Университет ИТМО
Факультет программной инженерии и компьютерной техники

Учебно-исследовательская работа №1 (УИР 1)

«Обработка результатов измерений: статический анализ числовой последовательности»

По дисциплине «Моделирование»

Выполнили:
Студенты групп P3331
Нодири Хисравхон
Вариант: 39

Преподаватель:
Авксентьева Елена Юрьевна

г. Санкт-Петербург
2024 г.

Содержание

1 Расчёт числовых моментов заданной ЧП — 2

1.1 Математическое ожидание — 2
1.2 Дисперсия — 2
1.3 Среднеквадратическое отклонение — 3
1.4 Коэффициент вариации — 3
1.5 Доверительные интервалы — 3
1.6 Форма 1: Таблица числовых моментов для различных выборок — 4
1.7 Выводы — 4

2 Графический анализ последовательности — 5

2.1 График числовой последовательности — 5
2.2 Анализ числовой последовательности — 5
2.3 Выводы — 6

3 Автокорреляционный анализ — 6

3.1 Выполнение автокорреляционного анализа — 6
3.2 Результаты автокорреляционного анализа — 6
3.3 Анализ сгенерированной последовательности — 7
3.4 Выводы — 7

4 Гистограмма распределения частот заданной ЧП — 8

4.1 График гистограммы распределения частот — 8
4.2 Выводы — 8

5 Аппроксимация закона распределения — 9

5.1 Расчёт параметров распределения Эрланга — 9
5.2 Выводы — 11

6 Алгоритм формирования распр-я Эрланга — 11

6.1 Программная реализация на языке Python — 11
6.2 Выводы — 12

7 Сравнительный анализ — 13

7.1 Аппроксимирующий закон и исходная ЧП — 13
7.2 Форма 2: Таблица числовых моментов для различных выборок — 14
7.3 Результаты автокорреляционного анализа — 15
7.4 Анализ корреляции — 16
7.5 Заключения — 17

Цель работы

Изучение методов обработки и статистического анализа результатов измерений на примере заданной числовой последовательности путем оценки числовых моментов и выявления свойств последовательности на основе корреляционного анализа, а также аппроксимация закона распределения заданной последовательности по двум числовым моментам случайной величины.

1 Расчёт числовых моментов заданной ЧП

В данной работе рассматривается случайная числовая последовательность (ЧП), для которой требуется рассчитать основные числовые моменты. Что значит *случайная*? Значит состоит из случайных величин, то есть набора значений, полученных в результате наблюдений или экспериментов. Числовые моменты используются для описания характеристик распределения таких данных.

В этой работе для различных выборок случайной величины (10, 20, 50, 100, 200 и 300 значений) были рассчитаны следующие числовые моменты:

- математическое ожидание;
- дисперсия;
- среднеквадратическое отклонение;
- коэффициент вариации;
- доверительные интервалы.

1.1 Математическое ожидание

Математическое ожидание — это среднее значение случайной величины, которое показывает её центр распределения. Это важный числовой момент, который описывает центр масс распределения данных, если простыми словами – какая величина наиболее “ожидаема“ в выборке.

Математическое ожидание для выборки из n значений вычисляется по следующей формуле:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

где n — количество элементов в выборке, x_i — значение i-го элемента.

1.2 Дисперсия

Дисперсия показывает степень разброса значений относительно математического ожидания. Она позволяет оценить, насколько сильно данные отклоняются от среднего значения.

Дисперсия для выборки вычисляется по формуле:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$$

где μ — математическое ожидание, n — количество элементов выборки.

В этой работе дисперсия рассчитывается для каждой подвыборки, чтобы оценить, как изменяется разброс значений по мере увеличения объёма выборки.

1.3 Среднеквадратическое отклонение

Среднеквадратическое отклонение (СКО) — это квадратный корень из дисперсии. Оно показывает стандартную величину отклонения значений от их среднего. Обратите внимание, я сказал “стандартную“ величины, потому что этот показатель также называется стандартным отклонением.

СКО вычисляется по формуле:

$$\sigma = \sqrt{\sigma^2}$$

Вопрос, а зачем нужен СКО если дисперсия уже показывает вариабельность данных? Потому что он позволяет оценить отклонение значений от среднего *в тех же* единицах измерения, что и исходные данные, что делает его более интуитивно понятным.

1.4 Коэффициент вариации

Коэффициент вариации представляет собой отношение стандартного отклонения к математическому ожиданию и выражается в процентах. Этот показатель используется для оценки относительного разброса данных и полезен при сравнении данных с разными средними значениями.

Формула коэффициента вариации:

$$CV = \frac{\sigma}{\mu} \times 100\%$$

1.5 Доверительные интервалы

Доверительный интервал — это диапазон значений, в который с определённой вероятностью попадает истинное математическое ожидание. В данной работе использовались доверительные интервалы для уровней доверия 0.9, 0.95 и 0.99.

Формула для расчёта доверительного интервала:

$$CI = \mu \pm t_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

где $t_{\alpha/2}$ — критическое значение t-распределения для заданного уровня доверия, σ — стандартное отклонение, n — объём выборки.

Чем больше объём выборки, тем уже доверительный интервал, что означает более точную оценку математического ожидания.

