

Федеральное государственное автономное образовательное учреждение высшего  
образования

«Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

**Учебно-исследовательская работа №1 (УИР1)**

**“ Обработка результатов измерений: статистический анализ  
числовой последовательности”**

по дисциплине «Моделирование»

Вариант: 15

**Выполнил:**

Векшин Арсений Р3316

**Преподаватель:**

Алиев Т.И

## 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.

## Рассчитать значения числовых моментов заданной числовой последовательности

Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. Ожидание	знач	11,020	14,577	24,193	25,751	25,563	25,946
	%	-57,53%	-43,82%	-6,76%	-0,75%	-1,47%	
Дисперсия	знач	308,524	364,398	1075,192	1489,086	1413,830	1480,35
	%	-79,16%	-75,38%	-27,37%	0,59%	-4,49%	
СКО	знач	17,565	19,089	32,790	38,589	37,601	38,475
	%	-54,35%	-50,39%	-14,78%	0,29%	-2,27%	
Коэффициент вариации	знач	1,594	1,310	1,355	1,499	1,471	1,483
	%	7,48%	-11,69%	-8,60%	1,05%	-0,81%	
Доверительный интервал (0.9)	знач	±10,182	±7,381	±7,775	±6,407	±4,394	±3,665
	%	177,80%	101,38%	112,12%	74,81%	19,88%	
Доверительный интервал (0.95)	знач	±12,565	±8,934	±9,319	±7,657	±5,243	±4,371
	%	187,44%	104,37%	113,18%	75,15%	19,94%	
Доверительный интервал (0.99)	знач	±18,051	±12,212	±12,428	±10,133	±6,915	±5,758
	%	213,47%	112,07%	115,81%	75,97%	20,08%	

- **Математическое ожидание**

При малом количестве случайных величин наблюдается значительное смещение оценок математического ожидания относительно его истинного значения, что приводит к большим относительным отклонениям. Это указывает на недостаточную точность оценки при малых выборках.

По мере увеличения выборки математическое ожидание стабилизируется и отклонения становятся минимальными. Это свидетельствует о том, что для получения достоверной оценки среднего значения требуется выборка не менее 100 элементов.

- **Дисперсия**

Дисперсия ведет себя схоже с математическим ожиданием, сильно расходясь на малом размере выборки, стабилизируясь по мере ее увеличения.

- **СКО**

Поведение схоже с математическим ожиданием и дисперсией.

- **Коэффициент вариации**

Коэффициент вариации, измеряющий относительный разброс данных, на малых выборках отклоняется от предполагаемого значения, указывая на изменчивость данных. По мере увеличения выборки он стабилизируется около 1,5, что говорит о более предсказуемом и точном поведении выборки на больших объемах данных

- **Доверительные интервалы**

Для малых выборок доверительные интервалы практически соответствуют ожидаемым, при увеличении числа случайных величин они расширяются, а затем опять сужаются. Это может указывать на особенность распределения данных. Возможно, с увеличением выборки начинают проявляться пограничные значения, что приводит к повышению разброса значения и росту доверительных интервалов.

Все рассмотренные уровни доверия ведут себя схожим образом.

**Построить график значений для заданной числовой последовательности и определить ее характер**



- Последовательность демонстрирует неравномерное поведение, не имея явной тенденции к росту или снижению. Наблюдаются несколько значительных пиков, намекающих на возможность повторяющихся событий, но их неравномерное распределение делает периодичность данных маловероятной
- Большая часть значений находится в нижнем диапазоне, что позволяет предположить, что последовательность преимущественно состоит из небольших значений.

**Выполнить автокорреляционный анализ и определить, можно ли считать заданную числовую последовательность случайной.**

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. ЧП	-0,013	0,007	-0,097	-0,095	-0,011	0,005	0,000	0,000	0,085	0,023



Коэффициенты автокорреляции для всех сдвигов находятся в пределах  $\pm 0.1$ , что указывает на отсутствие выраженной автокорреляции. Сдвиги не оказывают значительного влияния на следующие значения в последовательности, а значит, заданная числовая последовательность является случайной

**Построить гистограмму распределения частот для заданной числовой последовательности.**



На полученной гистограмме заметна сильная асимметрия, большая часть значений сосредоточена в первом интервале. По мере увеличения значений ЧП частота резко снижается. Большинство

интервалов в диапазоне [70,794, 273,304] имеют очень низкую частоту. Данные замечания свидетельствуют о редких высоких значениях и доминировании показателей, близких к минимальным.

**Выполнить аппроксимацию закона распределения заданной случайной последовательности по двум начальным моментам, используя, в зависимости от значения коэффициента вариации, одно из распределений.**

Так как распределение имеет коэффициент вариации больше 1, то для аппроксимации следует использовать гиперэкспоненциальное распределение.

