

УНИВЕРСИТЕТ ИТМО

Факультет программной инженерии и компьютерной техники

Направление подготовки 09.03.04 «Программная инженерия»

Дисциплина «Моделирование»

Учебно-исследовательская работа №1

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

Вариант №57

Выполнили:

Тучин Артём Евгеньевич

Герасимов Артём Кириллович

Емельянов Дмитрий Сергеевич

Группа: Р34111

Преподаватель:

Алиев Тауфик Измаилович

Санкт-Петербург, 2024 г.

Содержание

Содержание	2
Цель работы	3
Задание	3
Ход работы	5
Вычисление характеристик заданной ЧП.....	5
График значений заданной ЧП; определение характера ЧП	6
Автокорреляционный анализ заданной ЧП.....	7
Гистограмма распределения частот для заданной ЧП	8
Аппроксимацию закона распределения заданной ЧП.....	9
Генерация случайной ЧП в соответствии с аппроксимирующим законом распределения.....	10
Вычисление характеристик сгенерированной ЧП.....	10
График значений сгенерированной ЧП; определение характера ЧП	12
Автокорреляционный анализ сгенерированной ЧП.....	13
Гистограмма распределения частот для сгенерированной ЧП	14
Корреляционный анализ сгенерированной и заданной ЧП.....	15
Вывод.....	16

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.

Задание

В процессе исследований необходимо выполнить обработку заданной числовой последовательности (ЧП) для случаев, когда путем измерений получено 10, 20, 50, 100, 200 и 300 значений случайной величины, а именно:

- **рассчитать значения** следующих числовых моментов *заданной* числовой последовательности:
 - математическое ожидание;
 - дисперсию;
 - среднеквадратическое отклонение;
 - коэффициент вариации;
 - доверительные интервалы для оценки математического ожидания с доверительными вероятностями 0,9; 0,95 и 0,99;
 - относительные отклонения (в процентах) полученных значений от наилучших значений, полагая, что наилучшими (эталонными) являются значения, рассчитанные для наиболее представительной выборки из трехсот случайных величин;
- построить **график значений** для *заданной* числовой последовательности и определить ее характер, а именно: является эта последовательность *возрастающей/убывающей, периодичной* (при наличии периодичности оценить по графику длину периода);
- выполнить **автокорреляционный анализ** и определить, можно ли *заданную* числовую последовательность считать *случайной*;
- построить **гистограмму распределения частот** для *заданной* числовой последовательности;
- выполнить **аппроксимацию закона распределения** *заданной* случайной последовательности по двум начальным моментам, используя, в зависимости от значения коэффициента вариации, одно из следующих распределений:
 - равномерный;
 - экспоненциальный;
 - нормированный Эрланга k -го порядка или гипоекспоненциальный с заданным коэффициентом вариации;

- гиперэкспоненциальный с заданным коэффициентом вариации;
- реализовать **генератор случайных величин** в соответствии с *полученным аппроксимирующим законом* распределения (в EXCEL или программно) и проиллюстрировать на защите его работу;
- сгенерировать последовательность случайных величин с использованием реализованного **генератора** и рассчитать значения числовых моментов по аналогии с *заданной* числовой последовательностью;
- выполнить **автокорреляционный анализ** *сгенерированной* последовательности случайных величин;
- выполнить сравнительный анализ *сгенерированной* последовательности случайных величин с *заданной* последовательностью, построив соответствующие зависимости на **графике значений и гистограмме распределения частот**;
- оценить корреляционную зависимость *сгенерированной* и *заданной* последовательностей случайных величин.

Результаты проводимых исследований представить в виде таблиц и графиков.

На основе полученных промежуточных и конечных результатов следует сделать **обоснованные выводы** об исследуемой числовой последовательности, предложить **закон распределения** для её описания и **оценить качество аппроксимации** этим законом.

Ход работы

Расчёты выполнялись в сервисе Google Таблицы:

https://docs.google.com/spreadsheets/d/1S8jQQ47VdDHuLqPL3t7FcRm8MzWs5YyZqyltfhYX2_4

Вычисление характеристик заданной ЧП

Для заданной ЧП рассчитаны значения математического ожидания, дисперсии, среднеквадратическое отклонение, коэффициент вариации, доверительные интервалы для оценки МО, результаты представлены в таблице 1.

