

Университет ИТМО
Факультет ПИиКТ

Учебно-исследовательская работа 1 (УИР 1)
**«Обработка результатов измерений: статистический
анализ числовой последовательности»**
Вариант 21

Выполнила:

Батомункева Виктория Жаргаловна, Р34101

Преподаватель:

Алиев Тауфик Измайлович

Санкт-Петербург
2024

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.

Выполнение:

1. Оценки математического ожидания, дисперсии, среднеквадратического отклонения, коэффициента вариации заданной числовой последовательности и доверительные интервалы для оценки математического ожидания с доверительными вероятностями 0,9; 0,95 и 0,99, сведенные в таблицу (форма 1);

Характеристики заданной ЧП (вариант 21)

Характеристика	Количество случайных величин						
		10	20	50	100	200	
Мат. ож.	Знач	159,926	195,376	230,812	198,655	202,242	206,063
	%	22,39%	5,19%	10,72%	3,59%	1,85%	
Дов. инт. (0.9)	Знач	±44,298	±31,152	±29,620	±20,574	±14,058	±11,413
	%	74,23%	63,36%	61,47%	44,52%	18,81%	
Дов. инт. (0.95)	Знач	±52,844	±37,163	±35,335	±24,543	±16,770	±13,615
	%	74,23%	63,36%	61,47%	44,52%	18,81%	
Дов. инт. (0.99)	Знач	±69,453	±48,843	±46,440	±32,257	±22,041	±17,895
	%	74,23%	63,36%	61,47%	44,52%	18,81%	
Дисперсия	Знач	7269,183	7190,132	16250,169	15680,372	14641,636	14476,843
	%	49,79%	50,33%	10,91%	7,68%	1,13%	
С.к.о	Знач	85,260	84,795	127,476	125,221	121,003	120,320
	%	29,14%	29,53%	5,61%	3,91%	0,56%	
К-т вариации	Знач	0,533	0,434	0,552	0,630	0,598	0,584
	%	8,70%	25,67%	5,41%	7,37%	2,41%	

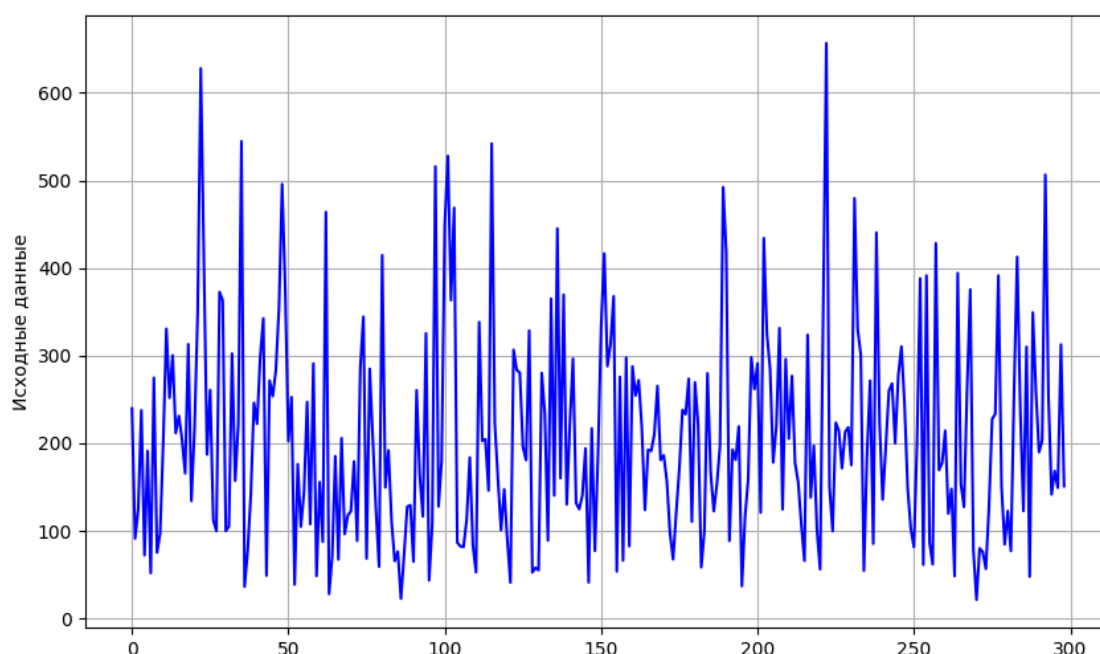
Выводы по пункту 1:

