1. Министерство образования и науки Российской Федерации
2. Санкт-Петербургский Политехнический Университет Петра Великого
3. —
4. Институт кибербезопасности и защиты информации

**КУРСОВАЯ РАБОТА**

1. «**Распределение Ломакса**»
2. по дисциплине «Теория вероятностей и математическая статистика»
3. Выполнили
4. студенты гр. 4851003/10001 Алькин М.Р.

Маркова А.М.

<*подпись*>

1. Преподаватель
2. асс. преподавателя Штыркина А.А.

<*подпись*>

Санкт-Петербург

1. 2023

Содержание

[Содержание 2](#_Toc138063353)

[Цель работы 3](#_Toc138063354)

[Постановка задач 3](#_Toc138063355)

[Ход работы 7](#_Toc138063356)

[1 Исследования распределения ломакса 7](#_Toc138063357)

[2 Понятие эмпирической функции распределения 8](#_Toc138063358)

[3 Понятие эмпирической функции распределения 14](#_Toc138063359)

[4 Понятие гистограммы 18](#_Toc138063360)

[5 Описание параметров распределения 21](#_Toc138063361)

[6 Понятие точечных оценок 25](#_Toc138063362)

[6.1 Метод моментов: 25](#_Toc138063363)

[6.2 Метод максимального правдоподобия: 26](#_Toc138063364)

[6.3 Свойства полученных оценок: 29](#_Toc138063365)

[7 Понятие интервальных оценок 31](#_Toc138063366)

[7.1 Для параметра смещения: 31](#_Toc138063367)

[7.2 Для параметра формы: 32](#_Toc138063368)

[8 Понятие статистических критериев 35](#_Toc138063369)

[8.1 Гипотезы о параметрах распределениях 35](#_Toc138063370)

[8.2 Гипотезы о виде распределения. 42](#_Toc138063371)

[8.2.1 Критерий Колмогорова 42](#_Toc138063372)

[8.2.2 Критерий Пирсона 43](#_Toc138063373)

[8.3 Гипотезы об однородности выборок. 45](#_Toc138063374)

[8.3.1 Критерий Смирнова. 47](#_Toc138063375)

[8.3.2 Мощность критерия Смирнова: 48](#_Toc138063376)

[Выводы 49](#_Toc138063377)

[Список используемых источников 50](#_Toc138063378)

Цель работы

Изучение распределения Ломакса с параметром location: построение эмпирической функции распределения, гистограммы, точечных и интервальных оценок.

