Санкт-Петербургский Политехнический
Университет им. Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

Курсовая работа

3 курс, группа 3630102/70301

Студент Лебедев К.С.

Преподаватель Баженов А. Н.

Санкт-Петербург
2020 г.

Содержание

1 Список таблиц

2 Постановка задачи

Для трех выборок 50, 200 и 1000 элементов, сгенерированных согласно закону распределения Фишера с параметрами $\mu = 4$ и $\nu = 2$ и Рэлея с параметром $\sigma = 0.7$ проверить гипотезы о согласии распределения смоделированной выборки с заданным законом распределения по критерию χ^2 для группирования выбирать интервалы равной длины, уровень значимости $\alpha = 0.05$. Проверить гипотезы о согласии распределения смоделированной выборки с заданным законом распределения по непараметрическому критерию Мизеса-Смирнова; уровень значимости $\alpha = 0.05$.

3 Теория

3.1 Законы распределения

Функция распределения Фишера:

$$F = \frac{Y_1/d_1}{Y_2/d_2}, \tag{1}$$

где Y_1,Y_2 - две независимые случайные величины, имеющие распределение χ^2, а d_1 и d_2 - их степени свободы соотвественно.

Функция распределения Рэлея:

$$f(x;\sigma) = \frac{x}{\sigma^2} exp\left(-\frac{x^2}{2\sigma^2}\right), x \geqslant 0, \sigma > 0. \tag{2}$$

Результаты моделирования и последующего анализа с использованием программной системы [7] показали, что законы распределения статистик непараметрических критериев достаточно хорошо описываются одним из двух законов распределения: *логарифмически нормальным* или *гамма-распределением*, поэтому и были выбраны распределения Фишера и Рэлея, которые можно свести к таковым.

3.2 Непараметрический критерий согласия

Классический непараметрический критерий согласия Крамера — Мизеса — Смирнова предназначен для проверки простых гипотез о принадлежности анализируемой выборки полностью известному закону, то есть для проверки гипотез вида: $\mathbf{H_0} : \mathbf{F_n}(\mathbf{x}) = \mathbf{F}(\mathbf{x}, \theta)$ с известным вектором параметров теоретического закона.

В критерии ω^2 Крамера — Мизеса — Смирнова используется статистика вида:

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(F(x_i, \theta) - \frac{2i-1}{2n}\right)^2, \tag{3}$$

где n – объем выборки, $x_1, x_2, ..., x_n$ – упорядоченные по возрастанию элементы выборки.

При справедливости простой проверяемой гипотезы статистика критерия подчиняется распределению вида $a1(S)$[4] .

При проверке простых гипотез критерий является свободным от распределения, то есть не зависит от вида закона, с которым проверяется согласие.

Проверяемая гипотеза отклоняется при больших значениях статистики. Процентные точки распределения $a1(S)$ приведены в таблицах математической статистики[4].

4 Пример

Проверяют простую гипотезу о принадлежности выборки экспоненциальному закону. Упорядоченная выборка объемом 100 наблюдений имеет вид: [0,0041 0,0051 0,0058 0,0074 0,0082 0,0110 0,0160 0,0191 0,0263 0,0279 0,0294 0,0323 0,0411 0,0452 0,0688 0,0741 0,0805 0,0809 0,1026 0,1124 0,1220 0,1226 0,1233 0,1317 0,1323 0,1368 0,1379 0,1475 0,1515 0,1598 0,1710 0,1789 0,2010 0,2014 0,2072 0,2102 0,2194 0,2205 0,2297 0,2300 0,2302 0,2373 0,2375 0,2397 0,2415 0,2492 0,2869 0,2908 0,2976 0,3058 0,3060 0,3073 0,3096 0,3278 0,3553 0,3620 0,3679 0,3833 0,3921 0,3985 0,4078 0,4080 0,4119 0,4169 0,4208 0,4568 0,4707 0,4880 0,4942 0,5214 0,5277 0,5878 0,6146 0,6180 0,6263 0,6415 0,6757 0,7156 0,7157 0,7207 0,7351 0,7485 0,7535 0,7541 0,7728 0,8875 0,9021 0,9581 0,9868 1,0440 1,2226 1,2402 1,2641 1,3034 1,3328 1,3553 1,4006 1,5586 1,6296 2,5018]

Проверяемая гипотеза имеет вид H_0: $f(x) = \frac{1}{\theta_0} e^{\frac{-x}{\theta_0}}$ при значении параметра $\theta_0 = 0,5$.

Теперь вычисляем значение статистики ω^2 по формуле (3). $S_\omega^* = 0,1272$.

При этом значении статистики вычисляют вероятность $P(S_\omega > S_\omega^*) = 1 - a1(S_\omega^*) = 0,4673$.

Как видно, при задании уровня значимости $\alpha < 0,4673$ нет оснований для отклонения проверяемой гипотезы по всем критериям согласия.

