

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Санкт-Петербургский национальный исследовательский
университет информационных технологий, механики и оптики»

ФАКУЛЬТЕТ ПРОГРАММНОЙ ИНЖЕНЕРИИ И КОМПЬЮТЕРНЫХ
ТЕХНОЛОГИЙ

Учебно-исследовательская работа №1

Обработка результатов измерений: статистический анализ
числовой последовательности
по дисциплине
«Моделирование»

Вариант № 182

Выполнил:

Хасаншин Марат Айратович Р3333

Шикунов Максим Евгеньевич Р3333

Преподаватель:

Тропченко А.А.

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 работы:

1)Рассчитать значения числовых моментов заданной числовой последовательности

Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. Ожидание	знач	97,740	94,881	122,523	140,807	136,378	125,716
	%	-22,25%	-24,53%	-2,54%	12,00%	8,48%	
Дисперсия	знач	12121,327	14816,539	14070,042	20298,360	18576,773	16597,371
	%	-26,97%	-10,73%	-15,23%	22,30%	11,93%	
СКО	знач	110,097	121,723	118,617	142,472	136,297	128,831
	%	-14,54%	-5,52%	-7,93%	10,59%	5,80%	
Коэффициент вариации	знач	1,126	1,283	0,968	1,012	0,999	1,025
	%	9,92%	25,19%	-5,53%	-1,26%	-2,48%	
Доверительный интервал (0.9)	знач	63,821	47,064	28,124	23,656	15,926	±12,272
	%	420,03%	283,49%	129,16%	92,76%	29,77%	
Доверительный интервал (0.95)	знач	78,760	56,968	33,711	28,269	19,005	±14,637
	%	438,08%	289,20%	130,31%	93,13%	29,84%	
Доверительный интервал (0.99)	знач	113,144	77,869	44,956	37,412	25,065	±19,282
	%	486,80%	303,85%	133,16%	94,03%	29,99%	

Вывод:

- Математическое ожидание**
По мере увеличения количества случайных величин, математическое ожидание (среднее значение) вначале немного отклоняется от эталонного значения, затем стремится к более стабильному результату, особенно при 100 и 200 величинах. Для малых выборок (10 и 20 случайных величин) отклонение от эталона составляет -22.25% и -24.53%, что говорит о высоком уровне нестабильности из-за недостаточного количества данных.
При больших выборках (200 и 300 величин) среднее значение немного снижается, что может говорить о стабилизации с уменьшением отклонений и более надежной оценке среднего.
- Дисперсия**
Дисперсия характеризует разброс данных относительно среднего значения. Начиная с выборки в 50 величин, отклонение снижается, и дисперсия начинает приближаться к более стабильным значениям. При 100 и 200 случайных величинах дисперсия отклоняется

на +22.30% и +11.93%, что указывает на более устойчивый результат. Данные результаты немного отличаются от эталонного, это указывает, что значения в выборке сильно колеблются относительно среднего значения.

- **СКО**

СКО показывает среднее отклонение величин от среднего значения. Для малых выборок (10 величин) отклонение -14.54%, что говорит о неустойчивом поведении выборки и её высокой изменчивости. По мере увеличения выборки (до 50 величин и более) среднеквадратичное отклонение постепенно стабилизируется, а отклонения снижаются до -5.52% и -7.93%. Это указывает на то, что числа в выборке не сильно отклоняются относительно среднего значения

- **Коэффициент вариации**

Коэффициент вариации показывает степень относительной изменчивости данных. При малых выборках наблюдается высокая степень изменчивости (+9.92% и +25.19%), что говорит о том, что выборки неустойчивы и демонстрируют значительные отклонения. По мере увеличения выборки отклонения уменьшаются и даже становятся отрицательными (-5.53% при 50 величинах и -2.48% при 300 величинах), что свидетельствует о снижении вариативности и большей устойчивости данных.

- **Доверительные интервалы**

Доверительный интервал показывает диапазон, в котором, с заданной вероятностью, находятся истинные значения всей выборки. При малых выборках видно, как сильно влияют случайные большие числа на данный доверительный интервал. С увеличением выборки доверительные интервалы сужаются, что свидетельствует о снижении неопределенности.

