Университет ИТМО

Факультет программной инженерии и компьютерной техники

Учебно-исследовательская работа №1 (УИР 1)

“Обработка результатов измерений: статистический анализ числовой последовательности”

по дисциплине “Моделирование”

Выполнили:

Студенты группы Р3334

Баянов Р. Д.

Кузнецов Д. А.

Вариант: 38

Преподаватель:

Авксентьев Е. Ю.

Санкт-Петербург

2024 г.

Оглавление

Цель работы....3
Расчёт статистических характеристик заданной числовой последовательности....4
График значений заданной ЧП....6
Автокорреляционный анализ....7
Гистограмма распределения....8
Аппроксимация закона распределения....9
Генератор случайных чисел по заданным параметрам....10
Анализ сгенерированной последовательности....11
Автокорреляционное сравнение заданной и сгенерированной ЧП....12
Сравнения графиков двух ЧП....13
Корреляционная зависимость заданной и сгенерированной ЧП....14
Вывод....15

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.

Расчёт статистических характеристик заданной числовой последовательности

Для расчёта оценки математического ожидания, оценки дисперсии, оценки среднеквадратического отклонения, коэффициента вариации, доверительных интервалов были использованы формулы:

$$\widetilde{m} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \widetilde{D} = \frac{\sum_{i=1}^{n} (X_i - \widetilde{m})^2}{n-1}.$$

$\widetilde{\sigma} = \sqrt{\widetilde{D}}$ – среднеквадратическое отклонение.

$$\widetilde{\sigma}_m = \frac{\widetilde{\sigma}}{\sqrt{n}}$$

$\varepsilon_p = t_p \widetilde{\sigma}_m$, где t_p - t-оценка доверительного уровня. Определять этот параметр мы будем с помощью таблицы из методички.

Таблица 1

p	t_p	p	t_p
0,80	1,282	0,91	1,694
0,81	1,310	0,92	1,750
0,82	1,340	0,93	1,810
0,83	1,371	0,94	1,880
0,84	1,404	0,95	1,960
0,85	1,439	0,96	2,053
0,86	1,475	0,97	2,169
0,87	1,513	0,98	2,325
0,88	1,554	0,99	2,576
0,89	1,597	0,9973	3,000
0,90	1,643	0,999	3,290

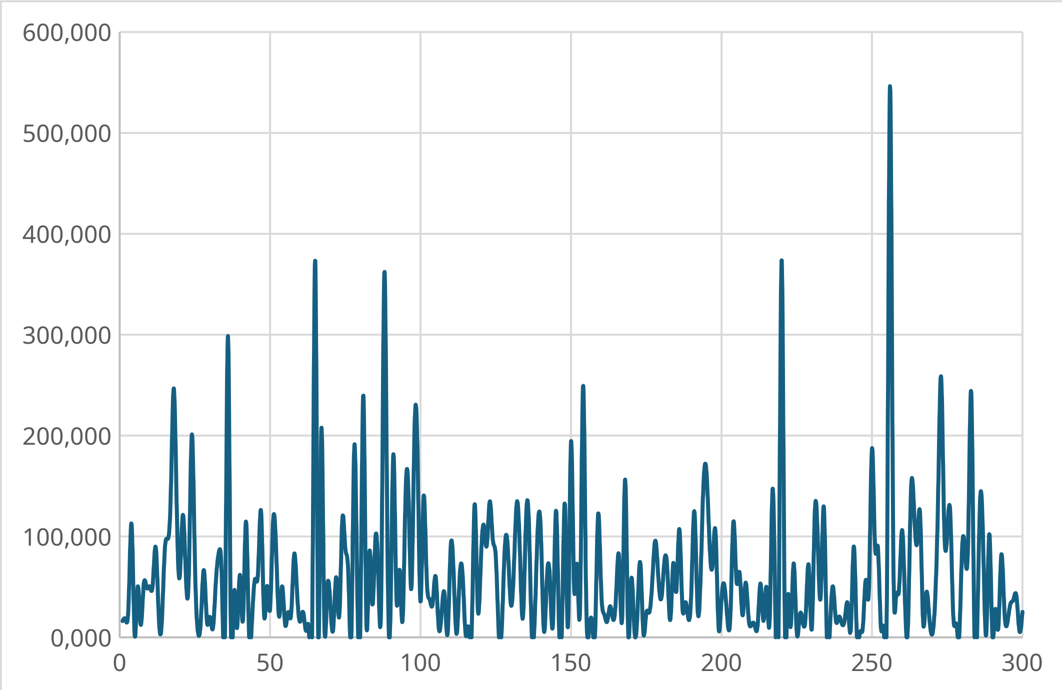
$v = \frac{\widetilde{\sigma}}{\widetilde{m}}$ – коэффициент вариации.