Таблица 1 – Характеристики заданной ЧП (вариант 57)

Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат.ож.	Знач.	17,11677	16,70282	21,85050	16,97765	14,94481	15,19542
	%	12,64%	9,92%	43,80%	11,73%	-1,65%	
Дов. инт. (0,9)	Знач.	±16,85660	±9,11560	±8,87386	±5,54867	±3,06370	±2,70951
	%	522,13%	236,43%	227,51%	104,78%	13,07%	
Дов. инт. (0,95)	Знач.	±20,10891	±10,87437	±10,58598	±6,61922	±3,65481	±3,23228
	%	522,13%	236,43%	227,51%	104,78%	13,07%	
Дов. инт. (0,99)	Знач.	±26,42885	±14,29203	±13,91300	±8,69955	±4,80347	±4,24814
	%	522,13%	236,43%	227,51%	104,78%	13,07%	
Дисперсия	Знач.	1052,6035 4	615,63857	1458,5454 2	1140,5173 4	695,42141	815,88164
	%	29,01%	-24,54%	78,77%	39,79%	-14,76%	
С.к.о.	Знач.	32,44385	24,81207	38,19091	33,77155	26,37084	28,56364
	%	13,58%	-13,13%	33,70%	18,23%	-7,68%	
К-т вариации	Знач.	1,89544	1,48550	1,74783	1,98918	1,76455	1,87975
	%	0,83%	-20,97%	-7,02%	5,82%	-6,13%	

С увеличением выборки оценки становятся более приближенны к истинным значениям, но так как это выборки и достаточно небольшие по размерам, то прямую зависимость

характеристик от размера выборки увидеть сложно. Однозначно можем увидеть, что с увеличением выборки уменьшается доверительный интервал, а с увеличением доверительной вероятности он увеличивается.

График значений заданной ЧП; определение характера ЧП

График значений ЧП представлен на рисунке 1.



Рисунок 1 – График значений заданной ЧП

По графику можем сказать, что числовая последовательность неубывающая и невозрастающая, а также не имеет периода.

Автокорреляционный анализ заданной ЧП

Результаты автокорреляционного анализа заданной ЧП представлены в таблице 2 и на рисунке 2.

Таблица 2 – Коэффициенты автокорреляции

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. ЧП	0,05691	-0,07125	-0,02902	-0,04272	-0,04757	-0,02014	0,17891	-0,04692	-0,00907	0,00766

Так как все коэффициенты близки к 0, то можем сделать вывод о том, что последовательность является случайной: нет взаимосвязи как между ближайшими соседями в последовательности, так и между значениями на расстоянии 1, 2,... значения.



Рисунок 2 – График значений коэффициентов автокорреляции заданной ЧП

Гистограмма распределения частот для заданной ЧП

Распределение частот заданной ЧП представлено в виде гистограммы на рисунке 3.



Рисунок 3 – Гистограмма распределения частот

По гистограмме можем увидеть, что большинство значений последовательности сосредоточены в левой части, а именно в интервале от 0,05 до 16,70. Примерно такое же значение будет и у математического ожидания, что мы можем увидеть в Таблице 1. Также исходя из распределения частот, так как есть немногочисленные значения в других интервалах, то становится понятно изменение доверительных интервалов с увеличением выборки (чем больше выборка, тем распределение частот больше похоже на данное, и больше значений находится в левой части, из-за чего будет сужаться влево и доверительный интервал).

Аппроксимацию закона распределения заданной ЧП

Для аппроксимации нам понадобятся вычисленные оценки математического ожидания (15,19542) и коэффициента вариации (1,87975). Так как коэффициент вариации больше 1, то используем гиперэкспоненциальное распределение.

Для начала необходимо определить вероятность q по формуле (1), с которой будут выбираться значения одного из двух экспоненциальных распределений:

$$q \leq \frac{2}{1+v^2}, q = 0,4 \quad (1)$$

Далее рассчитываем математические ожидания двух экспоненциальных распределений t_1 и t_2 по формулам (2) и (3):

$$t_1 = \left(1 + \sqrt{\frac{1-q}{2q}}(v^2 - 1)\right) \cdot t = 36,14144 \quad (2)$$

$$t_2 = \left(1 - \sqrt{\frac{q}{2(1-q)}}(v^2 - 1)\right) \cdot t = 1,23140 \quad (3)$$

И определяем параметры для генераторов экспоненциально распределённых случайных величин a_1 и a_2 по формулам (4) и (5):

$$a_1 = \frac{1}{t_1} = 0,02767 \quad (4)$$

$$a_2 = \frac{1}{t_2} = 0,81208 \quad (5)$$

Таким образом, мы выполнили аппроксимацию закона распределения с коэффициентом вариации $v > 1$ двухфазным гиперэкспоненциальным распределением. Мы получили три параметра q, a_1, a_2 , с помощью которых можем сделать генератор случайных величин.