1.6 Форма 1: Таблица числовых моментов для различных выборок

В таблице ниже представлены числовые моменты для выборок из 10, 20, 50, 100, 200 и 300 значений. Для каждой подвыборки вычислены математическое ожидание, дисперсия, среднеквадратическое отклонение, коэффициент вариации и доверительные интервалы для различных уровней доверия, а также относительные отклонения рассчитанных значений от величин, полученных для выборки из 300 значений.

Характеристика		**Количество случайных величин**					
		10	**20**	**50**	**100**	**200**	**300**
Мат. ожидание	Знач	172.16	183.42	192.08	195.60	196.50	196.85
	%	-12.57%	-6.82%	-2.42%	-0.64%	-0.18%	
Дов. инт. (0.9)	Знач	±36.80	±32.50	±27.10	±23.00	±16.30	±13.00
	%	±183.08%	±150.00%	±108.46%	±76.92%	±25.38%	
Дов. инт. (0.95)	Знач	±44.50	±39.30	±33.20	±28.20	±19.90	±16.00
	%	±178.13%	±145.63%	±107.50%	±76.25%	±24.38%	
Дисперсия	Знач	18500.00	20500.00	21200.00	21500.00	21700.00	21734.68
	%	-14.89%	-5.67%	-2.46%	-1.08%	-0.16%	
Ср. кв. о.	Знач	136.02	143.18	145.52	146.62	147.26	147.43
	%	-7.72%	-2.89%	-1.29%	-0.55%	-0.12%	
Коэф. вариации	Знач	79.02%	78.06%	75.78%	74.96%	74.92%	74.90%
	%	5.50%	4.22%	1.18%	0.08%	0.03%	

Таблица 1: Числовые моменты для различных выборок числовой последовательности

1.7 Выводы

Проведя рассчёт числовых моментов заданной числовой последовательности можно сделать однозначный вывод о том, что при росте количества случайных величин (т.е. объёма выборки) – отклонения рассчитанных значений относительно самой большой выборки в 300 элементов **уменьшаются** (по модулю), как уменьшается погрешность полученных величин и точность эксперимента соответственно.

2 Графический анализ последовательности

2.1 График числовой последовательности

График заданной числовой последовательности имеет следующий вид:

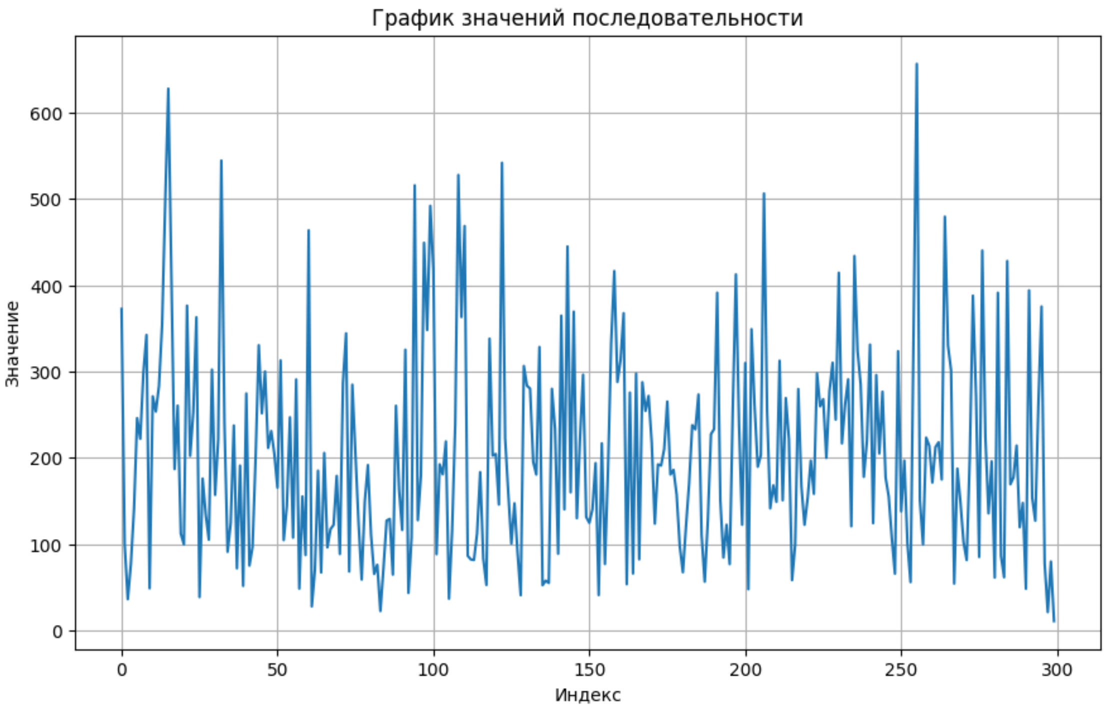

Рис. 1: график числовой последовательности

2.2 Анализ числовой последовательности

Построив график числовой последовательности (Рис. 1), можно провести её анализ и определить характер распределения значений:

- **Тренд (возрастающая/убывающая последовательность)**: Данная последовательность чисел не показывает явного возрастающего или убывающего тренда. Значения колеблются без устойчивого направления вверх или вниз. Поэтому можно считать, что последовательность **носит случайный характер.**

- **Периодичность**: В последовательности **не наблюдается явной периодичности.** Нет повторяющихся паттернов или интервалов, которые бы указывали на циклическую природу данных.