Характеристики заданной ЧП (вариант 38)							
Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат.ож.	Знач.	39,641	66,556	62.094	70.269	64.800	63.561
	%	37.633	4.712	2.308	10.554	1.949	
Дов. инт. (0,9)	Знач.	1,296	1,439	1,620	1,740	1,595	1,766
	%	26.614	18.516	8.267	1.472	9.683	
Дов. инт. (0,95)	Знач.	1,546	1,717	1,933	2,075	1,903	2,107
	%	26.626	18.510	8.258	1.519	9.682	
Дов. инт. (0,99)	Знач.	2,032	2,256	2,540	2,727	2,501	2,770
	%	26.643	18.556	8.303	1.552	9.711	
Дисперсия	Знач.	978.190	3398.003	3750.034	5535.453	3959.493	4669.721
	%	79.052	27.233	19.695	18.539	15.209	
С.к.о	Знач.	31.276	58.292	61.238	74.401	62.925	68.335
	%	54.231	14.697	10.386	8.877	7.917	
К-т вариации	Знач.	0.789	0.876	0.986	1.059	0.971	1.075
	%	26.605	18.512	8.279	1.488	9.674	

% - относительные отклонения рассчитанных значений от значений, полученных для выборки из трехсот величин

Заметим, что при увеличении количества значений в последовательности мы видим, что коэффициент корреляции стремится к значению выше 1, так как значения для последовательности из 300 элементов, мы считаем эталонной.

График значений заданной ЧП



Проанализировав данный график значений для заданной числовой последовательности, мы можем сделать вывод, что числовая последовательность не является: убывающей, периодической, возрастающей.

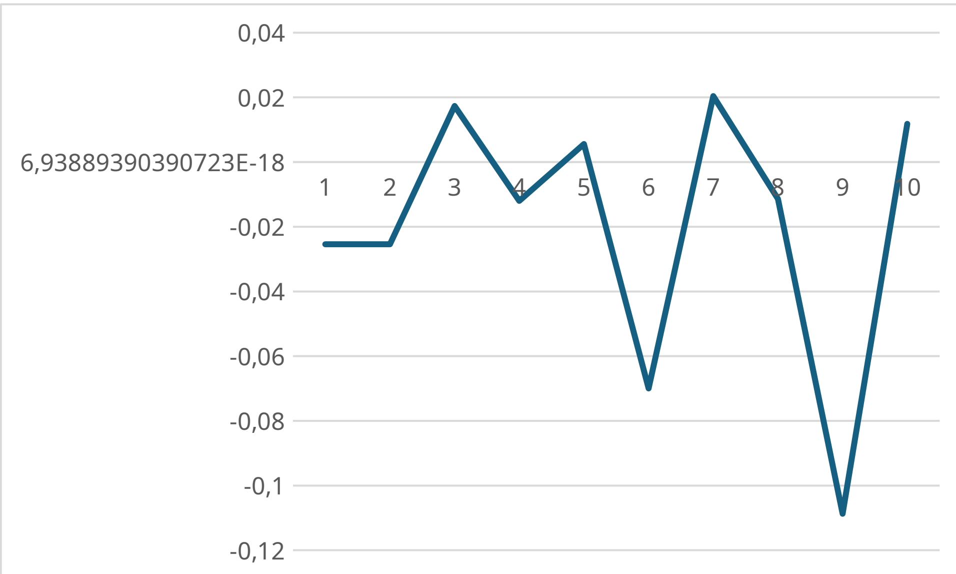
Автокорреляционный анализ

Формула для вычисления коэффициента автокорреляции для заданной ЧП:

$r_{X X_k} = \frac{cov(X, X_k)}{\sigma_X^2}$; где, k – сдвиг ЧП.