Генерация случайной ЧП в соответствии с аппроксимирующим законом распределения

Для генерации нашей последовательности нам понадобится два равномерно распределённых генератора от 0 до 1. Первый будем использовать для определения того, какой из параметров использовать a_1 или a_2 в соответствии с вероятностью q . Далее используя второй генератор, получим значение соответствующего экспоненциального распределения с нужным параметром следующим образом:

$$x_e = -\frac{\ln(x)}{a}, \text{ где } x - \text{сгенерированное значение от } 0 \text{ до } 1$$

Для генерации значений от 0 до 1 использовалась функция СЛЧИС() в Google Таблицах.

Вычисление характеристик сгенерированной ЧП

Для сгенерированной ЧП рассчитаны значения математического ожидания, дисперсии, среднеквадратическое отклонение, коэффициент вариации, доверительные интервалы для оценки МО, результаты представлены в таблице 3.

Таблица 3 – Характеристики сгенерированной случайной ЧП

<u>Закон распределения: гиперэкспоненциальный с заданным коэффициентом вариации</u>							
Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат.ож.	Знач.	3,21194	8,50440	13,97435	12,57601	16,59989	15,25757
	%	-78,86%	-44,03%	-8,04%	-17,24%	9,24%	0,41%
Дов. инт. (0,9)	Знач.	±2,46661	±5,80102	±6,18954	±4,31581	±3,57137	±2,68185
	%	-8,96%	114,10%	128,44%	59,28%	31,81%	-1,02%
Дов. инт. (0,95)	Знач.	±2,94251	±6,92027	±7,38375	±5,14850	±4,26043	±3,19929
	%	-8,96%	114,10%	128,44%	59,28%	31,81%	-1,02%
Дов. инт. (0,99)	Знач.	±3,86730	±9,09521	±9,70436	±6,76661	±5,59942	±4,20478
	%	-8,96%	114,10%	128,44%	59,28%	31,81%	-1,02%
Дисперсия	Знач.	22,53847	249,3239 3	709,5972 2	690,0014 6	944,9846 5	799,3119 9
	%	-97,24%	-69,44%	-13,03%	-15,43%	15,82%	-2,03%
С.к.о.	Знач.	4,74747	15,78999	26,63827	26,26788	30,74060	28,27211
	%	-83,38%	-44,72%	-6,74%	-8,04%	7,62%	-1,02%
К-т вариации	Знач.	1,47807	1,85669	1,90623	2,08873	1,85186	1,85299
	%	-21,37%	-1,23%	1,41%	11,12%	-1,48%	-1,42%

Так же, как и в случае заданной ЧП какой-либо прямой зависимости не наблюдается, так как это выборки. Можем сравнить наиболее большие выборки - они отличаются примерно на 1%, также мат. ожидание попадает в доверительный интервал и сами значения мат. ожидания очень близки. Можем сделать вывод, что характеристики числовых последовательностей похожи, и аппроксимация сделана корректно.

График значений сгенерированной ЧП; определение характера ЧП

Графики значений заданной и сгенерированной ЧП представлены на рисунке 4.



Рисунок 4 – График значений ЧП

Сами графики немного отличаются (разные значения в числовых последовательностях), но о характере последовательности можно сказать то же самое: невозрастающая, неубывающая и непериодическая.

Автокорреляционный анализ сгенерированной ЧП

Результаты автокорреляционного анализа сгенерированной ЧП представлены в таблице 4 и на рисунке 5.

Таблица 4 – Коэффициенты автокорреляции

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. ЧП	0,05691	-0,07125	-0,02902	-0,04272	-0,04757	-0,02014	0,17891	-0,04692	-0,00907	0,00766
К-т АК для сгенерир. ЧП	-0,05689	-0,05358	0,01738	0,01417	0,07164	0,0233	0,07856	-0,03964	0,00029	-0,00981
%	-199,97%	-24,79%	-159,90%	-133,18%	-250,61%	-215,67%	-56,09%	-15,50%	-103,21%	-228,20%

Для сгенерированной последовательности коэффициенты автокорреляции также близки к 0, что говорит о случайности последовательности. Относительное отклонение между коэффициентами двух последовательностей достаточно большое, но так как сами значения малы, то обе последовательности случайны. Разница между коэффициентами возможна из-за того, что это выборки и значения в них различаются.