Постановка задач

1. Провести исследование распределения из варианта: история его появления, его назначение, применение в научных исследованиях. Для каждого факта необходимо привести ссылку, оформленную по ГОСТ, на научную публикацию.
2. Знакомство с jupyter notebook.
   1. Установить интерпретатор языка Python, согласно инструкции, указанной на официальном сайте (<https://www.python.org/>) для операционной системы.  
      *В качестве дополнительного задания дальнейшие действия можно проводить в отдельно созданном виртуальном окружении (через virtualenv).*
   2. Установить среду интерактивной разработки jupyter notebook (<https://jupyter.org/>).
   3. Открыть среду интерактивной разработки, в открывшемся окне создать новый Notebook.
   4. Выполнить считывание выборки из файла варианта с помощью стандартных средств языка Python.
   5. Написать функции для вычисления статистик выборки из варианта (без использования готовых функций):
      1. сумма элементов выборки;
      2. выборочное среднее;
      3. медиана;
      4. мода;
      5. размах выборки;
      6. смещенная дисперсия;
      7. несмещенная дисперсия;
      8. выборочный начальный момент k-ого порядка (k подается как аргумент функции вместе с выборкой);
      9. выборочный центральный момент k-го порядка (k подается как аргумент функции вместе с выборкой);
   6. Выполнить функции, занести результаты в отчет. В качестве значений k для выборочных начального и центрального моментов можно взять любой, кроме первого и второго.
3. Понятие эмпирической функции распределения.
   1. Занести в отчет определение эмпирической функции распределения и алгоритм ее построения.
   2. Написать функцию для построения эмпирической функции распределения на языке Python (без использования готовых реализаций).
   3. Сделать случайные подвыборки размера 10, 100, 200 элементов из выборки.
   4. Для каждой из подвыборок построить эмпирическую функцию распределения и соответствующий этой функции график. Построенный график должен содержать наименование, подпись осей, легенду.
   5. Построенные графики занести в отчет.
   6. Сделать вывод о виде эмпирической функции распределения:
      1. насколько она похожа на теоретическую функцию распределения из варианта?
      2. какие можно сделать предположения о параметрах распределения?
4. Понятие гистограммы:
   1. Занести в отчет определение гистограммы и алгоритм ее построения.
   2. Написать функцию для построения гистограммы на языке Python (без использования готовых реализаций).
   3. Сделать случайные подвыборки размера 10, 100, 200 элементов из выборки.
   4. Для каждой из подвыборок построить гистограмму и соответствующий график. Построенный график должен содержать наименование, подпись осей, легенду.
   5. Построенные графики занести в отчет.
   6. Сделать вывод о виде гистограммы:
      1. насколько она соответствует распределению из варианта?
      2. какие можно сделать предположения о параметрах распределения?
5. Описание параметров распределения:
   1. Определить вид теоретической функции распределения из варианта, включающей в себя все заданные в варианте параметры.
   2. Занести формулу для вычисления теоретической функции распределения в отчет.
   3. Занести в отчет описание каждого из параметров распределения.
   4. Написать функцию построения теоретической функции распределения для распределения из варианта. Аргументами написанной функции должны быть параметры распределения.
   5. Выполнить построение графика теоретической функции распределения со случайно заданными параметрами распределения.
   6. Оценить, как изменяется график теоретической функции распределения при изменении параметров распределения. Чтобы это оценить для каждого неизвестного параметра распределения выполнить следующие шаги:
      1. фиксируются все неизвестные параметры распределения кроме одного;
      2. для незафиксированного параметра выбрать два сильно отличающихся друг от друга значения;
      3. для выбранных значений построить теоретические функции распределения и визуально оценить, чем они отличаются.
   7. Занести в отчет построенные графики и выводы о влиянии каждого параметра распределения на вид теоретической функции распределения.
6. Понятие точечных оценок.
   1. Оценить параметры распределения выборки методом моментов.
   2. Оценить параметры распределения выборки методом максимального правдоподобия.
   3. Привести и обосновать свойства полученных оценок: несмещенность, состоятельность, эффективность, R-эффективность.
   4. На одном графике построить:
      1. теоретическую функцию распределения с оценками параметров распределения, найденными методом моментов;
      2. теоретическую функцию распределения с оценками параметров распределения, найденными методом максимального правдоподобия;
      3. эмпирическую функцию распределения (по всей выборке).

Построенный график должен содержать наименование, подпись осей, легенду.

* 1. Занести соответствующие вычисления и график в отчет.
  2. Сделать вывод о значении оценок параметров распределения и схожести эмпирической и теоретических функций распределения.

1. Понятие интервальных оценок.
   1. Оценить параметры распределения выборки с помощью интервальной оценки с уровнями доверия
   2. Занести соответствующие вычисления в отчет.
   3. Сделать вывод о близости значений интервальных и точечных оценок.
2. Понятие статистических критериев.
   1. Гипотезы о параметрах распределениях.
      1. Построить наиболее мощный критерий для проверки нулевой гипотезы вида для каждого параметра распределения . Здесь оценка параметра распределения , полученная в результате применения метода моментов (), метода максимального правдоподобия ().

В качестве уровня значимости принять .

* + 1. Вычислить ошибки 1-го и 2-го рода.
    2. Занести соответствующие вычисления в отчет.
    3. Сделать вывод о результате работы критерия.
  1. Гипотезы о виде распределения.
     1. Используя критерий Колмогорова и критерий , проверить гипотезу о принадлежности выборки распределению из варианта.

Примечание: в качестве уровня значимости принять .

* + 1. Занести соответствующие вычисления в отчет.
    2. Сделать вывод о результате работы критериев.
  1. Гипотезы об однородности выборок.
     1. Провести поиск распределений, близких по виду к распределению из варианта. Утвердить вид найденного распределения у преподавателя.
     2. Методом обратного преобразования сгенерировать выборку из распределения .
     3. Построить критерий для проверки гипотезы вида: против альтернативы . Вид критерия утвердить у преподавателя.

Примечание: в качестве уровня значимости принять .

* + 1. Оценить мощность критерия.
    2. Занести соответствующие вычисления в отчет.
    3. Сделать вывод о результате работы выбранного критерия.

Ход работы

# Исследования распределения ломакса

Распределение Ломакса, условно называемое также распределением Парето типа II (Со смещением равным нулю), — это распределение с тяжелым хвостом, часто используется в бизнесе, экономике и бухгалтерском моделировании.

Изначально оно было получено и использовано Ломаксом К.С. в 1954 году для подгонки данных в случае неудач в бизнесе [1]. По сути, как уже говорилось ранее, это распределение Парето типа II [2, C.3], сдвинутое так, что оно начинается с нуля, но очень часто они используются как синонимы, в нашем задании, например дано распределение Ломакса, однако используются параметр location.