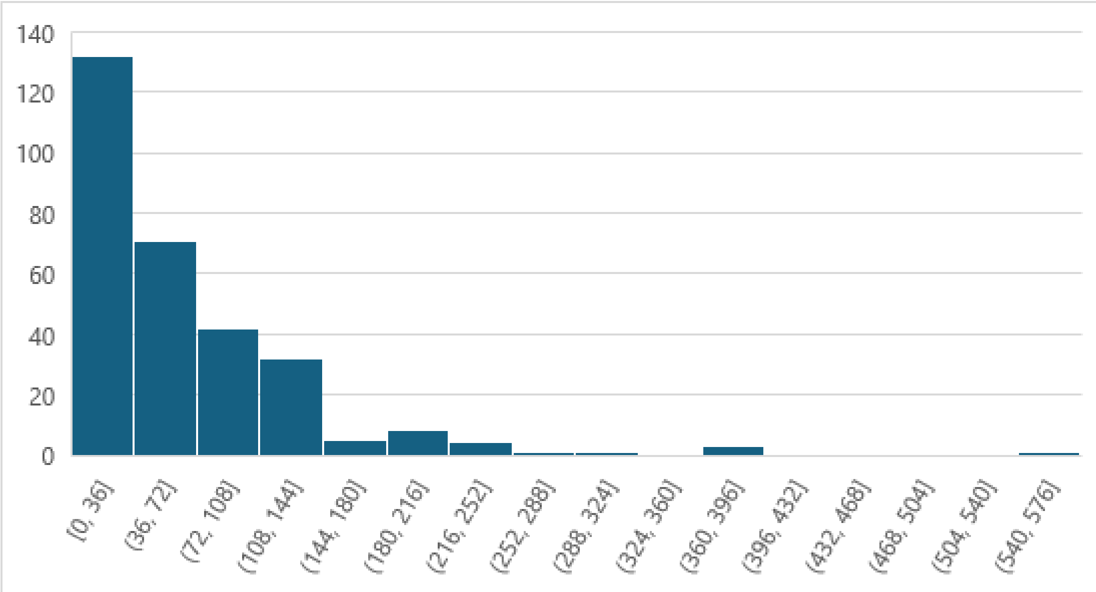
Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для заданной ЧП	0,008699	0,008699	0,067114	0,00196	0,06185	0,01759	0,116157	0,01881	0,02142	0,063837

График, значений коэффициентов и сдвигов:



Как видно из графика и таблицы, между коэффициентами автокорреляции нет никакой тенденции или периодичности. Поэтому последовательность можно считать случайной.

Гистограмма распределения



Заметим, что основная масса значений в ЧП находится в диапазоне маленьких значений. Чаще всего встречаются значения от 0 до 36 и от 36 до 72. Чуть меньше от 72 до 216, и очень редко встречаются большие значения.

Аппроксимация закона распределения

Так как мы имеем при выборке из заданной ЧП в 300 элементов коэффициент вариации больше 1, то для аппроксимации нашего набора чисел возьмём гиперэкспоненциальное распределение.

Рассчитаем три момента, так как аппроксимация с помощью гиперэкспоненциального распределения может осуществляться через три момента. Рассчитаем их по следующим формулам, $v=1.075$:

$q \le \frac{2}{1+v^2} = 0.9278$

Возьмём q = 0.9 и t = 63.561 – мат. ожидание

$$t_1 = \left[1 + \sqrt{\frac{1-q}{2q}\left(v^2-1\right)}\right] * t = 69.471$$

$$t_2 = \left[1 - \sqrt{\frac{q}{2\left(1-q\right)}\left(v^2-1\right)}\right] * t = 10.370$$

Мы получили вероятность генерации числа q и мат. ожидания первой и второй экспоненциальных фаз.