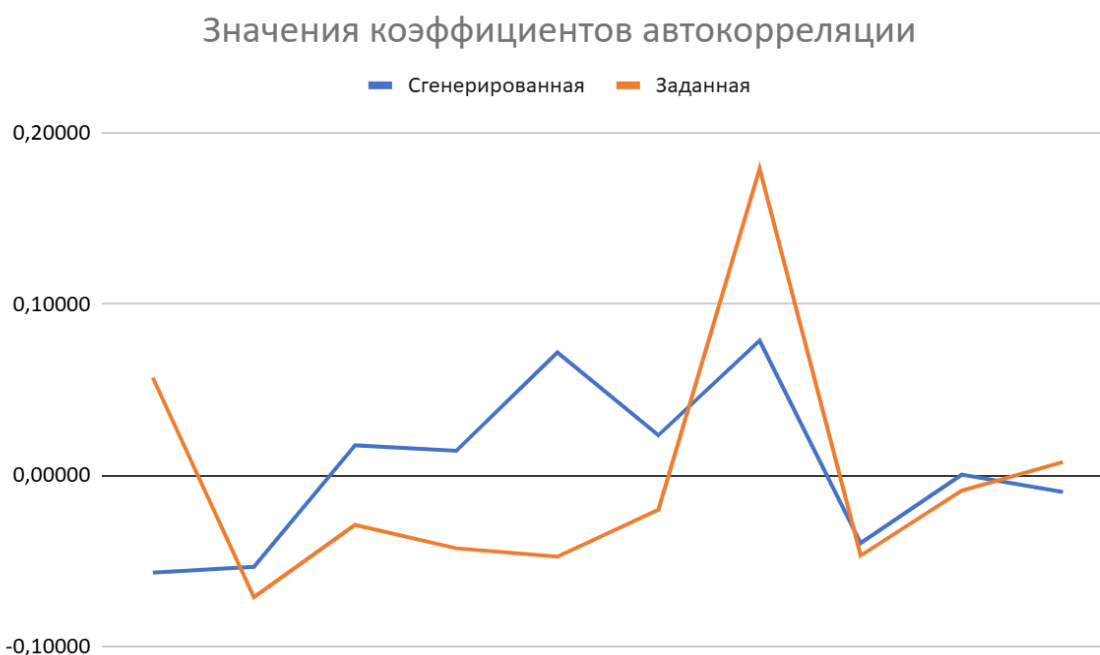


Рисунок 5 – Значения коэффициентов автокорреляции ЧП

Гистограмма распределения частот для сгенерированной ЧП

Распределение частот заданной и сгенерированной ЧП представлены в виде гистограммы на рисунке 6.

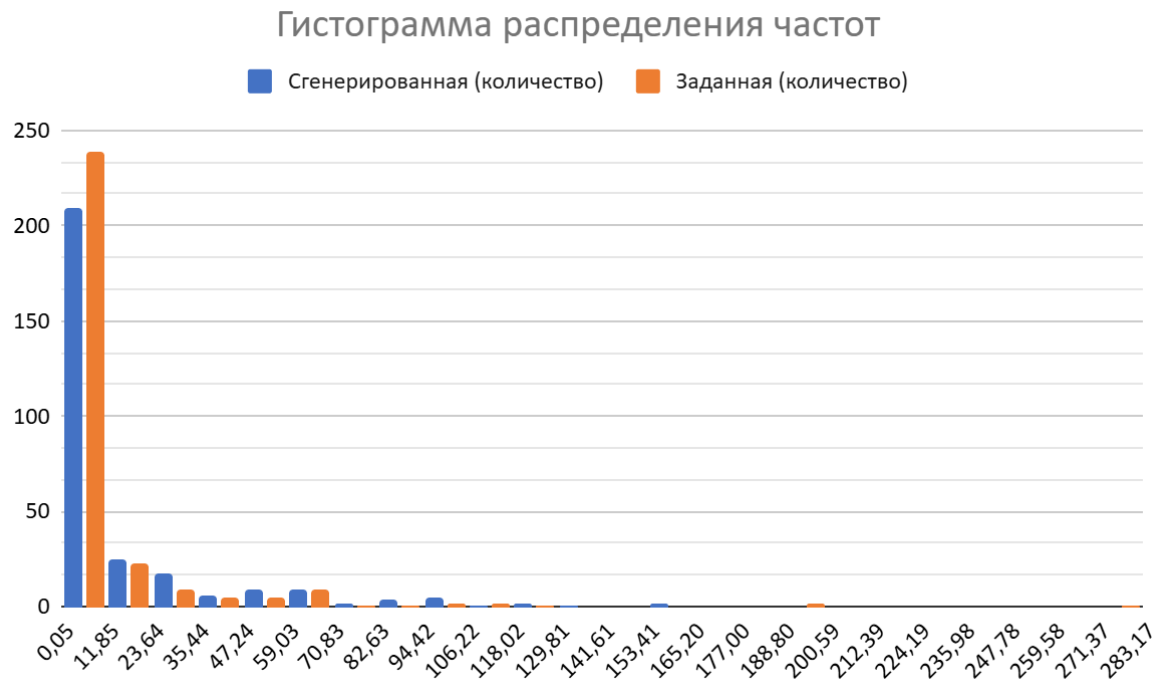


Рисунок 6 – Гистограмма распределения частот

Гистограмма для сгенерированной ЧП похожа на гистограмму для заданной ЧП, также большинство значений находится в левой стороне.

