Министерство науки и высшего образования Российской Федерации

Санкт-Петербургский Политехнический Университет Петра Великого

—

Институт компьютерных наук и кибербезопасности

Высшая школа кибербезопасности

**КУРСОВАЯ РАБОТА**

«Распределение Накагами»

по дисциплине «Теория вероятностей и математическая статистика»

Выполнили

студенты гр. 5151003/20002 Тимофеев И.А.

*<подпись>*

Окунев Д.Д.

*<подпись>*

Преподаватель профессор, д. т. н. Лаврова Д.С.

*<подпись>*

Санкт-Петербург

2024

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc167805063)

[1 ПОСТАНОВКА ЗАДАЧИ 5](#_Toc167805064)

[2 ХОД РАБОТЫ 6](#_Toc167805065)

[2.1 Исследование распределения Накагами 6](#_Toc167805066)

[2.2 Знакомство с Jupyter Notebook 6](#_Toc167805067)

[2.2.1 Считывание выборки 6](#_Toc167805068)

[2.2.2 Сумма элементов выборки 6](#_Toc167805069)

[2.2.3 Выборочное среднее 6](#_Toc167805070)

[2.2.4 Медиана 7](#_Toc167805071)

[2.2.5 Мода 7](#_Toc167805072)

[2.2.6 Размах выборки 7](#_Toc167805073)

[2.2.7 Смещенная дисперсия 7](#_Toc167805074)

[2.2.8 Несмещенная дисперсия 7](#_Toc167805075)

[2.2.9 Выборочный начальный момент k-ого порядка 8](#_Toc167805076)

[2.2.10 Выборочный центральный момент k-ого порядка 8](#_Toc167805077)

[2.2.11 Вычисление статистик 8](#_Toc167805078)

[2.3 Понятие эмпирической функции распределения 9](#_Toc167805079)

[2.3.1 Определение ЭФР 9](#_Toc167805080)

[2.3.2 Алгоритм построения ЭФР 9](#_Toc167805081)

[2.3.3 Построение графиков ЭФР 10](#_Toc167805082)

[2.3.4 Выводы 12](#_Toc167805083)

[2.4 Понятие гистограммы 13](#_Toc167805084)

[2.4.1 Определение гистограммы 13](#_Toc167805085)

[2.4.2 Алгоритм построения гистограммы 13](#_Toc167805086)

[2.4.3 Построение графиков гистограмм 14](#_Toc167805087)

[2.4.4 Выводы 16](#_Toc167805088)

[2.5 Описание параметров распределения 16](#_Toc167805089)

[2.5.1 Определение теоретической функции распределения 16](#_Toc167805090)

[2.5.2 Описание параметров распределения 17](#_Toc167805091)

[2.5.3 Функция построения теоретической функции распределения 17](#_Toc167805092)

[2.5.4 Построение графика теоретической функции распределения 17](#_Toc167805093)

[2.5.5 Анализ влияния параметров 18](#_Toc167805094)

[2.6 Понятие точечных оценок 21](#_Toc167805095)

[2.6.1 Оценка параметров распределения выборки методом моментов 21](#_Toc167805096)

[2.6.2 Оценка параметров распределения выборки методом максимального правдоподобия 21](#_Toc167805097)

[2.6.3 Свойства полученных оценок 23](#_Toc167805098)

[2.6.4 Построение графика и вывод 24](#_Toc167805099)

[2.7 Понятие интервальных оценок 25](#_Toc167805100)

[2.8 Понятие статистических критериев 26](#_Toc167805101)

[2.8.1 Гипотезы о параметрах распределениях 26](#_Toc167805102)

[2.8.2 Гипотезы о виде распределения 28](#_Toc167805103)

[2.8.3 Гипотезы об однородности выборок 28](#_Toc167805104)

[ЗАКЛЮЧЕНИЕ 31](#_Toc167805105)

[СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 32](#_Toc167805106)

ВВЕДЕНИЕ

Целью этой курсовой работы является изучение распределения Накагами, включая его характеристики и особенности. Особое внимание уделяется анализу применения этого распределения к конкретной выборке данных. Для достижения поставленных целей будут использоваться методы теории вероятностей и математической статистики, а также современные программные инструменты для симуляции и обработки данных. В частности, для анализа выборки будет применяться Jupyter Notebook. Это исследование поможет не только освоить теоретические аспекты распределения Накагами, но и получить практический опыт работы с реальными данными.