Генератор случайных чисел по заданным параметрам

```
from scipy.stats import expon
import random

t1 = 69.471
t2 = 10.370

# Генерация выборок из экспоненциального распределения
data1 = expon.rvs(scale=t1, loc=0, size=300)
data2 = expon.rvs(scale=t2, loc=0, size=300)

# Инициализация пустого списка для результата
result = []

# Проходим по диапазону 300 значений
for _ in range(300):
    q = random.uniform(0, 1)  # Генерация случайного числа от 0 до 1
    index = random.choice(range(300))  # Выбираем случайный индекс от 0 до
299
    value = data1[index] if q < 0.9 else data2[index]

    # Добавляем выбранное значение в результат
    result.append(value)

for val in result:
  print(val)
```

Данный код генерирует числовую последовательность по гиперэкспоненциальному распределению, основываясь на параметрах, которые мы высчитали ранее. Первые два массива создают по 300 чисел отталкиваясь от параметров t1 и t2. А дальше значения будут перераспределяться в результирующий массив на основе вероятности q.

Анализ сгенерированной последовательности

Характеристики сгенерированной ЧП (вариант 38)							
Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат.ож.	Знач.	102,997	84,189	70,719	68,529	69,921	64,895
	%	58,714	29,731	8,974	5,600	7,745	
Дов. инт. (0,9)	Знач.	1,534	1,669	1,750	1,919	1,839	1,811
	%	15,315	7,861	3,390	5,934	1,517	
Дов. инт. (0,95)	Знач.	1,830	1,991	2,088	2,289	2,194	2,161
	%	15,315	7,861	3,390	5,934	1,517	
Дов. инт. (0,99)	Знач.	2,405	2,617	2,744	3,009	2,883	2,840
	%	15,315	7,861	3,390	5,934	1,517	
Дисперсия	Знач.	9247,973	7314,359	5674,054	6406,238	6124,568	5119,189
	%	80,653	42,881	10,839	25,142	19,639	
С.к.о	Знач.	96,166	85,524	75,326	80,039	78,260	71,549
	%	34,407	19,533	5,280	11,867	9,380	
К-т вариации	Знач.	0,934	1,016	1,065	1,168	1,119	1,103
	%	15,315	7,861	3,390	5,934	1,517	

% - относительные отклонения рассчитанных значений от значений, полученных для выборки из трехсот величин

Отсюда мы видим, что сгенерированная последовательность не существенно отличается от оригинальной при выборке в 300 чисел, так как отклонения крайне малы. Это свидетельствует о правильном выборе распределения и о верном расчёте параметров для генерации ЧП.

Автокорреляционное сравнение заданной и сгенерированной ЧП

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для заданной ЧП	0,008699	0,008699	0,067114	0,00196	0,06185	0,01759	0,116157	0,01881	0,02142	0,063837
К-т АК для заданной ЧП	-0,02543	-0,02543	0,017347	-0,012	0,005582	-0,06996	0,020393	-0,01139	-0,10874	0,011829
%	192.32	192.32	74.153	712.244%	90.968	497.726	99.980	160.553	607.656	81.522

