боксплотов Тьюки и оценка выбросов**

По дисциплине «Теория вероятностей и Математическая статистика»

Выполнил

Студент гр. 3630201/80101

\_\_\_\_\_

М.Д. Маляренко

Руководитель

к.ф.-м.н., доцент

\_\_\_\_\_

А.Н. Баженов

« \_\_\_\_ » \_\_\_\_\_ 2020г.

Санкт-Петербург  
2020

# Содержание

<b>1</b>	<b>Постановка задачи</b>	<b>4</b>
<b>2</b>	<b>Теория</b>	<b>5</b>
2.1	Эмпирическая функция распределения . . . . .	5
2.2	Ядерные оценки . . . . .	5
<b>3</b>	<b>Реализация</b>	<b>6</b>
<b>4</b>	<b>Результаты</b>	<b>7</b>
4.1	Эмпирическая функция распределения . . . . .	7
4.2	Ядерные оценки . . . . .	8
	<b>Заключение</b>	<b>10</b>
	<b>Список литературы</b>	<b>11</b>
	<b>Приложение А. Репозиторий с исходным кодом</b>	<b>12</b>

## Список иллюстраций

1	Эмпирическая функция нормального распределения $N(x, 0, 1)$ . . . . .	7
2	Эмпирическая функция распределения Лапласа $L(x, 0, 1/\sqrt{2})$ . . . . .	7
3	Эмпирическая функция распределения Коши $C(x, 0, 1)$ . . . . .	7
4	Эмпирическая функция распределения Пуассона $P(k, 10)$ . . . . .	8
5	Эмпирическая функция равномерного распределения $U(x, -\sqrt{3}, \sqrt{3})$ . . . . .	8
6	Ядерная оценка плотности нормального распределения $N(x, 0, 1)$ . . . . .	8
7	Ядерная оценка плотности распределения Лапласа $L(x, 0, 1/\sqrt{2})$ . . . . .	9
8	Ядерная оценка плотности распределения Коши $C(x, 0, 1)$ . . . . .	9
9	Ядерная оценка плотности распределения Пуассона $P(k, 10)$ . . . . .	9
10	Ядерная оценка плотности равномерного распределения $U(x, -\sqrt{3}, \sqrt{3})$ . . . . .	9

# 1 Постановка задачи

Заданы 5 распределений случайных величин:

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

Для каждого распределения необходимо сгенерировать выборки размером в 20, 60 и 100 элементов. На этих выборках построить эмпирические функции распределения. Также для выборок построить ядерные оценки на нормальном (гауссовом) ядре с тремя параметрами сглаживания:  $h_n/2$ ,  $h_n$  и  $2h_n$ , где  $h_n$  – полоса пропускания, рассчитанная по эмпирическому правилу Сильвермана.

Все графики необходимо строить на промежутке  $[-4, 4]$  для непрерывных распределений и на  $[6, 14]$  для дискретного распределения Пуассона.

## 2 Теория

### 2.1 Эмпирическая функция распределения

Эмпирической (выборочной) функцией распределения называется относительная частота события  $X < x$ , полученная по данной выборке:[1]

$$F_n^*(x) = P^*(X < x)$$

Для расчета эмпирической функции распределения необходимо составить по выборке статистический ряд и просуммировать все частоты  $n_i$ , для которых элементы  $x_i < x$ :

$$F_n^*(x) = \frac{1}{n} \sum_{x_i < x} n_i \quad (1)$$

### 2.2 Ядерные оценки

Оценка плотности вероятности  $f(x)$  – функция  $\hat{f}_n(x)$ , построенная на основе выборки  $x_1, \dots, x_n$ , представимая суммой:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h_n}\right) \quad (2)$$

В формуле (2)  $K(u)$  – ядерная функция,  $\{h_n\}$  – последовательность положительных чисел, для которой

$$h_n \xrightarrow{n \rightarrow \infty} 0 \quad \frac{h_n}{n^{-1}} \xrightarrow{n \rightarrow \infty} \infty$$

В данной лабораторной работе в качестве ядерной функции  $K(u)$  используется гауссово ядро:

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

Для расчета полосы пропускания используется эмпирическое правило Сильвермана:[2]

$$h_n = 1.06\hat{\sigma}^{-1/5} \quad (4)$$

, где  $\hat{\sigma}$  – стандартное выборочное отклонение,  $n$  – размер выборки.