Считается, что распределение Ломакса не является достаточно гибким для моделирования и соответствия реальным данным [3]. Поэтому в последнее время его обобщают и изменяют путем добавления дополнительных (форма, масштаб, расположение или порог) параметров. Примером может быть:

* Распределение Рэлея-Ломакса [2]
* Распределение Skew Lomax [3]
* Five-parameter transmuted generalization of the Lomax distribution [4]

Самое простое распределение Ломакса обычно используется исследователями, как уже говорилось, для моделирования данных с тяжелым хвостом. Например, оно использовалось в распределении доходов и изучения личного благосостояния в Британии Харрисом и Аткинсоном [5], для изучения распределения компьютерных файлов на сервере Холланом [6], для анализа надежности и определения оптимального времени изменения уровня напряжения Хассаном [7].

# Понятие эмпирической функции распределения

**2.5.1** Сумма элементов выборки:

Сумма элементов выборки представляет собой сумму всех значений, входящих в выборку. Для вычисления суммы элементов выборки нужно просто сложить все значения. (Реализация представлена на Рисунке 1)

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

Рисунок 1. Реализация Суммы элементов выборки

**2.5.2** Выборочное среднее:

Выборочное среднее — это приближение теоретического среднего распределения, основанное на выборке из него. Оно вычисляется путем деления суммы элементов выборки на количество элементов в выборке. (Реализация представлена на Рисунке 2)

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

Рисунок 2. Реализация Выборочного среднего

**2.5.3** Медиана:

Медиана — число, которое находится в середине вариационного ряда. Для вычисления медианы нужно упорядочить выборку по возрастанию и выбрать серединный элемент (если количество элементов нечетное) или среднее арифметическое двух серединных элементов (если количество элементов четное). (Реализация представлена на Рисунке 3)

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 3. Реализация Медианы

**2.5.4** Мода:

Мода — это значение, которое наиболее часто встречается в выборке. В выборке может быть одна мода (унимодальная выборка), несколько мод (мультимодальная выборка) или отсутствие моды (амодальная выборка). Для вычисления моды нужно посчитать количество вхождений каждого значения и выбрать значение с наибольшим количеством вхождений. (Реализация представлена на Рисунке 4)

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 4. Реализация Моды

**2.5.5** Размах выборки:

Размах выборки — это разница между наибольшим и наименьшим значением в выборке. Для вычисления размаха нужно вычесть из наибольшего значения наименьшее значение. (Реализация представлена на Рисунке 5)

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Рисунок 5. Реализация Размаха выборки

**2.5.6** Смещенная дисперсия:

Смещенная дисперсия — это мера разброса значений в выборке относительно их среднего значения. Для вычисления смещенной дисперсии нужно вычислить среднее арифметическое квадратов отклонений каждого элемента от выборочного среднего. Затем сумма квадратов отклонений делится на количество элементов в выборке. (Реализация представлена на Рисунке 6)

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Рисунок 6. Реализация Смещенной дисперсии

**2.5.7** Несмещенная дисперсия:

Несмещенная дисперсия — это такая же мера разброса значений, как и смещенная дисперсия, но деление происходит не на количество элементов в выборке, а на количество элементов минус единица. Это поправка, которая используется для получения несмещенной оценки дисперсии. (Реализация представлена на Рисунке 7)

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Рисунок 7. Реализация Несмещенной дисперсии

**2.5.8** Выборочный начальный момент k-ого порядка:

Выборочный начальный момент k-ого порядка — это среднее арифметическое каждого элемента выборки, которые возведены в k-ую степень. Для вычисления выборочного начального момента нужно возвести каждый элемент в выборке в k-ую степень, затем найти среднее арифметическое полученных значений. (Реализация представлена на Рисунке 8)

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

Рисунок 8. Реализация Выборочного начального момента k-ого порядка

**2.5.9** Выборочный центральный момент k-го порядка:

Выборочный центральный момент k-го порядка — это среднее арифметическое степеней отклонений каждого элемента выборки от выборочного среднего, возведенных в k-ую степень. Для вычисления выборочного центрального момента нужно вычесть выборочное среднее из каждого элемента выборки, возвести полученные значения в k-ую степень, затем найти среднее арифметическое этих степеней. (Реализация представлена на Рисунке 9)

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Рисунок 9. Реализация Выборочного центрального момента k-го порядка

**Результаты выполнения функций на заданной выборке:**

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Рисунок 10. Демонстрация работы функций

# Понятие эмпирической функции распределения

Слово эмпирическая означает, что функция вычисляется по данным опыта (эмпирическим данным), то есть, по выборке.