Для аппроксимации выберем значение вероятности  $q$ :  $q \leq \frac{2}{1+v^2} = \frac{2}{1+1.483^2} \approx 0,62$

Рассчитаем значения  $t_1$  и  $t_2$

$$t_1 = \left( 1 + \sqrt{\frac{1-q}{2q}}(v^2 - 1) \right) * t = \left( 1 + \sqrt{\frac{1-0,62}{2 * 0,62}}(1.483^2 - 1) \right) * 25,946 \approx 41.67544$$
$$t_2 = \left( 1 - \sqrt{\frac{q}{2(1-q)}}(v^2 - 1) \right) * t = \left( 1 - \sqrt{\frac{0,62}{2(1-0,62)}}(1.483^2 - 1) \right) * 25,946 \approx 0.2821$$

**Реализовать генератор случайных величин в соответствии с полученным аппроксимирующим законом распределения (в EXCEL или программно) и проиллюстрировать на защите его работу**

```
import numpy as np
```

```
# из расчетов
```

```
size = 300
```

```
v = 1.483
```

```
t = 25.946
```

```
q = 0.62
```

```
t1 = (1 + np.sqrt((1 - q) / (2 * q) * (v ** 2 - 1))) * t
```

```
t2 = (1 - np.sqrt(q / (2 * (1 - q)) * (v ** 2 - 1))) * t
```

```
print("t1 = ", t1, "t2 = ", t2)
```

```
type = [1 / t1, 1 / t2] # интенсивность
```

```
prob = [q, 1 - q] # вероятность
```

```
for _ in range(size):
```

```
    chosen = np.random.choice(2, p=prob)
```

```
    print('{0:10f}'.format(np.random.exponential(1 / type[chosen])).replace('.', ','))
```

Описание алгоритма:

- 1) Расчет  $t_1$  и  $t_2$  по заданным параметрам
- 2) Выбираем компоненту распределения при помощи `np.random.choice`, на основе ранее рассчитанных вероятностей
- 3) На основе выбранной компоненты генерируется искомая экспоненциальная случайная величина (`np.random.exponential`)

**Сгенерировать последовательность случайных величин с использованием реализованного генератора и рассчитать значения числовых моментов по аналогии с заданной числовой последовательностью**

Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. Ожидание	знач	25,650	31,662	27,223	26,641	28,053	<b>28,232</b>
	%	-9,14%	12,15%	-3,57%	-5,63%	-0,63%	
Дисперсия	знач	368,829	1233,438	1706,104	1552,704	1631,780	<b>1645,749</b>
	%	-77,59%	-25,05%	3,67%	-5,65%	-0,85%	
СКО	знач	19,205	35,120	41,305	39,404	40,395	<b>40,568</b>
	%	-52,66%	-13,43%	1,82%	-2,87%	-0,43%	
Коэффициент вариации	знач	0,749	1,109	1,517	1,479	1,440	<b>1,437</b>
	%	-47,89%	-22,81%	5,59%	2,93%	0,21%	
Доверительный интервал (0.9)	знач	11,133	13,579	9,793	6,543	4,720	<b>3,865</b>
	%	188,07%	251,38%	153,42%	69,30%	22,14%	
Доверительный интервал (0.95)	знач	13,739	16,437	11,739	7,819	5,633	<b>4,609</b>
	%	198,07%	256,61%	154,68%	69,63%	22,21%	
Доверительный интервал (0.99)	знач	19,736	22,467	15,655	10,347	7,429	<b>6,072</b>
	%	225,06%	270,04%	157,83%	70,42%	22,35%	

- **Математическое ожидание**

При малом количестве случайных величин наблюдается значительное смещение оценок математического ожидания относительно его истинного значения, что приводит к большим относительным отклонениям. Это указывает на недостаточную точность оценки при малых выборках.

На выборке в 50 элементов наблюдается аномально малое отклонение, что можно объяснить особенностями выборки. По мере увеличения размера до «среднего» наблюдается ожидаемый рост неточности, вероятно из-за содержащегося в данных «шума».

При приближении выборки к максимальному размеру заметна стабилизация значения, с приближением к среднему

- **Дисперсия**

При малом количестве случайных величин наблюдается значительное смещение оценок дисперсии относительно его истинного значения, что приводит к большим относительным отклонениям. Это указывает на недостаточную точность оценки при малых выборках. По мере увеличения выборки дисперсия стабилизируется и отклонения становятся минимальными.

- **СКО**

Поведение схоже с дисперсией.

