Санкт-Петербургский Политехнический Университет Петра Великого

Физико-механический институт

Кафедра прикладной математики и вычислительной физики

Отчёт по лабораторной работе №1 по дисциплине

«Многомерный статистический анализ»

Выполнил студент гр. 5030102/90401: Саськов Л.К.

Преподаватель: к.ф.-м.н., доцент, Павлова Л.В.

Санкт-Петербург

2023

1. Постановка задачи:

Построить и обосновать модель распределения исследуемой случайной величины.

1. По данному в файле «Number\_23.txt» набору чисел были найдены выборочные несмещенные статистики:

Среднее: 1.786

Дисперсия: 4.593

Коэффициент асимметрии: 3.289

Коэффициент эксцесса: 14.126

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

1. Построены э.ф.р. и нормированная гистограмма:

Прежде всего выборка сортируется в целях повышения производительности далее описанных алгоритмов.

Гистограмма строится следующим образом:

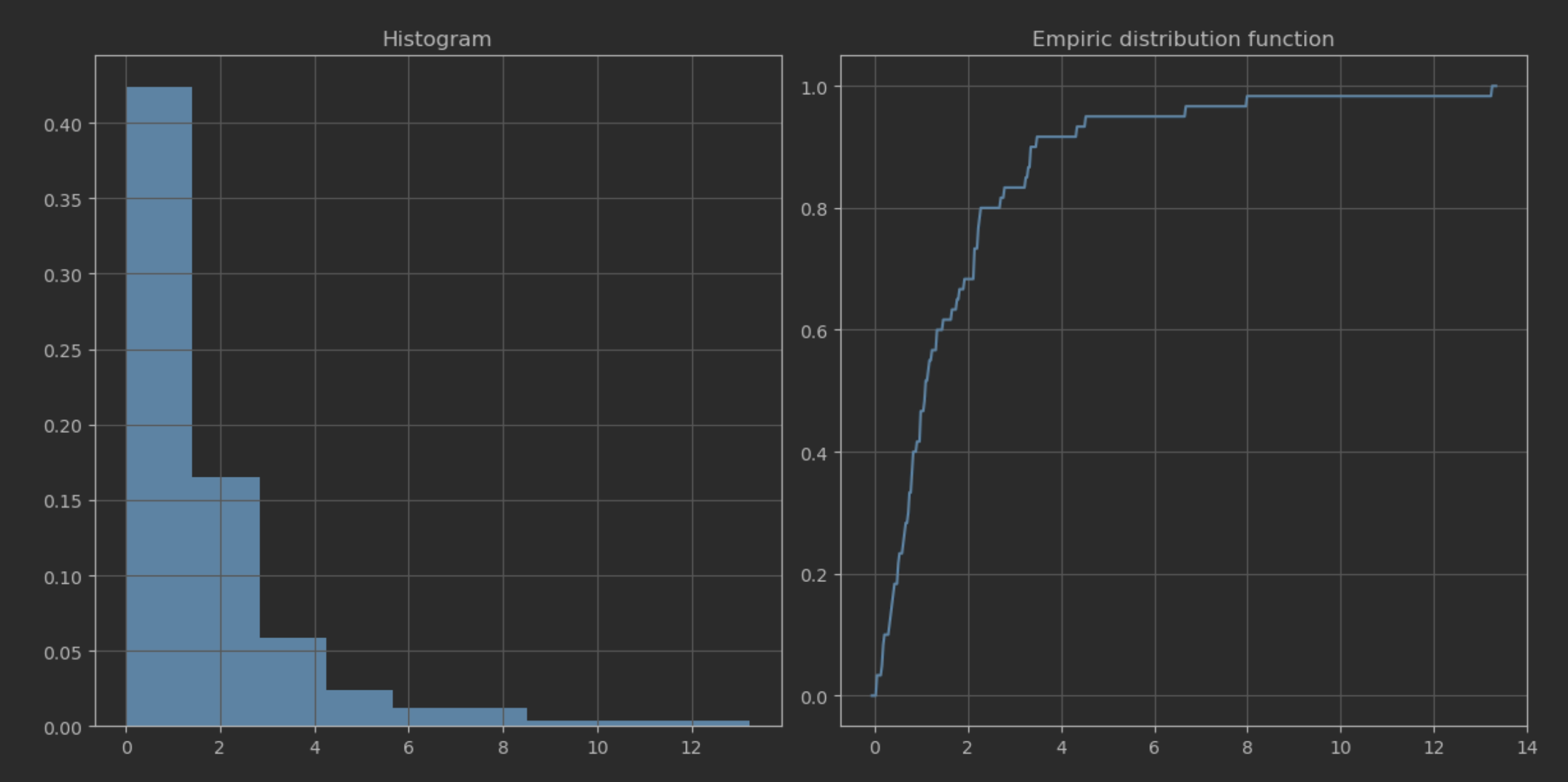
* Выбираются полуинтервалы. В данной работе используются интервалы равной длины, за исключением последнего – левая граница последнего полуинтервала выбирается таким образом, чтобы избежать “пробелов” столбцов гистограммы.
* Левая граница первого, а точнее нулевого, интервала на малое меньше минимального элемента выборки, а правая граница последнего интервала равна максимальному элементу выборки.
* Строятся столбцы гистограммы. Высота каждого столбца согласуется с условием того, что площадь столбца пропорциональна относительной частоте подвыборки, попавшей в полуинтервал, а сумма площадей всех интервалов равна 1:

