

Федеральное государственное автономное образовательное
учреждение высшего образования
«Национальный исследовательский университет ИТМО»
Факультет программной инженерии и компьютерной техники
Направление подготовки 09.03.04 «Программная инженерия» –
Системное и прикладное программное обеспечение

Отчёт

По лабораторной работе №1

**«Обработка результатов измерений:
статистический анализ числовой последовательности»**

По моделированию

Вариант: 19

Выполнил:

студент 3 курса

Саранча Павел Александрович

Группа: P3309

Принял:

Алиев Тауфик Измайлович

Отчёт принят «__»____2025 г.

Оценка: _____

г. Санкт-Петербург, 2025

Задание

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.

Содержание отчета

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);
2. график (график 1) значений заданной числовой последовательности с результатами анализа характера числовой последовательности (возрастающая, убывающая, периодичная и т.п.);
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);
4. гистограмма распределения частот для заданной числовой последовательности (график 2);
5. параметры, рассчитанные по двум начальным моментам и определяющие вид аппроксимирующего закона распределения заданной случайной последовательности (равномерный; экспоненциальный; нормированный Эрланга; гипоекспоненциальный; гиперэкспоненциальный);
6. описание алгоритма (программы) формирования аппроксимирующего закона распределения и расчета значений всех числовых характеристик с иллюстрацией (при защите отчета) его работоспособности;
7. выводы по результатам сравнения сгенерированной в соответствии с полученным аппроксимирующим законом распределения последовательности случайных величин и заданной числовой последовательности, а именно:
 1. сравнения плотности распределения аппроксимирующего закона с гистограммой распределения частот для исходной числовой последовательности (график 3);

2. расчета числовых характеристик *сгенерированной* 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;
 3. проведения *корреляционного анализа* сгенерированной в соответствии с аппроксимирующим законом распределения последовательности случайных величин и заданной числовой последовательности на основе *коэффициента корреляции*.
8. *по каждому из перечисленных выше пунктов отчета должны быть сформулированы результативные выводы и заключения.*

Ход работы

Этап 1. Форма №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.

Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. ож.	Знач.	125.718	129.050	101.401	104.779	99.451	104.923
	%	19.820	22.995	-3.356	-0.137	-5.215	
Дов. инт. (0,9)	Знач.	±68.833	±46.126	±22.139	±15.256	±10.510	±9.500
	%	624.568	385.538	133.046	60.588	10.635	
Дов. инт. (0,95)	Знач.	±84.944	±55.833	±26.537	±18.231	±12.542	±11.331
	%	649.684	392.759	134.205	60.900	10.688	
Дов. инт. (0,99)	Знач.	±122.031	±76.317	±35.389	±24.131	±16.541	±14.926
	%	717.578	411.306	137.099	61.675	10.819	
Дисперсия	Знач.	14099.940	14231.729	8718.834	8441.928	8089.901	9945.086
	%	41.778	43.103	-12.330	-15.115	-18.654	
С. к. о.	Знач.	118.743	119.297	93.375	91.880	89.944	99.725
	%	19.071	19.626	-6.368	-7.867	-9.808	
К-т вариации	Знач.	0.945	0.924	0.921	0.877	0.904	0.950
	%	-0.625	-2.739	-3.116	-7.740	-4.846	

% – относительные отклонения полученных значений от наилучших значений, полагая, что наилучшими (эталонными) являются значения, рассчитанные для наиболее представительной выборки из трехсот случайных величин.

(*пояснение*, отрицательное отклонение говорит о том, что исследуемое значение выборки оказалось **меньше** эталонного)

Вывод из 1 этапа:

Коэффициент вариации находится в диапазоне от 0.877 до 0.950, что говорит об очень высокой вариабельности данных.