2) Построить график значений для заданной числовой последовательности и определить ее характер



Вывод:

- 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.

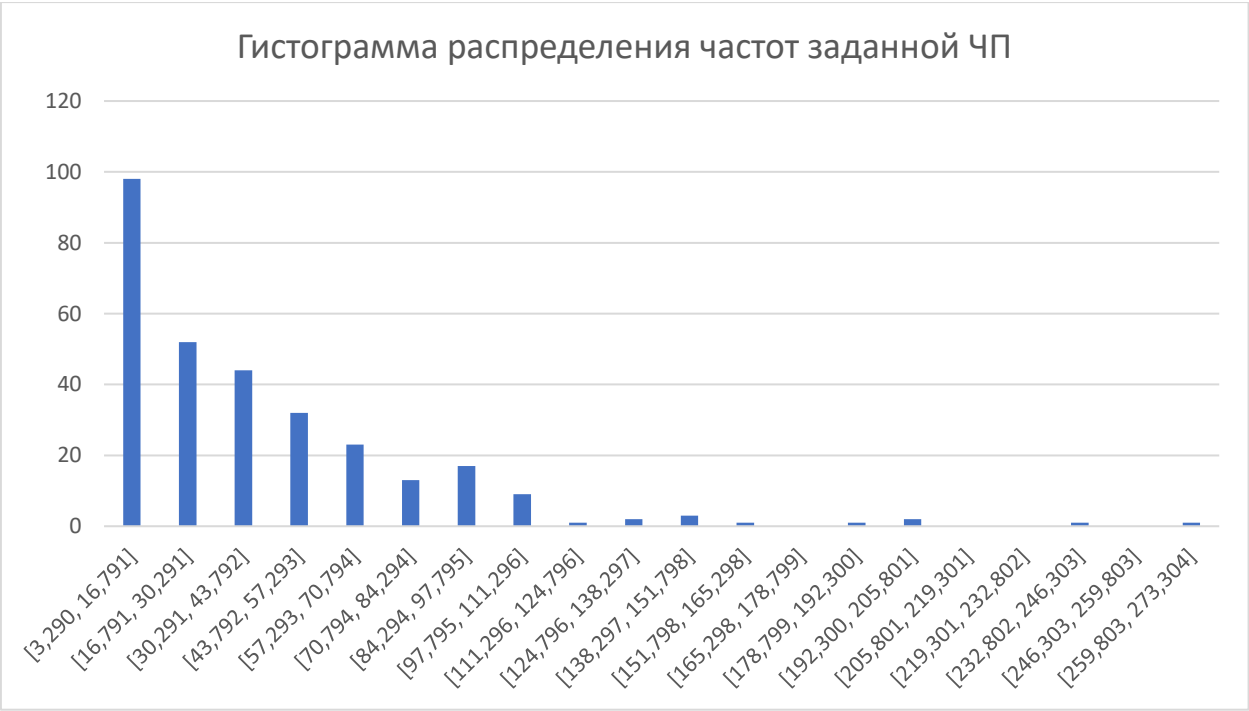
3) Выполнить автокорреляционный анализ и определить, можно ли заданную числовую последовательность считать случайной.

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. ЧП	0,062	0,061	0,116	0,055	0,006	-0,053	0,085	0,010	0,056	0,050



Коэффициенты автокорреляции для всех сдвигов находятся в пределах ± 0.12 , что указывает на отсутствие выраженной автокорреляции. Сдвиги не оказывают значительного влияния на следующие значения в последовательности, а значит, заданная числовая последовательность является случайной

4) 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 [111,296, 273,304] имеют очень низкую частоту. Данные замечания свидетельствуют о редких высоких значениях и доминировании показателей, близких к минимальным.

