

Университет ИТМО

Факультет программной инженерии и компьютерной техники

Учебно-исследовательская работа No1 (УИР 1)

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

по дисциплине «Моделирование»

Выполнили:

Студентки группы Р34312

Морозова А.А.

Шаповалова Д.С.

Вариант: 52

Преподаватель:

Алиев Т.И.

Санкт-Петербург

2023

Цель	2
Выполнение	2
1. Расчет статистических характеристик заданной числовой последовательности	2
2. Построение графика значений для заданной числовой последовательности	3
3. Выполнение автокорреляционного анализа	4
4. Построение гистограммы распределения	5
5. Выполнение аппроксимации закона распределения	5
6. Реализация генератора случайных величин	6
7. Генерация последовательности случайных величин	7
8. Автокорреляционный анализ сгенерированной последовательности	8
9. Сравнительный анализ заданной и сгенерированной последовательностей	9
10. Оценка корреляционной зависимости заданной и сгенерированной последовательностей	11
Выводы	11

Цель

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. Расчет статистических характеристик заданной числовой последовательности

Для расчета оценки математического ожидания, оценки дисперсии, оценки СКО, доверительных интервалов, коэффициента вариации были использованы следующие формулы расчета:

$$\tilde{m} = \frac{\sum_{i=1}^n X_i}{n} \quad \tilde{D} = \frac{\sum_{i=1}^n (X_i - \tilde{m})^2}{n-1}$$

Среднее квадратическое отклонение (часто просто называемое стандартным отклонением) — это статистический показатель, который измеряет степень разброса значений в наборе данных относительно среднего значения. Среднее квадратическое отклонение дает представление о том, насколько близко значения в наборе данных сгруппированы вокруг среднего значения.

$$\tilde{\sigma}_m = \sqrt{\frac{\tilde{D}}{n}}$$

Доверительный интервал — это интервал оценок для параметра генеральной совокупности, который строится на основе выборочных данных. Этот интервал задается таким образом, что с определенной заранее выбранной вероятностью (например, 95%) истинное значение параметра генеральной совокупности попадает в этот интервал.

$$\varepsilon = (1 - \alpha) * \frac{\bar{\sigma}}{\sqrt{n}}$$

Коэффициент вариации — это статистический показатель, который измеряет относительную изменчивость (разброс) данных по отношению к их среднему значению. Он часто используется для сравнения стабильности двух или более наборов данных, когда средние значения сильно отличаются.

$$v = \frac{\bar{\sigma}}{\bar{m}}$$

Характеристики заданной ЧП (вариант 52)							
Характеристика		10	20	50	100	200	300
Мат. ож.	Значение	31,12	26,98	22,00	19,33	14,44	16,39
	%	89,92	64,66	34,24	17,97	-11,86	
Дов. инт. (0,9)	Значение	29,99	16,57	9,24	6,29	3,43	3,19
	%	840,14	419,45	189,58	97,07	7,64	
Дов. инт. (0,95)	Значение	35,73	19,74	11,01	7,49	4,09	3,80
	%	840,14	419,45	189,58	97,07	7,64	
Дов. инт. (0,99)	Значение	46,96	25,95	14,46	9,84	5,38	4,99
	%	840,14	419,45	189,58	97,07	7,64	
Дисперсия	Значение	3323,53	2029,21	1576,59	1460,34	871,26	1128,06
	%	194,62	79,88	39,76	29,46	-22,76	
С. К. О.	Значение	57,65	45,05	39,71	38,21	29,52	33,59
	%	71,65	34,12	18,22	13,78	-12,12	
К-т вариации	Значение	1,85	1,67	1,80	1,98	2,04	2,05
	%	-9,62	-18,55	-11,93	-3,56	-0,29	

% - относительные отклонения рассчитанных значений от значений, полученных для выборки из трехсот величин

Приведенная выше таблица отражает полученные характеристики случайной последовательности. Полагается, что выборка из 300 чисел и рассчитанные для нее моменты являются эталонными.

Если брать случайную выборку, то с увеличением размера выборки моменты будут меньше отклоняться от эталонных значений. Таким образом, увеличение размера выборки делает оценки статистик более надежными и близкими к истинным параметрам генеральной совокупности.