Сходимость математического ожидания

У малых выборок ($n=10, 20$) наблюдается значительное отклонение от эталонного значения

$n = 10$ отклонение составило 19.82%,

$n = 20$ отклонение равно 22.995%

У средних выборок ($n=50, 100$) удовлетворительная сходимость, стремящаяся к нулю с увеличением числа значений

$n = 50$ отклонение составило -3.356%

$n = 100$ отклонение равно -0.137%

У крупной выборки $n = 200$ произошёл неожиданный скачок относительного отклонения по мат.ожиданию, и отклонение составило -5.215% , это произошло из-за высоковариабельности данных и влияния выбросов.

Доверительные интервалы разной точности показали одну и ту же закономерность - при увеличении выборки, относительное отклонение полуширины ДИ стало уменьшаться. Наиболее заметными являются изменения в выборках ($n=10,20,50$), где, например, для ДИ 95% отклонение отличаются от эталона на 649.684% потом на 392.758% и на 134.205% соответственно.

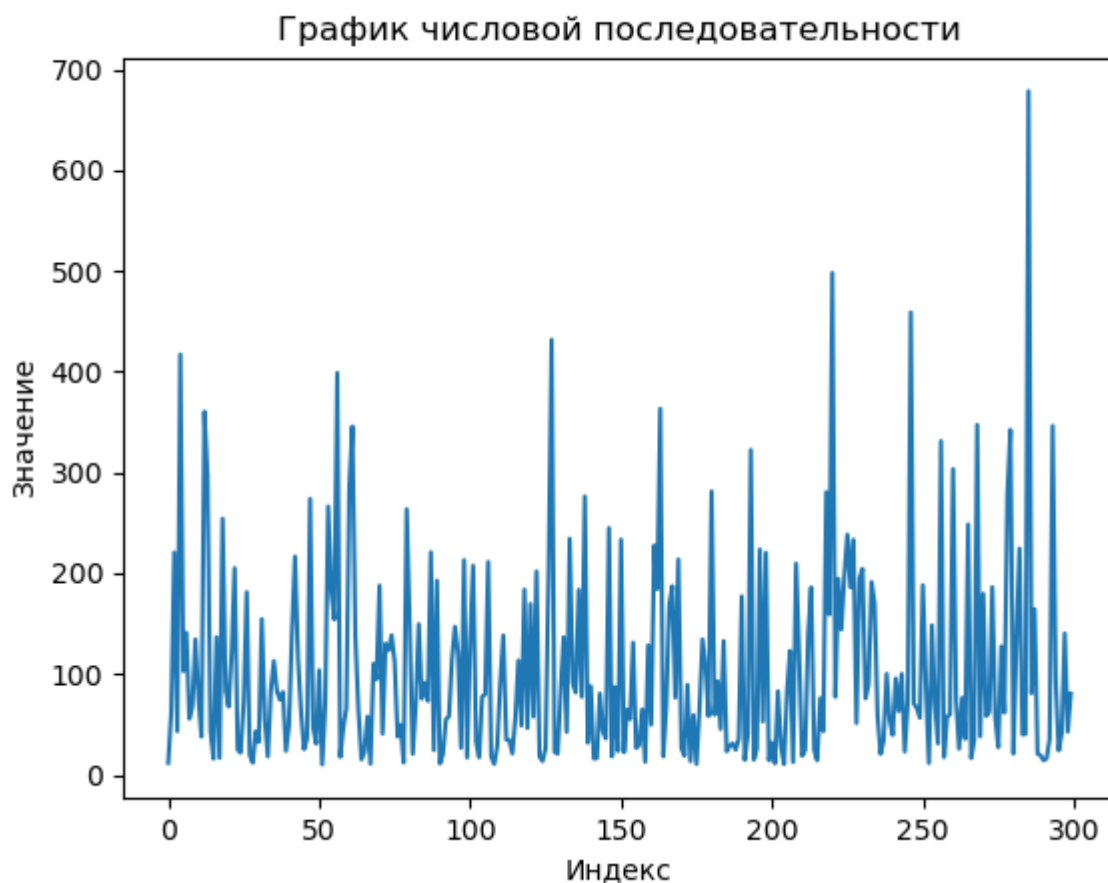
Дисперсия имеет значительные колебания от 8089.901 до 14231.729 в зависимости от взятой выборки, данные не однородны, присутствуют выбросы, но дисперсия имеет тенденцию к уменьшению, с увеличением значений выборки.

