

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧЕРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
“НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО”
Факультет ПИИКТ



УНИВЕРСИТЕТ ИТМО

ОТЧЁТ

По лабораторной работе:

“Обработка результатов измерений: статистический анализ числовой последовательности”

По предмету: Моделирование

Вариант 20

Студент:

Андрейченко Леонид Вадимович

Группа Р34301

Преподаватель:

Алиев Тауфик Измайлович

Санкт-Петербург

2023

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 EXEL или программно) и проиллюстрировать на защите его работу;
- сгенерировать последовательность случайных величин в соответствии с полученным законом распределения 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 ее описания и оценить качество аппроксимации этим законом.

Выполнение

Пункт 1

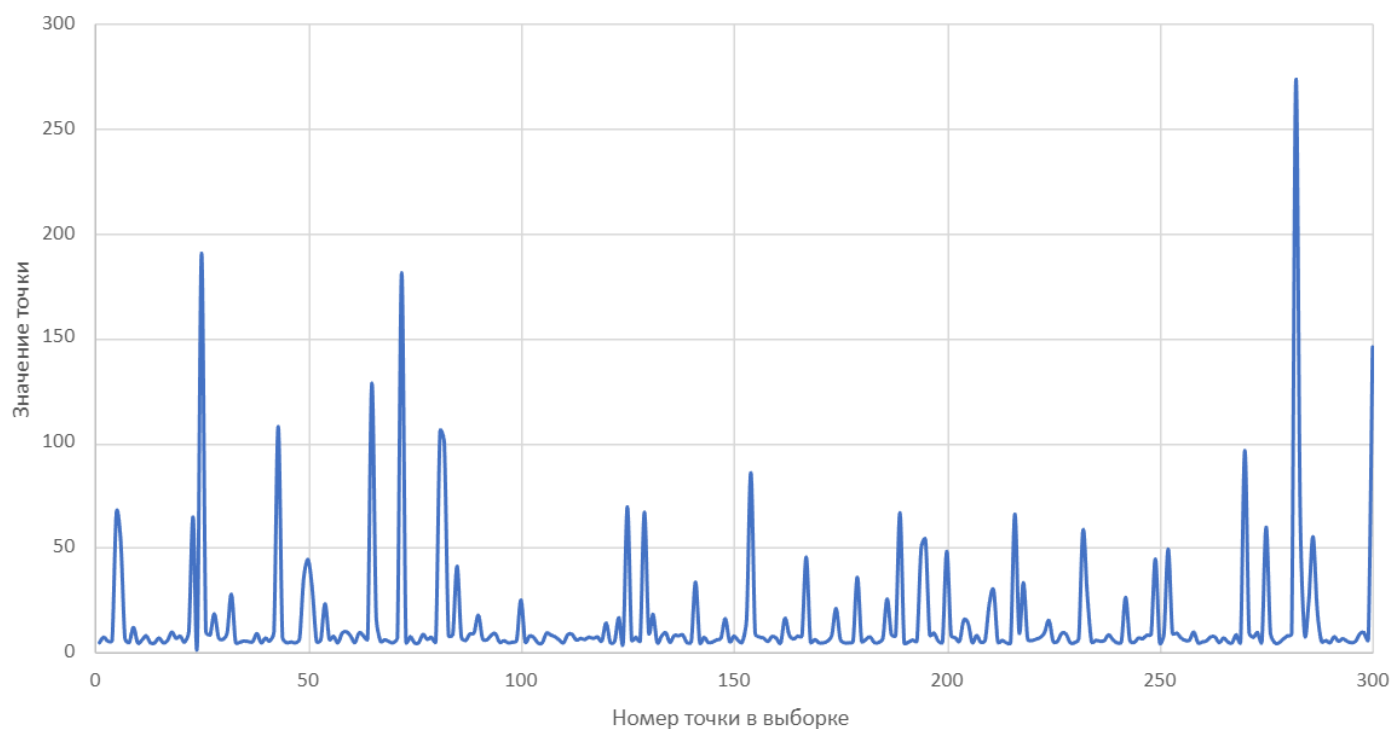
Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. Ож.	Знач	17,15	11,82	17,50	18,12	15,19	15,70
	%	9	25	12	15	3	
Дов. Инт. (0.9)	Знач	11,98	6,18	7,46	5,47	3,04	2,72
	%	340	127	174	101	12	
Дов. Инт. (0.95)	Знач	14,27	7,37	8,89	6,52	3,62	3,25
	%	340	127	174	101	12	
Дов. Инт. (0.99)	Знач	18,75	9,68	11,68	8,57	4,76	4,27
	%	340	127	174	101	12	
Дисперсия	Знач	530,15	282,43	1028,68	1107,91	683,24	823,30
	%	36	66	25	35	17	
С.К.О.	Знач	23,02	16,81	32,07	33,29	26,14	28,69
	%	20	41	12	16	9	
К-т вариации	Знач	1,34	1,42	1,83	1,84	1,72	1,83
	%	27	22	0	0	6	

Чем большее количество значений мы берем, тем более точные числовые характеристики.