# ПОСТАНОВКА ЗАДАЧИ

Задачи данной курсовой работы:

1. Исследование истории возникновения распределения, указанного в варианте, анализ его функций и роли.
2. Изучение функционала Jupyter Notebook как инструмента для выполнения научных расчётов и анализа данных.
3. Освоение концепции эмпирической функции распределения.
4. Знакомство с понятием гистограммы и её применением для визуализации распределений.
5. Детальное описание параметров изучаемого распределения, их математического смысла и методов оценки.
6. Изучение точечных оценок как метода статистического вывода для определения значений параметров распределения.
7. Введение в концепцию интегральных оценок и их использование в статистическом анализе.
8. Понимание и применение статистических критериев для проверки гипотез и оценки адекватности моделей.

# ХОД РАБОТЫ

## Исследование распределения Накагами

Распределение Накагами, также известное как m-распределение Накагами, является вероятностным распределением, которое обычно используется для описания амплитуды каналов связи или радиосигналов после прохождения через множество слабых отражений. Оно было предложено как математическая модель Минору Накагами для мелкомасштабного замирания при распространении радиоволн высокой частоты на большие расстояниях [1]. Распределение было названо в честь японского статистика Минору Накагами. Применяется для описания амплитуды каналов связи или радиосигналов после прохождения через множество слабых отражений [2]. В сфере информационной безопасности может быть использовано для обеспечения безопасности в физической природе беспроводного канала связи.

## Знакомство с Jupyter Notebook

Для проведения исследования была установлена интерактивная среда разработки Jupyter Notebook, позволяющая в режиме реального времени выполнять код и визуализировать данные.

### Считывание выборки

Для считывания выборки из файла используется функция *read\_csv()* модуля *pandas*. Выборка сохраняется в переменной *nakagami* в виде списка.

nakagami = pd.read\_csv("Выборка.csv", header=None).squeeze().tolist()

### Сумма элементов выборки

def summation(sample):

"""сумма элементов выборки"""

summation = 0

for value in sample:

summation += value

return summation

### Выборочное среднее

def average(sample):

"""выборочное среднее"""

return summation(sample) / len(sample) if len(sample) > 0 else 0

### Медиана

def median(sample):

"""медиана"""

n = len(nakagami)

sorted\_sample = sorted(sample)

if n % 2 == 1:

return sorted\_sample[n // 2]

else:

return (sorted\_sample[n // 2 - 1] + sorted\_sample[n // 2]) / 2

### Мода

def mode(sample):

"""мода"""

frequency = {}

for value in sample:

if value in frequency:

frequency[value] += 1

else:

frequency[value] = 1

max\_frequency = max(frequency.values())

mode = [key for key, val in frequency.items() if val == max\_frequency]

return None if len(mode) == len(sample) else mode

### Размах выборки

def range(sample):

"""размах выборки"""

return max(sample) - min(sample)

### Смещенная дисперсия

def biased\_variance(sample):

"""смещенная дисперсия"""

avrg = average(sample)

return sum((x - avrg) \*\* 2 for x in sample) / len(sample)

### Несмещенная дисперсия

def unbiased\_variance(sample):

"""несмещенная дисперсия"""

avrg = average(sample)

return sum((x - avrg) \*\* 2 for x in sample) / (len(sample) - 1)

### Выборочный начальный момент k-ого порядка

def initial\_moment(nakagami, k):

"""выборочный начальный момент k-ого порядка"""

return sum(x \*\* k for x in nakagami) / len(nakagami)

### Выборочный центральный момент k-ого порядка

def central\_moment(sample, k):

"""выборочный центральный момент k-го порядка"""

avrg = average(sample)

return sum((x - avrg) \*\* k for x in sample) / len(sample)

### Вычисление статистик

Было произведено вычисление статистик для исходной выборки с использованием разработанных функций, результаты которой представлены в таблице 1.

Таблица 1 – Результаты вычисления статистик рассматриваемой выборки

|  |  |
| --- | --- |
| Статистика выборки | Значение |
| Сумма элементов выборки | -573.4518790000004 |
| Выборочное среднее | -1.9115062633333346 |
| Медиана | -1.9356339999999999 |
| Мода | — |
| Размах выборки | 1.4257630000000001 |
| Смещенная дисперсия | 0.05865071974614067 |
| Несмещенная дисперсия | 0.05884687599947224 |
| Выборочный начальный момент  6-го порядка | 60.121218583858415 |
| Выборочный центральный момент 6-го порядка | 0.003553360661384235 |

## Понятие эмпирической функции распределения

Эмпирическая функция распределения (ЭФР) представляет собой статистический инструмент, который оценивает функцию распределения случайной величины на основе наблюдаемой выборки данных.