- Большие значения доверительных интервалов говорят о том, что последовательность содержит значения с большой разницей друг другу.
- Большие значения дисперсии указывают на то, что последовательность содержит данные с высокими колебаниями или редкими, но экстремальными значениями (возможно, выбросами)

- Среднеквадратичное отклонение также стабилизируется при больших n , колеблясь вокруг значения **120–127**. Высокое значение СКО подтверждает, что в последовательности присутствует значительный разброс вокруг среднего значения.

2. График (график 1) значений заданной числовой последовательности с результатами анализа характера числовой последовательности (*возрастающая, убывающая, периодичная и т.п.*);

График заданной числовой последовательности



Выводы по пункту 2:

На графике видно, что последовательность:

- не является периодической, так как частота появления «пиков» нерегулярна
- не возрастает, поскольку также присутствуют эти «пики»
- не убывает, так как имеются «пики» — значения, которые значительно превышают соседние

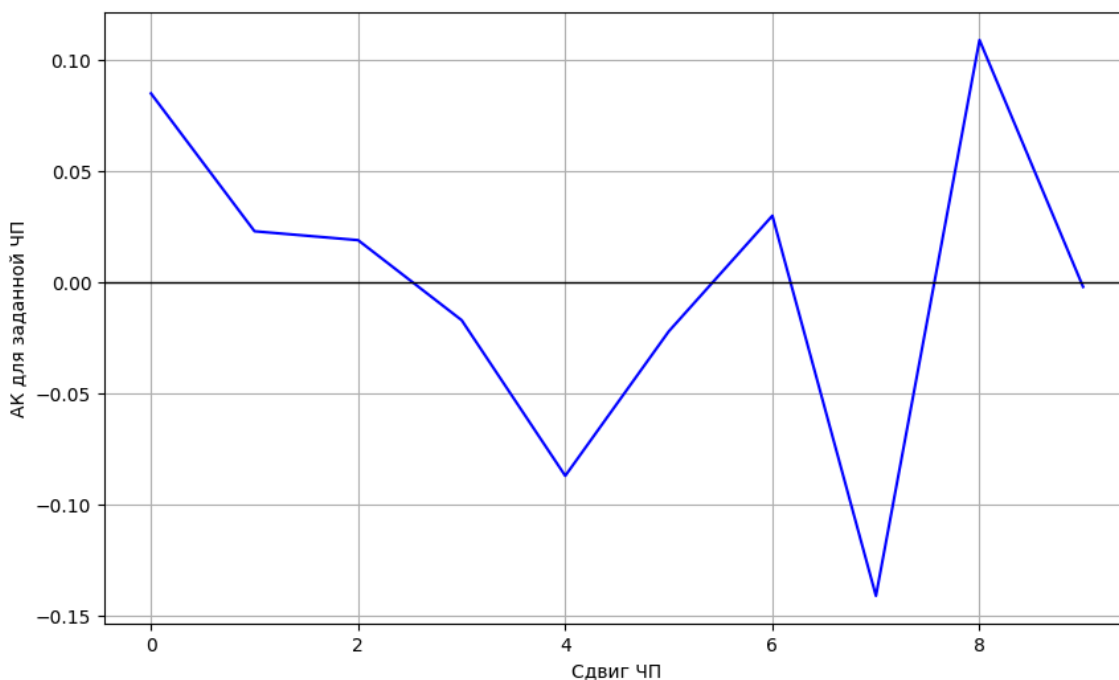
Также график показывает результаты прошлого этапа, а именно большой разброс значения.