Оценка мат. ожидания показывает ожидаемое среднее значение случайной величины.

Доверительный интервал характеризует диапазон значений, в котором с определенной вероятностью находится случайная величина. Видно, что с увеличением доверительной вероятности доверительный интервал увеличивается — чем больше требование к вероятности попадания случайной величины в доверительный интервал, тем шире требуется интервал.

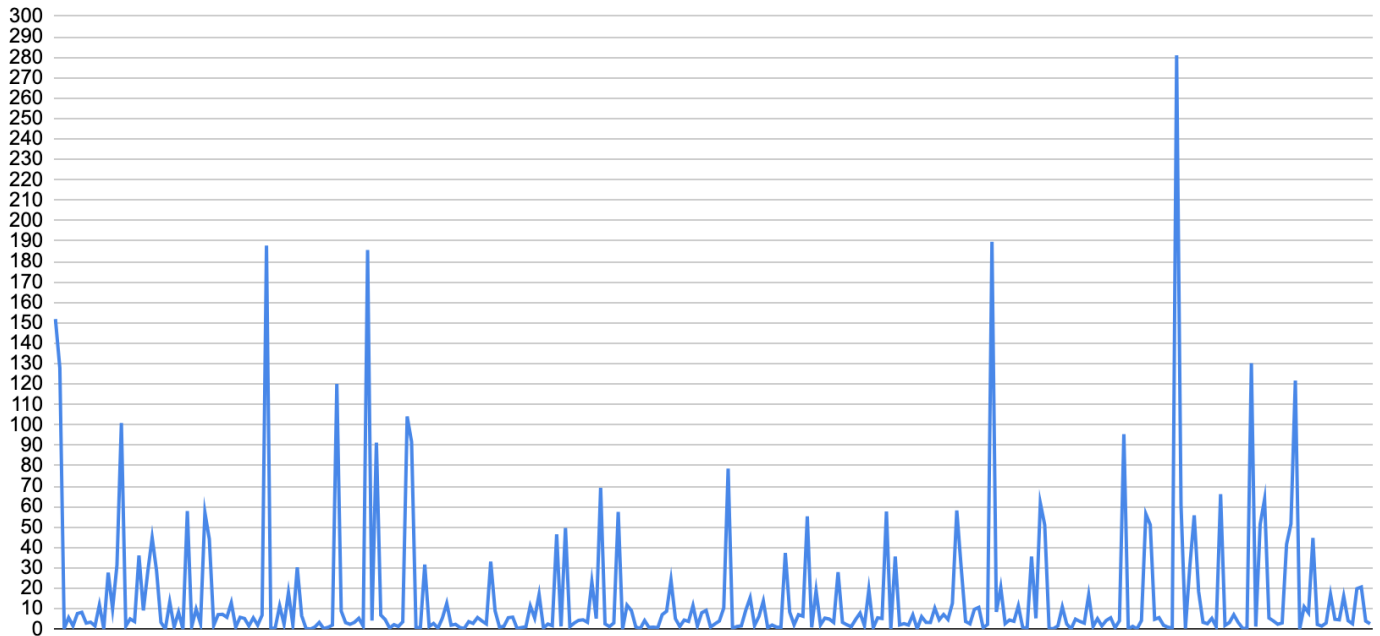
Дисперсия и СКО - характеризуют математическое ожидание отклонений от математического ожидания.

Коэффициент вариации - показывает отношение между мат. ожиданием отклонений и мат. ожиданием, то есть то, насколько отклонение близко или далеко от мат. ожидания

2. Построение графика значений для заданной числовой последовательности

Обратимся к построенному графику значений заданной числовой последовательности

График значений заданной ЧП



Проанализировав данный график, можно сделать выводы о характере заданной числовой последовательности:

- Не является возрастающей/невозрастающей
- Не является убывающей/неубывающей
- Не является периодической

Поскольку график значений для заданной числовой последовательности не имеет выраженных черт для данных характеристик