Значения **С.К.О.** колеблются от 89.944 до 119.297 , максимальное по модулю относительное отклонение не превышает 7.74% .

Наибольшие скачки в значениях относительного отклонения видны на переходе от выборки $n=20$ к выборке $n=50$, значения стали резко меньше в выборке $n=50$, что говорит об экстремальных выбросах в выборке размера $n=20$.

Основной вывод заключается в том, что опасно строить предположения о свойствах всей выборки по её малой части. Как показали исследования, выборки с малым количеством элементов едва ли могут считаться репрезентативными.

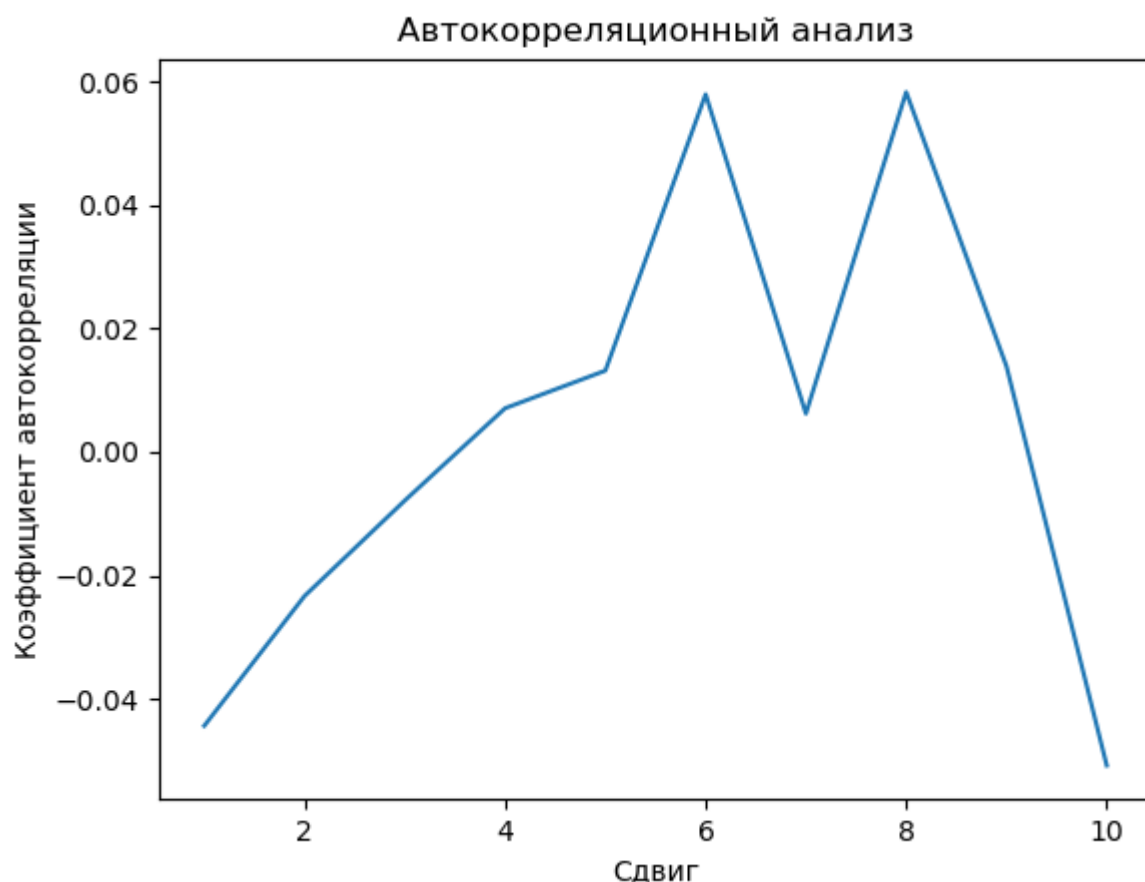
Этап 2. График №1. Значений заданной числовой последовательности с результатами анализа характера числовой последовательности.