Если говорить более формально, Эмпирической функцией распределения, построенной по выборке   объема n, называется случайная функция  , при каждом   равная

где

**Алгоритм построения:**

Можно построить эмпирическую функцию распределения по вариационному ряду:

* Составляем вариационный ряд.
* Подсчитываем количество элементов меньших X с заданным шагом (использовался шаг = 0.1)
* Строим эмпирическую функцию по полученным данным.

Получившиеся эмпирические функции распределения представлены на рисунках:

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

Рисунок 11. Эмпирическая функция для подвыборки из 10 элементов

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

Рисунок 12. Эмпирическая функция для подвыборки из 100 элементов

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

Рисунок 13. Эмпирическая функция для подвыборки из 200 элементов

Теперь рассмотрим отдельно функцию распределения Ломакса:

Изображение выглядит как линия, График, диаграмма, текст

Автоматически созданное описание

Рисунок 14. Функция распределения Ломакса

Видно, что функция напоминает нашу эмпирическую функцию, но заметно наличие смещения. Попробуем уменьшить шаг эмпирической функции и наложить функцию распределения Ломакса со смещением и посмотреть на сколько они идентичны друг другу:

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 15. Две функции рядом.

Поигравшись с параметрами, для альфы = 8 и лямбды = 1 можем видеть, что эмпирическая функция очень близка к функции распределения Ломакса. Однако однозначные выводы сделать сложно, так как эмпирическая функция также очень похожа на распределение Парето.

# Понятие гистограммы

Гистограмма – это кусочно-постоянная функция, принимающая на интервале значение , где *ni* - число элементов выборки, попавших в ; ; *i* = 1, 2, …, k; -

Алгоритм построения гистограммы по имеющейся выборке, следующий:

1. Сортировка выборки, то есть составление вариационного ряда;
2. Подсчитываем размах полученного вариационного ряда, определяем размер интервалов на оси X делением полученного размаха на заданное количество интервалов (в нашем случае каждый интервал имеет длину 0.1);
3. Построение гистограммы путем размещения по оси Y значений, являющихся количеством элементов ряда, попавших в интервал, заданный на оси X;

В нашем случае из исходной выборки составлялись подвыборки размером 10, 100 и 200 элементов. Для каждой из них были построены соответствующие гистограммы, графики которых представлены на рисунках ниже:Изображение выглядит как текст, снимок экрана, диаграмма, График

Автоматически созданное описание

Рисунок 16. Гистограмма для подвыборки из 10 элементов

Изображение выглядит как текст, снимок экрана, диаграмма, График

Автоматически созданное описаниеРисунок 17. Гистограмма для подвыборки из 100 элементов

Изображение выглядит как текст, снимок экрана, График, диаграмма

Автоматически созданное описание

Рисунок 18. Гистограмма для подвыборки из 200 элементов

Кроме того, функция плотности распределения Ломакса имеет следующий вид:

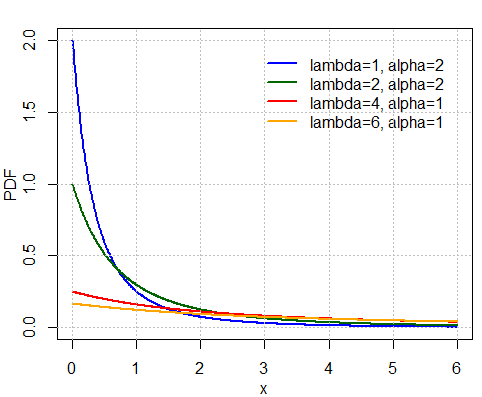


Рисунок 19. Плотность распределения Ломакса

Заметно, что полученные графики сильно напоминают функцию плотности распределения Ломакса. Тем не менее, невозможно однозначно утверждать, что гистограммы соответствуют распределению из нашего варианта, так как они также очень похожи на распределение Парето.

# Описание параметров распределения

Теоретическая функция распределения Ломакса имеет следующий вид:

Видно, что распределение имеет два основных параметра, но также существует еще и третий – смещение.

Параметры распределения:

* – Параметр масштаба. Чем больше значение данного параметра, тем больше разброс распределения (т.е. при помощи данного параметра можно растягивать или сжимать форму распределения);
* – Параметр формы. При помощи данного параметра нельзя растянуть или сжать график, при помощи него, например, можно увеличить или уменьшить угол наклона графика в нашем распределении.
* Смещение – Параметр, позволяющий сдвигать график по оси X

Далее была написана функция, реализующая построение теоретической функции распределения Ломакса. Было получено 6 графиков, иллюстрирующих изменение формы кривой в зависимости от трех описанных параметров:

* При фиксированных значениях параметров масштаба и смещения:

Изображение выглядит как линия, График, диаграмма, текст

Автоматически созданное описание

Рисунок 20. Функция распределения при

Изображение выглядит как линия, График, текст, диаграмма

Автоматически созданное описание

Рисунок 21. Функция распределения при

Здесь использовались следующие значения зафиксированных параметров: , offset = 0;

Видно, что с увеличением значения параметра формы , все острее становился угол кривой.

