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: 309

Выполнил:

студент 3 курса

Поленов Кирилл Александрович

Группа: P3313

Принял:

Тропченко Андрей Александрович

г. Санкт-Петербург, 2024

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; гипоэкспоненциальный; 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

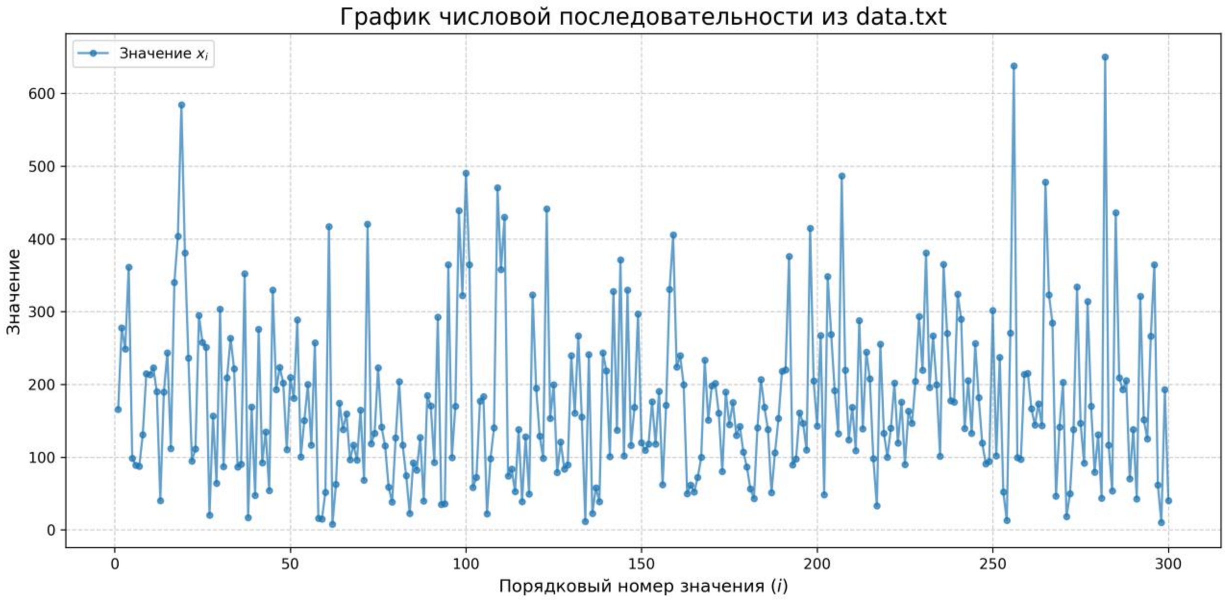
Этап 1. Форма №1. Оценки математического ожидания, дисперсии, среднеквадратического отклонения, коэффициента вариации заданной числовой последовательности и доверительные интервалы для оценки математического ожидания с доверительными вероятностями 0,9; 0,95 и 0,99, сведенные в таблицу.

Характеристика		**Количество случайных величин**					
		10	**20**	**50**	**100**	**200**	**300**
Мат. ож.	Знач,	188.73	229.76	195.08	174.63	169.83	176.87
	%	6.71	29.90	10.30	1.27	-3.98	
Дов. инт. (0,9)	Знач,	± 52.87	± 51.41	± 27.16	± 19.46	± 12.93	± 10.90
	%	± 385.23	± 371.83	± 149.22	± 78.57	± 18.64	
Дов. инт. (0,95)	Знач,	± 65.24	± 62.23	± 32.55	± 23.25	± 15.43	± 13.00
	%	± 402.05	± 378.85	± 150.46	± 78.92	± 18.70	
Дов. инт. (0,99)	Знач,	± 93.73	± 85.06	± 43.41	± 30.78	±20.34	±17.12
	%	± 447.52	± 396.8697	± 153.56	± 79.78	± 18.84	
Дисперсия	Знач,	8318.92	17680.07	13117.60	13732.65	12238.14	13083.12
	%	-36.41	35.14	0.26	4.96	-6.46	
С. к. о.	Знач,	91.21	132.97	114.53	117.19	110.63	114.38
	%	-20.26	16.25	0.13	2.45	-3.28	
К-т вариации	Знач,	0.48	0.58	0.59	0.67	0.65	0. 64.67
	%	-25.27	-10.51	-9.21	3.77	0.72	

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

Вывод из 1 этапа: Дисперсия и среднеквадратическое отклонение возрастают с увеличением выборки, что говорит о большей вариативности в данных, но при больших объемах выборки наблюдается стабилизация. Коэффициент вариации показывает умеренные изменения, что указывает на относительную стабильность отношения стандартного отклонения к среднему при увеличении выборки.