Плотность распределения полученного аппроксимирующего закона представлена на рисунке 7.

Плотность распределения аппроксимирующего закона

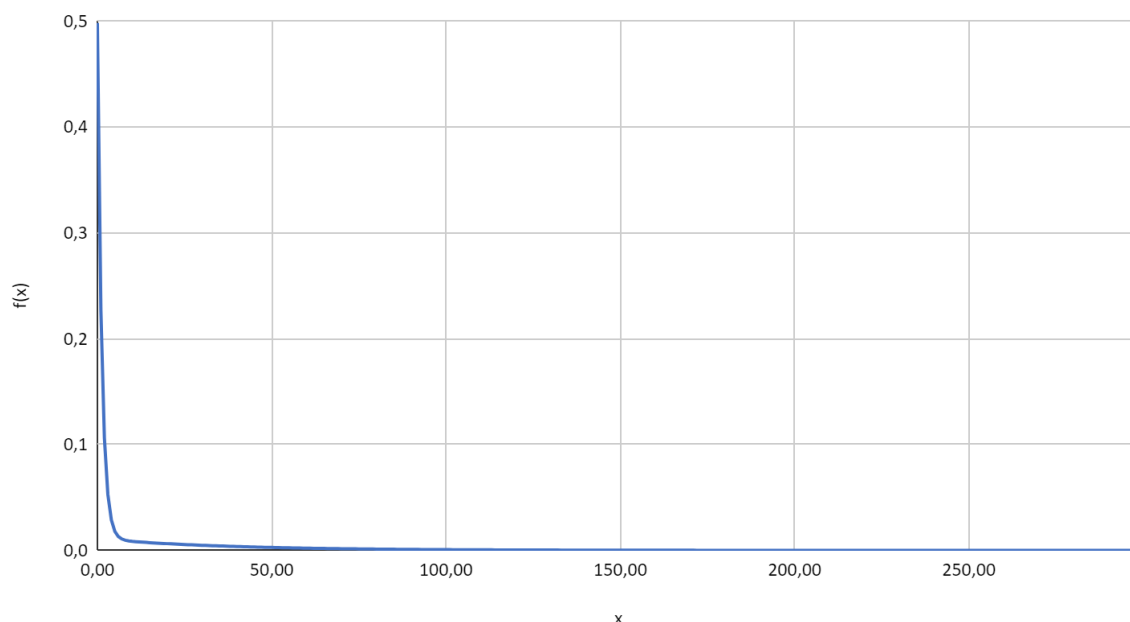


Рисунок 7 – График плотности распределения аппроксимирующего закона

По графику плотности распределения аппроксимирующего закона видно, что большинство значений расположены от 0 до 50, что соответствует полученным гистограммам числовых последовательностей.

Корреляционный анализ сгенерированной и заданной ЧП

Для проведения анализа отсортируем полученные последовательности и вычислим для них коэффициент корреляции по формуле (6): 0,9173535587. Коэффициент корреляции получился близким к 1, что говорит о схожем распределении значений в последовательностях.

Также проведём анализ без сортировки. В таком случае получаем коэффициент корреляции по формуле (6): 0,0614500403. Значение очень близко к нулю, что показывает отсутствие связи между значениями в последовательностях, что верно для последовательностей значений случайных величин.

$$r_{XY} = \frac{cov_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n (x_i - M[X])(y_i - M[Y])}{\sqrt{\sum_{i=1}^n (x_i - M[X])^2 \sum_{i=1}^n (y_i - M[Y])^2}} \quad (6)$$

Вывод

Выполнив данную работу, мы провели статистический анализ заданной числовой последовательности путём оценки числовых моментов и определили случайность последовательности на основе корреляционного анализа. Также выполнили аппроксимацию закона распределения заданной числовой последовательности по двум числовым моментам, сгенерировали числовую последовательность по аппроксимирующему закону и сравнили её с заданной последовательностью. Полученная последовательность оказалось схожа с заданной, характеристики находятся в пределах доверительных интервалов.