

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский университет ИТМО»

ФАКУЛЬТЕТ ПРОГРАММНОЙ ИНЖЕНЕРИИ И КОМПЬЮТЕРНОЙ ТЕХНИКИ

УЧЕБНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА №1

«Обработка результатов измерений: статистический анализ
числовой последовательности »

по дисциплине

«Моделирование »

Вариант №28

Выполнил работу:

Студент группы Р3318

Рамеев Тимур Ильгизович

Преподаватель:

Авксентьева Елена

Юрьевна

Содержание

Цель	2
Выполнение	2
Расчет статистических характеристик заданной числовой последовательности .	2
Построение графика значений для заданной числовой последовательности . . .	4
Построение гистограммы частот	5
Выполнение автокорреляционного анализа	6
Выполнение аппроксимации закона распределения	6
Реализация генератора случайных величин	7
Выводы	8

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

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.

$$\varepsilon_p = t_\alpha \cdot \frac{\tilde{\sigma}_m}{\sqrt{n}}$$

Коэффициент вариации - это статистический показатель, который измеряет относительную изменчивость (разброс) данных по отношению к их среднему значению.

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

Таблица 1: Характеристики заданной ЧП

Характеристика		10	20	50	100	200	300
Мат. ожидание	Значение	11,02	14,58	24,19	25,75	25,56	25,95
	%	57,53	43,82	6,76	0,75	1,47	
Дов. интервал (0,9)	Значение	8,12	11,10	15,22	12,68	8,74	7,30
	%	11,18	52,11	108,45	73,71	19,68	
Дов. интервал (0,95)	Значение	9,68	13,25	18,15	15,13	10,42	8,71
	%	11,18	52,11	108,45	73,71	19,68	
Дов. интервал (0,99)	Значение	12,72	17,41	23,86	19,88	13,70	11,44
	%	11,18	52,11	108,45	73,71	19,68	
Дисперсия	Значение	61,00	228,35	1072,06	1489,05	1413,68	1480,35
	%	95,88	84,57	27,58	0,59	4,50	
Ср. кв. отклонение	Значение	7,81	15,11	32,74	38,58	37,59	38,48
	%	79,70	60,72	14,90	0,29	2,28	
К-т вариации	Значение	0,71	1,04	1,35	1,50	1,47	1,48
	%	52,21	30,09	8,73	1,05	0,81	

% - 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. Однако, стоит отметить, что бывают исключения - не всегда при росте выборки мы будем наблюдать более точные значения статистик. Наша таблица это наглядно демонстрирует.

Оценка математического ожидания показывает ожидаемое среднее значение случайной величины.

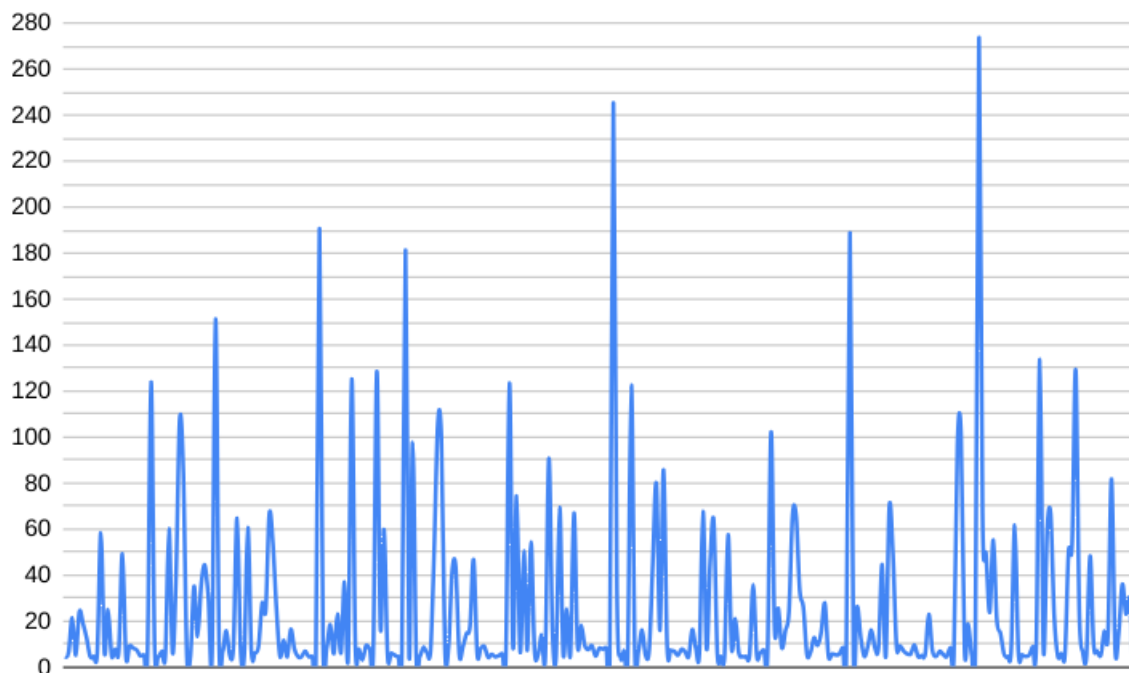
Доверительный интервал характеризует диапазон значений, в котором с определенной вероятностью находится случайная величина. Видно, что с увеличением доверительной вероятности доверительный интервал увеличивается - чем больше требование к вероятности попадания случайно величин в доверительный интервал, 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.