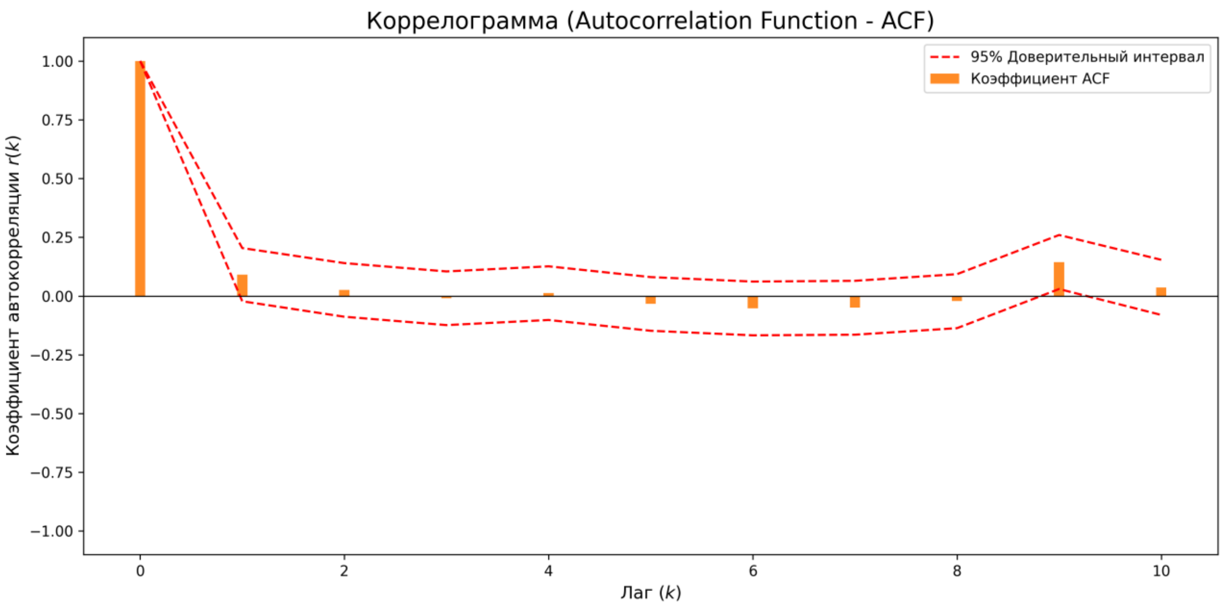
Этап 2. График №1. Значений заданной числовой последовательности с результатами анализа характера числовой последовательности.



Вывод из 2 этапа: Изучив график, можно сделать вывод, что исходная последовательность не является периодической, возрастающей или убывающей.

<u>**Этап 3. Форма 3.**</u> 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	0.0908	0.0259	-0.0096	0.0122	-0.0337	-0.0529	-0.0500	-0.0219	0.1447	0.0368



Вывод из 3 этапа: Последовательность можно считать случайной так как данные коэффициенты указывают на то, что между числами не было выявлено ни линейной, ни циклической зависимости, нет тенденции и периодичности. Значения коэффициентов автокорреляции колеблются в окрестности нуля.

<u>Этап 4. График 2.</u> Гистограмма распределения частот для заданной числовой последовательности (график 2).

Количество интервалов (по правилу Стерджеса): k = 9

Ширина интервала: h ≈ 71.3833

Интервалы									
№	1	2	3	4	5	6	7	8	9
Левая граница	7.9300	79.3133	150.6967	222.0800	293.4633	364.8467	436.2300	507.6133	578.9967
Правая граница	79.3133	150.6967	222.0800	293.4633	364.8467	436.2300	507.6133	578.9967	650.3800
Частота	53	95	73	35	23	11	7	0	3

Вывод из 4 этапа: Исходя из гистограммы мы можем видеть, что большая часть значений располагается в промежутке от 79.3133 до 150.6967, еще часть располагается до 222.0800 и наименьшая часть значений располагается в диапазоне больше от 507.6133 до 578.9967. По гистограмме можно предположить, что ЧП имеет нормированное распределение Эрланга. Это мы увидим позже при расчете коэффициента вариации.