Вывод из 2 этапа: На графике видны частые выбросы. Выбросы - это наблюдения, существенно отклоняющиеся от основной массы данных. Видно, что данные высоковариабельны, что соответствует значениям коэффициентов вариации, полученных в ходе исследования (от 0.877 до 0.950).

Этап 3. Форма 3. Результаты автокорреляционного анализа (значения коэффициентов автокорреляции со сдвигом 1, 2, 3, ...), представленные как в числовом (форма 3), так и графическом виде.

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК	-0.0434	-0.0209	-0.0090	0.0066	0.0163	0.0482	0.0003	0.0490	0.0200	-0.0543



Вывод из 3 этапа: Автокорреляционный анализ показал, что все коэффициенты корреляции между последовательными значениями статистически незначимы и близки к нулю (в диапазоне от -0.054 до +0.049). Поскольку все значения значительно меньше критического порога ± 0.113 , можно сделать вывод об отсутствии временной зависимости в данных. Это свидетельствует о случайном характере последовательности, где каждое следующее значение не зависит от предыдущих. Для проверки значимости коэффициентов автокорреляции использован стандартный статистический подход с критическим значением $\pm 1.96/\sqrt{n}$. Данный метод корректен для выборки объемом $n=300$, так как основан на асимптотической нормальности выборочных коэффициентов автокорреляции, что гарантируется центральной предельной теоремой независимо от распределения исходных данных.

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

Интервалы																		
№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Левая граница	10.18	47.32	84.45	121.59	158.73	195.86	233.00	270.14	307.27	344.41	381.55	418.68	455.82	492.95	530.09	567.23	604.36	641.50
Правая граница	47.32	84.45	121.59	158.73	195.86	233.00	270.14	307.27	344.41	381.55	418.68	455.82	492.95	530.09	567.23	604.36	641.50	678.64
Частота	109	68	26	27	22	17	9	8	3	5	2	1	1	1	0	0	0	1



Вывод из 4 этапа: Исходя из гистограммы мы можем видеть, что большая часть значений располагается в промежутке от 10.18 до 84.45, еще часть располагается до 381.55 и наименьшая часть значений располагается в диапазоне больше 381.55. Количество разбиений было взято по формуле $k = \sqrt{n}$, где k - кол-во разбиений, n - это число значений выборки. Также существует правило Стерджиса, Райса, но если считать по ним, кол-во интервалов уменьшится, потеряется часть

информации на графике, для Стерджиса было бы 10 интервалов для Райса - 14. Из коэффициента вариации (близок к 1), можно предположить, что закон распределения ЧП – Эрланга.

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

Для данной по варианту выборки коэффициент вариации практически равен единице, следовательно, для аппроксимации последовательности будем использовать закон: Эрланга. Значения математического ожидания ($t = M(X) = 104.923$) и коэффициента вариации ($v = 0.95$) были определены ранее.

$$k = \left\lceil \frac{1}{v^2} \right\rceil = 2$$

где $\lceil x \rceil$ означает ближайшее целое, большее x

$$\lambda = \frac{k}{t} \text{ параметр интенсивности эксп. распределения}$$

Соответственно, получаем следующий аппроксимирующий закон распределения:

$$F(r1, r2) = \left(-\frac{1}{\lambda} * \ln(1 - r1)\right) + \left(-\frac{1}{\lambda} * \ln(1 - r2)\right)$$

$$F(r1, r2) = -\frac{t}{k} * \ln((1 - r1) * (1 - r2))$$

где t – математическое ожидание

Вывод из 5 этапа: Исходя из прошлого этапа и вычислений в данном этапе, можем сказать, что аппроксимирующий закон распределения данной ЧП: Эрланга 2-го порядка.

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

Описание:

Для генерации случайной последовательности используем Excel. В листе Генератор создадим 300 пар случайных чисел (r1, r2) с помощью =СЛМАССИНВ(300; 2).