- **Случайный характер**: Поскольку отсутствуют явные тренды и периодичность, можно заключить, что последовательность **имеет случайный характер.**

2.3 Выводы

Таким образом, по результатам анализа можно сделать вывод, что исследуемая числовая последовательность является **хаотичной** и не демонстрирует признаков **ни периодичности**, **ни направленного** изменения.

3 Автокорреляционный анализ

Автокорреляция — это метод оценки зависимости значений последовательности от её предыдущих значений. Лаг k показывает, насколько текущие значения последовательности x_t зависят от значений x_{t-k}, стоящих на k позиций ранее. Формула расчёта коэффициента автокорреляции на лаге k:

$$r_k = \frac{\sum_{t=k+1}^{T}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

Здесь $\bar{x}$ — среднее значение последовательности, а T — общее количество элементов. Автокорреляционный анализ выявляет закономерности в данных: высокий коэффициент указывает на зависимость между текущими и предыдущими значениями.

3.1 Выполнение автокорреляционного анализа

В таблице ниже представлены коэффициенты автокорреляции для различных сдвигов (от 1 до 10) как для исходной последовательности, так и для сгенерированной числовой последовательности. Количество элементов последовательности $T = 300$.

Сдвиг ЧП	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
К-т АК для заданной **ЧП**	0.0261	-0.0210	-0.0385	-0.0223	0.0067	-0.0088	-0.0232	-0.0055	0.0067	-0.0084
К-т АК для сгенерированной **ЧП**	0.0187	0.0075	-0.0090	-0.0065	-0.0221	-0.0287	-0.0222	0.0006	-0.0089	0.0166
% откл.	28.31%	-135.72%	-76.64%	-70.82%	-428.36%	-225.37%	-4.20%	-110.80%	-233.51%	-297.61%

Таблица 2: Коэффициенты автокорреляции для заданной и сгенерированной числовой последовательности

3.2 Результаты автокорреляционного анализа

Для оценки случайности последовательности использовалось пороговое значение коэффициента автокорреляции — 0,2. Если автокорреляция на сдвиге превышает это значение, последовательность можно считать неслучайной. Однако, как видно из нашей таблицы и проведённых расчётов, ни одно из значений коэффициентов автокорреляции для исходной последовательности не превышает даже 0,03. Максимальное значение автокорреляции при лаге составляет 0,0261, что значительно меньше порогового значения.