3. Результаты автокорреляционного анализа (значения коэффициентов автокорреляции со сдвигом 1, 2, 3, ...), представленные как в числовом (форма 3), так и графическом виде, с обоснованным выводом о характере заданной числовой последовательности (можно ли ее считать случайной);

Коэффициенты автокорреляции для **заданной** последовательности

Сдвиг ЧП	1,0000	2,0000	3,0000	4,0000	5,0000	6,0000	7,0000	8,0000	9,0000	10,0000
К-т АК для задан. ЧП	0,0844	0,0229	0,0206	-0,0189	-0,0897	-0,0177	0,0305	-0,1351	0,0944	-0,0006

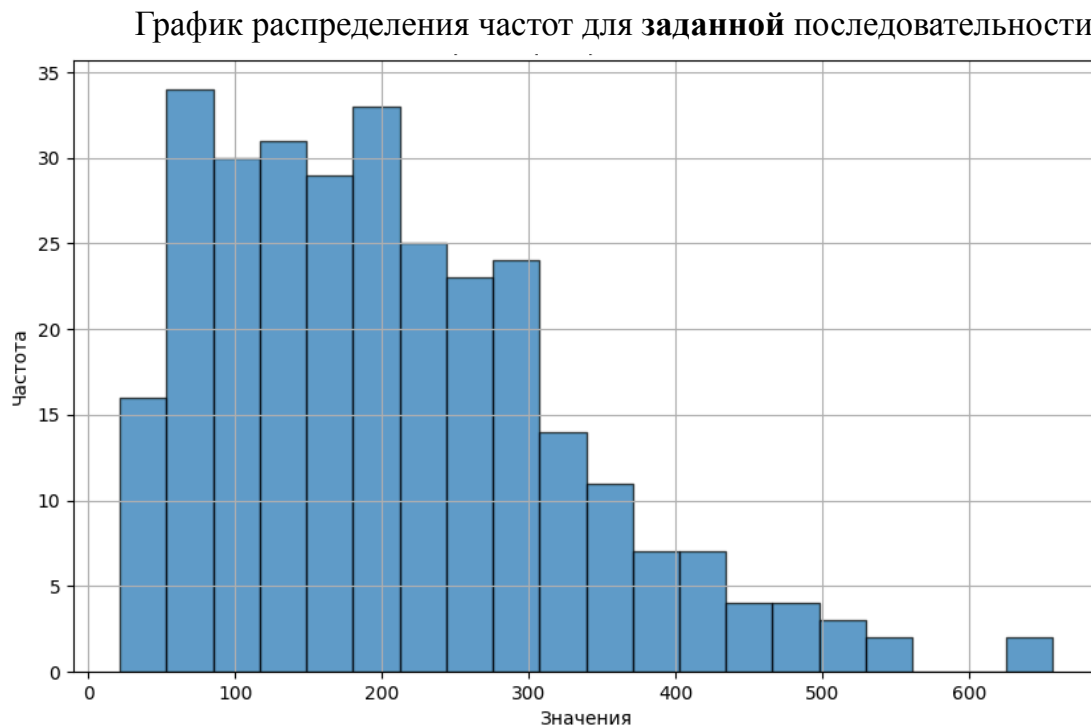
График зависимости к-тов АК для **заданной** последовательности от “Сдвига ЧП”



Выводы по пункту 3:

- Заданную последовательность можно считать случайной, так как коэффициенты автокорреляции не превышают 0.15

4. Гистограмма распределения частот для заданной числовой последовательности (график 2);



Выводы по пункту 4:

- Большинство значений сосредоточено в начале (от 0 до 300, с более редкими более крупными значениями в правой части (от 300 до 600))
- Значения в последовательности распределены на большом диапазоне, от 0 до более 600. Это объясняет высокие значения дисперсии и среднеквадратического отклонения.
- В правой части гистограммы, от 400 до 600, частота встречаемости значений резко падает, что подтверждает редкость высоких значений. Эти значения сильно влияют на среднее и дисперсию, увеличивая их, что также объясняет значительные доверительные интервалы и высокие значения среднеквадратического отклонения.