### Определение ЭФР

Пусть {*xi*} — наблюдаемая выборка из некоторого распределения. Эмпирическая функция распределения *Fn(x)* определяется как доля наблюдений, которые меньше или равны *x*:

где *nx* – кол-во элементов выборки, которые меньше или равны *x*; *n* – объём выборки.

### Алгоритм построения ЭФР

1. Сортировка выборки в возрастающем порядке.
2. Для каждого значения в отсортированной выборке рассчитывается доля элементов, которые меньше или равны текущему значению, что является значением ЭФР в этой точке.

Код, реализующий этот алгоритм:

def build\_edf(sample, size):

"""построение эмпирической функции распределения для подвыборки из size элементов"""

values = np.random.choice(sample, size=size)

data\_sorted = np.sort(values)

# Вычисляем значения ЭФР в точках, соответствующих отсортированным данным

n = len(data\_sorted)

y\_values = np.arange(1, n + 1) / n

# Визуализация эмпирической функции распределения

plt.step(data\_sorted, y\_values, where='post', label='ЭФР')

plt.title(f'График эмпирической функции для подвыборки из {size} элементов')

plt.xlabel('x')

plt.ylabel('Fn(x)')

plt.grid(True)

plt.legend()

plt.show()

### Построение графиков ЭФР

Были построены графики эмпирической функции распределения для случайных подвыборок из 10, 100 и 200 элементов из исходной выборки, представленные на рисунке 1.

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

Рисунок 1 – График эмпирической функции для подвыборок из 10, 100 и 200 и теоретической функции

### Выводы

Построенные графики эмпирической функции распределения совпадают по форме с графиком теоретической функции распределения Накагами, представленной на рисунке 4. Можно заметить, что с увеличением размера выборки эмпирическая функция становится все более похожей по форме на теоретическую функцию распределения.

Можно заметить, что построенному графику для подвыборки из 200 элементов наиболее соответствуют параметры μ = 2, ω = 2, график теоретической функции распределения с этими параметрами представлен на рисунке 4.

Изображение выглядит как текст, диаграмма, линия, снимок экрана

Автоматически созданное описание

Рисунок 4 – График эмпирических функций, наложенный на график теоретической функции распределения Накагами

## Понятие гистограммы

### Определение гистограммы

Гистограмма — это графическое представление распределения числовых данных, которое позволяет оценить форму, разброс и центральную тенденцию данных. Гистограмма делит весь диапазон значений выборки на серию интервалов и подсчитывает, сколько значений попадает в каждый интервал. Каждый интервал обычно представлен вертикальной столбиком, высота которого соответствует числу элементов в этом интервале.

### Алгоритм построения гистограммы

1. Определение и расчет границ интервалов.
2. Подсчет числа значений в каждом интервале.
3. Рисование графика гистограммы.

Код, реализующий алгоритм построения гистограммы:

def build\_hist(sample, size):

"""построение гистограммы для подвыборки из size элементов"""

values = np.random.choice(sample, size=size)

data\_sorted = np.sort(values)

plt.hist(data\_sorted, bins='auto', edgecolor='black')

plt.title(f'Гистограмма для подвыборки из {size} элементов')

plt.xlabel('Значения')

plt.ylabel('Частота')

plt.show()

### Построение графиков гистограмм

Были построены графики гистограмм для случайных подвыборок из 10, 100 и 200 элементов из исходной выборки, представленные на рисунках 5–7.

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

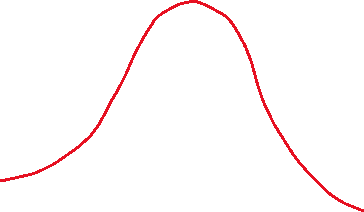


Рисунок 5 – Гистограмма для подвыборки из 10 элементов

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

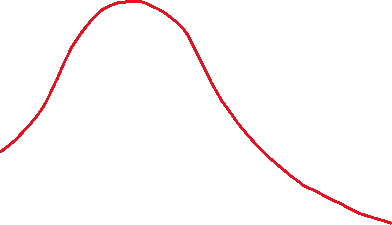


Рисунок 6 – Гистограмма для подвыборки из 100 элементов

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

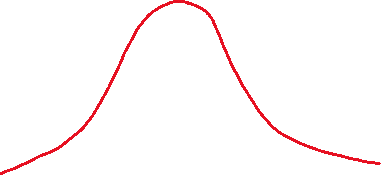


Рисунок 7 – Гистограмма для подвыборки из 200 элементов

### Выводы

Построенные гистограммы совпадают по форме с графиком плотности вероятности распределения Накагами, представленной на рисунке 8.