Используем эти данные для генерации распределения Эрланга 2-го порядка:

$$=-(\$H\$1/\$E\$1)*LN((1-Генератор!A1)*(1-Генератор!B1))$$

Получим числовую последовательность.

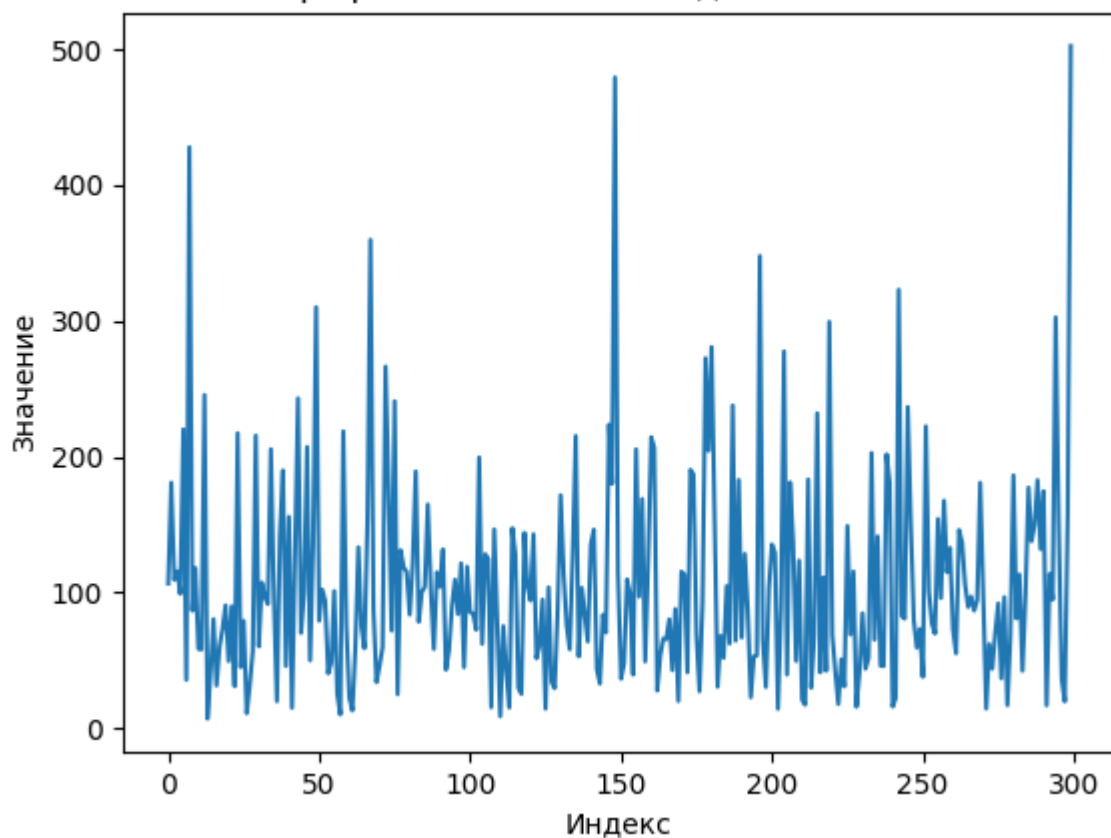
Вывод из 6 этапа: Удалось сформировать ЧП по аппроксимирующему закону в Excel и описать алгоритм формирования ЧП.

7 этап. График 3. Форма 2. Выводы по результатам сравнения сгенерированной в соответствии с полученным аппроксимирующим законом распределения последовательности случайных величин и заданной числовой последовательности.