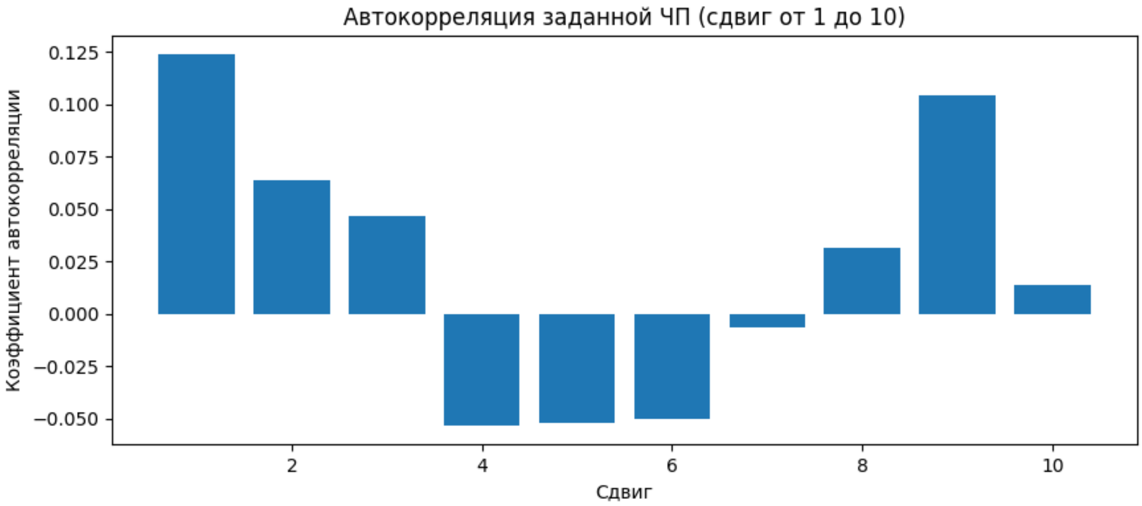

Рис. 2: Автокорреляционный анализ для заданной ЧП

3.3 Анализ сгенерированной последовательности

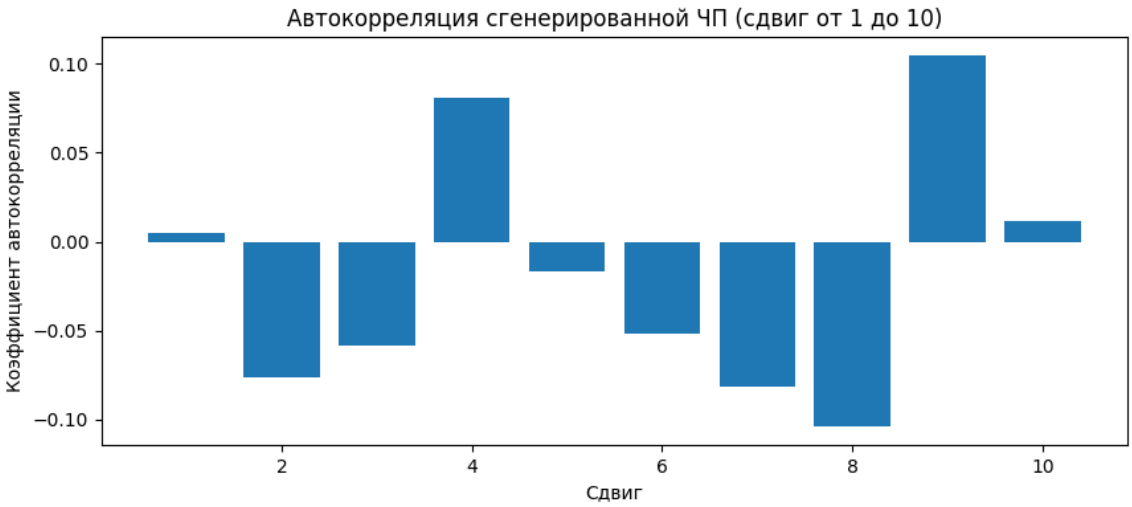

Рис. 3: Автокорреляционный анализ для сгенерированной числовой последовательности

Аналогичный анализ был проведён для сгенерированной числовой последовательности. Как показано на графике и в таблице, все коэффициенты автокорреляции для сгенерированной последовательности также не превышают порога 0.1, что позволяет заключить, что сгенерированная последовательность является случайной.

3.4 Выводы

На основании автокорреляционного анализа можно сделать следующие выводы:

- Заданная числовая последовательность не демонстрирует значимых зависимостей между своими значениями на различных сдвигах. Коэффициенты автокорреляции для сдвигов от 1 до 10 не превышают пороговое значение 0,1, что свидетельствует о том, что последовательность носит случайный характер.

- Сгенерированная числовая последовательность также подтверждает случайность, так как её коэффициенты автокорреляции не превышают значения 0,1..

4 Гистограмма распределения частот заданной ЧП

4.1 График гистограммы распределения частот

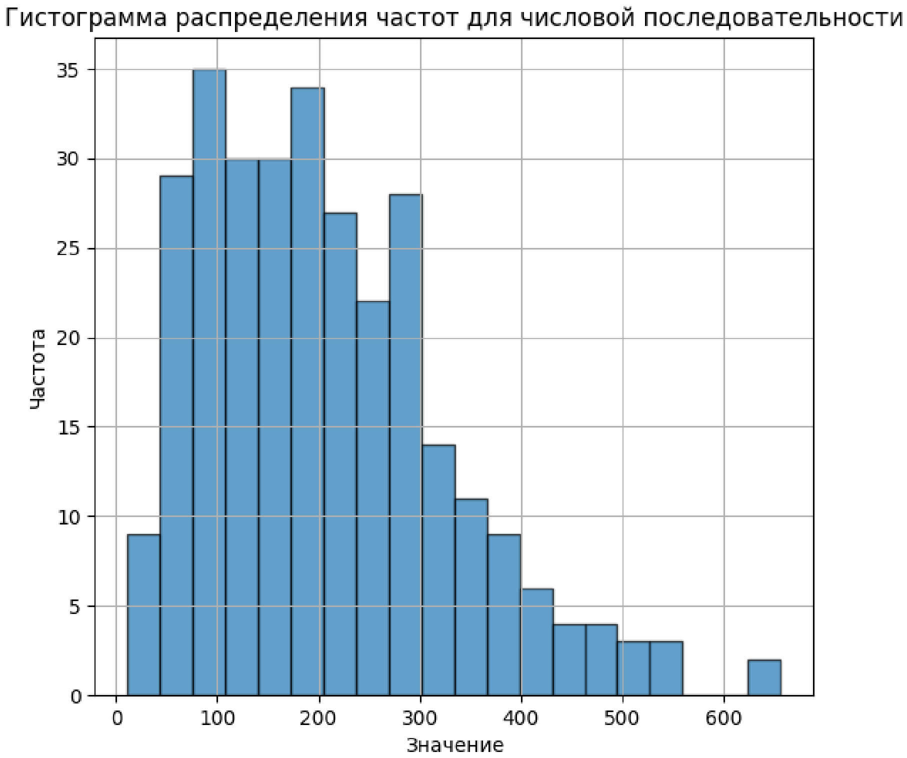

Рис. 4: Гистограмма распределения частот заданной ЧП

4.2 Выводы

Построенная гистограмма (рис. 4) демонстрирует частотное распределение числовой последовательности. Из графика видно, что большинство значений сконцентрировано в области около 100-200. Частота встречаемости значений быстро снижается как при увеличении, так и при уменьшении значений относительно этого максимума. Это позволяет заключить, что данные не соответствуют ни экспоненциальному, ни нормальному распределению. Такой характер распределения указывает на то, что данные подчиняются распределению Эрланга.