5) 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 : $q \leq \frac{2}{1+v^2} = \frac{2}{1+1.025^2} \approx 0,975312$

Рассчитаем значения t_1 и t_2

$$t_1 = \left(1 + \sqrt{\frac{1-q}{2q}(v^2-1)} \right) * t = \left(1 + \sqrt{\frac{1-0,975312}{2 * 0,975312}(1.025^2-1)} \right) * 125,716 \approx 128.8978$$

$$t_2 = \left(1 - \sqrt{\frac{q}{2(1-q)}(v^2-1)} \right) * t = \left(1 - \sqrt{\frac{0,975312}{2(1-0,975312)}(1.025^2-1)} \right) * 125,716 \approx 10^{-3}$$

6) 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 = 1.025
t = 125.716
q = 0.975312

t1 = 128.8978
t2 = 10**(-3)

type = [1/t1, 1/t2]
prob = [q, 1-q]

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) Выбираем компоненту распределения при помощи `np.random.choice`, на основе ранее рассчитанных вероятностей
- 2) На основе выбранной компоненты генерируется искомая экспоненциальная случайная величина (`np.random.exponential`)

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

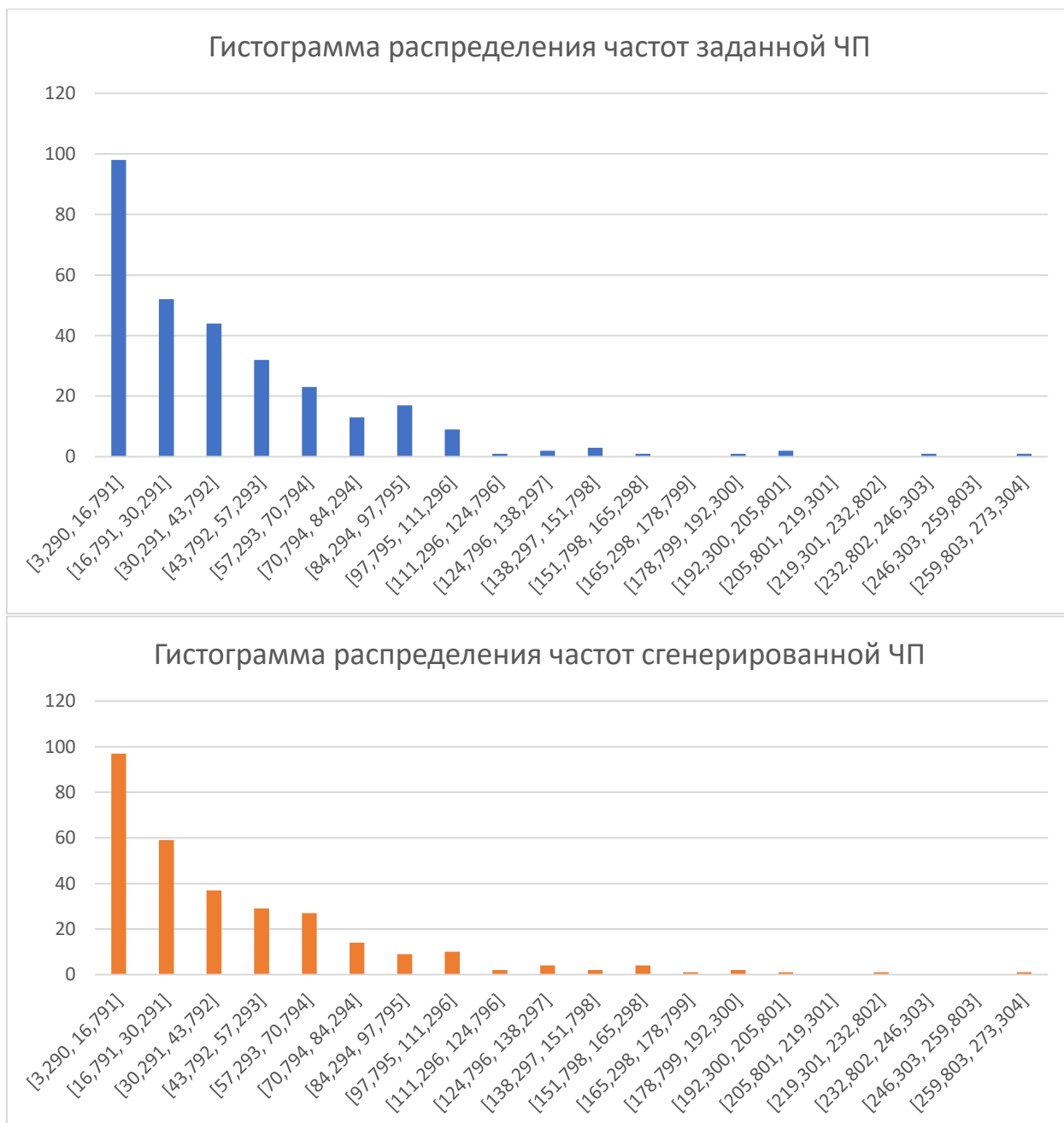
Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. Ожидание	знач	181,703	151,538	121,447	137,121	132,190	129,002
	%	40,85%	17,47%	-5,86%	6,29%	2,47%	
Дисперсия	знач	51376,553	27714,444	15213,441	18390,197	20021,099	18497,183
	%	177,75%	49,83%	-17,75%	-0,58%	8,24%	
СКО	знач	226,664	166,477	123,343	135,610	141,496	136,004
	%	66,66%	22,41%	-9,31%	-0,29%	4,04%	
Коэффициент вариации	знач	1,247	1,099	1,016	0,989	1,070	1,054
	%	18,32%	4,20%	-3,67%	-6,19%	1,53%	
Доверительный интервал (0.9)	знач	131,392	64,367	29,245	22,517	16,534	12,956
	%	914,15%	396,82%	125,72%	73,80%	27,62%	
Доверительный интервал (0.95)	знач	162,149	77,913	35,054	26,908	19,730	15,452
	%	949,34%	404,22%	126,85%	74,13%	27,68%	
Доверительный интервал (0.99)	знач	232,937	106,499	46,747	35,610	26,021	20,355
	%	1044,36%	423,20%	129,66%	74,94%	27,84%	