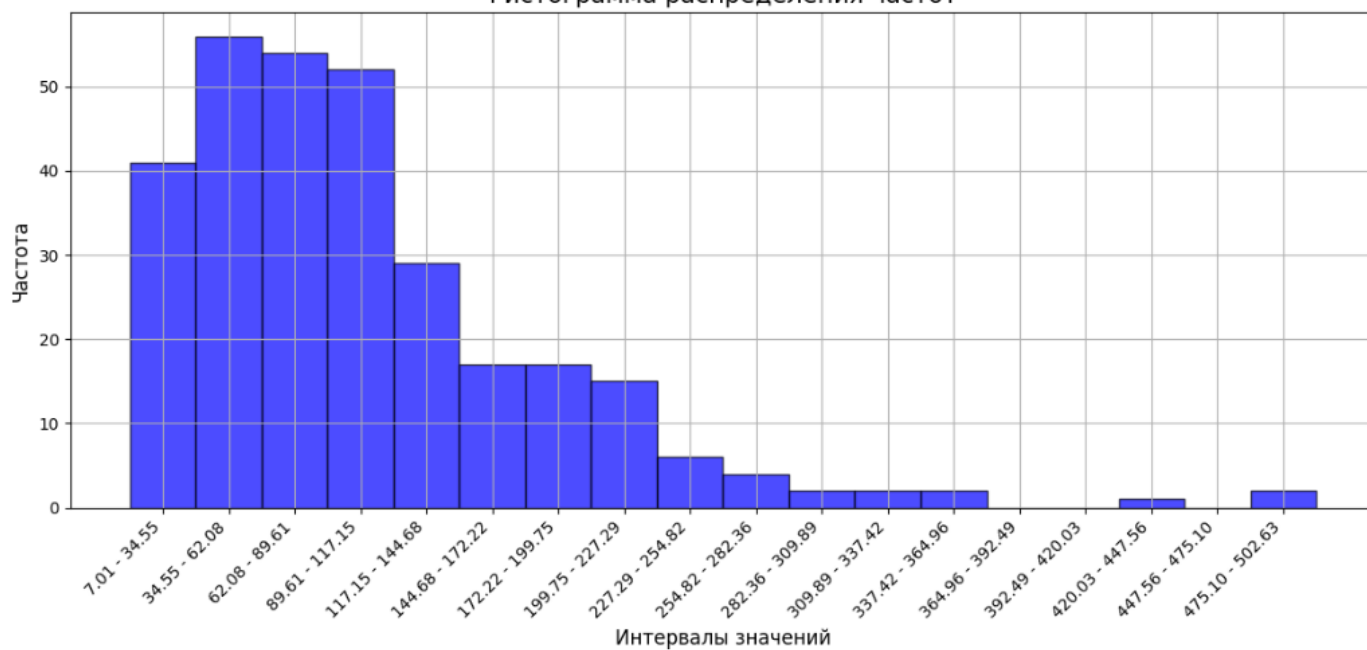
Закон распределения: Эрланга 2-го порядка							
Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. ож.	Знач,	149.955	112.201	110.872	106.077	104.807	105.582
	%	42.026	6.268	5.010	0.468	-0.735	0.629%
Дов. инт. (0,9)	Знач,	63.697	37.113	20.007	12.615	8.879	7.352
	%	766.439	404.828	172.141	71.595	20.771	-22.614%
Дов. инт. (0,95)	Знач,	78.605	44.923	23.981	15.075	10.595	8.768
	%	796.472	412.336	173.494	71.929	20.829	-22.614%
Дов. инт. (0,99)	Знач,	112.925	61.405	31.980	19.954	13.973	11.551
	%	877.659	431.619	176.874	72.756	20.973	-22.614%
Дисперсия	Знач,	12074.152	9213.393	7120.075	5772.283	5773.105	5955.673
	%	102.734	54.699	19.551	-3.079	-3.065	-40.114%
С. к. о.	Знач,	109.882	95.986	84.381	75.976	75.981	77.173
	%	42.385	24.378	9.339	-1.552	-1.545	-22.614%
К-т вариации	Знач,	0.733	0.855	0.761	0.716	0.725	0.731
	%	0.252	17.041	4.123	-2.010	-0.816	-23.098%