5 Реализация

Работы была выполнена на языке $Python3.7$. Для генерации выборок использовался модуль [1]. Для построения графиков использовалась библиотека matplotlib [2]. Функции распределения обрабатывались при помощи библиотеки scipy.stats [3]

6 Результаты

6.1 функция распределения Фишера

Рис. 1: Функция распределения Фишера с $n = 50$

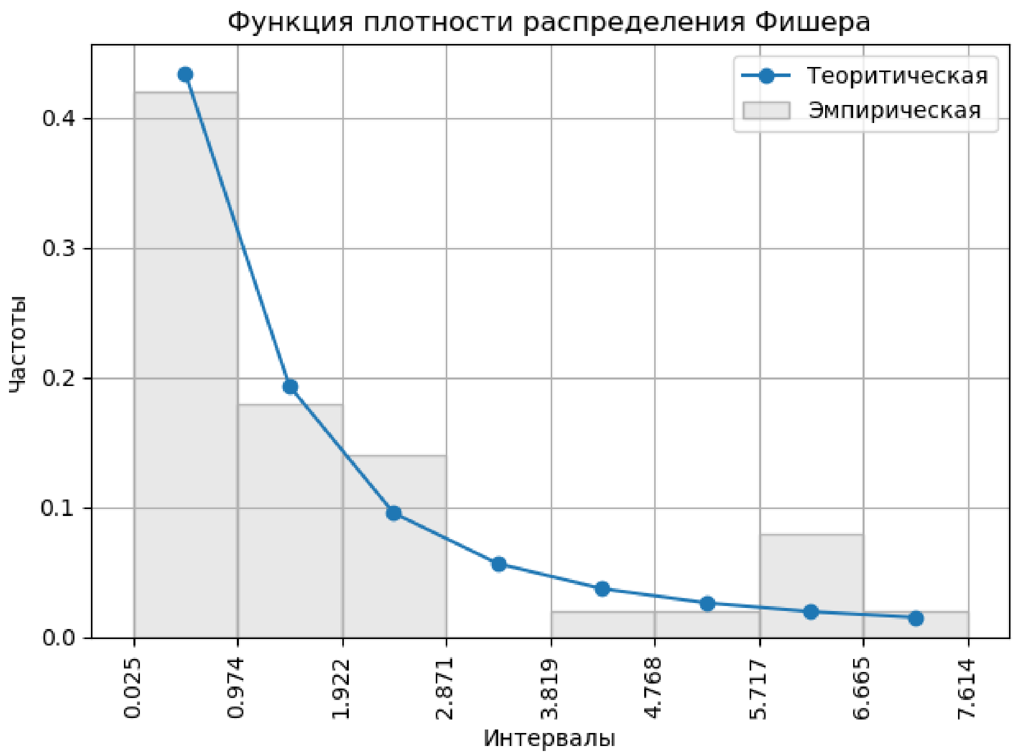


Рис. 2: Функция нормального распределения с $n = 50$

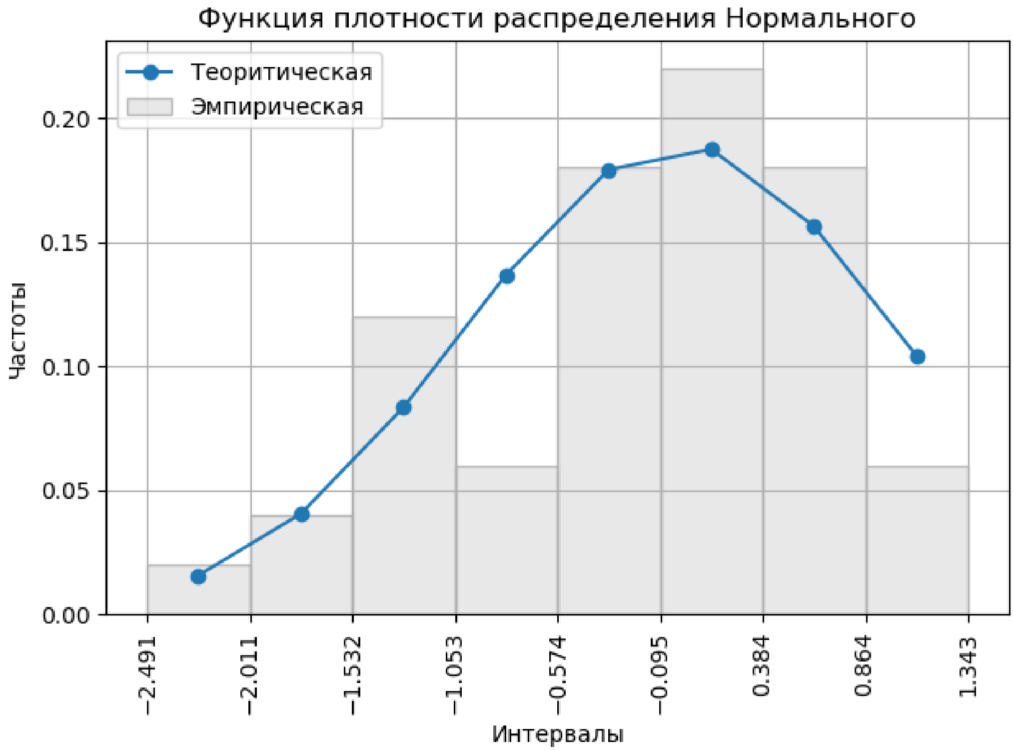


Рис. 3: Функция распределения Фишера с $n = 200$

Рис. 4: Функция нормального распределения с $n = 200$