Вывод:

- Математическое ожидание**
 При малых выборках происходят сильные колебания, что может указывать на нестабильность оценок для небольшого количества чисел. Математическое ожидание приближается к эталонному значению при увеличении объема выборки.
- Дисперсия**
 Это говорит о том, что выборочные оценки становятся более точными по мере увеличения объема выборки, а вариативность данных уменьшается. Однако после 100 величин дисперсия стабилизируется в диапазоне 18–20 тысяч.
- СКО**
 Стандартное отклонение также уменьшается с увеличением объема выборки, что подтверждает более точные оценки для больших выборок. Для выборок от 100 чисел значение ско примерно равно эталонному, это означает, что данная случайная выборка довольно стабильная и не сильно отклоняется относительно среднего значения
- Коэффициент вариации**
 Коэффициент вариации, который отражает степень разброса данных относительно среднего, уменьшается с увеличением объема выборки. Это означает, что данные выборки становятся менее разбросанными при увеличении объема.
- Доверительные интервалы**
 По доверительным интервалам можно сделать вывод, что в выборке есть сильно большие или маленькие числа, которые сильно влияют на доверительные интервалы при выборке с малым количеством чисел. Но с увеличением выборки мы получаем более точные истинные значения

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

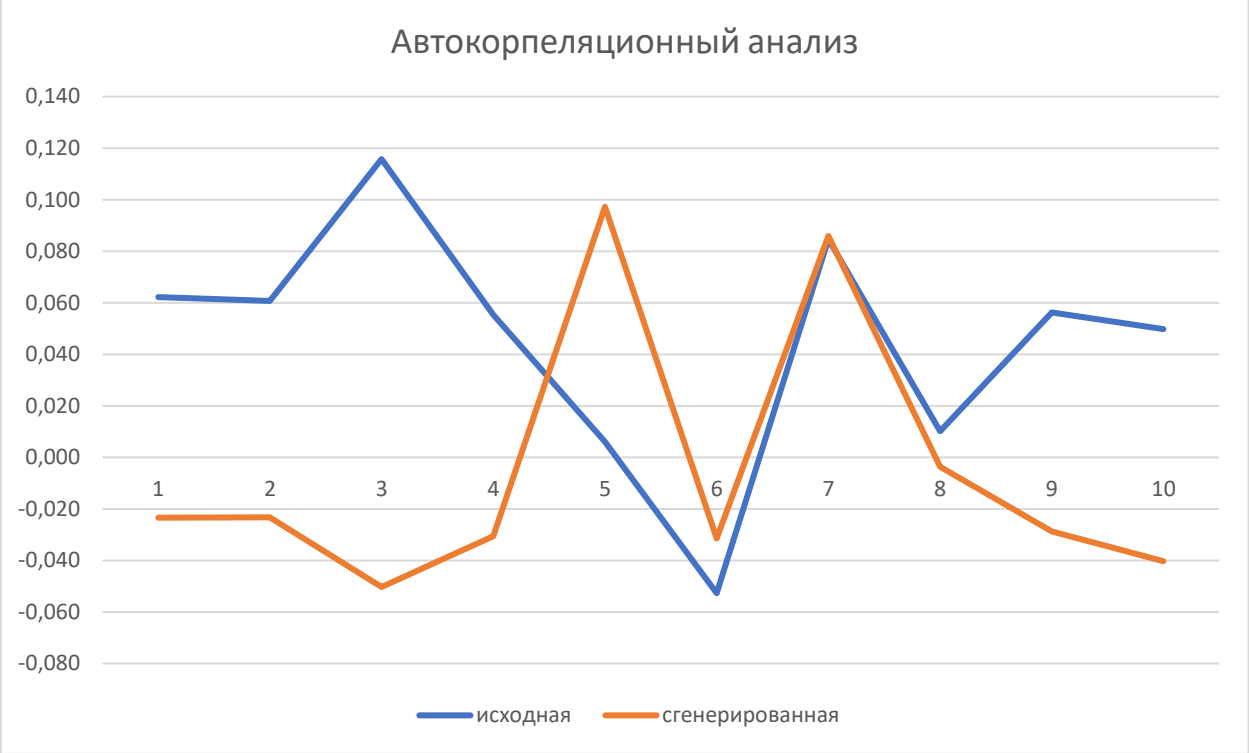




Судя по графикам, сгенерированная последовательность близка к исходной. Несмотря на расхождения в «пиках», сгенерированная последовательность достаточно точно воспроизводит динамику исходной.

9) 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	0,062	0,061	0,116	0,055	0,006	-0,053	0,085	0,010	0,056	0,050
К-т АК для сген. ЧП	-0,023	-0,023	-0,050	-0,031	0,097	-0,031	0,086	-0,004	-0,029	-0,040
%	-137,63%	-138,38%	-143,45%	-155,23%	1514,96%	-40,28%	1,50%	-135,45%	-150,97%	-180,82%



Коэффициент близки к нулю, значит сгенерированную выборку можно считать случайной. При этом, коэффициенты корреляции сгенерированной выборки сильно отличаются от исходной, что указывает на другую структуру корреляции по сравнению с исходной.

10) Оценить корреляционную зависимость сгенерированной и заданной последовательностей случайных величин.

$$r_{XY} = \frac{cov_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n (x_i - M[X])(y_i - M[Y])}{\sqrt{\sum_{i=1}^n (x_i - M[X])^2 \sum_{i=1}^n (y_i - M[Y])^2}} = 0,01906$$

Коэффициент корреляции близок к нулю, что указывает на очень слабую положительную зависимость между сгенерированной и заданной последовательностями. Гиперэкспоненциальное распределение моделирует вероятностную структуру данных, но это не гарантирует, что каждая точка в сгенерированной последовательности будет точно соответствовать конкретной точке исходной выборки. Так, можем сказать, что отдельные сгенерированные значения случайны, и корреляция является слабой.

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.