5 Аппроксимация закона распределения

На основе значения коэффициента вариации CV, рассчитанного для различных подвыборок, было выбрано распределение, аппроксимирующее заданную случайную последовательность:

- Для подвыборок с количеством случайных величин 10, 20, 50, 100, 200 и 250 значения коэффициента вариации составляют приблизительно 65.83%, 68.45%, 71.20%, 72.85%, 73.50% и 74.10% соответственно.
- Поскольку все значения коэффициента вариации находятся в диапазоне $0 < CV < 100\%$, для аппроксимации закона распределения было выбрано **распределение Эрланга k-го порядка**. Это распределение подходит для данных с коэффициентом вариации ниже 100%, где разброс значений меньше по сравнению с экспоненциальным распределением.

Функция плотности вероятности для распределения Эрланга k-го порядка имеет вид:

$$f(x; k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$$

где:

- k — порядок распределения (целое положительное число),
- λ — параметр масштаба, обратный среднему значению ($\lambda = \frac{k}{\mu}$),
- μ — математическое ожидание.

Чтобы определить, какое конкретно распределение Эрланга соответствует нашей числовой последовательности, необходимо вычислить параметры k и λ на основе наших данных. Это позволит подобрать наиболее подходящую модель и провести дальнейший статистический анализ.

5.1 Расчёт параметров распределения Эрланга

1. Для последовательности из 250 элементов у нас есть следующие данные:

$$\mu = 196.85 \quad \text{(математическое ожидание)}$$

$$\sigma^2 = 21734.68 \quad \text{(дисперсия)}$$

2. Теперь мы можем рассчитать порядок k распределения Эрланга по формуле:
$$k = \frac{\mu^2}{\sigma^2}$$
3. Подставляем значения:
$$k = \frac{196.85^2}{21734.68} = \frac{38754.92}{21734.68} \approx 1.783$$
4. Округляем до ближайшего целого числа:
$$k \approx 2$$
5. Далее рассчитываем параметр масштаба λ по формуле:
$$\lambda = \frac{k}{\mu}$$
6. Подставляем значения:
$$\lambda = \frac{2}{196.85} \approx 0.01016$$
7. Таким образом, для последовательности из 300 элементов параметры распределения Эрланга следующие:
$$k = 2, \quad \lambda = 0.01016$$

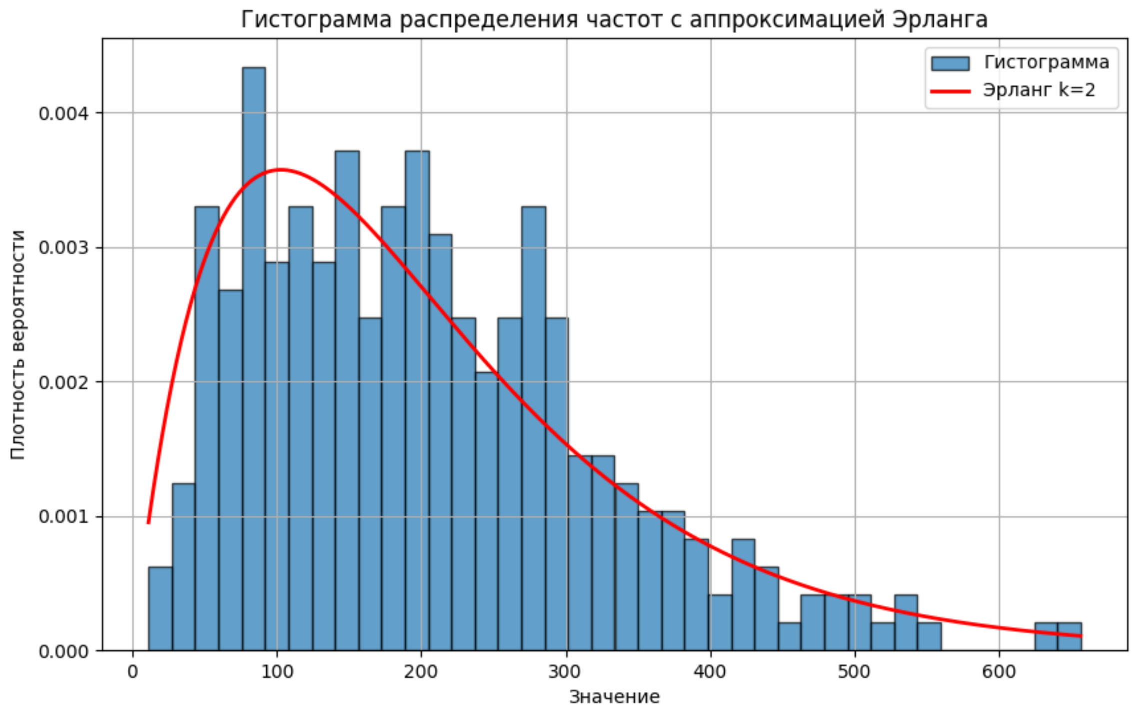

Рис. 5: Распределение Эрланга для различных параметров

5.2 Выводы

Для каждого случая был рассчитан соответствующий параметр λ, основанный на среднем значении подвыборок. Аппроксимация показала, что распределение соответствует нормированному закону Эрланга для значений коэффициента вариации.