### 3 Реализация

Расчёты и построение графиков производились в среде аналитических вычислений *Mathematica*. Для построения ядерной оценки были написаны вспомогательные функции, параметрически задающие оценку плотности вероятности. Для построения графиков использовалась интегрированная в среду утилита *gnuplot*. Полный текст скрипта для среды *Mathematica* представлен в репозитории *GitHub*.

## 4 Результаты

### 4.1 Эмпирическая функция распределения

На рисунках 1-5 представлены графики, по которым можно визуальнo оценить отклонение эмпирической функции распределения (1) с теоретической функций распределения.

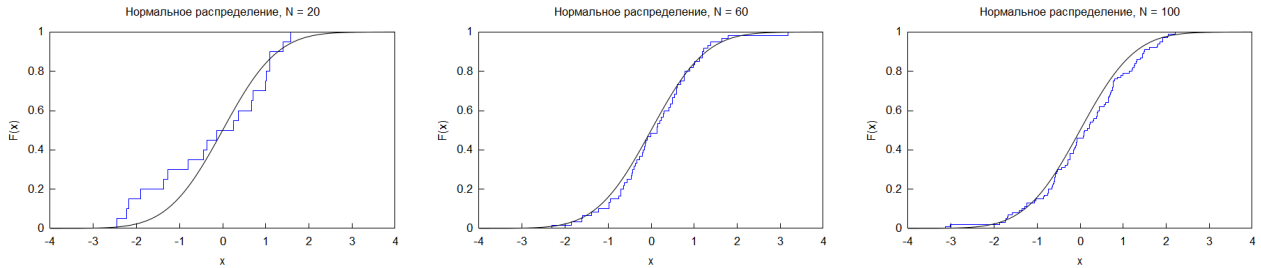


Рис. 1: Эмпирическая функция нормального распределения  $N(x, 0, 1)$

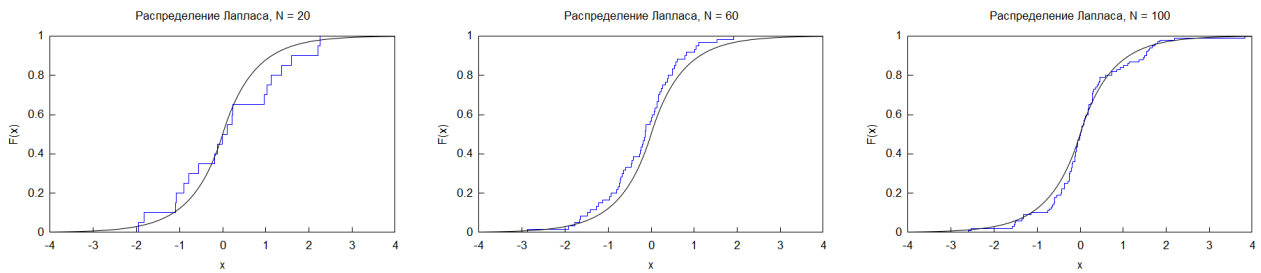


Рис. 2: Эмпирическая функция распределения Лапласа  $L(x, 0, 1/\sqrt{2})$