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

Начальные моменты: 176.87 и 114.3815 (мат ожидание и СКО соответственно).

Коэффициент вариации: $V_k = \frac{114.3815}{176.87} = 0.6467$.

КВ < 1, следовательно имеем дело либо с распределением Эрланга либо с Гипоэкспоненциальным.

$0.6467 > 0.57735$, следовательно распределение не является нормальным, а значит мы имеем дело с распределением Эрланга.

Вычислим его порядок (параметр формы):

$$k = \frac{1}{{V_k}^2} = \frac{1}{0.6467^2} = 2.39 \approx 3$$

Вычислим его интенсивность (параметр масштаба):

$$\lambda = \frac{k}{M(x)} = \frac{3}{176.87} = 0.0170$$

Вывод из 5 этапа: Исходя из прошлого этапа и вычислений в данном этапе, можем сказать, что аппроксимирующий закон распределения данной ЧП: Эрланга 3-его порядка.

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

Описание:

Используем Гамма-распределение, так как распределение Эрланга является частным случаем Гамма-распределения с целочисленным параметром формы.

Распределение Эрланга k-го порядка можно представить как сумму k независимых и одинаково распределенных случайных величин, каждая из которых подчиняется Экспоненциальному распределению с одной и той же интенсивностью λ.

Таким образом, алгоритм генерации одного значения X следующий:

Сгенерировать k случайных величин τ1, τ2, …, τk, где каждая τi распределена по Экспоненциальному закону с параметром λ.

Сложить эти k значений: X=τ1+τ2+⋯+τk.

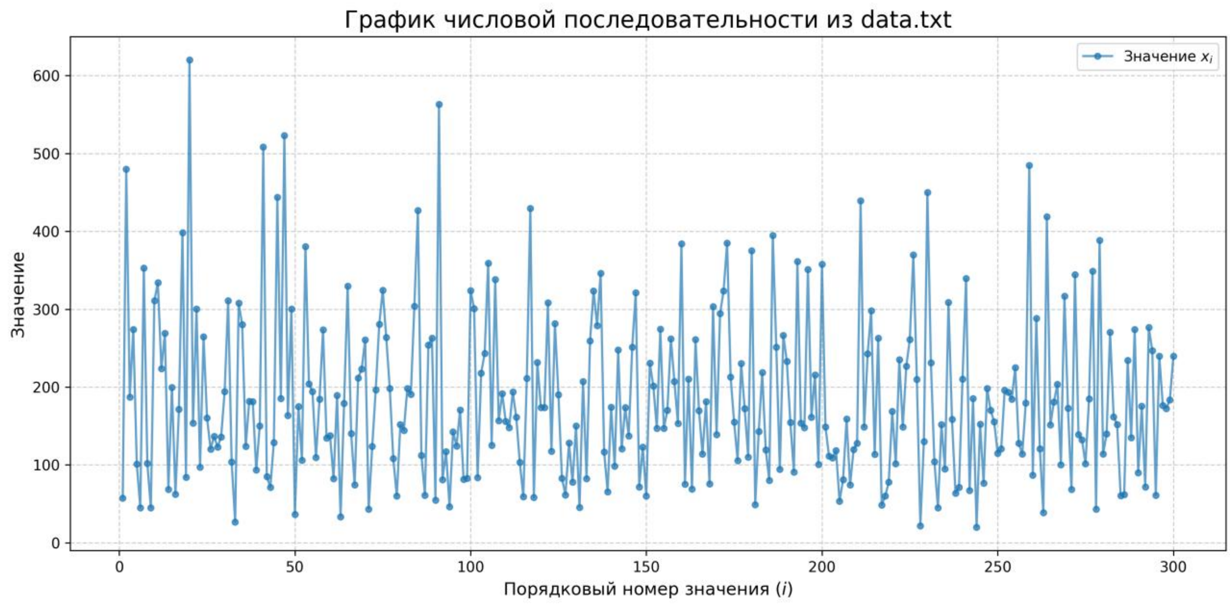
```
import numpy as np

k_order = 3
lambda_rate = 0.0170  # Параметр масштаба (scale) в NumPy - это 1/lambda
output_filename='generated_sequence.txt'

# 1. Параметр масштаба (scale) для NumPy:
scale_param = 1.0 / lambda_rate

# 2. Генерация 300 случайных чисел по закону Эрланга (k=3, lambda=0.0170)
N_values = 300
erlang_sequence = np.random.gamma(
    shape=k_order,   # Параметр формы (k)
    scale=scale_param,   # Параметр масштаба (1/lambda)
    size=N_values
)

np.savetxt(output_filename, erlang_sequence, fmt='%.4f', delimiter=',')

print(f"Сгенерировано {N_values} значений, среднее:
{np.mean(erlang_sequence):.4f}")
```