- Дисперсия и СКО резко увеличились после того, как мы взяли выборку > 50 , это значит, что в ней присутствуют числа с большим “разбросом”.
- К-т вариации с самого начала был больше 1, и на протяжении всех выборок таким и остался.

Пункт 2

График значений



Изучив график, можно сделать вывод, что исходная последовательность не является периодической, возрастающей или убывающей.

Пункт 3

Коэффициенты автокорреляции

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. ЧП	0,06	-0,03	-0,04	-0,02	0,00	-0,02	0,09	-0,07	-0,01	-0,02
К-т АК для сгенерир. ЧП	-0,01	0,01	-0,06	-0,09	0,02	-0,03	-0,05	0,07	0,06	0,04
%	657%	307%	55%	23%	23%	55%	194%	98%	14%	39%



Последовательность можно считать случайной так как данные коэффициенты указывают на то, что между числами не было выявлено зависимости, нет тенденции и периодичности. Коэффициент автокорреляции всегда был меньше чем 0.1.

Пункт 4



По гистограмме мы можем видеть, что большая часть значений располагается в промежутке от 0 до 31, ещё часть располагается до 112 и всего пара значений располагается в диапазоне более 112.

Пункт 5

Для данной выборки коэффициент вариации получился больше единицы. Это значит, что для аппроксимации данного ряда, мы будем использовать гиперэкспоненциальное распределение. Значения мат ожидания и дисперсии были получены ранее. $t = 15.7$, $v = 1.83$

Данное распределение имеет три независимых параметра q α_1 α_2 ($0 < q < 1$; $\alpha_1 \geq 0$; $\alpha_2 \leq 0$)

$$q \leq \frac{2}{1 + v^2} \approx 0.46$$

$q = 0.2$

$$t_1 = t \cdot \left[1 + \sqrt{\frac{1-q}{2q}} (v^2 - 1) \right] = 15.7 \cdot \left(1 + \sqrt{\frac{1-0.2}{2 \cdot 0.2}} (1.83^2 - 1) \right) = 49.7$$

$$t_2 = t \cdot \left[1 - \sqrt{\frac{q}{2(1-q)}} (v^2 - 1) \right] = 15.7 \cdot \left(1 - \sqrt{\frac{0.2}{2 \cdot (1-0.2)}} (1.83^2 - 1) \right) = 7.2$$

Пункт 6

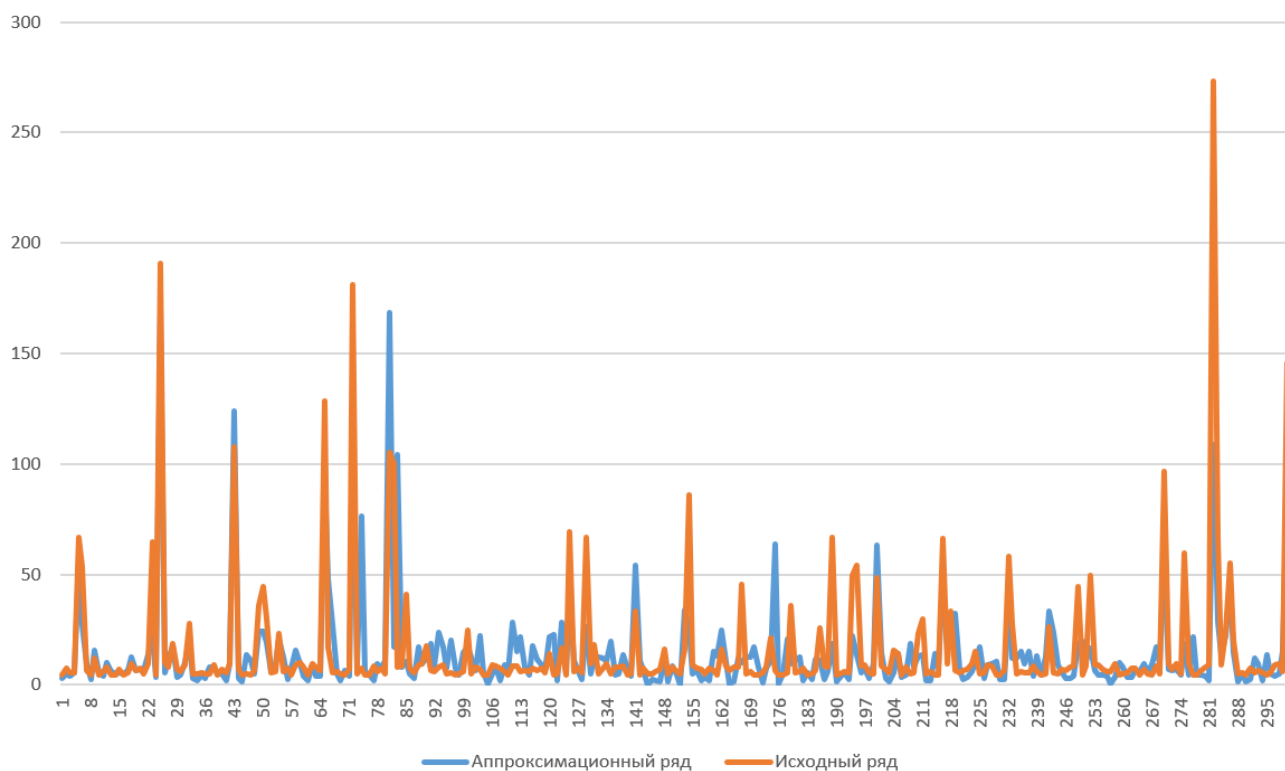
Для построения гиперэкспоненциальной функции было принято решение реализовывать ее в Excel.