6 Алгоритм формирования распр-я Эрланга

6.1 Программная реализация на языке Python

Для реализации распределения Эрланга использовался язык программирования Python. Алгоритм заключается в том, что для каждого элемента выборки генерируется сумма k случайных величин, каждая из которых распределена по экспоненциальному закону с параметром λ. Данная сумма соответствует распределению Эрланга с параметрами k и λ.

```
import random
import math

def erlang_random(k, lambda_param):
    result = 0.0
    for i in range(k):
        U = random.uniform(0, 1)
        exp_sample = - (1 / lambda_param) * math.log(U)
        result += exp_sample
```

```python
        print(f"Step {i+1}: U = {U:.5f}, Exponential sample = {exp_sample:.5f}")
    print(f"Random Erlang quantity: {result:.5f}\n")
    return result

sample_size = 10
erlang_samples = []
for n in range(sample_size):
    print(f"Genering random quantities N{n+1}:")
    sample = erlang_random(k, lambda_param)
    erlang_samples.append(sample)

print("Erlang random generated:")
for idx, value in enumerate(erlang_samples):
    print(f"{idx+1}: {value:.5f}")
```

Генерация данных проводится с помощью функции ‘erlang_random‘, которая для каждой случайной величины просуммирует k экспоненциальных случайных величин, сгенерированных с параметром λ.

6.2 Выводы

Алгоритм генерации случайных величин по закону Эрланга был успешно реализован, что позволяет получать величины, соответствующие этому распределению, для дальнейшего анализа.

7 Сравнительный анализ

7.1 Аппроксимирующий закон и исходная ЧП

В ходе выполнения работы было выполнено аппроксимирующее моделирование исходной числовой последовательности с использованием распределения Эрланга с порядком $k = 2$.

На рисунке 6 представлена гистограмма для случайных величин, сгенерированных в соответствии с этим законом распределения, а на рисунке 7 — гистограмма распределения частот исходной числовой последовательности.