3. Выполнение автокорреляционного анализа

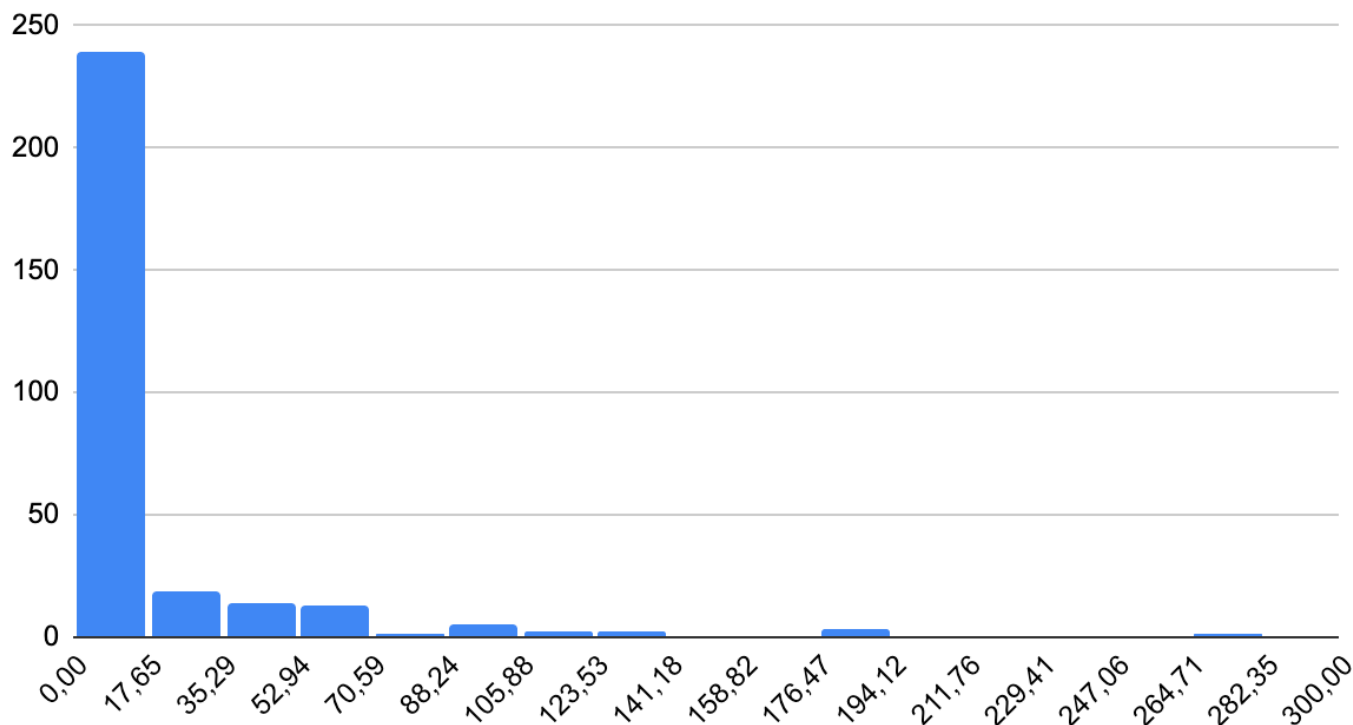
Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для заданн ой ЧП	0,0605834	-0,02056738	-0,05948580	-0,022853 29	-0,0633 7704	-0,03202 964	0,11681 756	0,00937 591	0,03757 042	0,05759 160

Автокорреляция - это способ измерить, есть ли закономерность в последовательности. Она показывает, насколько текущее значение данных связано с предыдущими значениями, можно ли зная текущее значение предсказать предыдущее/следующее.

Коэффициенты автокорреляции заданной числовой последовательности для 9 сдвигов < 0.1 , для одного случая < 0.2 , что является вариантом нормы для коэффициента корреляции, что позволяет сказать, что последовательность случайна.

4. Построение гистограммы распределения

Гистограмма распределения частот для заданной ЧП



Гистограмма дает представление функции плотности вероятности случайной величины, построенное по выборке. Можно сделать предположение, что заданная ЧП соответствует закону гиперэкспоненциального распределения.

Гистограмма имеет длинный хвост, что означает наличие редких экстремальных значений в последовательности, что характерно для гиперэкспоненциального распределения. Это означает, что при гиперэкспоненциальном распределении вероятность появления больших значений случайной величины значительно выше, чем, например, для экспоненциального распределения.