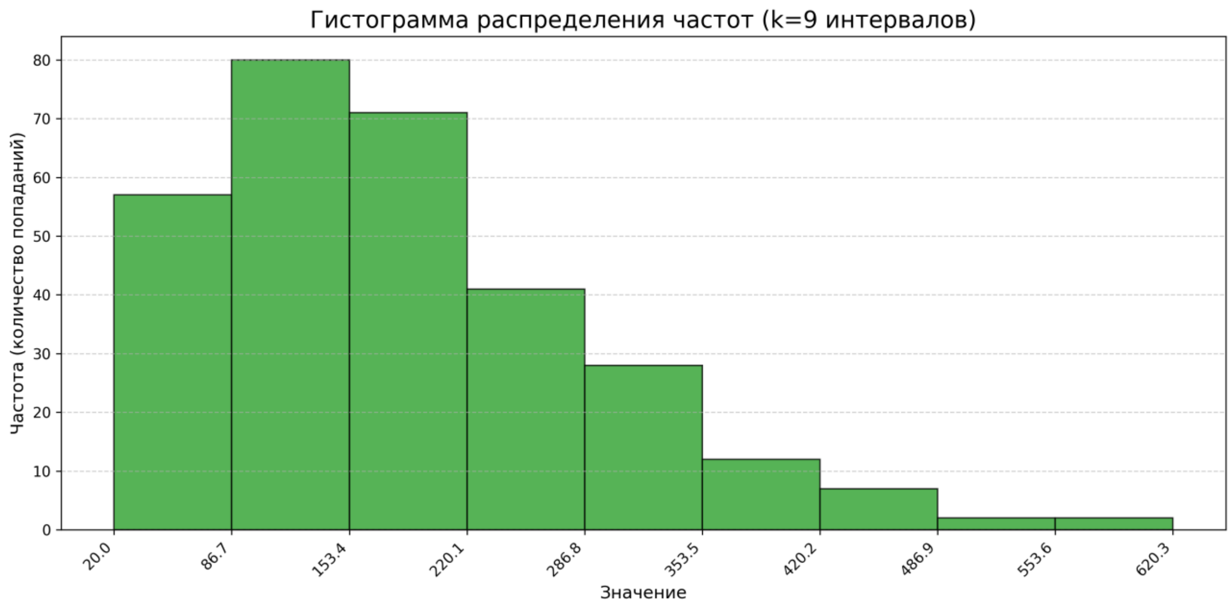
Вывод из 6 этапа: Мне удалось сформировать ЧП по аппроксимирующему закону в Python и описать алгоритм формирования ЧП.

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.