5. Параметры, рассчитанные по двум начальным моментам и определяющие вид аппроксимирующего закона распределения **заданной** случайной последовательности (*равномерный; экспоненциальный; нормированный Эрланга; гипоэкспоненциальный; гиперэкспоненциальный*);

Коэффициент вариации $v = \frac{\sigma}{t} = \frac{120,320}{206,063} = 0,584$, где σ – среднеквадратическое отклонение для заданной последовательности на 300 элементах, t — математическое ожидание.

Так как $v: 0 < 0,585 < 1$, то аппроксимирующим законом распределения заданной последовательности будет выбран **гипоэкспоненциальный** закон

Посчитаем необходимое количество фаз k :

$$k = \lceil \frac{1}{v^2} \rceil = \lceil \frac{1}{0,584^2} \rceil = \lceil 2,93 \rceil = 3$$

Выберем $k_1 = 1$, тогда $k_2 = 2$.

Рассчитаем t_1 и t_2 :

$$t_1 = \frac{t}{k} \left[1 + \sqrt{\frac{k_2}{k_1} (kv^2 - 1)} \right] = \frac{206,063}{3} \left[1 + \sqrt{\frac{2}{1} (3 \cdot 0,584^2 - 1)} \right] = 83,473$$

$$t_2 = \frac{t}{k} \left[1 - \sqrt{\frac{k_1}{k_2} (kv^2 - 1)} \right] = \frac{206,063}{3} \left[1 - \sqrt{\frac{1}{2} (3 \cdot 0,584^2 - 1)} \right] = 61,295$$

Гипоэкспоненциальное распределение моделируется как сумма нескольких независимых экспоненциальных распределений с разными параметрами λ . Параметры λ_k для каждой фазы можно найти через математическое ожидание t_k по формуле:

$$\lambda_k = \frac{1}{t_k}$$

$$\lambda_1 = \frac{1}{t_1} = \frac{1}{83,473} = 0,01198$$

$$\lambda_2 = \frac{1}{t_2} = \frac{1}{61,295} = 0,01631$$

6. Описание алгоритма (программы) формирования аппроксимирующего закона распределения и расчета значений всех числовых характеристик с иллюстрацией (при защите отчета) его работоспособности;

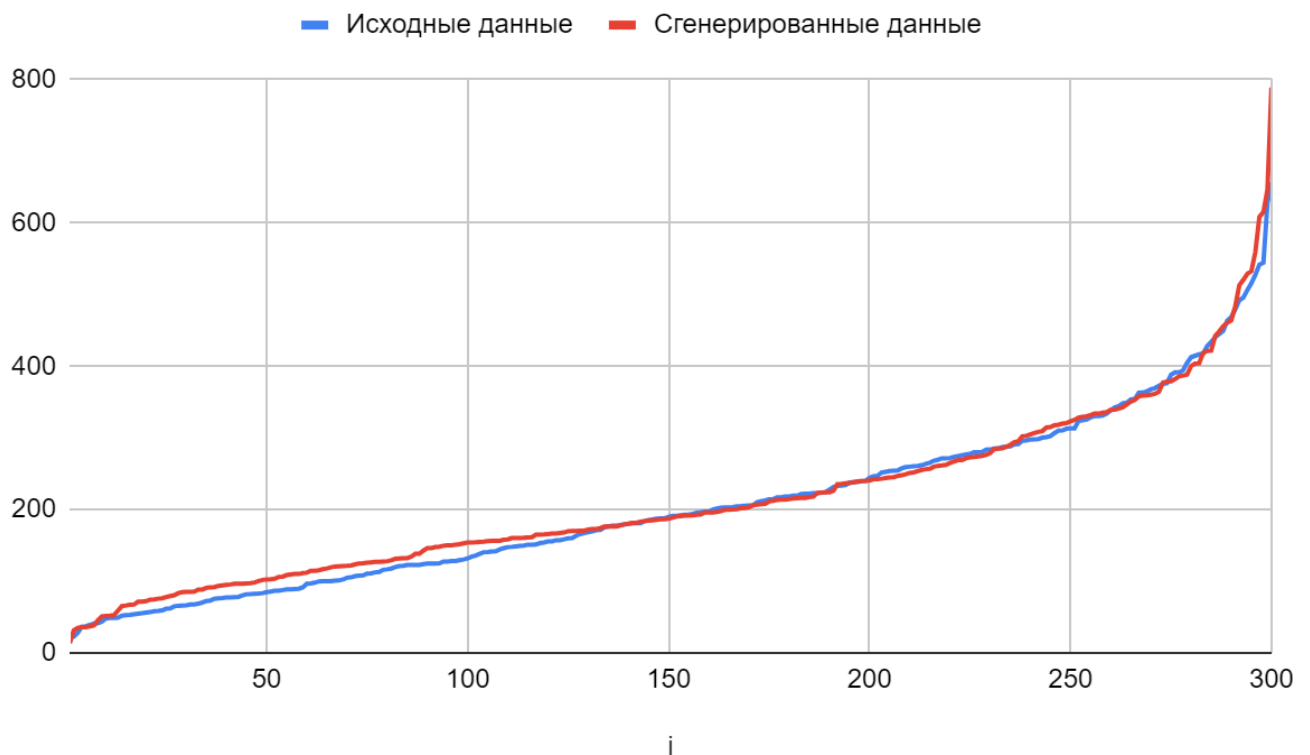
Описание алгоритма формирования аппроксимирующего закона распределения:

- В Excel генерируется по 300 случайных чисел y_i при помощи функции СЛЧИС() для 3 фаз
- Для каждого числа y_i вычисляется значение $x_{ij} = -\frac{1}{\lambda_j} \ln(1 - y_{ij})$, где
 - $\lambda_1 = 0,011833$ – для одной фазы
 - $\lambda_2 = 0,016469$ – для 2 оставшихся фаз
- Суммируется x_{ij} каждой фазы, получая x_i
- Полученные 300 значений x_i сформировывают последовательность, сгенерированную в соответствии с гипоекспоненциальным законом распределения

7. Выводы по результатам сравнения сгенерированной в соответствии с полученным аппроксимирующим законом распределения последовательности случайных величин и заданной числовой последовательности, а именно:

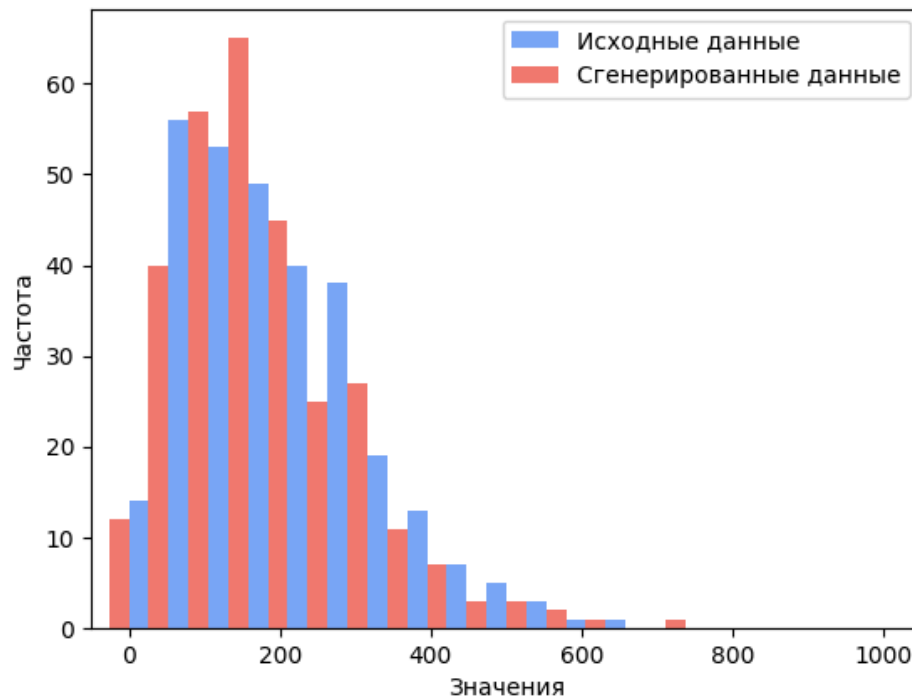
Сравнение значений полученных последовательностей