где – правая и левая границы полуинтервала, – высота столбца, – количество элементов выборки, попавших в интервал , – размер выборки.

Эмпирическая функция распределения в каждой точке рассчитывается как относительная частота элементов, расположенных на числовой прямой левее аргумента:

Изображение выглядит как текст

Автоматически созданное описание



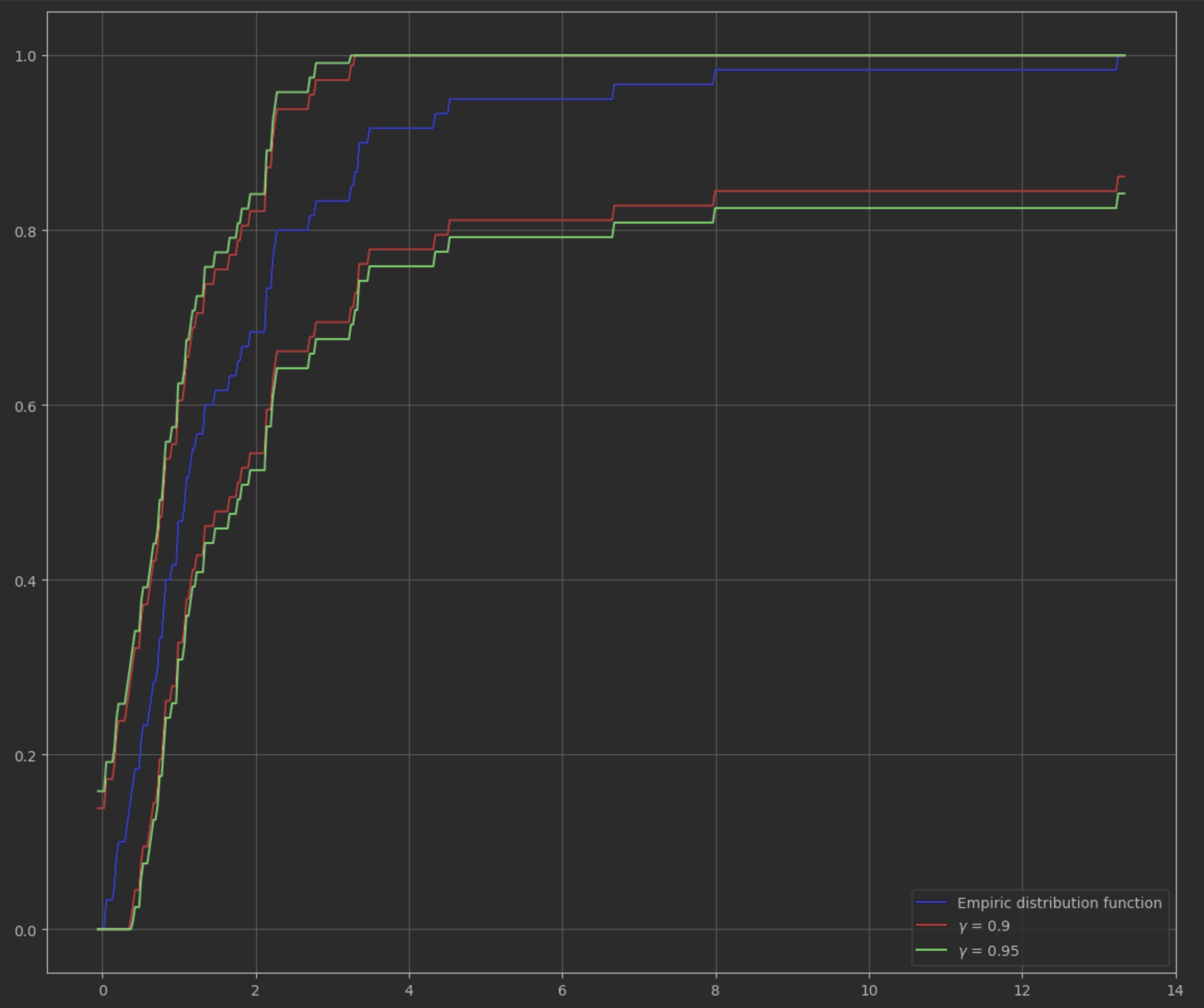
1. По эмпирической функции распределения были построены доверительные полосы для теоретической функции распределения (т.ф.р.) с доверительными вероятностями и .