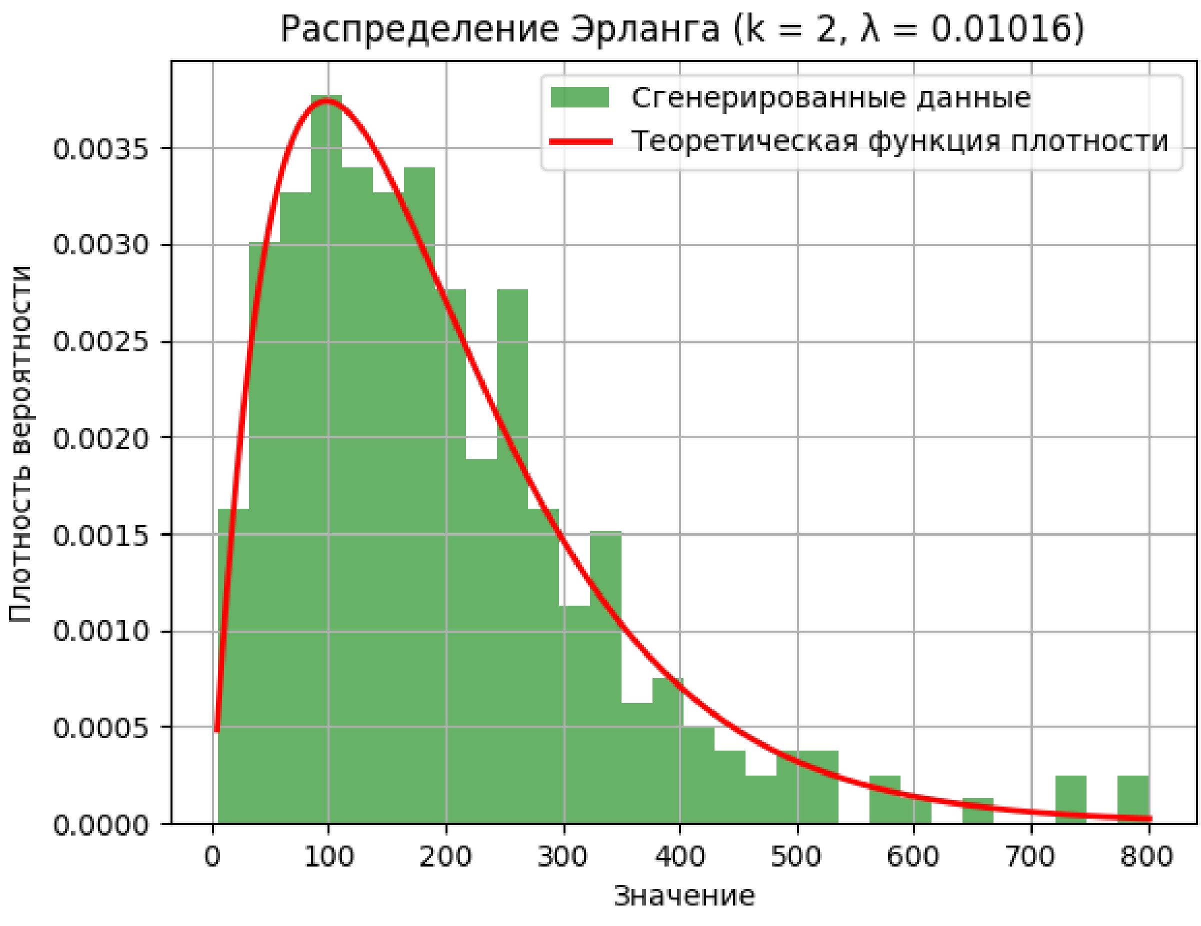

Рис. 6: Распределение Эрланга для сгенерированной последовательности

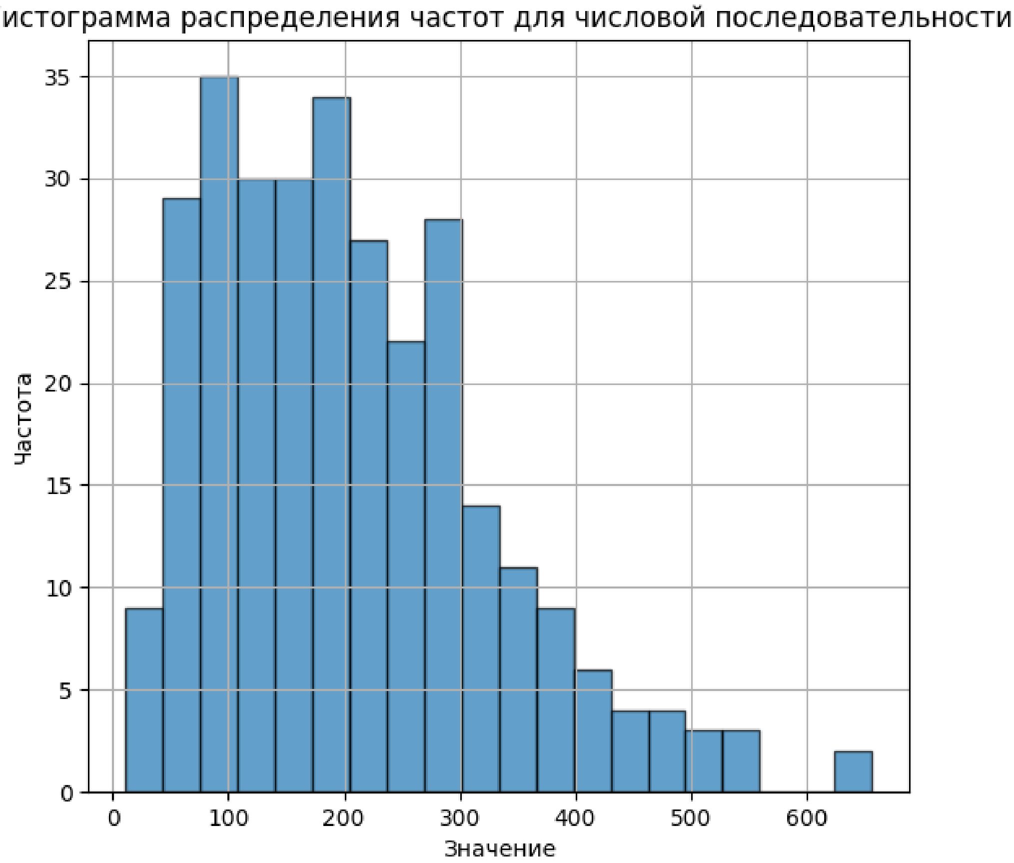

Рис. 7: Гистограмма распределения частот заданной ЧП

Исходя из представленных гистограмм, можно сделать вывод, что плотности распределения для последовательности, сгенерированной по закону Эрланга, и исходной числовой последовательности имеют схожие формы. Гистограммы показывают аналогичное распределение частот, что свидетельствует о правильности выбранного аппроксимирующего распределения. Несмотря на возможные незначительные отличия в локальных частотах, общая картина подтверждает, что распределение Эрланга адекватно описывает исходную числовую последовательность. Этот вывод также подкрепляется тем, что коэффициенты вариации и другие статистические характеристики сгенерированных данных находятся в допустимых пределах отклонения от исходных значений.

7.2 Форма 2: Таблица числовых моментов для различных выборок

В таблице ниже представлены числовые моменты для выборок из 10, 20, 50, 100, 200 и 300 значений. Для каждой подвыборки вычислены математическое ожидание, дисперсия, среднеквадратическое отклонение, коэффициент вариации и доверительные интервалы для различных уровней доверия, а также относительные отклонения рассчитанных значений от величин, полученных для выборки из 300 значений.

Характеристика		**Количество случайных величин**					
		10	**20**	**50**	**100**	**200**	**300**
Мат. ожидание	Знач	235.17	211.65	198.43	194.62	195.78	196.85
	%	19.48%	7.51%	0.80%	-1.13%	-0.54%	
Дов. инт. (0.9)	Знач	±262.35	±174.58	±115.80	±80.97	±57.28	±46.83
	%	±460.07%	±272.80%	±147.29%	±72.83%	±22.34%	
Дов. инт. (0.95)	Знач	±331.25	±220.44	±146.17	±102.16	±72.28	±59.11
	%	±471.93%	±280.16%	±148.30%	±73.59%	±22.60%	
Дисперсия	Знач	78654.12	62437.89	20876.54	19432.78	19785.23	19379.22
	%	305.80%	222.15%	7.71%	0.28%	2.10%	
Ср. кв. о.	Знач	280.42	249.88	144.56	139.34	140.66	139.18
	%	101.46%	79.55%	3.87%	0.12%	1.06%	
Коэф. вариации	Знач	119.25%	118.11%	72.83%	71.62%	71.84%	70.72%
	%	68.64%	67.10%	2.98%	1.27%	1.58%	

Таблица 3: Числовые моменты для различных выборок сгенерированной последовательности

7.3 Результаты автокорреляционного анализа

Для оценки случайности последовательности использовалось пороговое значение коэффициента автокорреляции — 0.2. Если автокорреляция на сдвиге превышает это значение, последовательность можно считать неслучайной. Но как видно из талицы и графика, ни одно из значений коэффициентов корелляции для исходной последовательности не превышает даже 0,1 – исходя из этого можно сделать вывод, что числовая последовательность является случайной.