График значений исходной и сгенерированной последовательностей



Сравнение гистограмм распределения частот

Гистограммы исходной и сгенерированной последовательностей



Расчет числовых характеристик **сгенерированной** последовательности в соответствии с аппроксимирующим законом распределения случайной последовательности: математического ожидания, дисперсии, среднеквадратического отклонения, коэффициента вариации (представленные в таблице по форме 2) и коэффициентов автокорреляции при разных значениях сдвигов (в таблице по форме 3), а также сравнения (в %) полученных значений со значениями, рассчитанными для заданной числовой последовательности;

Характеристики сгенерированной случайной ЧП

Характеристика	Количество случайных величин						
		10	20	50	100	200	300
Мат. ож.	Знач	191,077	221,917	216,620	218,418	218,243	212,615
	%	16,30%	11,96%	6,15%	9,05%	7,33%	3,08%
Дов. инт. (0.9)	Знач	±54,209	±46,857	±25,774	±19,481	±14,247	±11,303
	%	18,28%	33,52%	12,98%	5,31%	1,33%	0,97%
Дов. инт. (0.95)	Знач	±64,668	±55,898	±30,747	±23,240	±16,996	±13,484
	%	18,28%	33,52%	12,98%	5,31%	1,33%	0,97%
Дов. инт. (0.99)	Знач	±84,992	±73,466	±40,410	±30,544	±22,337	±17,721
	%	18,28%	33,52%	12,98%	5,31%	1,33%	0,97%
Дисперсия	Знач	10885,945	16267,011	12304,567	14059,023	15038,593	14198,057
	%	33,22%	55,80%	24,28%	10,34%	2,64%	1,93%
С.к.о	Знач	104,336	127,542	110,926	118,571	122,632	119,156
	%	18,28%	33,52%	12,98%	5,31%	1,33%	0,97%
К-т вариации	Знач	0,546	0,575	0,512	0,543	0,562	0,560
	%	2,37%	24,48%	7,28%	13,88%	6,08%	4,02%

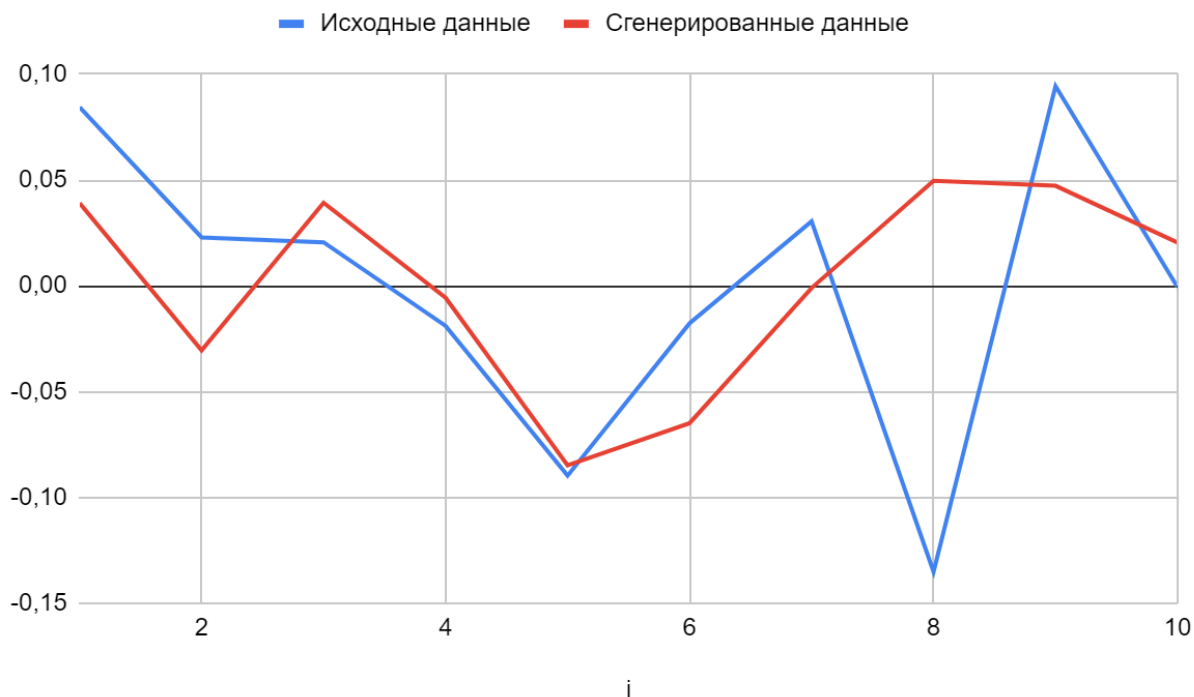
% - относительные отклонения характеристик **сгенерированной** случайной последовательности от одноименных значений **заданной** числовой последовательности