Также гистограмма гиперэкспоненциального распределения по сравнению с экспоненциальным распределением характеризуется более резким спадом в области малых значений случайной величины, причем чем больше коэффициент вариации случайной величины, тем круче эта зависимость. Это значит, что вероятность появления маленьких значений случайной величины для гиперэкспоненциального распределения намного больше вероятности появления больших значений.

5. Выполнение аппроксимации закона распределения

Рассчитанный коэффициент вариации случайной величины $v = 2.05 > 1$.

Для аппроксимации закона распределения такой случайной величины используют **гиперэкспоненциальное распределение**, представляющее собой композицию экспоненциальных распределений.

Стоит отметить, что аппроксимация гиперэкспоненциального распределения может осуществляться по трем моментам распределения, так как оно является трехпараметрическим, т.е. содержит три независимых параметра. Таким образом, имеется три параметра: q, t_1, t_2 .

По следующим формулам были рассчитаны три момента:

Для аппроксимации закона распределения с коэффициентом вариации > 1 двухфазным гиперэкспоненциальным распределением следует выбрать следующее значение вероятности:

$$q \leq \frac{2}{1+v^2} = 0,384578$$

Пусть $q = 0,3$ (вероятность формирования значения случайной величины в первой фазе), тогда математические ожидания первой и второй экспоненциальных фаз будут равны соответственно рассчитанным t_1, t_2 :

$$t_1 = \left[1 + \sqrt{\frac{1-q}{2q}} (v^2 - 1) \right] * t = 48,05401085$$

$$t_2 = \left[1 - \sqrt{\frac{q}{2(1-q)}} (v^2 - 1) \right] * t = 2,816309635$$

Где t - математическое ожидание.

6. Реализация генератора случайных величин

Для реализации случайных величин в соответствии с законом гиперэкспоненциального распределения было решено воспользоваться тремя генераторами. Первые два экспоненциальных генератора выдают по 300 случайных величин с рассчитанными для них математическими ожиданиями. Третий генератор в свою очередь генерирует значение вероятности от 0 до 1. Т.к. было рассчитано, что $q = 0.3$, то при попадании сгенерированной вероятности в отрезок от 0 до 0.3 будет выбрано число из последовательности первого генератора, а при попадании в полуинтервал от 0.3 до 1 - второго. Таким образом, получается последовательность из 300 случайных значений.

```
import random
from scipy.stats import expon
```

```
a1 = 48.05401085
a2 = 2.16309635
```

```

q1 = 0.3
q2 = 0.7

generator1 = expon.rvs(scale=a1,loc=0,size=300)
generator2 = expon.rvs(scale=a2,loc=0,size=300)
result = []

for i in range(1,301):
    number = random.random()
    idx = random.randint(0, 299)
    if number <= 0.3:
        result.append(generator1[idx])
    else:
        result.append(generator2[idx])
for i in result:
    print(i)

```

7. Генерация последовательности случайных величин

Была сгенерирована и проанализирована последовательность соответствии с законом гиперэкспоненциального распределения

Характеристики сгенерированной ЧП (вариант 52)							
Характеристика		10	20	50	100	200	300
Мат. ож.	Значение	13,28	25,92	18,77	12,91	13,96	14,61
	%	-57,32	-3,96	-14,69	-33,22	-3,34	-10,82
Дов. инт. (0,9)	Значение	10,40	14,20	8,46	4,76	3,62	3,01
	%	-65,31	-14,30	-8,44	-24,33	5,43	-5,51
Дов. инт. (0,95)	Значение	12,39	16,92	10,08	5,67	4,31	3,59
	%	-65,31	-14,30	-8,44	-24,33	5,43	-5,51
Дов. инт. (0,99)	Значение	16,29	22,23	13,24	7,45	5,67	4,72
	%	-65,31	-14,30	-8,44	-24,33	5,43	-5,51
Дисперсия	Значение	399,89	1490,24	1321,79	836,16	968,43	1007,20
	%	-87,97	-26,56	-16,16	-42,74	11,15	-10,71
С. К. О.	Значение	20,00	38,60	36,36	28,92	31,12	31,74
	%	-65,31	-14,30	-8,44	-24,33	5,43	-5,51
К-т вариации	Значение	1,51	1,49	1,94	2,24	2,23	2,17
	%	-18,72	-10,77	7,32	13,31	9,08	5,95

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

Поскольку отклонения имеют малые значения для выборки из 300 чисел, можно утверждать, что сгенерированная последовательность близка по характеристикам к исходной. Из-за того, что генерируются случайные величины, отклонение не может отсутствовать, что говорило бы о полном совпадении последовательностей.