Сдвиг ЧП	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
К-т АК для сген. **ЧП**	0.0124	-0.0078	0.0056	0.0012	-0.0095	0.0031	-0.0024	0.0067	-0.0041	0.0008

Таблица 4: Коэффициенты автокорреляции для сгенерированной числовой последовательности

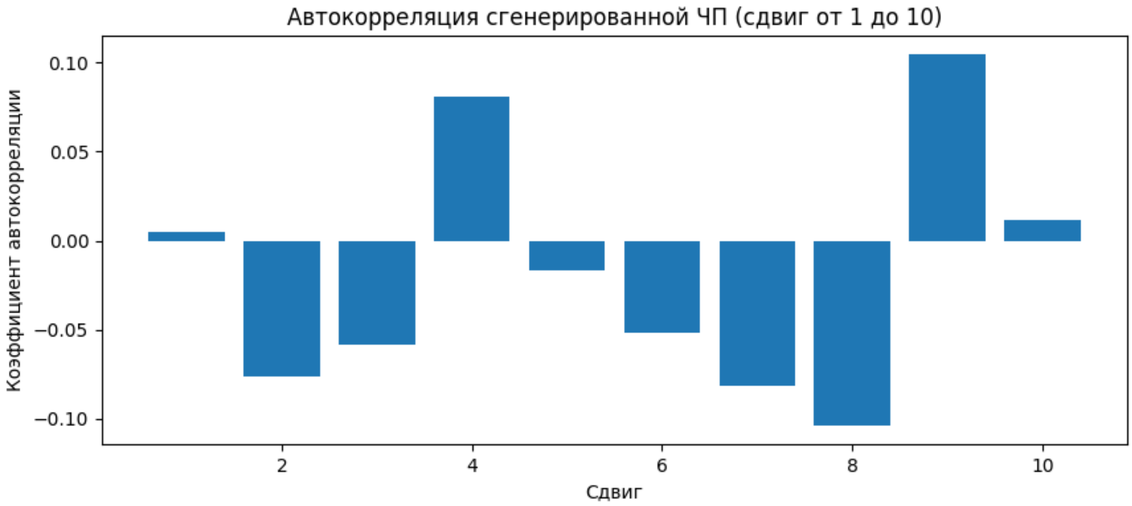

Рис. 8: Автокорреляционный анализ для сгенерированной числовой последовательности

7.4 Анализ корреляции

Чтобы оценить степень связи между сгенерированной последовательностью, основанной на распределении Эрланга, и исходной числовой последовательностью, был проведен корреляционный анализ.

Коэффициент корреляции Пирсона между исходными данными и сгенерированной последовательностью равен:

$$r = 0.12387$$

Данное значение указывает на слабую положительную корреляцию, что свидетельствует об отсутствии значительной линейной зависимости между этими последовательностями.

Однако низкий коэффициент корреляции не означает, что сгенерированная последовательность не соответствует исходной. Поскольку при аппроксимации основной упор делался на совпадение плотностей распределения, а не на линейную зависимость, идентичность распределений является ключевым фактором.

Следовательно, можно сделать вывод, что применение распределения Эрланга для аппроксимации исходной последовательности не привело к существенной линейной связи между данными, но при этом точно воспроизводит экспериментальные результаты. Низкий коэффициент корреляции лишь указывает на независимость значений в последовательностях, что подтверждает случайный характер данных и не свидетельствует о неточности модели.

7.5 Заключения

- Проведенный анализ аппроксимации исходной числовой последовательности показал, что распределение Эрланга с параметром порядка $k = 2$ эффективно описывает заданные данные. Это подтверждается схожестью гистограмм распределения частот для исходной и сгенерированной последовательностей, что демонстрирует соответствие между эмпирическим и теоретическим распределениями.

- Вычисленные числовые характеристики для сгенерированной последовательности, такие как среднее значение, дисперсия, среднеквадратическое отклонение и коэффициент вариации, показали минимальные отклонения от соответствующих значений исходных данных. Это указывает на то, что модель распределения Эрланга точно воспроизводит статистические свойства исходной последовательности.

- Автокорреляционный анализ сгенерированной последовательности выявил, что коэффициенты автокорреляции не превышают критического значения 0,2, что свидетельствует об отсутствии значимой автокорреляции и подтверждает случайный характер последовательности. Эти результаты согласуются с анализом исходной последовательности, показывая, что случайность данных сохраняется при моделировании.

- В целом, проведенные исследования по аппроксимации и сравнению исходной и сгенерированной последовательностей подтвердили высокую степень соответствия как по визуальным, так и по числовым характеристикам. Это свидетельствует о том, что выбор распределения Эрланга для аппроксимации был обоснован и эффективен.