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>*

Проверил Лаврова Д.С.

*<подпись>*

Санкт-Петербург

2024

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc167715830)

[1 ПОСТАНОВКА ЗАДАЧИ 5](#_Toc167715831)

[2 ХОД РАБОТЫ 6](#_Toc167715832)

[2.1 Исследование распределения Накагами 6](#_Toc167715833)

[2.2 Знакомство с Jupyter Notebook 6](#_Toc167715834)

[2.2.1 Считывание выборки 6](#_Toc167715835)

[2.2.2 Сумма элементов выборки 6](#_Toc167715836)

[2.2.3 Выборочное среднее 6](#_Toc167715837)

[2.2.4 Медиана 7](#_Toc167715838)

[2.2.5 Мода 7](#_Toc167715839)

[2.2.6 Размах выборки 7](#_Toc167715840)

[2.2.7 Смещенная дисперсия 7](#_Toc167715841)

[2.2.8 Несмещенная дисперсия 7](#_Toc167715842)

[2.2.9 Выборочный начальный момент k-ого порядка 8](#_Toc167715843)

[2.2.10 Выборочный центральный момент k-ого порядка 8](#_Toc167715844)

[2.2.11 Вычисление статистик 8](#_Toc167715845)

[2.3 Понятие эмпирической функции распределения 9](#_Toc167715846)

[2.3.1 Определение ЭФР 9](#_Toc167715847)

[2.3.2 Алгоритм построения ЭФР 9](#_Toc167715848)

[2.3.3 Построение графиков ЭФР 9](#_Toc167715849)

[2.3.4 Выводы 11](#_Toc167715850)

[2.4 Понятие гистограммы 12](#_Toc167715851)

[2.4.1 Определение гистограммы 12](#_Toc167715852)

[2.4.2 Алгоритм построения гистограммы 12](#_Toc167715853)

[2.4.3 Построение графиков гистограмм 12](#_Toc167715854)

[2.4.4 Выводы 15](#_Toc167715855)

[2.5 Описание параметров распределения 15](#_Toc167715856)

[2.5.1 Определение теоретической функции распределения 15](#_Toc167715857)

[2.5.2 Описание параметров распределения 16](#_Toc167715858)

[2.5.3 Функция построения теоретической функции распределения 16](#_Toc167715859)

[2.5.4 Построение графика теоретической функции распределения 16](#_Toc167715860)

[2.5.5 Анализ влияния параметров 17](#_Toc167715861)

[2.6 Понятие точечных оценок 20](#_Toc167715862)

[2.6.1 Оценка параметров распределения выборки методом моментов 20](#_Toc167715863)

[2.6.2 Оценка параметров распределения выборки методом максимального правдоподобия 20](#_Toc167715864)

[2.6.3 Свойства полученных оценок 22](#_Toc167715865)

[2.6.4 Построение графика 23](#_Toc167715866)

[2.6.5 Вывод 24](#_Toc167715867)

[ЗАКЛЮЧЕНИЕ 25](#_Toc167715868)

[СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 26](#_Toc167715869)

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.

### Построение графиков ЭФР

Были построены графики эмпирической функции распределения для случайных подвыборок из 10, 100 и 200 элементов из исходной выборки, представленные на рисунках 1–3.

Изображение выглядит как текст, диаграмма, линия, График

Автоматически созданное описание

Рисунок 1 – График эмпирической функции для подвыборки из 10 элементов

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 2 – График эмпирической функции для подвыборки из 100 элементов

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 3 – График эмпирической функции для подвыборки из 200 элементов

### Выводы

Построенные графики эмпирической функции распределения совпадают по форме с графиком теоретической функции распределения Накагами, представленной на рисунке 4.

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, График, линия

Автоматически созданное описание

Рисунок 4 – График эмпирической функции, 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.

### 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

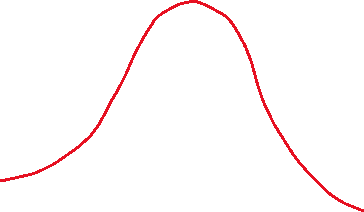


Рисунок 5 – Гистограмма для подвыборки из 10 элементов

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

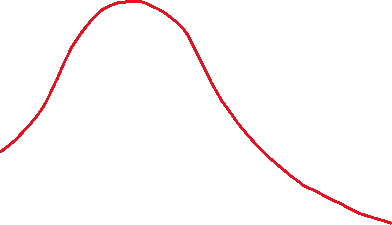


Рисунок 6 – Гистограмма для подвыборки из 100 элементов

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

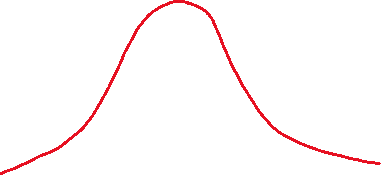


Рисунок 7 – Гистограмма для подвыборки из 200 элементов

### 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.

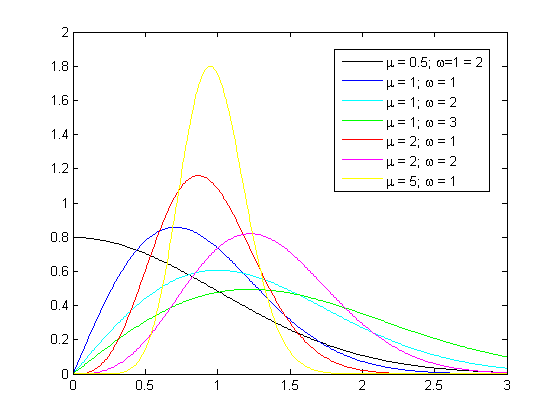


Рисунок 8 – Плотность вероятности распределения Накагами

## 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.

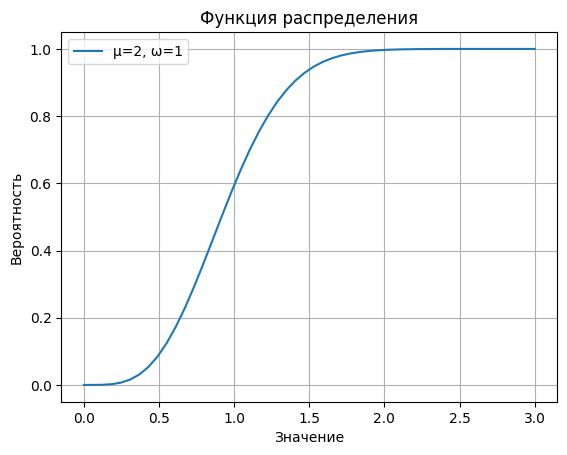


Рисунок 9 – График функции распределения с параметрами *μ* = 2, *ω* = 1

### 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.
2. Состоятельность: Оценка состоятельна, если она стремится к истинному значению параметра при увеличении размера выборки.
3. Эффективность: Оценка эффективна, если она имеет минимальную дисперсию среди всех несмещенных оценок.
4. R-эффективность: Оценка R-эффективна, если она приближает параметры распределения в рамках конкретного метода.

### Построение графика

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

Рисунок 14 – ММ

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

### Вывод

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.

Также в ходе курсовой работы были проведены различные оценки параметров: точечные (метод моментов и метод максимального правдоподобия) и интервальные.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