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. 3630102/80401: Данилов Н.А.

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

Санкт-Петербург

2022

1. Постановка задачи:

Построить и обосновать модель распределения исследуемой случайной величины.

1. По данному в файле «Number\_3.txt» набору чисел были найдены выборочные несмещенные статистики:

Среднее: 3.867

Дисперсия: 6.814

Коэффициент асимметрии: 1.41

Коэффициент эксцесса: 2.504

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:

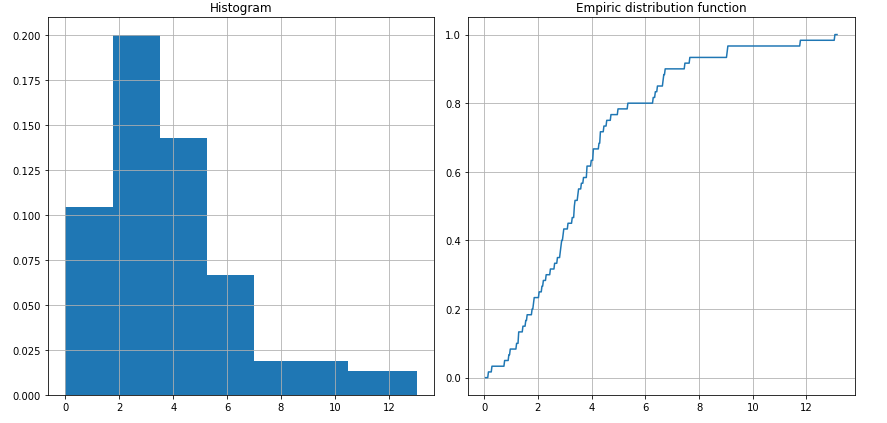
* Выбираются полуинтервалы. В данной работе используются интервалы равной длины, за исключением последнего – левая граница последнего полуинтервала выбирается таким образом, чтобы избежать “пробелов” столбцов гистограммы.
* Левая граница первого, а точнее нулевого, интервала на малое меньше минимального 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



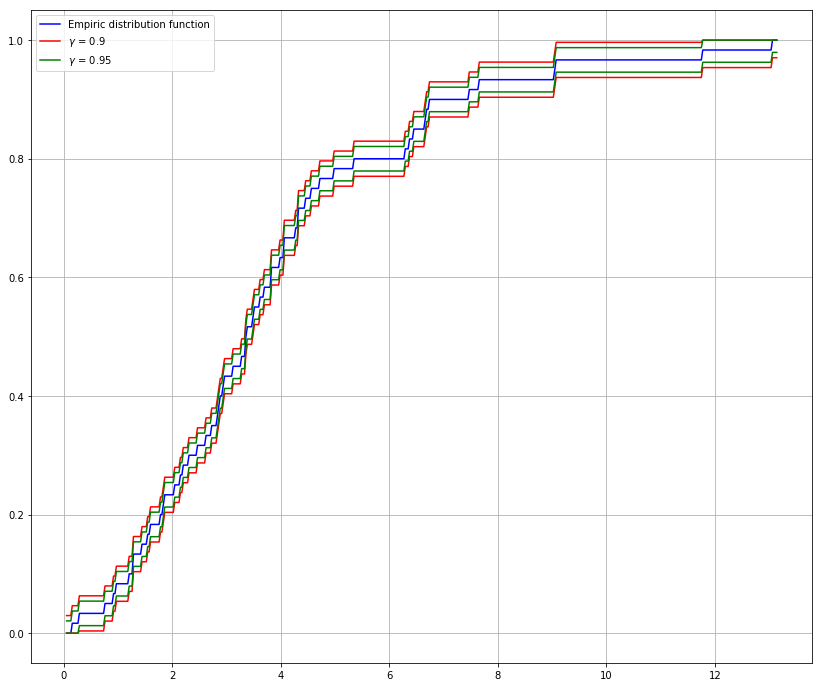
1. По эмпирической функции распределения были построены доверительные полосы для теоретической функции распределения (т.ф.р.) с доверительными вероятностями и .