8. Автокорреляционный анализ сгенерированной последовательности

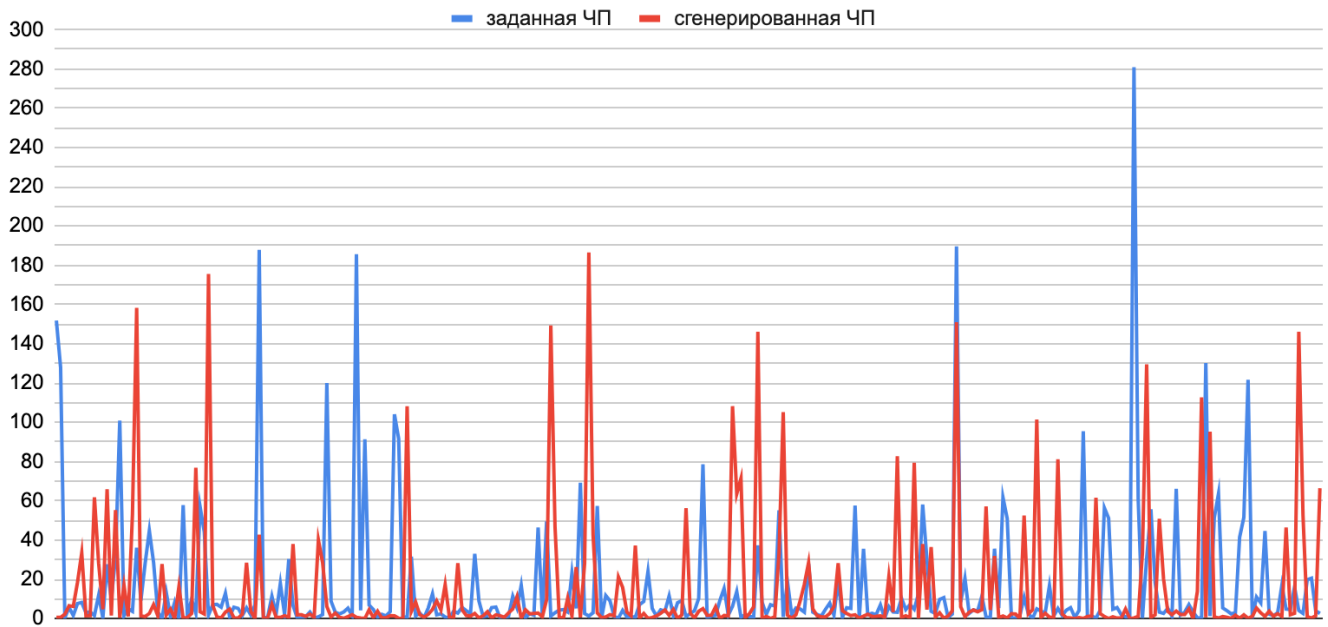
Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для заданн ой ЧП	0,06058	-0,02056738	-0,05948580	-0,02285	-0,06337	-0,03202	0,11681	0,00937	0,03757	0,05759
К-т АК для сгенер. ЧП	0,04542	-0,04996376	-0,01359873	-0,02485	0,00188	0,04642	-0,03605	-0,01815	0,05410	0,03113
%	-25,02	142,93	-77,14	8,74	-102,98	-244,96	-130,87	-293,63	44,00	-45,94

Коэффициенты автокорреляции сгенерированной числовой последовательности для всех сдвигов < 0.1, что является хорошим показателем для коэффициента корреляции и позволяет сказать, что последовательность случайна.