* При фиксированных значениях параметров формы и смещения:

Изображение выглядит как линия, текст, График, диаграмма

Автоматически созданное описание

Рисунок 22. Функция распределения при

Изображение выглядит как линия, График, диаграмма, текст

Автоматически созданное описание

Рисунок 23. Функция распределения при

Здесь использовались следующие значения зафиксированных параметров: , offset = 0;

Видно, что чем больше значение параметра масштаба , тем более растянутым становится график вдоль оси X.

* При фиксированных значениях параметров формы и масштаба:

Изображение выглядит как линия, График, диаграмма, текст

Автоматически созданное описание

Рисунок 24. Функция распределения при

Изображение выглядит как линия, График, диаграмма, текст

Автоматически созданное описание

Рисунок 25. Функция распределения при

Здесь использовались следующие значения зафиксированных параметров: , ;

Видно, что параметр смещения действительно просто отвечает за сдвиг графика по оси X.

# Понятие точечных оценок

## Метод моментов:

Для оценки двух параметров распределения необходимо два теоретических момента: первого и второго порядка соответственно. В распределении Ломакса со смещением они имеют следующий вид:

Выборочные моменты первого и второго порядка имеют следующий вид:

Приравнивая теоретический и выборочные моменты, получаем:

Для решения системы, в первую очередь необходимо решить второе уравнение, т.к. в нем фигурирует только один параметр, который необходимо оценить:

Если :

Учитывая, что = 0.03929:

Однако, еще в начале вычислений накладывалось условие, что , а значит

Тогда для того, чтобы оценить параметр смещения, выразим его из первого уравнения системы, подставив вместо параметра формы полученное значение:

При условии, что и учитывая, что :

Таким образом, получены оценки параметров формы и смещения распределения Ломакса методом моментов.

## Метод максимального правдоподобия:

В первую очередь, необходимо узнать функцию правдоподобия:

Тогда логарифм функции правдоподобия:

Тогда для того, чтобы оценить неизвестные параметры смещения и формы, необходимо взять две частные производные логарифма функции правдоподобия и приравнять их к нулю, чтобы найти максимальные значения параметров:

Далее, рассмотрим каждое из полученных уравнений по-отдельности:

1. Частная производная по параметру смещения:

Из данного уравнения можем получить систему:

Однако, значение параметра формы распределения Ломакса всегда должно быть положительным, а значит далее необходимо рассматривать только уравнение:

Решить данное уравнение обычными методами не получилось, поэтому использовался другой подход.

Данное уравнение может принять значение, равное 0 в двух случаях: либо знаменатель равен бесконечности, при делении на которую получаем 0, либо в сумме постоянно будет чередоваться знак полученного значения, что в итоге даст значение, близкое к 0 (т.е. результат, имеющий вид , либо, например, )

Однако, второй вариант точно невозможен, т.к. любой xi из выборки всегда должен быть больше, чем смещение, либо равен ему, то есть знаменатель никогда не примет отрицательное значение, а значит уравнение никогда не примет значение, равное 0. Первый вариант также невозможен, т.к. мы знаем, что размах нашей выборки составляет всего ~1.89, а значит вариант с бесконечным значением x сразу же отпадает. Параметр масштаба в нашем случае задан, как конечная константа. Остается только вариант, когда бесконечность достигается на разности между xi и параметром смещения, но опять же, x не может быть бесконечным, а значит и смещение тоже, т.к. любой x не меньше него в любом случае.

Таким образом, данное уравнение либо в принципе не имеет решений, либо для нас невозможно решить данное уравнение без применения более сложных методов, с которыми мы на данном этапе обучения не знакомы. Поэтому далее будет использована оценка, полученная в методе моментов.

Попробуем использовать оценку, равную минимальному элементу выборки, т.е**. 7.940066**

1. Частная производная по параметру формы:

Отсюда имеем:

Для решения данного уравнения использовалось значение смещения, полученное в методе моментов. Был написан скрипт на Python, при помощи которого было получено значение оценки параметра формы

либо, если используется оценка минимальным элементов выборки, то **7.416134581451691**

## Свойства полученных оценок:

* Несмещенность:

Оценка называется несмещенной, если математическое ожидание от оценки равняется истинному значению параметра, т.е. .