Можно заметить, что построенный график для подвыборки из 200 элементов наиболее похож на график функции распределения с параметрами μ = 2, ω = 1.

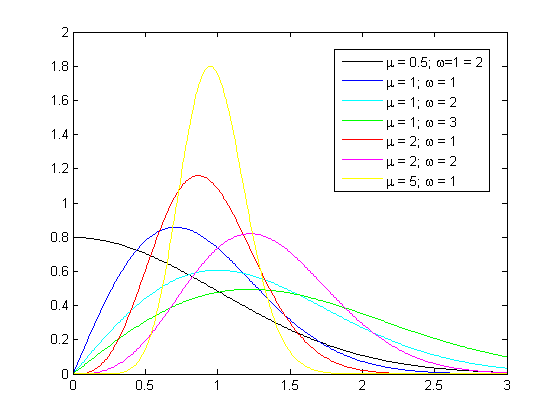


Рисунок 8 – Плотность вероятности распределения Накагами

## Описание параметров распределения

### Определение теоретической функции распределения

Функция распределения *F(x)* для распределения Накагами не выражается в элементарных функциях и обычно представляется через неполную гамма-функцию:

,

где *γ(a, x)* — нижняя неполная гамма-функция.

### Описание параметров распределения

1. Параметр формы *μ*≥0,5 определяет количество рассеивания и вариабельность амплитуды сигнала. Большие значения *μ* указывают на меньшее рассеивание.
2. Параметр масштаба *ω*>0 контролирует масштабирование переменной.

### Функция построения теоретической функции распределения

def nakagami\_cdf(x, nu, loc):

return gammainc(nu, (nu / loc) \* x\*\*2)

def build\_nakagami\_cdf(nu, loc):

x = np.linspace(0,3)

y = nakagami\_cdf(x, nu, loc)

plt.plot(x, y, label=f'μ={nu}, ω={loc}')

plt.title('Распределение Накагами')

plt.xlabel('Значение')

plt.ylabel('Вероятность')

plt.legend()

plt.grid(True)

plt.show()

### Построение графика теоретической функции распределения

Был построен график теоретической функции распределения со случайно заданными параметрами распределения, представленный на рисунке 9.

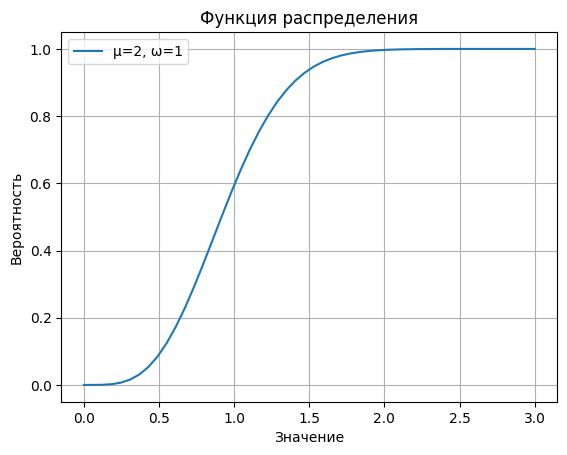


Рисунок 9 – График функции распределения с параметрами *μ* = 2, *ω* = 1

### Анализ влияния параметров

Для оценки влияния параметров распределения были построены графики теоретической функции распределения при изменении параметров, представленные на рисунках 10–13

При увеличении *μ* функция распределения становится более крутой, что указывает на уменьшение рассеивания сигнала и увеличение плотности вероятности около среднего значения.

При увеличении *ω* функция распределения смещается вправо, показывая увеличение среднеквадратичного значения амплитуды.

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

Рисунок 10 – График функции распределения с параметрами μ = 0.5, ω = 1

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 11 – График функции распределения с параметрами μ = 5, ω = 1

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 12 – График функции распределения с параметрами μ = 2, ω = 0.5

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 13 – График функции распределения с параметрами μ = 2, ω = 5

## Понятие точечных оценок

### Оценка параметров распределения выборки методом моментов

Метод моментов заключается в том, что теоретические моменты приравниваются к их эмпирическим аналогам, после чего из этих уравнений определяются оценки неизвестных параметров. Количество уравнений в системе соответствует числу оцениваемых параметров.