Построение доверительных полос обосновывается теоремой Дворецкого-Кифера-Вольфовица. Границы полос выражаются следующим образом:

где – доверительная вероятность, – размер выборки.

Изображение выглядит как текст

Автоматически созданное описание



1. После анализа гистограммы, эмпирической функции распределения и выборочных статистик была выдвинута гипотеза о принадлежности распределения случайной величины семейству гамма-распределения:

Параметры:

1. Гипотеза о виде распределения проверяется на основе критерия хи-квадрат Фишера

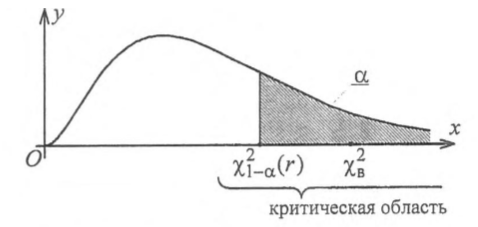
Разбивая допустимый для экспоненциального и гамма распределения (от нуля до бесконечности) интервал на 5 интервалов в ходе проверки гипотезы хи-квадрат методом Фишера, получаем величины,

– количество элементов выборки, попавших в -ый интервал; – теоретическая вероятность попасть в -ый интервал,

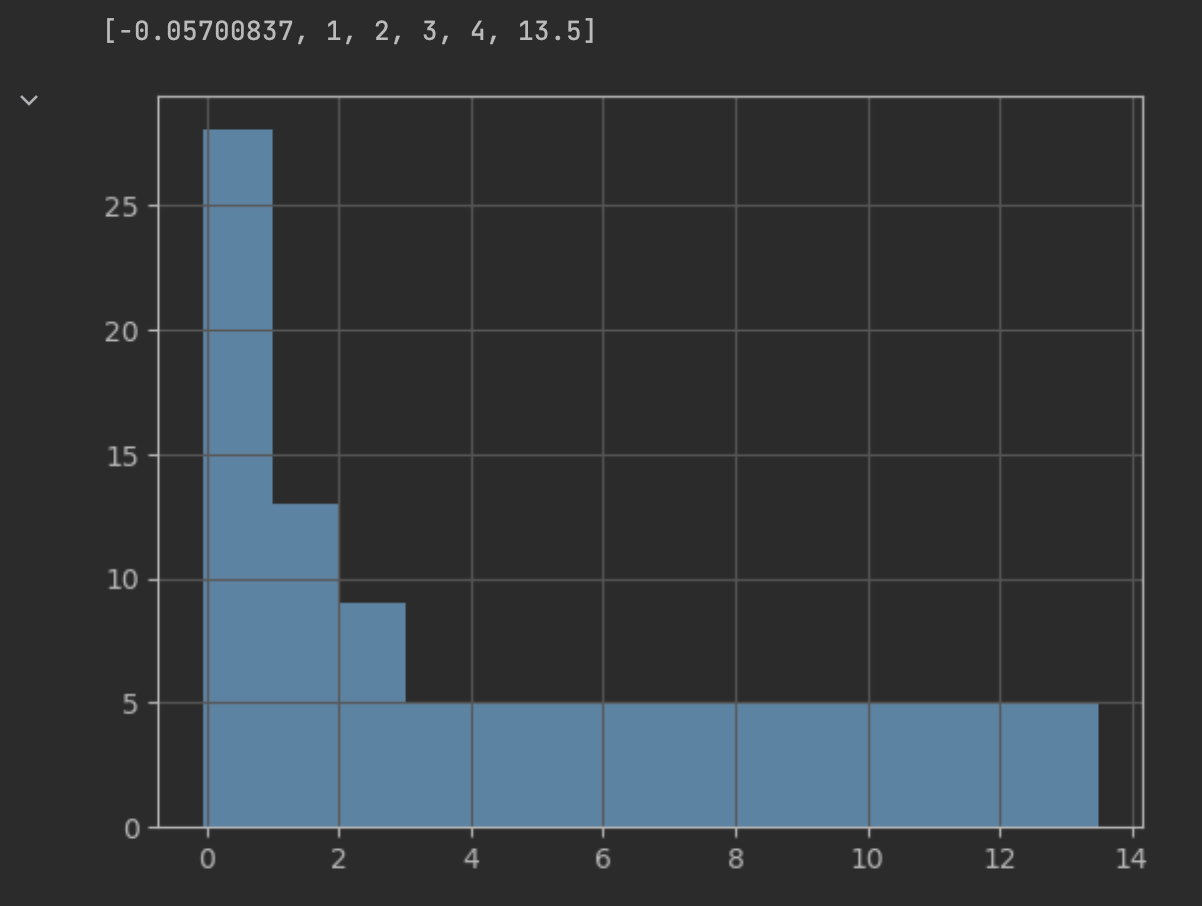