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 логнормального распределения, имеющее следующую форму функции плотности вероятности:

Параметры:

Также была выдвинута гипотеза о принадлежности распределения случайной величины семейству гамма-распределения:

Параметры:

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

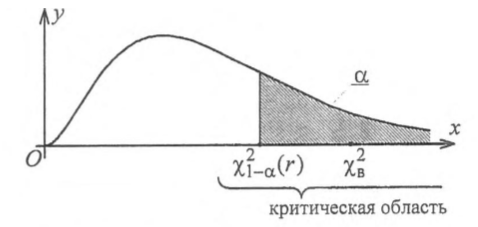
Разбивая допустимый для логнормального и гамма- распределений (в обоих случаях полуинтервал от нуля до бесконечности) интервал на 7, 10, 15 полуинтервалов в ходе проверки гипотезы хи-квадрат методом Фишера, получаем величины,

– 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.

Изображение выглядит как текст

Автоматически созданное описание

Figure 1 Функция вычисляющая хи-квадрат от перменного \Theta

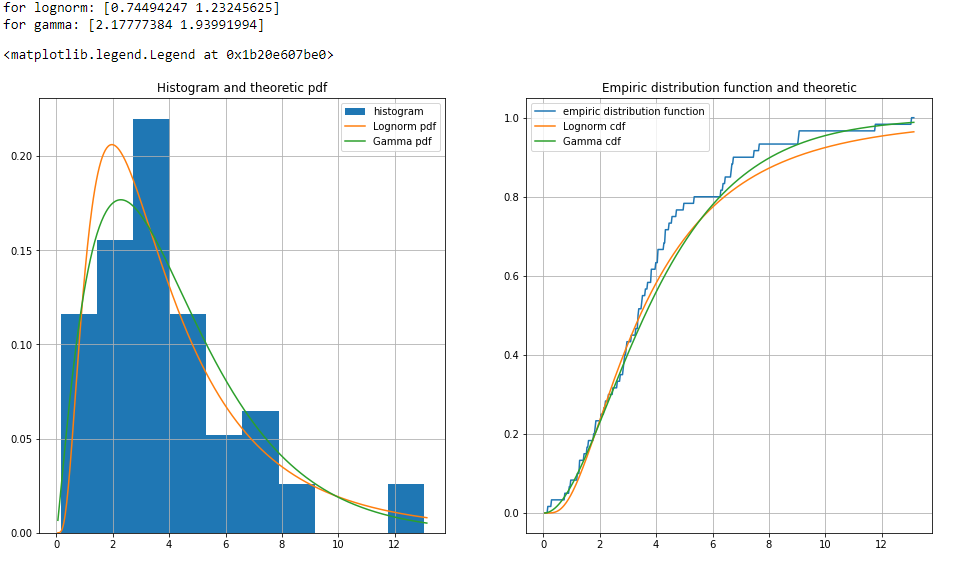
Изображение выглядит как текст

Автоматически созданное описание

Для обеих гипотез в случаях трех различных конфигураций интервалов гипотеза принимается, так как вычисленные значения меньше табличных значений – 0,95 квантилей.

1. В результате минимизации выражения

в ходе проверки гипотезы о принадлежности распределения с.в. получили три набора вероятных параметров. Возьмем среднее и построим теоретические функции плотности и распределения:



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 логнормального распределения. Далее гипотеза была принята на основании критерия хи-квадрат Фишера тремя независимыми исследованиями (7, 10, 15 интервалов).

Были получены оценки параметров :

для логнормального распределения

для гамма-распределения

Были получены оценки значений матожидания, дисперсии, коэффициента асимметрии и коэффициента эксцесса для логнормального распределения.

Изображение выглядит как текст

Автоматически созданное описание

Среднее: 4.526

Дисперсия: 15.198

Коэффициент асимметрии: 3.222

Коэффициент эксцесса: 22.876

Которые, однако, значительно расходятся со значениями выборочных статистик, упомянутых выше:

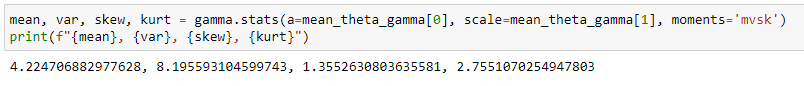
Среднее: 3.867

Дисперсия: 6.814

Коэффициент асимметрии: 1.41

Коэффициент эксцесса: 2.504

Также были получены оценки значений матожидания, дисперсии, коэффициента асимметрии и коэффициента эксцесса для гамма-распределения.



Среднее: 4.224

Дисперсия: 8.195

Коэффициент асимметрии: 1.355

Коэффициент эксцесса: 2.755

Отметим, что все оценки статистик гамма-распределения гораздо “ближе” к выборочным оценкам, нежели оценки логнормального распределения – особенно заметно это в оценках дисперсии и коэффициента эксцесса.