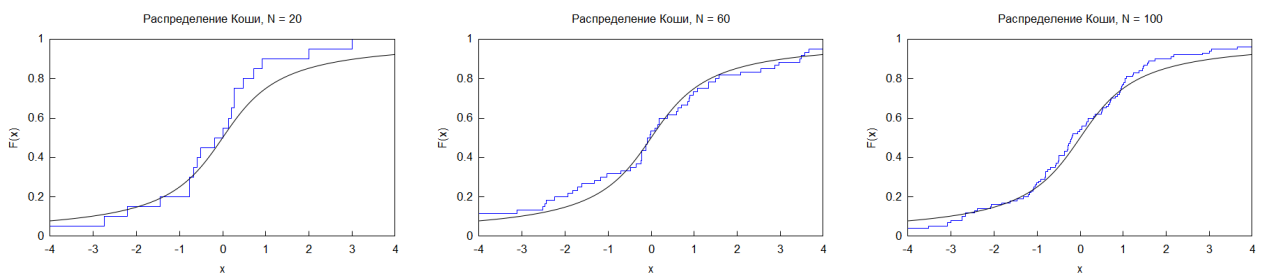


Рис. 3: Эмпирическая функция распределения Коши  $C(x, 0, 1)$

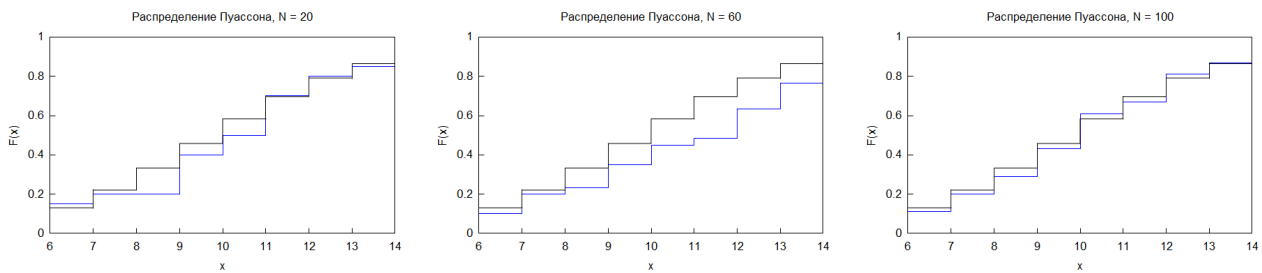


Рис. 4: Эмпирическая функция распределения Пуассона  $P(k, 10)$

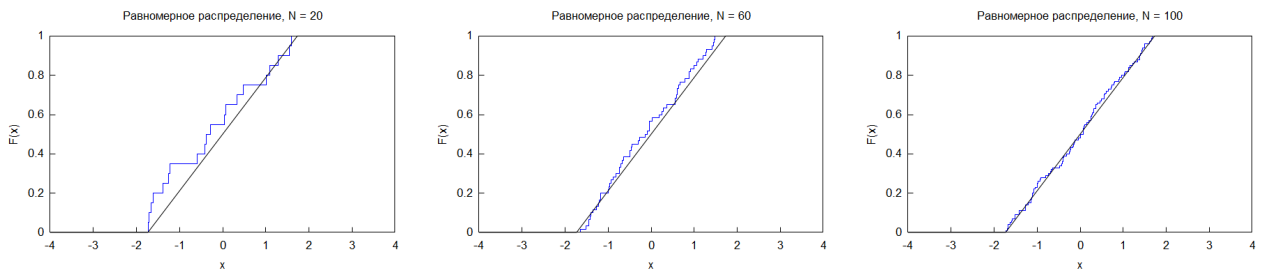


Рис. 5: Эмпирическая функция равномерного распределения  $U(x, -\sqrt{3}, \sqrt{3})$

## 4.2 Ядерные оценки

На рисунках 6-10 представлены графики ядерных оценок плотности для выборок заданных распределений. Ядерная функция имеет вид (3). Черной пунктирной линией изображается теоретическая плотность вероятности, красной линии соответствует оценка с параметром сглаживания  $h_n/2$ , зелёной – с параметром  $h_n$  и синей линии соответствует оценка с параметром сглаживания  $2h_n$ . Параметр  $h_n$  рассчитан согласно формуле (4).