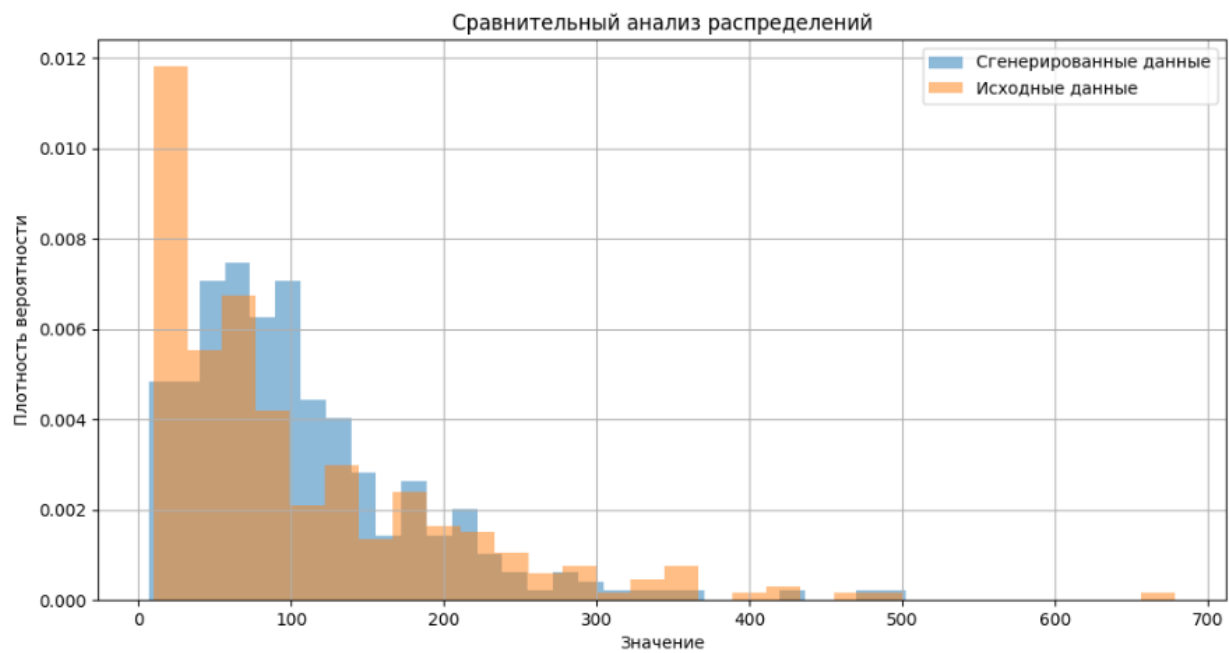
Математическое ожидание отличается от математического ожидания исходной выборки на величину, не превосходящую доверительные интервалы. Это говорит о том, что аппроксимация выполнена качественно.