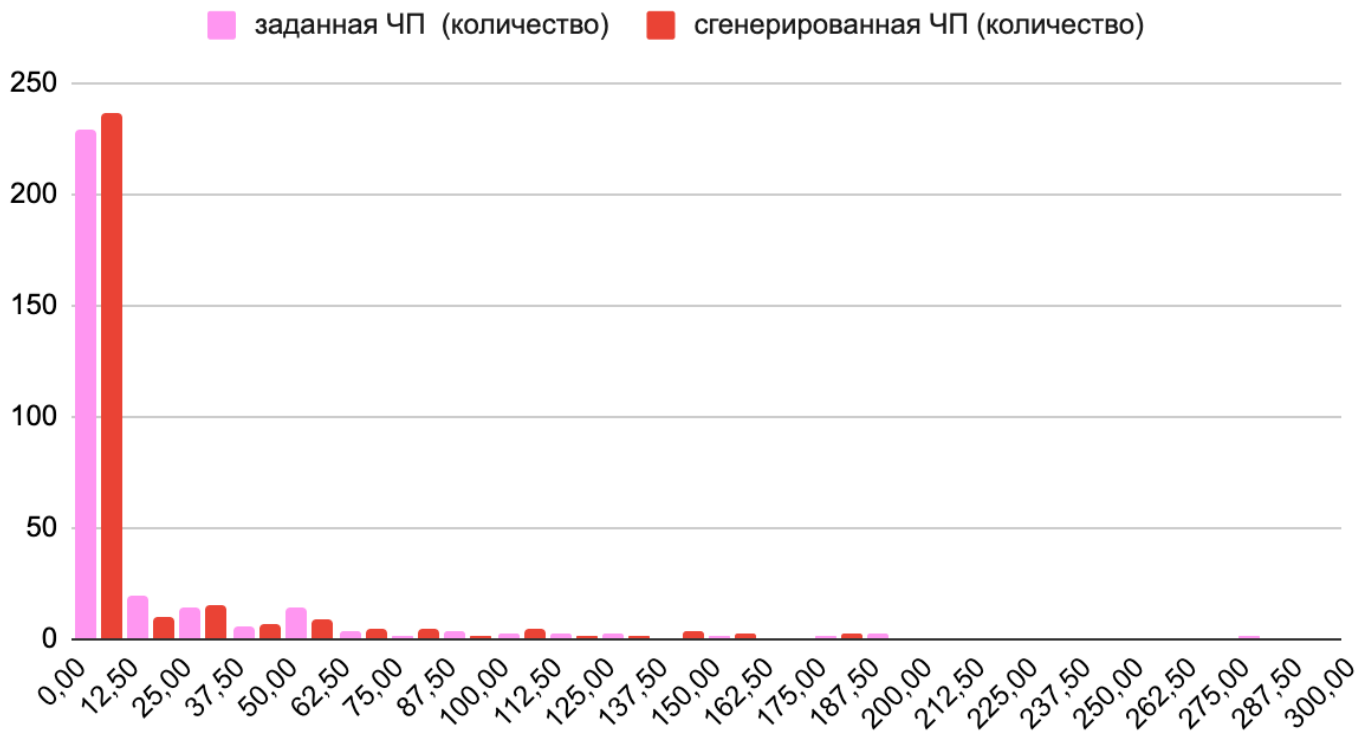
Автокорреляционный анализ — это инструмент для изучения зависимостей внутри временных рядов. Он позволяет выявлять, сохраняется ли зависимость между значениями временного ряда на различных лагах (задержках) во времени.