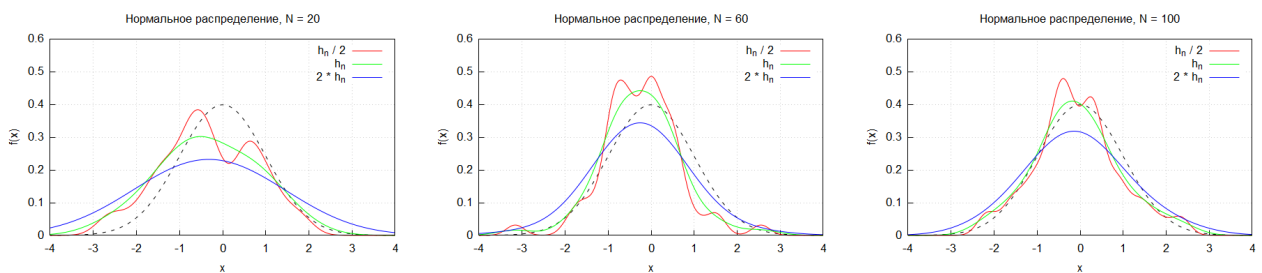


Рис. 6: Ядерная оценка плотности нормального распределения  $N(x, 0, 1)$



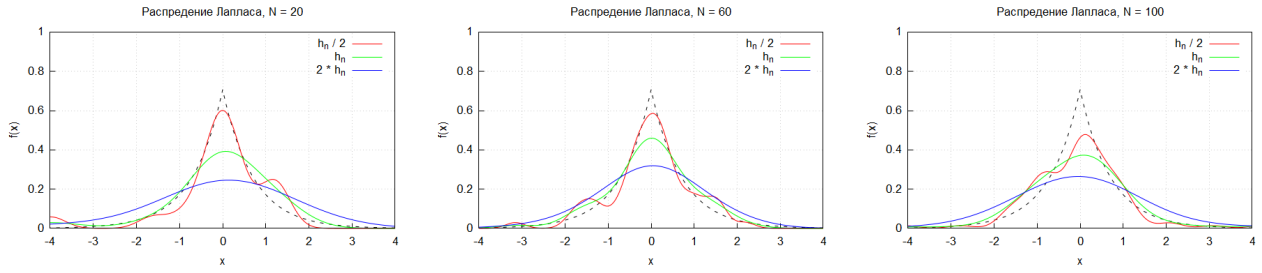


Рис. 7: Ядерная оценка плотности распределения Лапласа  $L(x, 0, 1/\sqrt{2})$

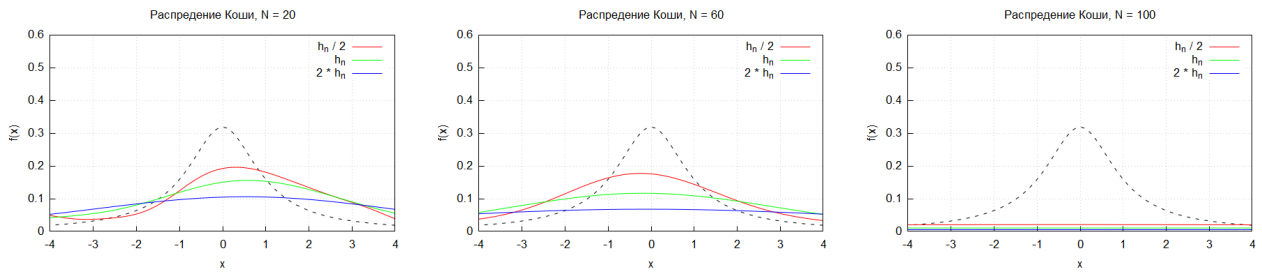


Рис. 8: Ядерная оценка плотности распределения Коши  $C(x, 0, 1)$

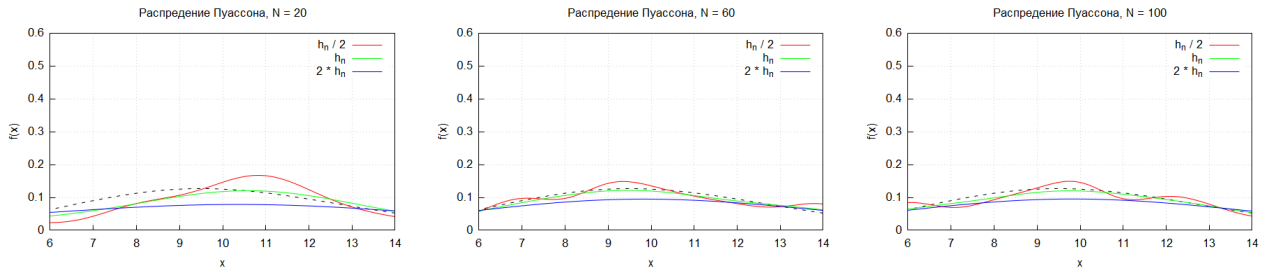


Рис. 9: Ядерная оценка плотности распределения Пуассона  $P(k, 10)$

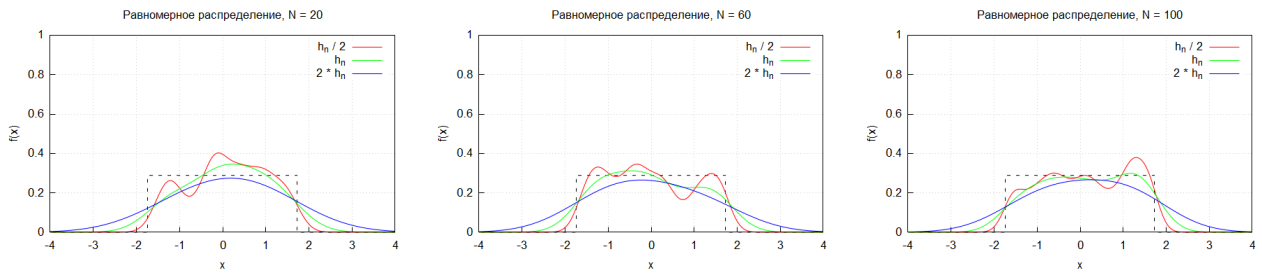


Рис. 10: Ядерная оценка плотности равномерного распределения  $U(x, -\sqrt{3}, \sqrt{3})$

## Заключение

В результате выполнения лабораторной работы были построены эмпирические функции распределения для выборок, соответствующих заданным распределениям. По графикам видно, что чем больше мощность выборки, тем больше эмпирическая функция соответствует теоретической функции распределения.

Также были построены ядерные оценки плотностей распределений. Относительно выбора полосы пропускания  $h_n$  можно сделать вывод, что чем больше полоса пропускания, тем более сглажен график и менее чувствителен к отдельным частотам выборки. Для нормального распределения и распределении Пуассона наиболее подходит полоса пропускания  $h_n$ , полоса пропускания  $h_n/2$  точнее приближает ядерную оценку к теоретической плотности распределения Лапласа и равномерного распределения. Выборки, соответствующие распределению Коши, оценивать ядерными оценками не эффективно ни при какой полосе пропускания.

## Список литературы

- [1] Теоретическое приложение к лабораторным работам №1-4 по дисциплине «Математическая статистика». – СПб.: СПбПУ, 2020. – 12 с
- [2] Амосова Н. Н., Куклин Б. А. и др. – Вероятностные разделы математики. Учебник для бакалавров технических направлений. Под общей редакцией Ю. Д. Максимова – СПб.: "Иван Федоров 2001. – 589 с

## **Приложение А. Репозиторий с исходным кодом**

Исходный код скрипта для среды аналитических вычислений *Maxima* находится в репозитории GitHub – URL <https://github.com/malyarenko-md/TeorVer>