Построение графика значений для заданной числовой последовательности

Обратимся к построенному графику значений заданной числовой последовательности.

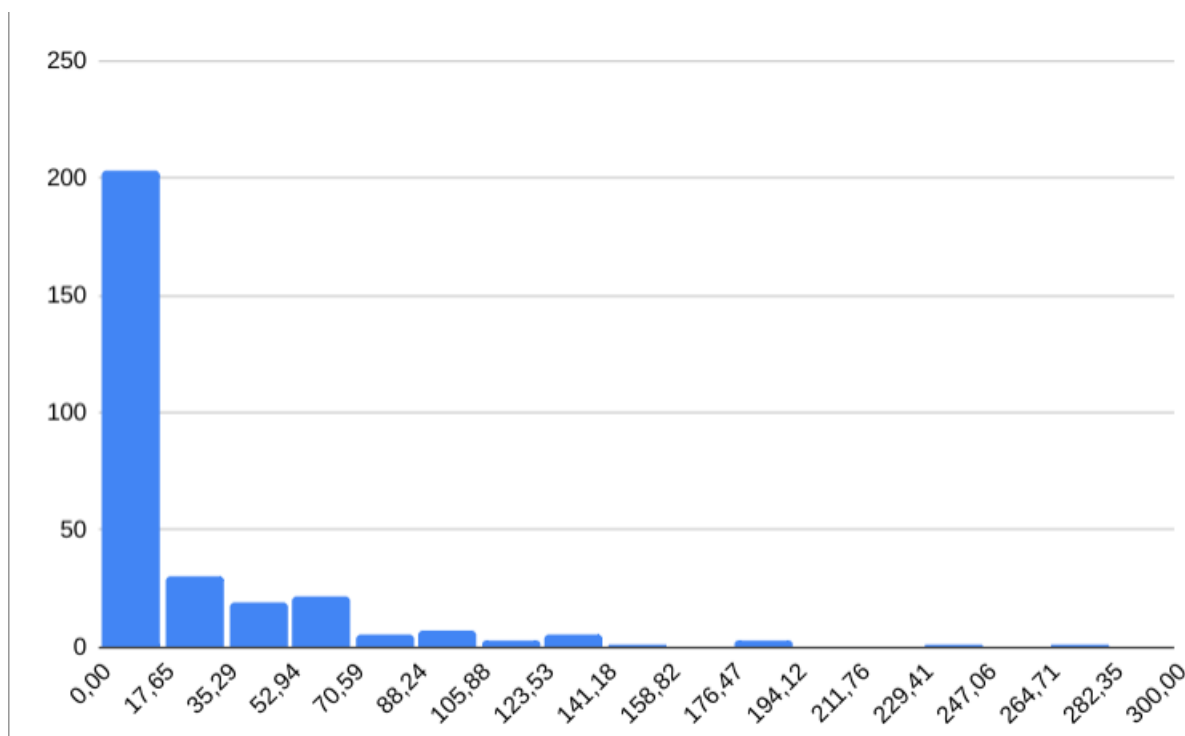


Проанализировав данный график, можно сделать выводы о характере заданной числовой последовательности:

- 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.

Построение гистограммы частот



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 спад. Это значит, что вероятность появления маленьких значений случайной величины для гиперэкспоненциального распределения намного больше вероятности появления больших значений.

Выполнение автокорреляционного анализа

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.

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для заданной ЧП	- 0,0131	0,0069	- 0,0969	- 0,0948	- 0,0109	0,0051	- 0,0003	- 0,0002	0,0846	0,0226

Коэффициенты автокорреляции заданной числовой последовательности для 9 сдвигов меньше 0.1, что позволяет сказать, что последовательность случайна.

Выполнение аппроксимации закона распределения

Рассчитанный коэффициент вариации случайной величины $v = 1,48 > 1$

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 больше 1 двухфазным гиперэкспоненциальным распределением следует выбрать следующее значение вероятности:

$$q \leq \frac{2}{1+v^2} = 0,62688$$

Пусть $q = 0.6$ (вероятность формирования значения случайной величины в первой фазе), тогда математические ожидания первой и второй экспоненциальных фаз будут равны соответственно рассчитанным t_1 и t_2 :

$$t_1 = \left[1 + \sqrt{\frac{1-q}{2q}}(v^2 - 1) \right] \cdot t = 42,348805$$

$$t_2 = \left[1 - \sqrt{\frac{q}{2(1-q)}}(v^2 - 1) \right] \cdot t = 1,341817457$$

Где t - математическое ожидание.

Выводы

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

Был написан код, где.... (Вывод дописать)