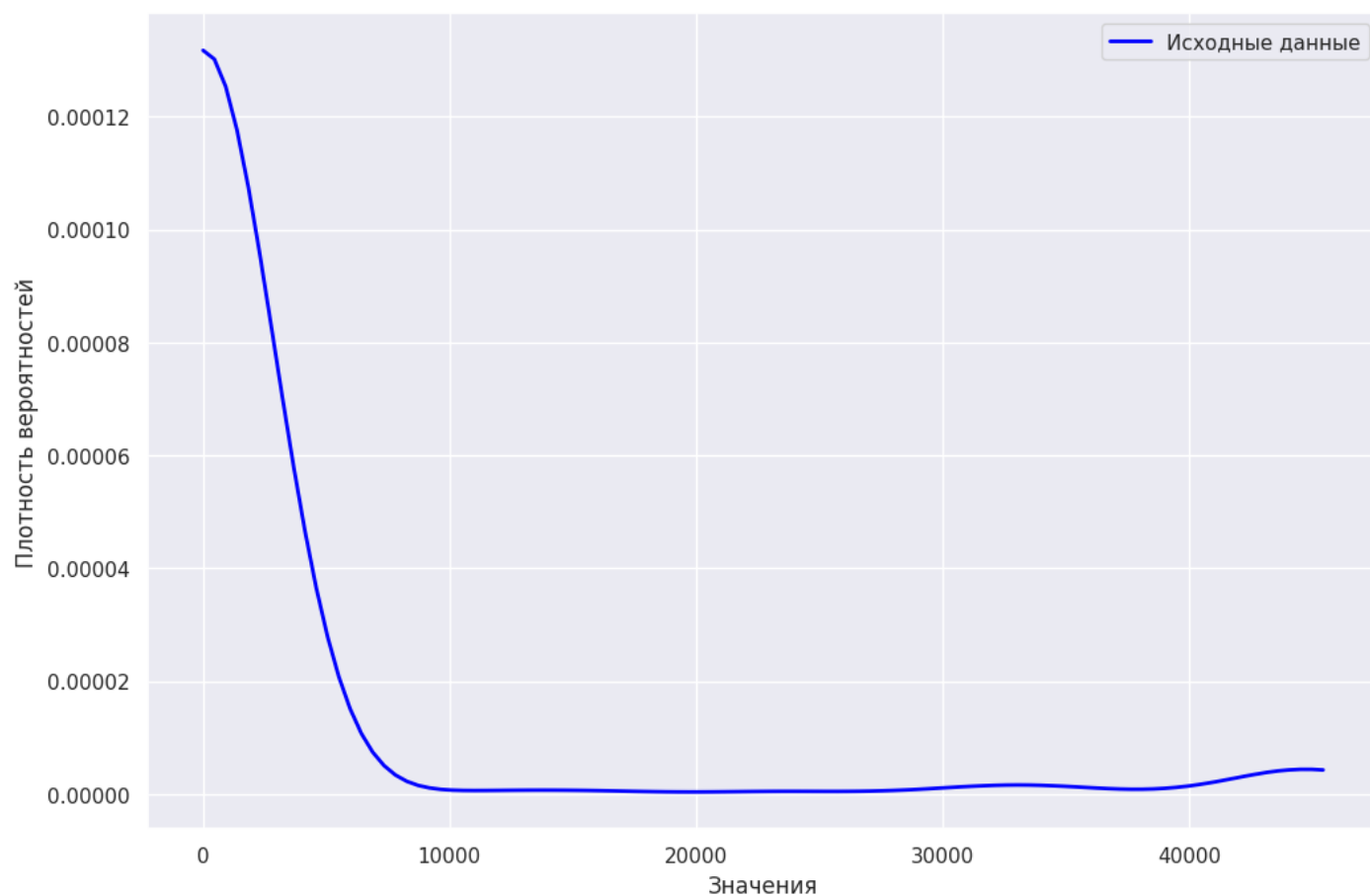
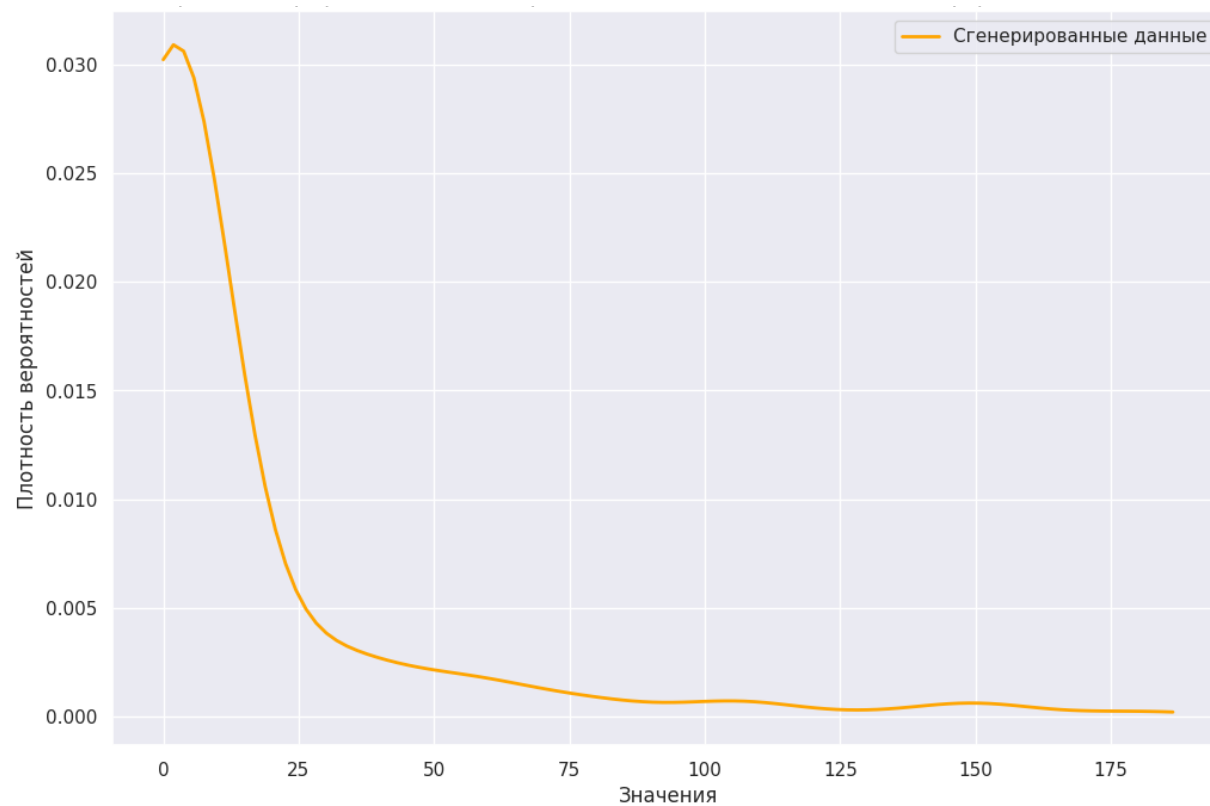
9. Сравнительный анализ заданной и сгенерированной последовательностей