График числовой последовательности

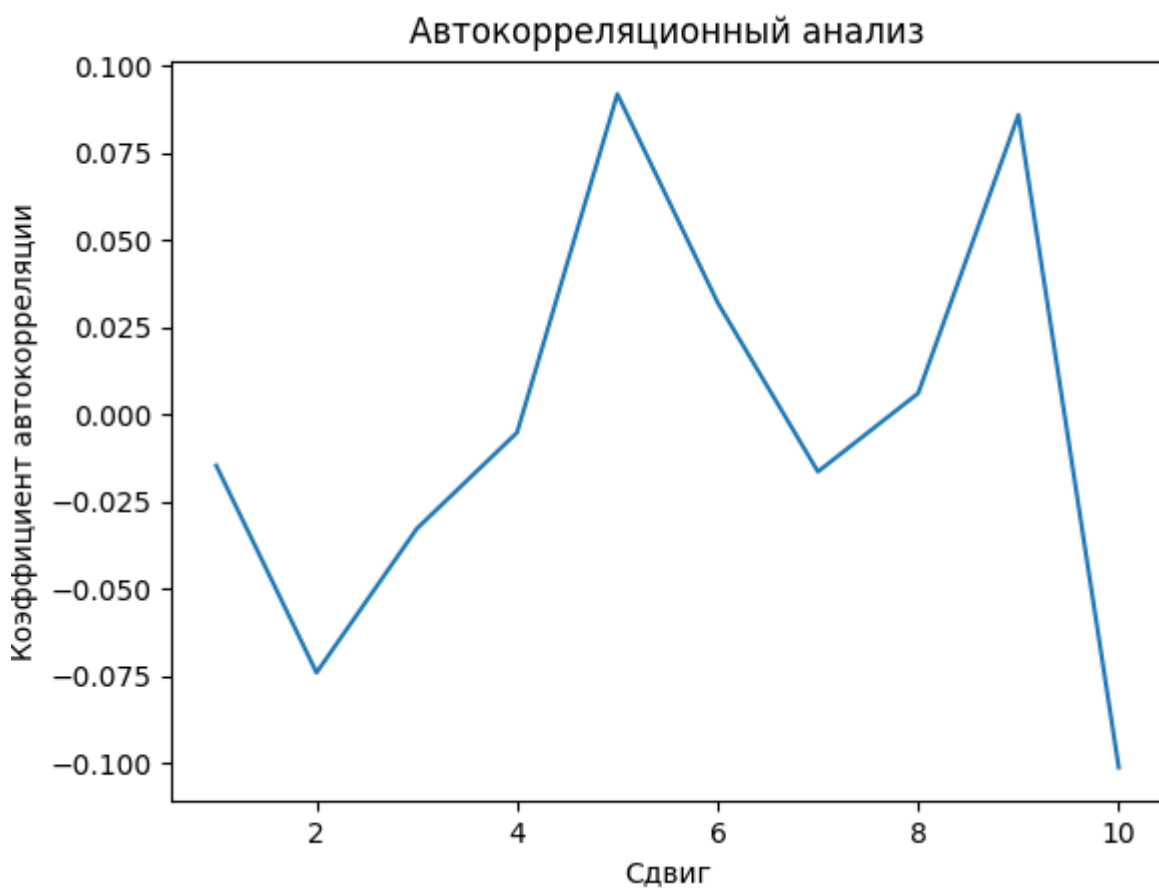


Гистограмма распределения частот





При сравнении полученных гистограмм видно, что полученная нами последовательность похожа на исходную. Тем самым, мы доказали, что выбранная нами аппроксимация подходит.



Коэффициент автокорреляции очень близок к нулю и находится в диапазоне от -0.1 до 0.1. Это свидетельствует о случайном характере сгенерированных данных.

Коэффициент корреляции между двумя числовыми последовательностями:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}} = 0.093$$

Для сгенерированной и полученной последовательности была рассчитана корреляционная зависимость. Корреляция между исходной и сгенерированной случайной последовательностями очень слабая, и это можно расценивать, как независимость.

Вывод из 7 этапа: Сравнение гистограммы распределения частот исходной числовой последовательности и плотности распределения закона Эрланга 2-го порядка показало, что действительно исходная ЧП соотносится с Эрлангом 2-го порядка. Сравнение числовых характеристик исходной и сгенерированной ЧП показало явное сходство характеристик, хотя нельзя не заметить особенности применения закона Эрланга, в частности из-за того что коэффициент вариации который рассматривался в нем отличается от того, который был в исходной последовательности (0.707 и 0.95). Проследить это можно, взглянув на разницу между итоговыми значениями различных параметров сгенерированной и исходной последовательности, дисперсия отличается на 40.114% по модулю, то есть сгенерированная дисперсия оказалась меньше исходной, как и остальные параметры (~20%), кроме разве что мат.ожидания. Разница сгенерированного мат.ожидания и исходного составила всего 0.629% . Распределение Эрланга с целочисленным параметром k обладает фундаментальным ограничением - оно может аппроксимировать только определённые значения коэффициента вариации (1, 0.707, 0.577, 0.5, ...). Для коэффициента вариации 0.95 наилучшим приближением является k=2, что и объясняет расхождение в дисперсии и других параметрах.

Выводы

В рамках лабораторной работы была проанализирована исходная последовательность случайных величин, и на основе этих данных был создан генератор случайных последовательностей на основе распределения Эрланга 2-го порядка, который повторяет основное поведение случайных величин в исходной последовательности и при этом никак от неё не зависит.