Составим систему уравнений так, чтобы математическое ожидание распределения соответствовало среднему значению выборки, а дисперсия распределения — выборочной дисперсии.

Решим приведенную систему с помощью Python и модуля scipy:

def equations(vars):

mu, omega = vars

eq1 = (gamma(mu + 0.5) / gamma(mu)) \* np.sqrt(omega / mu) - X

eq2 = omega \* (1 - 1/mu \* ((gamma(mu + 0.5) / gamma(mu))\*\*2)) - s

return [eq1, eq2]

initial\_guess = [0.5, 0.0001]

solution = fsolve(equations, initial\_guess)

В результате.

### Оценка параметров распределения выборки методом максимального правдоподобия

Метод максимального правдоподобия заключается в поиске такого значения параметра θ, при котором функция правдоподобия L(θ), заданная для конкретных значений выборки {*xi*}, достигает своего максимума. Суть метода состоит в максимизации вероятности того, что имеющиеся у нас наблюдения произошли именно при этом значении параметра.

Функция плотности вероятности распределения Накагами:

Тогда функция правдоподобия:

Для нахождения параметров необходимо решить следующую систему:

Найдем :

Найдем :

Найдем :

Тогда (1) принимает вид:

Решим второе уравнение относительно :

Решим первое уравнение с помощью Python и модуля scipy, подставив найденное значение . В результате .

### Свойства полученных оценок

1. Несмещенность: Оценка является несмещенной, если её математическое ожидание равно оцениваемому параметру. Для проверки несмещенности оценки, рассмотрим:

Таким образом, оценка ММ является несмещенной.

2. Состоятельность: Оценка состоятельна, если ее дисперсия стремится к нулю при увеличении размера выборки.

Таким образом, оценка ММ является состоятельной.

3. Эффективность: Оценка эффективна, если она имеет минимальную дисперсию среди всех несмещенных оценок.

Оценим параметр

Оценка эффективна.

### Построение графика и вывод

Был построен график с теоретическими функциями распределения с параметрами, полученными ММ и ММП, и эмпирической функцией распределения по всей выборке (рис. 14).

Можно заметить, что график теоретической функции с параметрами, полученными методом моментов, наиболее соответствует графику эмпирической функции по всей выборке. Таким образом, можно предположить, что оценка параметров методом моментов наиболее удачная.

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 14 – График теоретических и эмпирической функций распределения

## Понятие интервальных оценок

Интервальной оценкой называют числовой интервал , определяемый по результатам выборки, относительно которого можно утверждать с определенной, близкой к единице вероятностью, что он заключает в себе значение оцениваемого параметра генеральной совокупности, т.е. , где и называют также нижней и верхней границами доверительного интервала параметра . Вероятность принято называть доверительной вероятностью.

Оценим параметры распределения выборки с помощью интервальной оценки с уровнями доверия Согласно центральной предельной теореме:

*;*

Далее был оценен параметр формы с уровнем доверия

Получился интервал с уровнем доверия .

Далее был оценен параметр формы с уровнем доверия

Получился интервал с уровнем доверия .

В результате были оценены параметры с помощью интервальной оценки. Оценки параметров, полученные в предыдущем пункте, не попали в найденные доверительные интервалы.

## Понятие статистических критериев

### Гипотезы о параметрах распределениях

Сформулируем гипотезы:

,

,

где – оценка параметра распределения, полученная в результате применения метода моментов, – метода максимального правдоподобия.

В качестве уровня значимости принято .

Согласно лемме Неймана-Пирсона, наиболее мощным критерием будет являться критерий отношения правдоподобия, который и будем использовать.

Функция правдоподобия:

Отношение правдоподобия:

Логарифм отношения правдоподобия:

Распределение Накагами заменой переменной сводится к гамма-распределению.

можно найти из условия

где , а – плотность распределения случайной величины .

Отсюда = , где – квантиль порядка α функции (y).

Критическая функция =

Функция мощности будет представляться в следующем виде:

Построенный критерий – наиболее мощный, если гипотеза – простая, то есть : θ=. При построении критерия значение используется неявно, важно лишь, что >. Значит, построенный критерий – равномерно наиболее мощный.

При проверке гипотез возможны ошибки: первого рода или второго рода. Ошибка первого рода – ложноположительное срабатывание, верная гипотеза *H*0 отвергается и принимается неверная гипотеза *H*1, где вероятность ошибки – уровень значимости α. Ошибка второго рода – ложноотрицательное срабатывание, принимается ложная гипотеза *H*0, когда верна альтернативная гипотеза *H*1. Ошибка второго рода наиболее опасна.