Для оценок, полученных методом моментов:

Математическое ожидание оценки параметра смещения совпало со значением параметра смещения, а значит эта оценка несмещенная.

Для оценок, полученных методом максимального правдоподобия:

Известно, что , заменим на

Видно, что оценка параметра формы, полученная методом максимального правдоподобия, является смещенной.

* Состоятельность:

Оценки методом моментов являются состоятельными, если непрерывны моментные функции (т.е. функции, от которых берется математическое ожидание при расчете моментов). Функции непрерывны, а значит и оценки состоятельны.

Оценка параметра формы, полученная методом максимального правдоподобия, смещенная, а значит не является состоятельной.

* Эффективность:

# Понятие интервальных оценок

## Для параметра смещения:

Оценка неизвестного параметра offset, т.е. параметра смещения – минимальный элемент выборки:

Пусть , тогда:

Пусть , тогда:

Для :

Тогда:

В нашем случае рассматривается два случая, когда доверительная вероятность и . Также, нам известно, что минимальный элемент выборки , а параметр масштаба в нашем случае постоянен и равен 1. Тогда:

* При :

Возьмем параметр формы из точечной оценки методом моментов, он равен 8.65217, тогда:

* При :

Возьмем параметр формы из точечной оценки методом моментов, он равен 8.65217, тогда:

Видно, что полученный методом моментов параметр смещения только немного не входит в полученный интервал. (Если подставить вместо параметра формы результат, полученный методом максимального правдоподобия, получится примерно тот же результат).

## Для параметра формы:

Для построение доверительного интервала для параметра формы будет использоваться оценка, полученная методом максимального правдоподобия. Мы знаем, что эта оценка имеет вид:

Известно, что при оценка методом максимального правдоподобия сходится по распределению к , где – информация Фишера для параметра формы, равная

Отцентрируем и отнормируем:

Получили искомую функцию, зависящую от параметра, далее строим интервал:

Имеем:

Тогда:

* При :

Тогда:

* При :

Тогда:

Видно, что оценки параметра формы методом моментов и методом максимального правдоподобия входят в полученные интервалы при обоих доверительных вероятностях.

# Понятие статистических критериев

## Гипотезы о параметрах распределениях

В данном пункте необходимо было построить наиболее мощный критерий для проверки нулевой гипотезы вида , уровень значимости .

Во-первых, по лемме Неймана-Пирсона, среди всех критериев заданного уровня значимости, проверяющих две простые гипотезы, критерий отношения правдоподобия является наиболее мощным. Так что первой мыслью было попробовать использовать именно этот критерий. Однако, этот критерий не совсем подходит под наш случай, так как простой у нас является только основная гипотеза, а альтернативная должна быть сложной, то есть иметь вид: . То есть, необходимо было использовать многопараметрические гипотезы и метод отношения правдоподобия.

Далее, для каждого из параметров:

* Параметр формы:

1. Полученный методом моментов:

Имеем две гипотезы:

Строим функцию правдоподобия:

Для второй гипотезы в качестве значений параметров будут использованы оценки, полученные методом максимального правдоподобия.

То есть:

Для основной гипотезы необходимо найти такой вектор параметров (в нашем случае состоит из двух параметров: формы и смещения, параметр масштаба принят константой, равной 1), чтобы максимизировать функцию L на множестве , то есть при значении условии, что основная гипотеза верна. Для этого необходимо взять частную производную от логарифма функции правдоподобия по второму параметру.

Логарифм функции правдоподобия:

Частная производная:

В данном уравнении можно сразу же сократить множитель , так как параметр формы не может быть меньше нуля. Оставшаяся сумма решений не имеет, поэтому принимаем параметр смещения равным первому элементу полученного из выборки вариационного ряда, то есть

Тогда функция правдоподобия для первой гипотезы:

Отношение правдоподобия:

Однако, в рамках критерия нас интересует не просто отношение правдоподобия, а величина *.* Критерий гласит: если , то принимаем основную гипотезу, иначе отвергаем ее в пользу альтернативной.

Согласно асимптотическому свойству метода отношения правдоподобия, для выборок большого объема, какой и является наша выборка, эта величина имеет распределение хи-квадрат, а значит величину C можно найти, как квантиль распределения хи-квадрат.

Тогда:

Имеем распределение хи-квадрат с 1 степенью свободы, уровень значимости в нашем случае 0.05, тогда .

Был написан скрипт Python, при помощи которого была посчитана величина . Она равняется: .

Таким образом, видно, что полученное значение больше, чем C, а значит в данном случае основная гипотеза отвергается в пользу альтернативной.

Далее использовался тот же алгоритм, поэтому будем расписывать менее подробно.