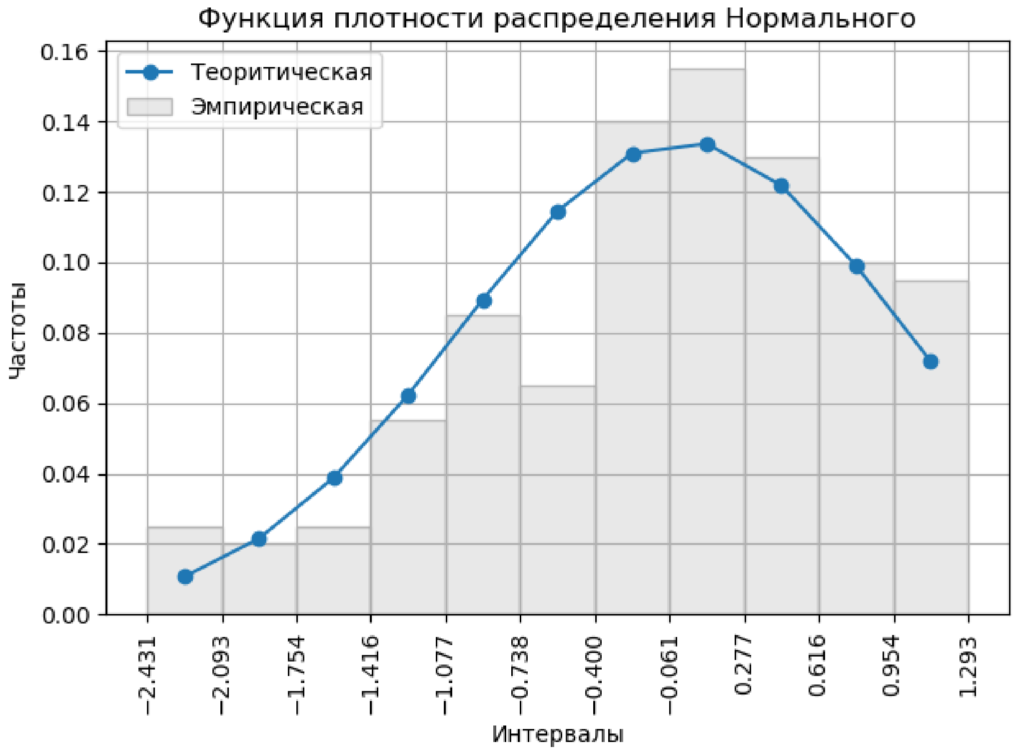


Рис. 5: Функция распределения Фишера с $n = 1000$

Рис. 6: Функция нормального распределения с $n = 1000$

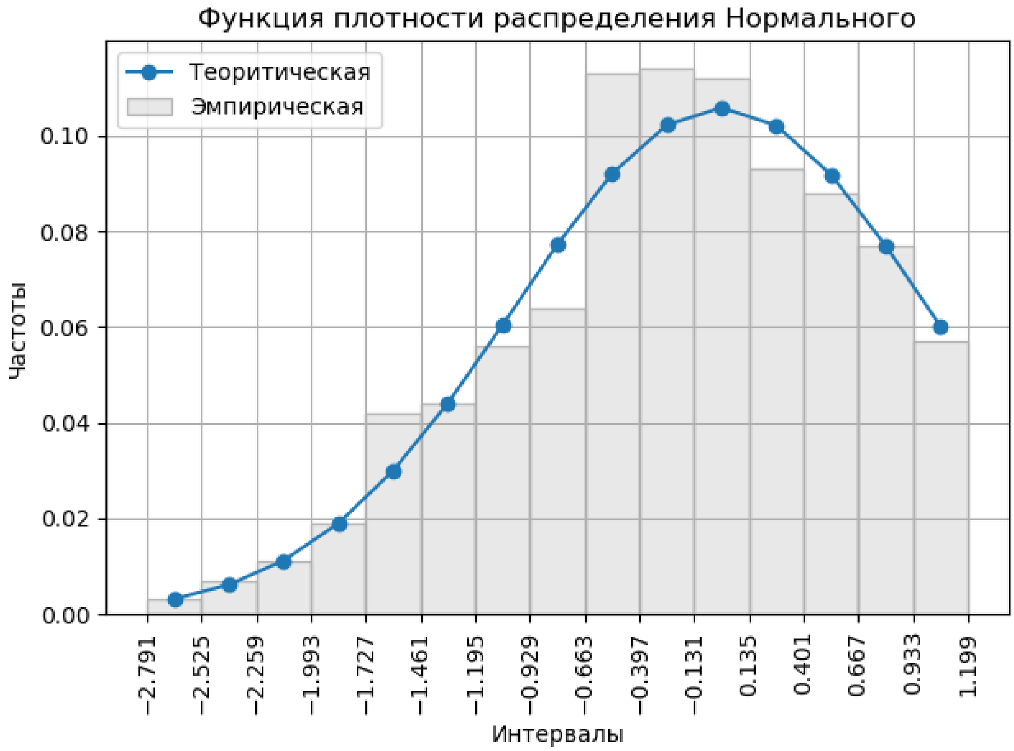


Рис. 7: График функции распределения Фишера

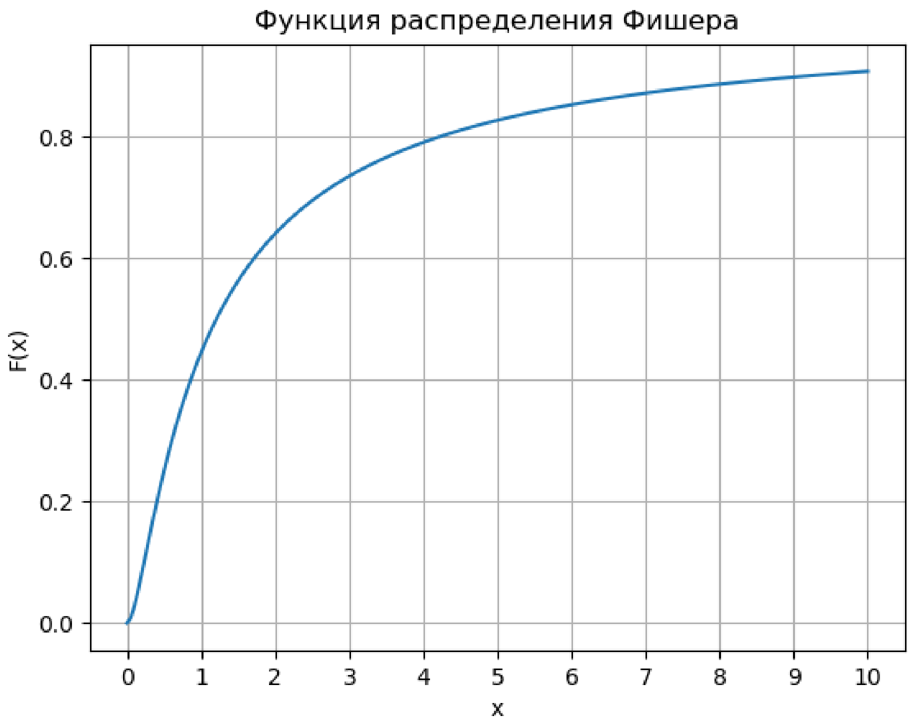
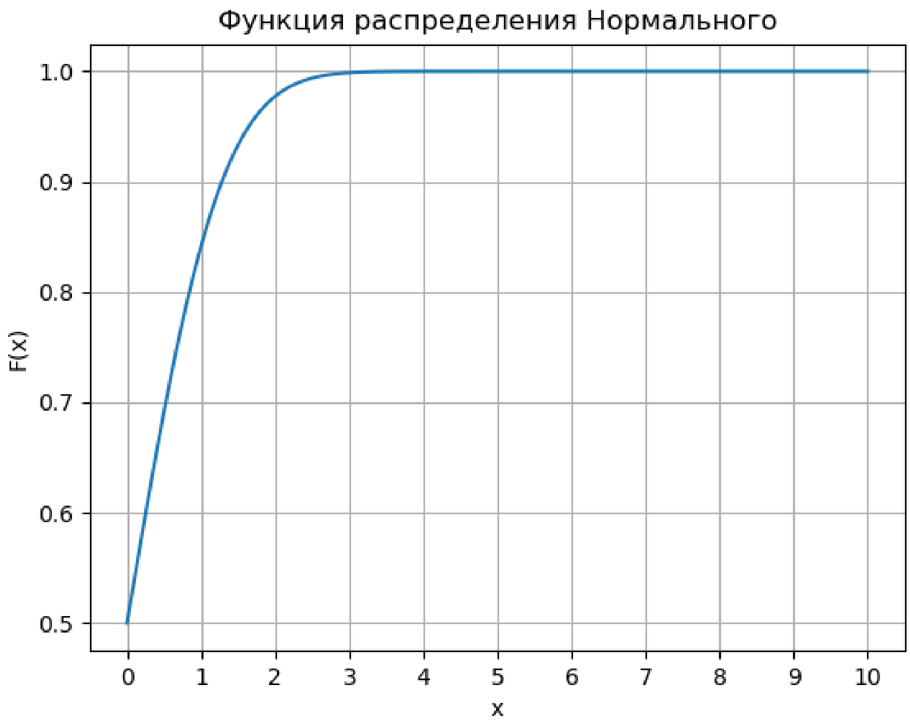