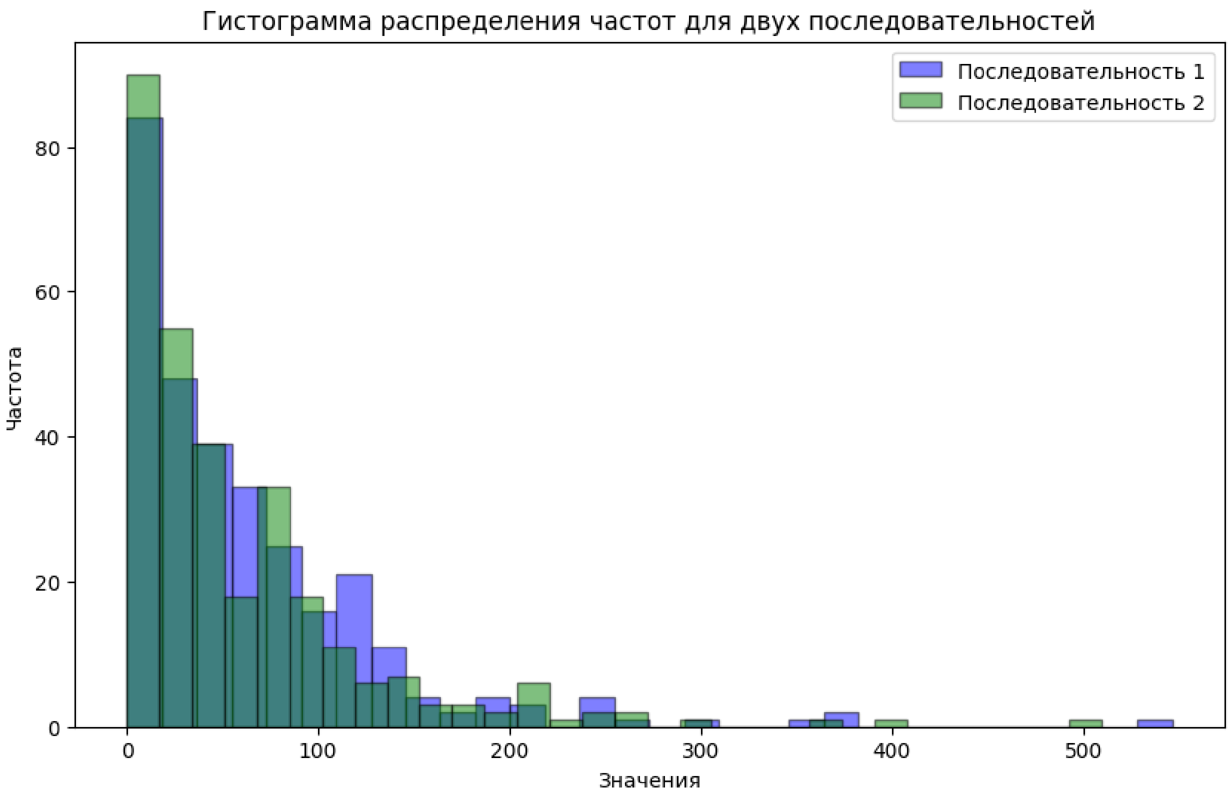
Коэффициенты автокорреляции крайне низкие и не совпадают друг с другом, что говорит о том, что и заданная, и сгенерированная ЧП являются случайными.

Сравнения графиков двух ЧП

Ряд1 – сгенерированная, ряд2 - заданная

Последовательность 1 – заданная, последовательность 2 – сгенерированная

тихдву



Из этих двух графиков мы чётко видим, что последовательности очень схожи, но ни в коем случае не одинаковые.

Корреляционная зависимость заданной и сгенерированной ЧП

Коэффициент корреляции двух ЧП равен k = 0.013841, это значение крайне невелико, что говорит о том, что эти ЧП никак друг на друга не влияют и никак друг от друга не зависят. Они не являются одинаковыми и никак не связаны, но при этом имеют похожий характер распределения значений.

Вывод

Выполнив данную лабораторную работу, мы вспомнили основы математической статистике и попробовали изучить заданную числовую последовательность. Выяснили, что по коэффициенту ковариации и с помощью аппроксимации можно построить ЧП очень похожую на изначальную по своим распределительным характеристикам. Сравнили эти две ЧП и поняли, что дисперсия и мат. ожидания у них хоть и отличаются, но эти значения не выйдут за пределы доверительных интервалов. К тому же эти значения отличаются не сильно.