1. Полученный методом максимального правдоподобия:

Имеем две гипотезы:

Строим функцию правдоподобия:

Для второй гипотезы в качестве значений параметров будут использованы оценки, полученные методом максимального правдоподобия.

То есть:

Для основной гипотезы:

Логарифм функции правдоподобия:

Частная производная:

Тогда

Тогда , так как в итоге в обеих функциях использовались одни и те же параметры. Таким образом, в данном случае критерий сработал, полученное значение меньше, чем C.

* Параметр смещения:

1. Полученный методом моментов:

Гипотезы:

Строим функцию правдоподобия:

Функция правдоподобия для альтернативной гипотезы:

Логарифм функции правдоподобия:

Оценка для параметра формы:

Тогда:

Тогда функция правдоподобия для первой гипотезы:

Тогда:

Такой результат показывает, что отношение правдоподобия имело значение меньше 1, кроме того, данное значение меньше, чем C, а значит в данном случае мы принимаем основную гипотезу.

1. Полученный методом максимального правдоподобия:

Гипотезы:

Строим функцию правдоподобия:

Функция правдоподобия для альтернативной гипотезы:

Логарифм функции правдоподобия:

Оценка для параметра формы:

Тогда:

Тогда функция правдоподобия для первой гипотезы:

Тогда:

Такой результат показывает, что отношение правдоподобия имело значение , равное 1 (в итоге использовались одинаковые параметры), кроме того, данное значение меньше, чем C, а значит в данном случае мы принимаем основную гипотезу.

## Гипотезы о виде распределения.

Предполагаем, что эмпирическая функция распределения подчиняется распределению Ломакса.

### Критерий Колмогорова

Используется только при условии, что с.в. принимают бесконечно много значений.

Выберем критическое значение используя табличные значения:

Изображение выглядит как текст, Шрифт, число, линия

Автоматически созданное описание

Рисунок 26. [Критические точки распределения Колмогорова](https://portal.tpu.ru/SHARED/s/SHAROVA/Ucheba/Tab5/Tablichnye_znacheniya_kriteriev.pdf).

Зададимся уровнем значимости следовательно:

В результате выполнения скрипта на python была получена статистика:

Критерий имеет правостороннюю критическую область, следовательно, т.к. , гипотеза о том, что генеральная совокупность имеет распределение Ломакса для полученных оценок отвергается.

### Критерий Пирсона

Процедура проверки гипотез с использованием критериев типа предусматривает группирование наблюдений. Область определения случайной величины разбивают на k непересекающихся интервалов граничными точками.

В соответствии с заданным разбиением подсчитывают число выборочных значений, попавших в i-й интервал, и вероятности попадания в интервал:

*, F – теоретическая ф-я распределения*

Статистика критерия согласия Пирсона определяется соотношением:

По аналогии с построением гистограммы, был написан скрипт на Python, который разбивает выборку на интервалы размером 0.1, результат разбиения на **Рисунке 27**:

Изображение выглядит как текст, Шрифт, снимок экрана

Автоматически созданное описание

Рисунок 27. Результат работы скрипта.

Можем заметить, что у хвоста в интервалах нет ни одного элемента. Объединим последние 11 интервалов в один большой [8.83-9.84):

Изображение выглядит как текст, Шрифт, снимок экрана

Автоматически созданное описание

Рисунок 28. Измененный скрипт.

Рассчитаем число степеней свободы:

Найдем табличное значение критической точки для степени свободы **6** и уровня значимости **0.05**:

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рисунок 29. [Табличные значения](https://portal.tpu.ru/SHARED/s/SHAROVA/Ucheba/Tab5/Tablichnye_znacheniya_kriteriev.pdf)

Следовательно:

В результате выполнения скрипта на python была получена статистика:

При гипотеза принимается, следовательно по критерию Пирсона гипотеза о том, что генеральная совокупность имеет распределение Ломакса для полученных оценок принимается.

## Гипотезы об однородности выборок.

Задача: верна ли гипотеза две независимые выборки получены из одной генеральной совокупности?

То есть проверяем гипотезы вида:

.

Близким по виду к распределению Ломакса является экспоненциальное распределение с параметром сдвига**.**

Изображение выглядит как текст, График, снимок экрана, линия

Автоматически созданное описание

Рисунок 28. Экспоненциальное распределение рядом с распределением Ломакса.

Метод обратного преобразования – способ генерации случайных величин с заданной функцией распределения, путём модификации работы генератора равномерно распределённых чисел.

Рассмотрим функцию распределения y=F(x) интересующего нас закона распределения и попробуем выразить обратную функцию:

F(x) всегда принимает значения из интервала . Тогда, если сгенерировать случайную величину , то через обратную функцию можно получить значение , принимающие все значения из области определения.