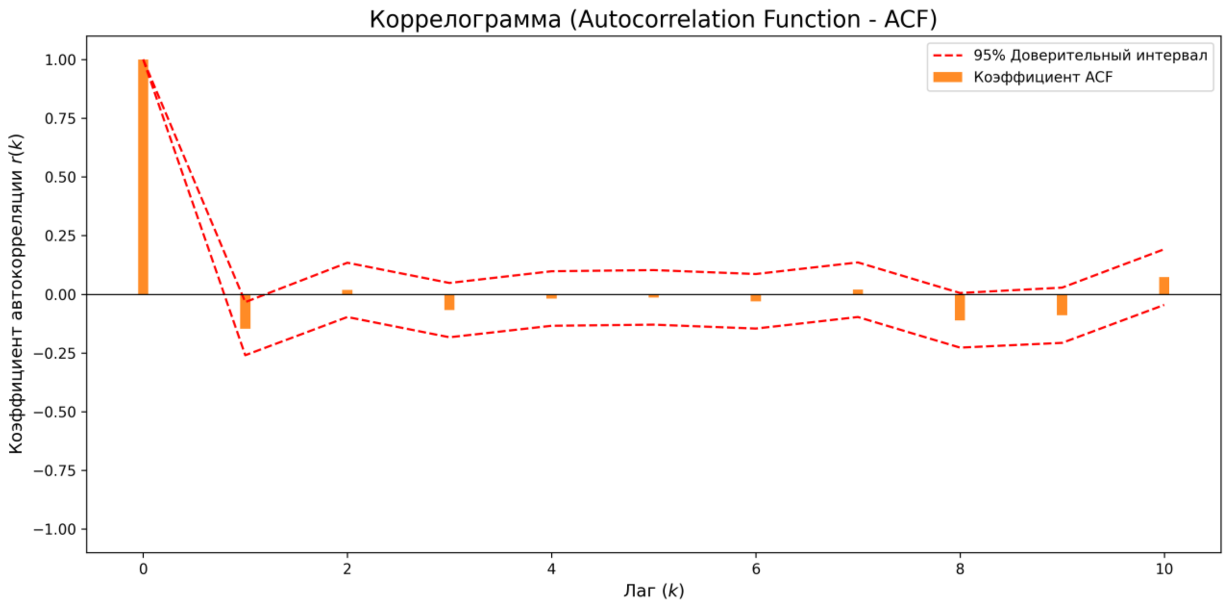
Закон распределения: Эрланга 3-его порядка							
Характеристика		**Количество случайных величин**					
		10	**20**	**50**	**100**	**200**	**300**
Мат. ож.	Знач,	195.4943	219.2969	205.5253	193.9062	192.4594	185.7245
	%	5.2603	18.0764	10.6613	4.4052	3.6263	
Дов. инт. (0,9)	Знач,	± 87.9866	± 62.0931	± 33.3159	± 20.7688	± 12.9665	± 10.2631
	%	757.3077	505.0121	224.6179	102.3637	26.3406	
Дов. инт. (0,95)	Знач,	± 108.5800	± 75.1605	± 39.9337	± 24.8194	± 15.4727	± 12.2409
	%	787.0246	514.0098	226.2311	102.7574	26.4014	
Дов. инт. (0,99)	Знач,	± 155.9870	± 102.7361	± 53.2552	± 32.8521	± 20.4065	± 16.1251
	%	867.3564	537.1201	230.2634	103.7331	26.5515	
Дисперсия	Знач,	23038.4993	25790.5514	19744.2505	15646.0066	12313.1080	11607.2824
	%	98.4831	122.1928	70.1023	34.7947	6.0809	
С. к. о.	Знач,	151.7844	160.5944	140.5142	125.0840	110.9644	107.7371
	%	40.8840	49.0613	30.4233	16.1011	2.9956	
К-т вариации	Знач,	77.64	73.23	68.37	64.51	57.66	58.01
	%	33.8435	26.2414	17.8580	11.2024	-0.6086	

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





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



Коэффициент автокорреляции интервалов от 1 до 10 приближены к нулю, следовательно, можно сказать, что выборка случайна.

```
# Расчет коэффициента корреляции Пирсона
# np.corrcoef возвращает матрицу корреляции. Нас интересует элемент [0, 1]
# (корреляция между первой и второй последовательностями).
correlation_matrix = np.corrcoef(data_original, data_generated)
correlation_coefficient = correlation_matrix[0, 1]
```

$$r = \frac{\sum_{i=1}^{n}(x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \underline{x})^2 \times \sum_{i=1}^{n}(y_i - \underline{y})^2}} = 0.0384$$

Коэффициент корреляции Пирсона (r): 0.0384

Корреляция отсутствует (или крайне слабая). Последовательности, по сути, независимы.

Вывод из 7 этапа: Сравнение гистограммы распределения частот исходной числовой последовательности и плотности распределения закона Эрланга показало, что действительно исходная ЧП соотносится с аппроксимирующим законом Эрланга. Сравнение числовых характеристик исходной и сгенерированной ЧП показало явное сходство характеристик.

Выводы

В рамках лабораторной работы была дана числовая последовательность, для которой определено математическое ожидание, дисперсия и другие параметры. Проанализирована гистограмма, по которой не было выявлено возрастания, убывания или периодичности последовательности. Исследуемую последовательность можно назвать случайной исходя из автокорреляционного анализа. Вычислены параметры аппроксимирующего закона и по ним сгенерирована новая последовательность. Коэффициент автокорреляции первой и второй последовательности варьируется около нуля, исходя из этого можно сделать вывод о том, что выборка случайна. Математическое ожидание и дисперсия отличаются, но отличие не выходит за пределы доверительных интервалов.