Графики значений ЧП



Гистограммы распределения частот





Полученные графики сгенерированной последовательности обладают теми же характеристиками, что и графики заданной последовательности. Это позволяет говорить о том, что сгенерированная последовательность удовлетворяет закону гиперэкспоненциального распределения.

10. Оценка корреляционной зависимости заданной и сгенерированной последовательностей

Мы можем использовать функцию =КОРРЕЛ(диапазон1; диапазон2), чтобы вычислить коэффициент корреляции между заданной и сгенерированной последовательностями данных. Результат расчета коэффициента корреляции для этих двух последовательностей составляет **-0,02406308721**.

Это значение коэффициента корреляции невелико, что позволяет сделать вывод о том, что между числовыми последовательностями нет значительной зависимости или влияния друг на друга.

Выводы

В рамках выполненной лабораторной работы было проведено исследование числовой последовательности. Анализ последовательности включал в себя оценку числовых моментов, а также корреляционный анализ для выявления её особенностей. Кроме того, предпринимались попытки аппроксимации данной последовательности гиперэкспоненциальным законом распределения, основанным на трех числовых моментах случайной величины.

Был написан код, генерирующий случайную последовательность чисел по закону гиперэкспоненциального распределения. В рамках сравнительного анализа статистических характеристик и графического представления данных были проведены сопоставления между исходной числовой последовательностью и сгенерированной последовательностью.

На основе полученных выводов можно сделать заключение о том, что был выбран верный закон распределения для аппроксимации данной случайной величины. Это подтверждается тем, что оценки математического ожидания, доверительных интервалов и коэффициентов автокорреляции практически совпадают между исходной и сгенерированной последовательностями.

Полученные результаты позволили углубить понимание основных статистических характеристик данных, что может быть полезным для более точного анализа и интерпретации данных в будущем.