Экспоненциальное распределение с параметром смещения задается как:

Тогда, можно вычислить обратную функцию, выглядеть она будет как:

Числа были сгенерированы с использованием генератора Mersenne Twister (numpy.random). Этот псевдослучайный генератор генерирует числа с равномерным распределением, следовательно можем его использовать.

Полученная выборка распределения из 300 элементов выглядит следующим образом:

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 29. Распределение по выборке

### Критерий Смирнова

По аналогии с критерием Колмогорова.

Для двух выборок размера k, n статистика критерия:

Был написан скрипт на Python, который считает эту статистику, результат:

Для выборок, где , критическая точка вычисляется как:

Где

Или по таблице в предыдущем пункте.

Для :

Если для заданного уровня значимости то, **гипотеза об однородности выборок** отвергается. Следовательно, в нашем случае отвергается.

### Мощность критерия Смирнова

Считается, что асимптотическая мощность критерия Смирнова **равна 1**. Оценим это самостоятельно.

Мощность это вероятность критерия верно отвергнуть основную гипотезу. В качестве экспериментальной оценки было сделано следующее: 100 раз проверили как отработает критерий в случае, когда основная гипотеза НЕ выполняется, и посчитали сколько раз он дал верный ответ.

Для этого так же сгенерируем выборку Ломакса, с параметрами:

И соответственно выборку Экспоненциального распределения:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Уровень значимости | Мощность критерия относительно альтернативы H1 | | |  |  |
| N=30 | N=50 | N=100 |  |  |
| 0.1 | 0.1118 | 0.1480 | 0.1766 |  |  |
| 0.05 | 0.0463 | 0.0569 | 0.0944 |  |  |

Выводы

**Гипотезы:**

Были рассмотрели гипотезы о виде распределения и использовали критерии Колмогорова и для проверки гипотезы о принадлежности выборки к распределению Ломакса. Критерия дали противоречивые результаты. Вероятно, это связано с тем, что критерий Пирсона изначально используется при условии, что с.в. – дискретная величина. Возможно, если бы выборка была разбита на большее количество интервалов, мы бы получили непротиворечивые результаты.

Список используемых источников

1. Lomax KS, 1954. Business failures: Another example of the analysis of failure data. Journal of the American Statistical Association. 1954; 49: 847–852. – URL: <https://doi.org/10.1080/01621459.1954.10501239>. – (дата обращения: 12.06.2023).
2. Alghamdi, A.S., 2018. Study of generalized Lomax distribution and change-point problem (Doctoral dissertation, Bowling Green State University). – URL: [OhioLINK ETD: Alghamdi, Amani Saeed](https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=bgsu1526387579759835) – (дата обращения: 12.06.2023).
3. Arjun Kumar Gaire, 2022. SKEW LOMAX DISTRIBUTION, PARAMETER ESTIMATION, ITS PROPERTIES, AND APPLICATIONS. URL: [(PDF) SKEW LOMAX DISTRIBUTION, PARAMETER ESTIMATION, ITS PROPERTIES, AND APPLICATIONS (researchgate.net)](https://www.researchgate.net/publication/366824883_SKEW_LOMAX_DISTRIBUTION_PARAMETER_ESTIMATION_ITS_PROPERTIES_AND_APPLICATIONS) – (дата обращения: 12.06.2023).
4. Ahmed Mostafa Khalil, 2021. Five-parameter transmuted generalization of the Lomax distribution. URL: https://www.hindawi.com/journals/cin/2021/5918511/ – (дата обращения: 12.06.2023).
5. Atkinson, M., Harrison, A. J. (1978). Personal Wealth in Britain. Cambridge university Press, Cambridge, UK. URL: https://wid.world/document/atkinsonharrison1978/ – (дата обращения: 12.06.2023).
6. Holland, O., Golaup, A., Aghvami, A. H. (2006). Traffic characteristics of aggregated module downloads for mobile terminal reconfiguration. IEE Proceedings-Communications, 153(5), 683-690. URL: https://kclpure.kcl.ac.uk/portal/en/publications/traffic-characteristics-of-aggregated-module-downloads-for-mobile – (дата обращения: 12.06.2023).
7. Hassan, A. S., Al-Ghamdi, A. S. (2009). Optimum step stress accelerated life testing for Lomax distribution. Journal of applied sciences research, 5(12),2153-2164.URL: https://www.researchgate.net/publication/265432570\_Optimum\_Step\_Stress\_Accelerated\_Life\_Testing\_for\_Lomax\_Distribution – (дата обращения: 12.06.2023).