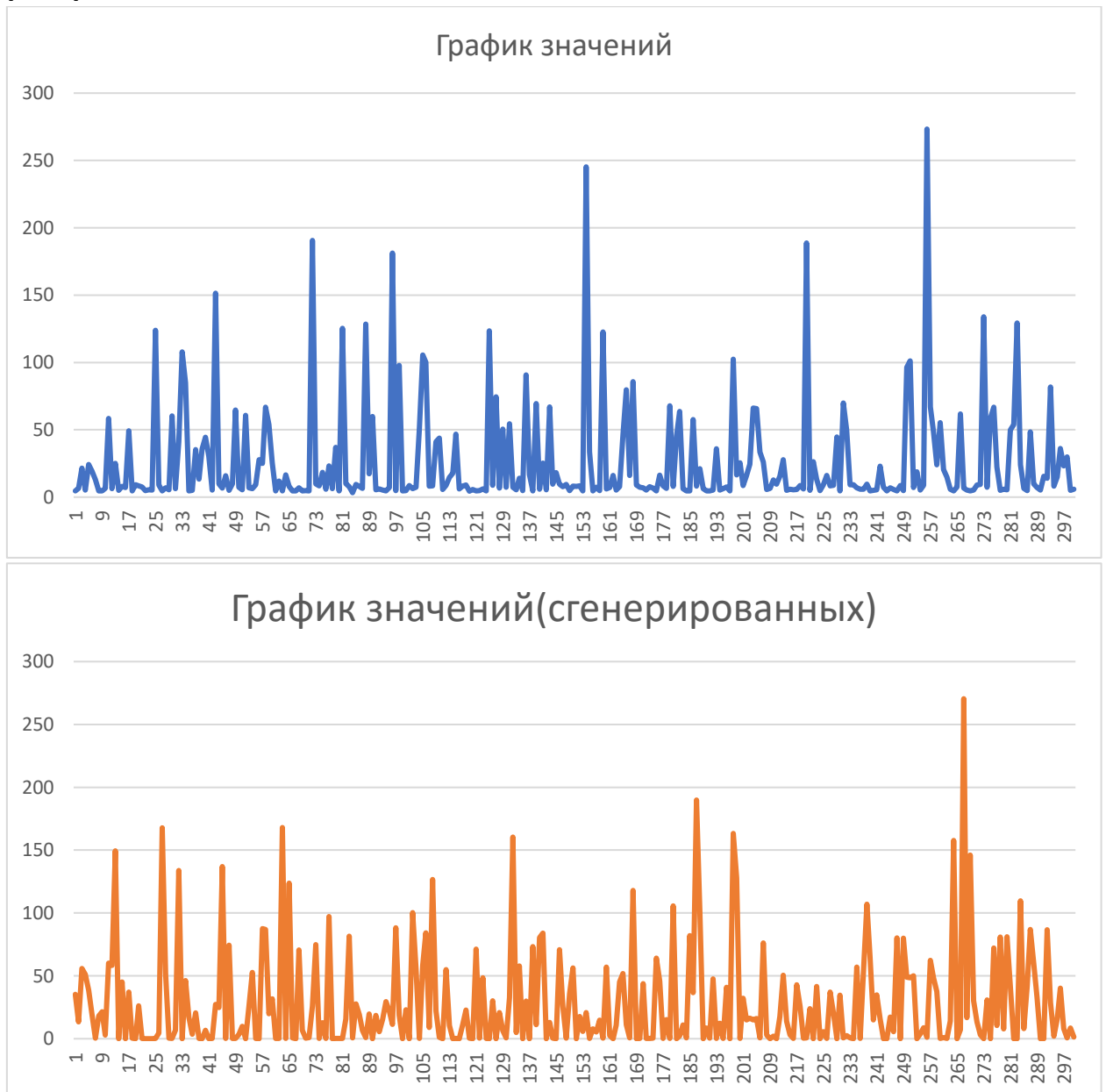
- **Коэффициент вариации**

Коэффициент вариации, измеряющий относительный разброс данных, на малых выборках отклоняется от предполагаемого значения, указывая на изменчивость данных. По мере увеличения выборки он стабилизируется около 1,5, что говорит о более предсказуемом и точном поведении выборки на больших объемах данных

- **Доверительные интервалы**

Для малого размера выборки доверительные интервалы значительно расширяются (225,06%), что говорит о высокой неопределенности в исходных данных. По мере увеличения размера выборки интервалы постепенно приближаются к расчетным.