Которые в предельном случае распределены по закону хи-квадрат с степенями свободы, где – число интервалов, r – число параметров:

Выражение минимизируется, вычисляются оптимальные значения вектора параметров

Далее строится критическая область:



Если вычисленное значение меньше , то гипотеза принимается, в ином случае попадает в критическую область – нулевая гипотеза отвергается.

Следует отметить, что при разделении точек на 5 интервалов, в каждом интервале располагается минимум 5 точек (границы интервалов: [-0.05700837, 1, 2, 3, 4, 13.5]): 

Функция вычисляющая хи-квадрат:

Изображение выглядит как текст

Автоматически созданное описание

Проверка гипотезы:

Изображение выглядит как текст

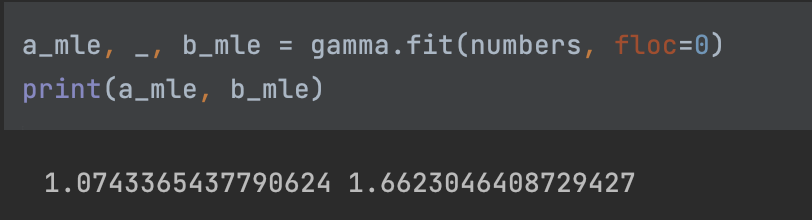
Автоматически созданное описание

Гипотеза принимается, так как вычисленное значение статистики 0,47 меньше табличного значения – 0,95 квантиля 6,0.