Проведение корреляционного анализа сгенерированной в соответствии с аппроксимирующим законом распределения последовательности случайных величин и заданной числовой последовательности на основе коэффициента корреляции.

Коэффициенты автокорреляции:

Сдвиг ЧП	1,0000	2,0000	3,0000	4,0000	5,0000	6,0000	7,0000	8,0000	9,0000	10,0000
К-т АК для задан. ЧП	0,0844	0,0229	0,0206	-0,0189	-0,0897	-0,0177	0,0305	-0,1351	0,0944	-0,0006
К-т АК для сгенерир. ЧП	0,0391	-0,0305	0,0393	-0,0056	-0,0848	-0,0649	-0,0010	0,0497	0,0473	0,0205
%	53,67%	233,19%	47,58%	70,37%	5,46%	72,73%	103,28%	136,79%	49,89%	3516,67%

График зависимости к-тов АК для исходной и сгенерированной последовательностей от
“Сдвига ЧП”



Коэффициент корреляции между числовыми последовательностями

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = -0,056$$

Так как мы получили $r = -0,056$, то между исходной и сгенерированной последовательностями присутствует крайне малая корреляция.

Выводы по пункту 7:

- Аппроксимация исходной последовательности гипоекспоненциальным распределением оказалась достаточно успешной для больших объемов выборки (100, 200, и 300 элементов). Статистические показатели при таких значениях, такие как мат. ожидание, дисперсия и среднеквадратичное отклонение, близки к исходной последовательности, с отклонениями менее 5–10%.
- На малых выборках (10 и 20 элементов) аппроксимация показывает более высокие дисперсию и доверительные интервалы. Это связано с тем, что гипоекспоненциальное распределение имеет значительную вариативность, которая усиливается на малых объемах данных.

- Низкое значение коэффициента корреляции указывает на то, что практически нет зависимости между сгенерированной и заданной последовательностями.
- Коэффициенты автокорреляции у исходной и у сгенерированной последовательностей близки к 0, это указывает на их случайность.

Выводы

Я провела статистический анализ числовой последовательности, оценила ее числовые моменты и определила основные свойства с помощью корреляционного анализа.

Для аппроксимации закона распределения заданной последовательности я использовала гипоекспоненциальное распределение, так как коэффициент вариации оказался больше нуля, но меньше единицы. Я рассчитала количество фаз k и математическое ожидание t , на основе которых определила параметры λ .

Коэффициент автокорреляции сгенерированной последовательности близок к нулю, что подтверждает её случайность. Небольшие отклонения числовых моментов от исходных значений говорят о правильном выборе распределения и корректной аппроксимации.