**Выполнить сравнительный анализ сгенерированной последовательности случайных величин с заданной последовательностью, построив соответствующие зависимости на графике значений и гистограмме распределения частот.**



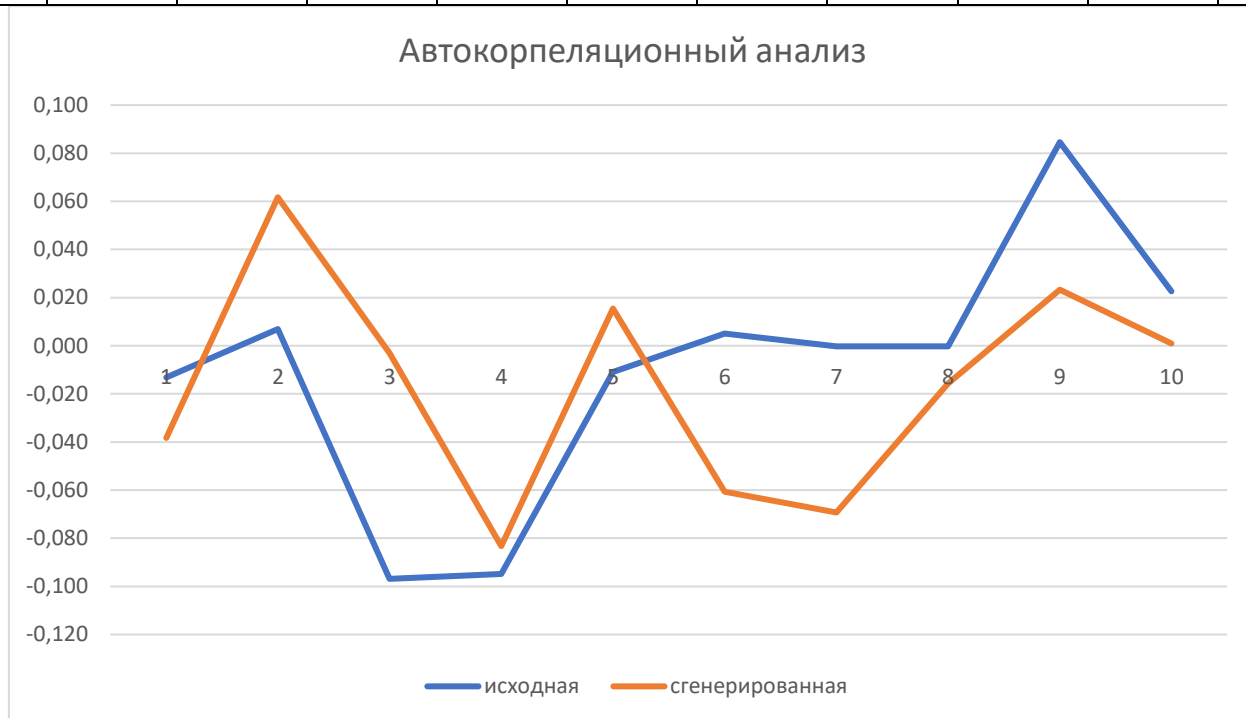


Судя по графикам, сгенерированная последовательность близка к исходной. Несмотря на расхождения в «пиках», сгенерированная последовательность достаточно точно воспроизводит динамику исходной.



## Выполнить автокорреляционный анализ сгенерированной последовательности случайных величин.

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. ЧП	-0,014	0,015	0,161	-0,084	-0,024	0,127	0,062	-0,053	0,111	-0,139
К-т АК для сген. ЧП	-0,038	0,062	-0,003	-0,083	0,015	-0,061	-0,069	-0,016	0,023	0,001
%	173,29%	311,27%	-101,87%	-0,87%	-164,25%	-147,74%	-211,73%	-70,35%	-79,04%	-100,64%



Коэффициенты близки к нулю, значит сгенерированную выборку можно считать случайной. При этом коэффициенты корреляции сгенерированной выборки отличаются от исходной, однако «повторяют рывки» за исходной выборкой, что указывает на, возможно, схожую с исходной структуру корреляции.

## Вывод после выполнения работы

В ходе выполнения данной работы были проведены оценки математического ожидания, дисперсии, среднеквадратического отклонения и коэффициента вариации заданной числовой последовательности. Также были рассчитаны доверительные интервалы для оценки математического ожидания с доверительными вероятностями 0.9, 0.95 и 0.99. Анализ характера числовой последовательности и автокорреляционный анализ позволили сделать вывод о её случайности. Построенная гистограмма распределения частот и определенные параметры аппроксимирующего закона распределения помогли подобрать соответствующую модель. Был описан алгоритм формирования аппроксимирующего закона распределения и продемонстрирована его работоспособность, а также проведено сравнение сгенерированной последовательности с заданной и вычислен коэффициент корреляции для дальнейшего анализа.