Рис. 8: График функции нормального распределения



6.2 функция распределения Рэлея

Рис. 9: Функция распределения Рэлея с $n = 50$

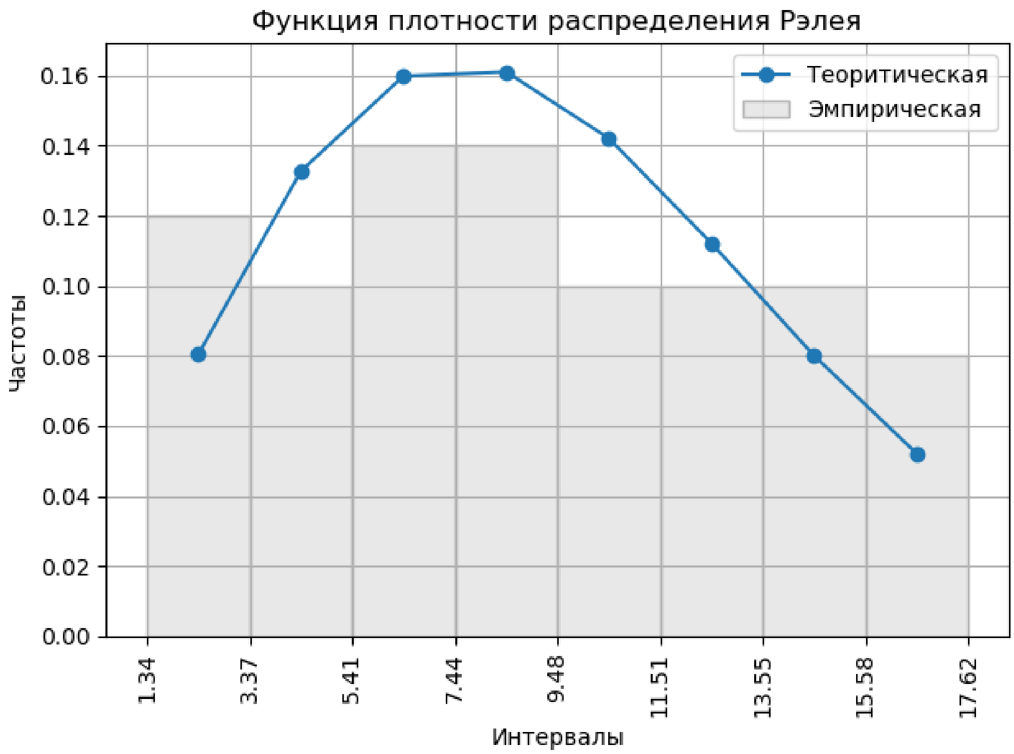


Рис. 10: Функция нормального распределения с $n = 50$

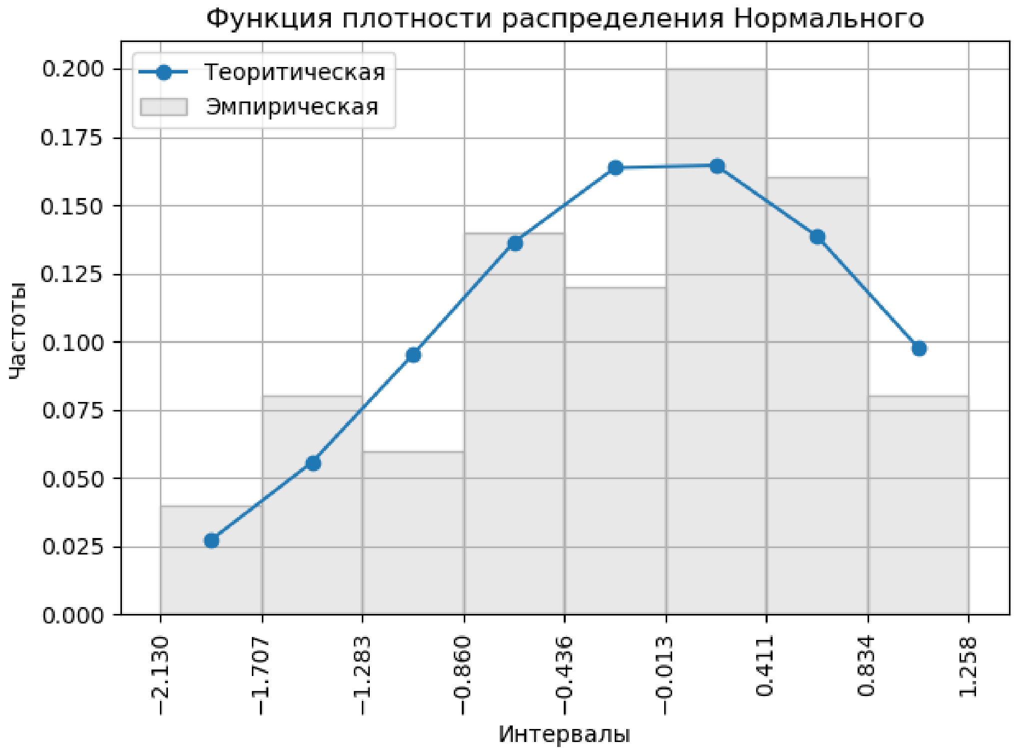


Рис. 11: Функция распределения Рэлея с $n = 200$

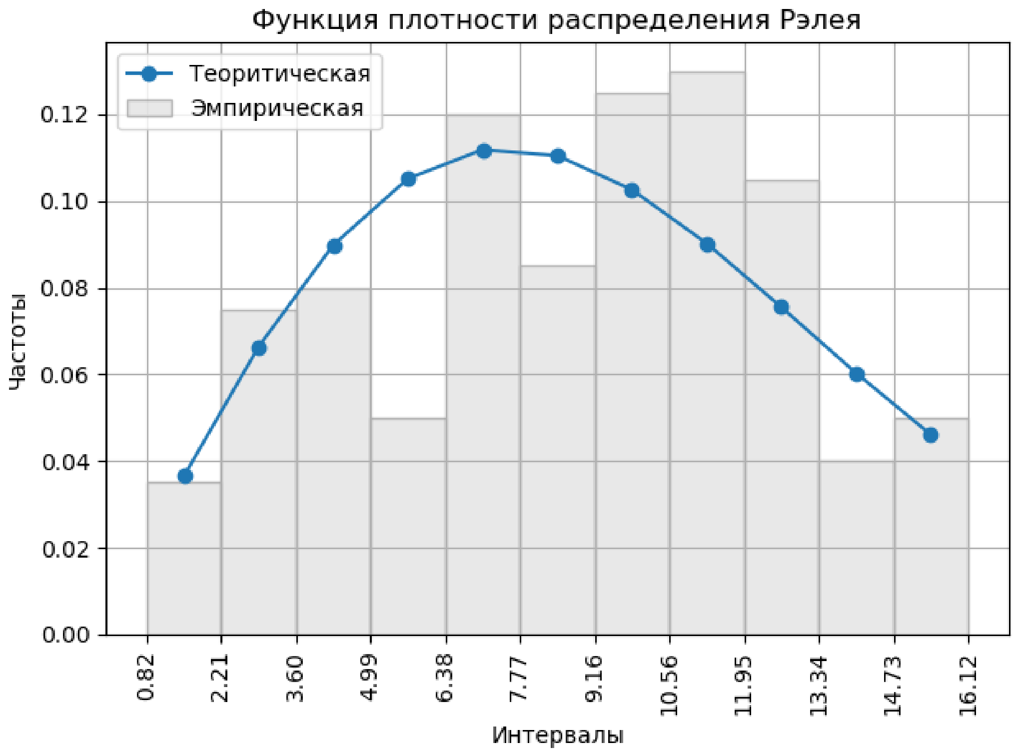


Рис. 12: Функция нормального распределения с $n = 200$

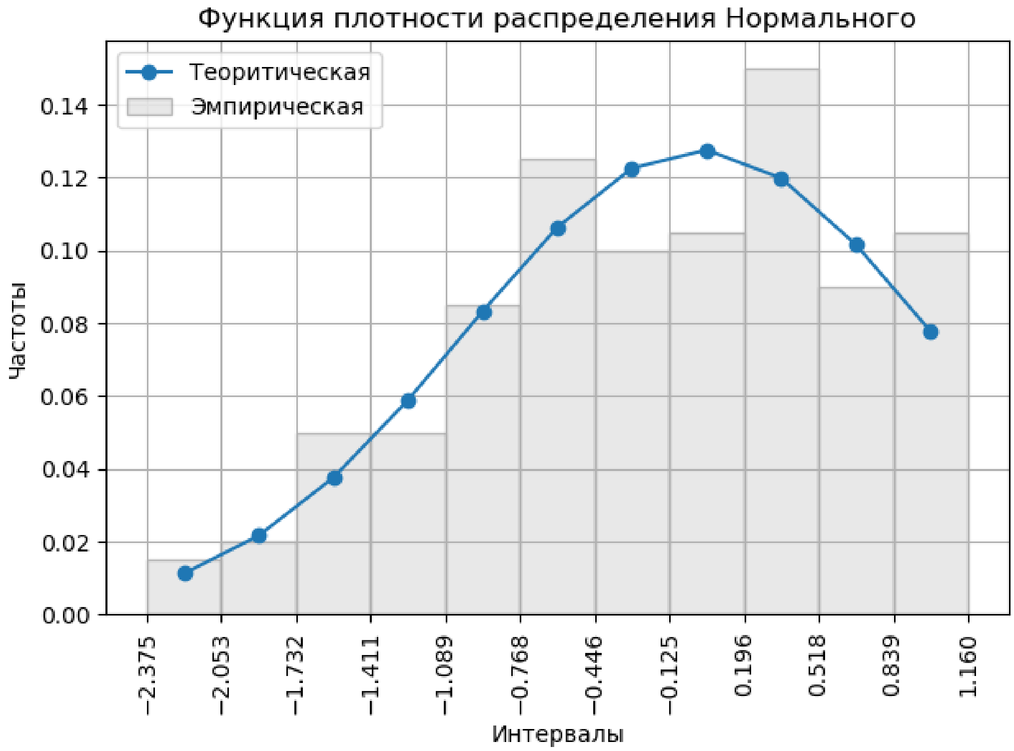