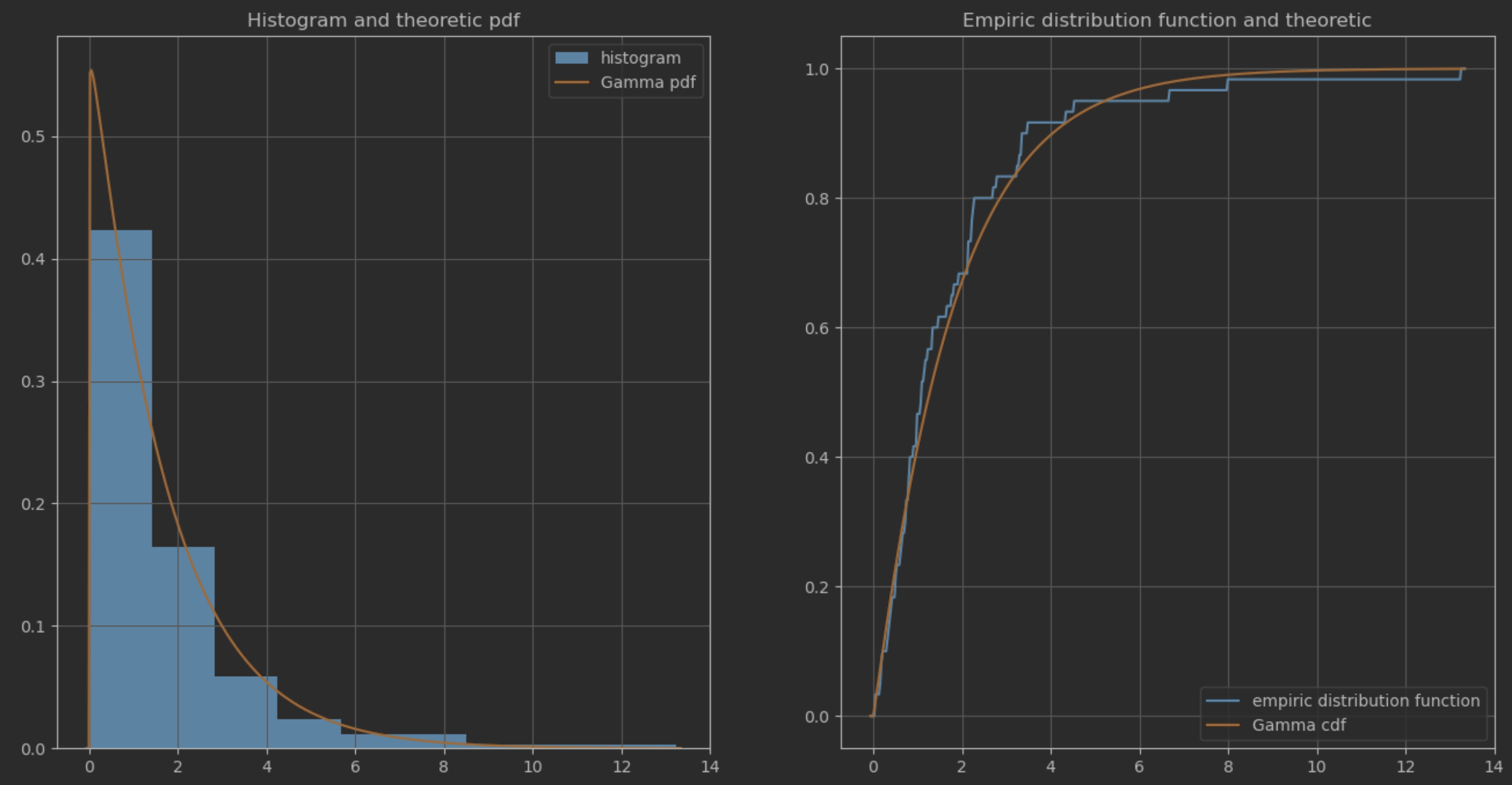
Критическое значение 6,0 – значение хи-квадрат для 2х степеней свободы (5 интервалов – 2 параметра – 1 = 2)

Гипотеза о виде распределения принята, теперь вычислим оценки максимального правдоподобия и построим теоретические кривые с этими оценками:

С помощью метода максимального правдоподобия были найдены оценки максимального правдоподобия:



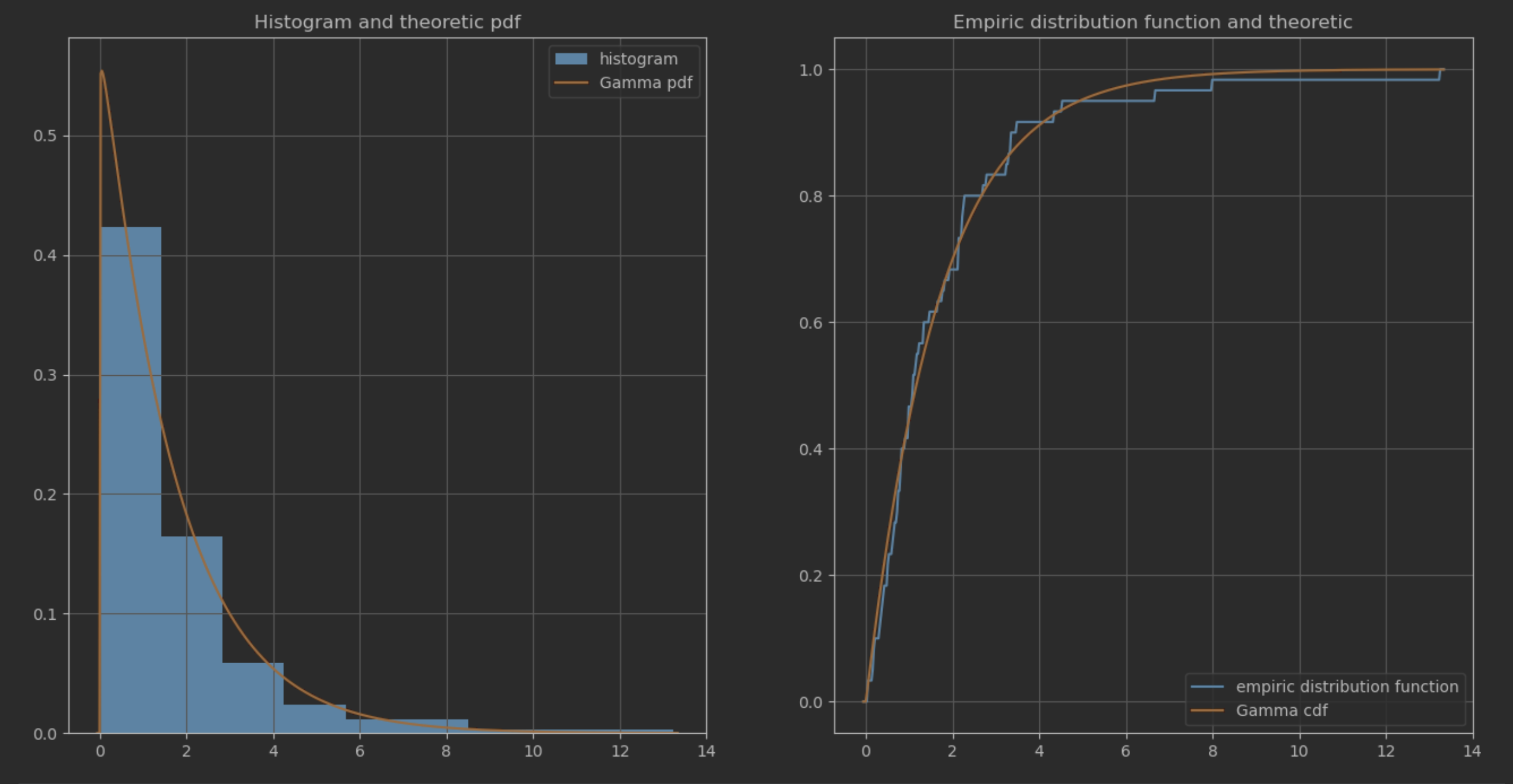
Сравним найденную теоретическую функцию и выборочную:



1. В результате минимизации выражения

Возьмем оценку параметра для интервалов образованных 5 точками и построим теоретические функции плотности и распределения:

Гамма распределение:



1. В результате исследования были получены выборочные статистики, построена эмпирическая функция распределения и гистограмма. Построены доверительные полосы э.ф.р. Была предложена гипотеза о принадлежности распределения случайной величины семейству экспоненциального распределения. Далее гипотеза была принята на основании критерия хи-квадрат Фишера.

Была получена оценка  параметров гамма распределения:

для гамма распределения

Посчитаем оценки среднего и дисперсии, коэффициента асимметрии, эксцесса для полученного распределения:

Изображение выглядит как текст

Автоматически созданное описание

При выборочных оценках из данной нам выборки:

Видим, что значения для среднего и дисперсии близки.