Для нашего критерия справедливо следующее:

### Гипотезы о виде распределения

Сформулируем гипотезы:

: Выборка принадлежит заданному распределению.

: Выборка не принадлежит заданному распределению.

В качестве уровня значимости принято .

Используя критерий Колмогорова, проверим гипотезу о принадлежности выборки распределению Накагами.

Был написан код на Python с использованием модуля scipy. Результат работы кода равен 0,035, что меньше заданного уровня значимости. Значит, гипотеза верна, и выборка действительно принадлежит распределению Накагами.

Используя критерий ꭓ2, проверим гипотезу о принадлежности выборки распределению Накагами.

Число степеней свободы равно k-m-1 = 3. Критическое значение при уровне значимости 0,5 и степеней свободы 3 равняется 7,82.

Наблюдаемое значение меньше критического, следовательно гипотеза *H0* верна, выборка соответствует распределению Накагами.

### Гипотезы об однородности выборок

Был проведен поиск распределений близкое по виду к распределению Накагами и найдено распределению Вейбулла. На рисунке 15 приведен график функции распределения Вейбулла с наложенным на него эмпирической функцией исходной выборки. Можно заметить, что графики функций похожи.

Функция распределения Вейбулла:

Тогда обратная функция:

Изображение выглядит как линия, График, диаграмма

Автоматически созданное описание

Рисунок 15 – График функции распределения Вейбулла и эмпирической функции подвыборки из 200 элементов

Сгенерируем выборку из стандартного непрерывного равномерного распределения на интервале (0, 1) и на основании обратной функции вычислим выборку из искомого распределения Вейбулла по формуле: .

Код для генерирования выборки:

def num(u):

return lambd\*(-1\*np.log(u))\*\*(1/k)

veybull = []

for i in range(300):

random\_number = random.uniform(0, 1)

veybull.append(num(random\_number))

Построенный график эмпирической функции для сгенерированной выборки представлен на рисунке 16.

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

Рисунок 16 – График эмпирической функции по сгенерированной выборке Вейбулла

Критерий Колмогорова-Смирнова.

0: Выборки из одной генеральной совокупности.

1: Выборки из разных генеральных совокупностей.

Эмпирическое значение критерия: , где количество элементов в первой выборке, во второй, .

Если , то принимается гипотеза 0, в другом случае 1.

Для уровня значимости 0,05 критическое значение (0,05) = 1,36.

Эмпирическое значение превосходит критическое, следовательно выборки не были произведены из одной генеральной совокупности

ЗАКЛЮЧЕНИЕ

В процессе выполнения курсовой работы было осуществлено детальное исследование распределения Накагами. Для анализа данных и выполнения расчётов активно использовалась среда Jupyter Notebook, которая позволила эффективно обрабатывать выборку, проводить все необходимые вычисления и визуализировать их. В частности, статистика выборки Накагами была вычислена в Jupyter Notebook, что включало расчёт основных статистических показателей без использования предварительно заготовленных функций.

В ходе курсовой работы были проведены различные оценки параметров: точечные (метод моментов и метод максимального правдоподобия) и интервальные. Вычисленные точечные оценки не попали в доверительные интервалы. Наиболее точная оценка была получена методом моментов: .

В результате курсовой работы были получены навыки по работе с распределением Накагами и его параметрами, а также по использованию различных методов оценки параметров распределения и проверки различных гипотез о распределении.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Hoffman C. W. The m-Distribution, a general formula of intensity of rapid fading //Statistical Methods in Radio Wave Propagation: Proceedings of a Symposium held June. – 1958. – С. 18-20.

2. Nakagami M. The m-distribution—A general formula of intensity distribution of rapid fading //Statistical methods in radio wave propagation. – Pergamon, 1960. – С. 3-36.

3. Parsons J. D. The Mobile Radio Propagation Channel. New York: Wiley, 1992.

4. Yacoub M. D., Bautista J. E. V., de Rezende Guedes L. G. On higher order statistics of the Nakagami-m distribution //IEEE Transactions on Vehicular Technology. – 1999. – Т. 48. – №. 3. – С. 790-794.

5. Yang Y. H., Wang L. F., Shen M. X. A Simulation Study of Nakagami-m Distribution //Applied Mechanics and Materials. – 2015. – Т. 713. – С. 1405-1408.