Рис. 13: Функция распределения Рэлея с $n = 1000$

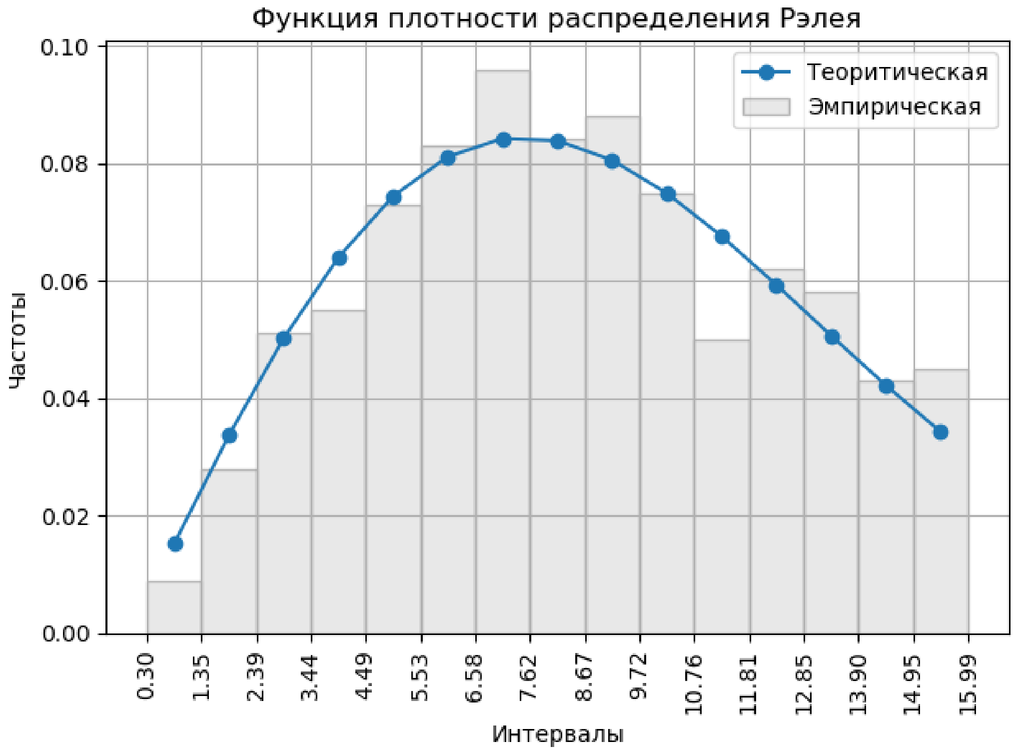


Рис. 14: Функция нормального распределения с $n = 1000$

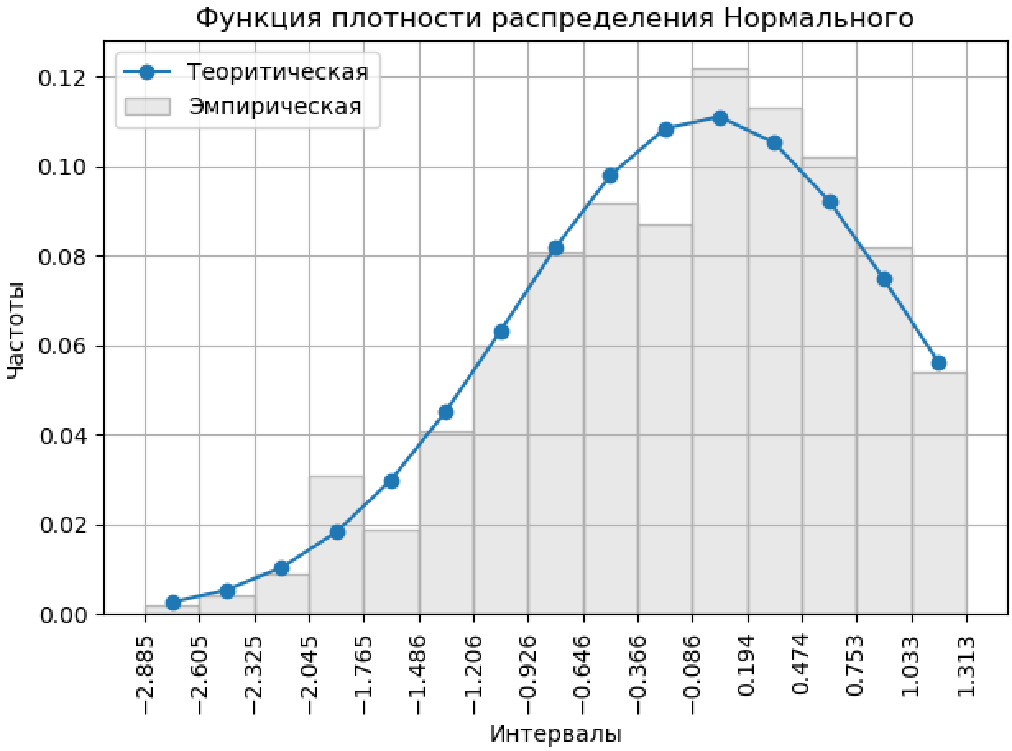


Рис. 15: График функции распределения Рэлея

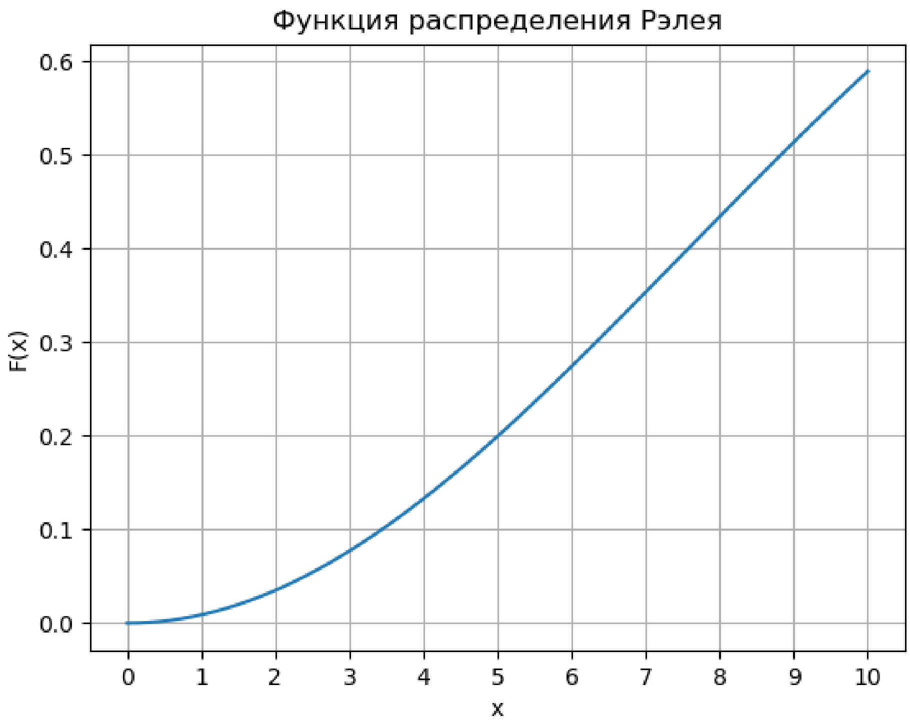


Рис. 16: График функции нормального распределения

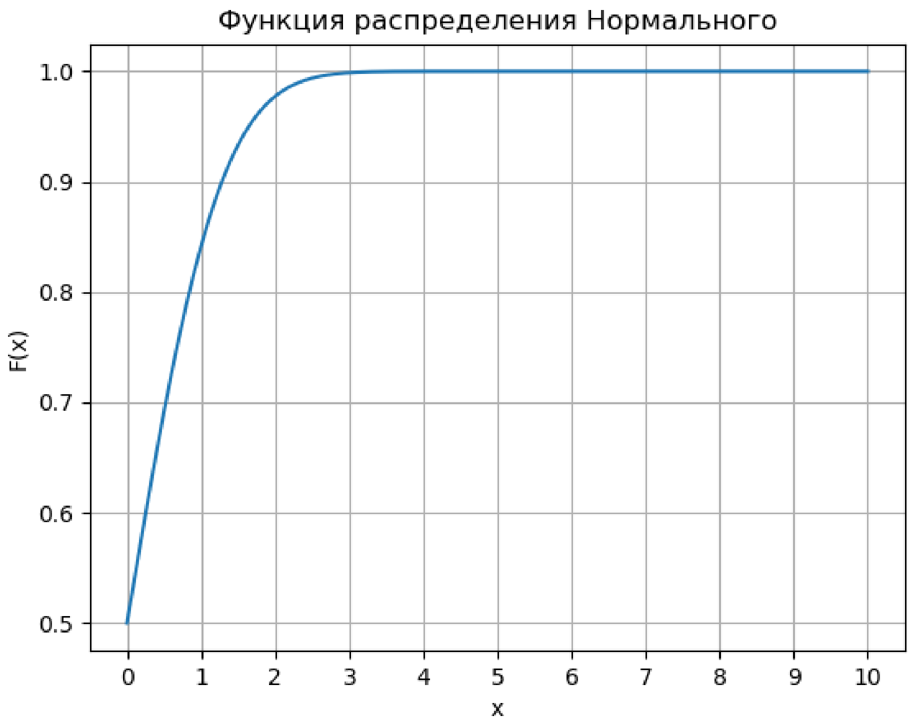


Таблица 1: Результаты

Распределение	n	Тест по критерию ω^2	Тест по критерию ω^2 для нормального распределения
Фишера	50	$True$	$True$
	200	$True$	$True$
	1000	$True$	$True$
Рэлея	50	$True$	$True$
	200	$False$	$True$
	1000	$True$	$True$

7 Выводы

По полученным результатам видно, что оба подхода дают лучший результат на выборках большого объема. Если рассматривать результаты для выборки объема n = 200 элементов, то видно, что при распределении Фишера тест на критерий Крамера — Мизеса — Смирнова пройден в отличии от Рэлея.

8 Список литературы

[1] Модуль numpy - https://physics.susu.ru/vorontsov/language/numpy.html

[2] Модуль matplotlib - https://matplotlib.org/users/index.html

[3] Модуль scipy - https://docs.scipy.org/doc/scipy/reference/

[4] Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983.

[5] http://www.machinelearning.ru/

[6] https://ru.wikipedia.org/

[7] Лемешко Б.Ю. Статистический анализ одномерных наблюдений случайных величин: Программная система. - Новосибирск: Изд-во НГТУ. - 1995. - 125 с.

9 Приложения

Код отчёта: `https://github.com/9Shikamaru/CourseProjMatStat`