- Для этого я использовал функцию слчисл(). Данная функция возвращает случайное рациональное число в диапазоне от 0 до 1.
- Далее применив эту функцию к двум столбцам каждый по 300 клеток в длину я получил две случайные выборки.
- Первый столбец будет использоваться для определения того, в какую функцию попадет число $q = 0.2$
- Далее я брал число из второго столбца и “приводил” его к выбранному мат ожиданию

E2					$\text{=ЕСЛИ(B2<H\$2;H\$3;H\$4)*(-LN(1-C2))}$			
	A	B	C	D	E	F	G	H
1	i			Исходный ряд	Аппроксимационный ряд			
2	1	0,497842	0,326621	4,66	$\text{=ЕСЛИ(B2<H\$2;H\$3;H\$4)*(-LN(1-C2))}$			0,2
3	2	0,227065	0,506818	7,35	5,089521491		t1 =	49,7
4	3	0,21	0,435669	5,39	4,119219263		t2 =	7,2
5	4	0,739744	0,55	5,7	5,749255413		alpha1 =	0,02
6	5	0,15	0,67	66,73	55,10053244		alpha2 =	0,14

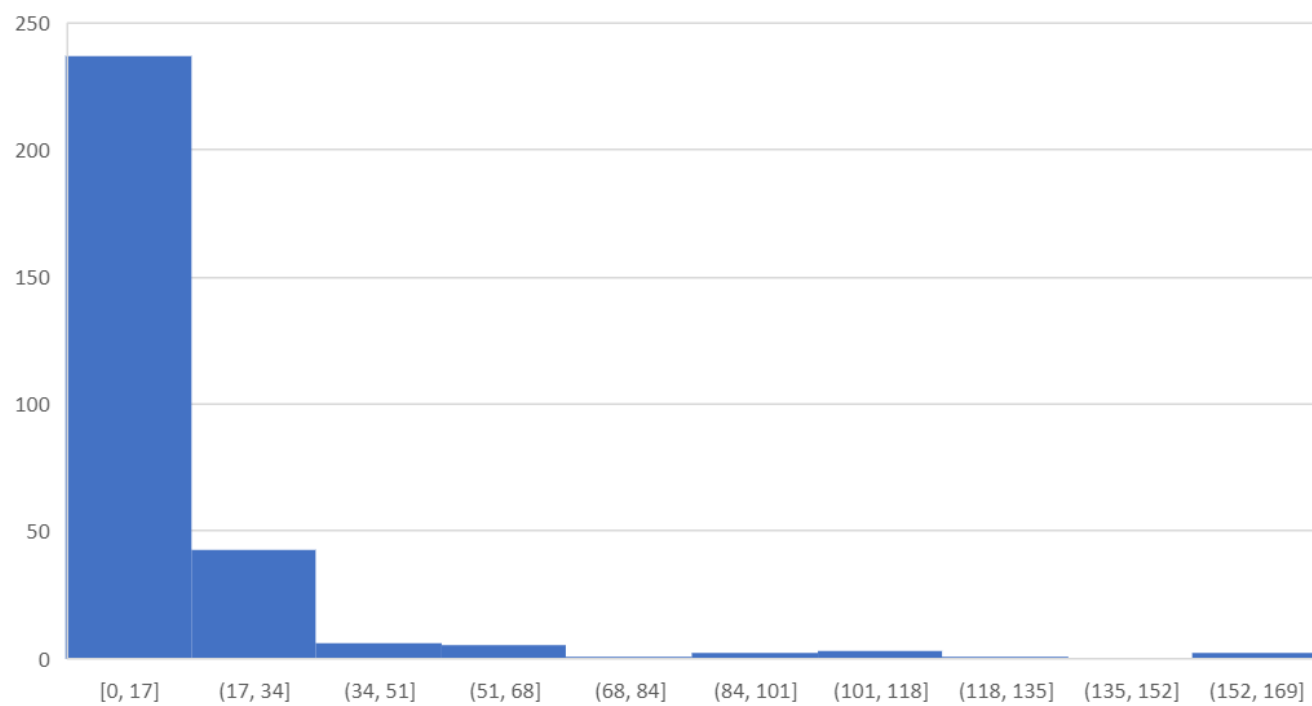
Пункт 7

Графики значений ф-й



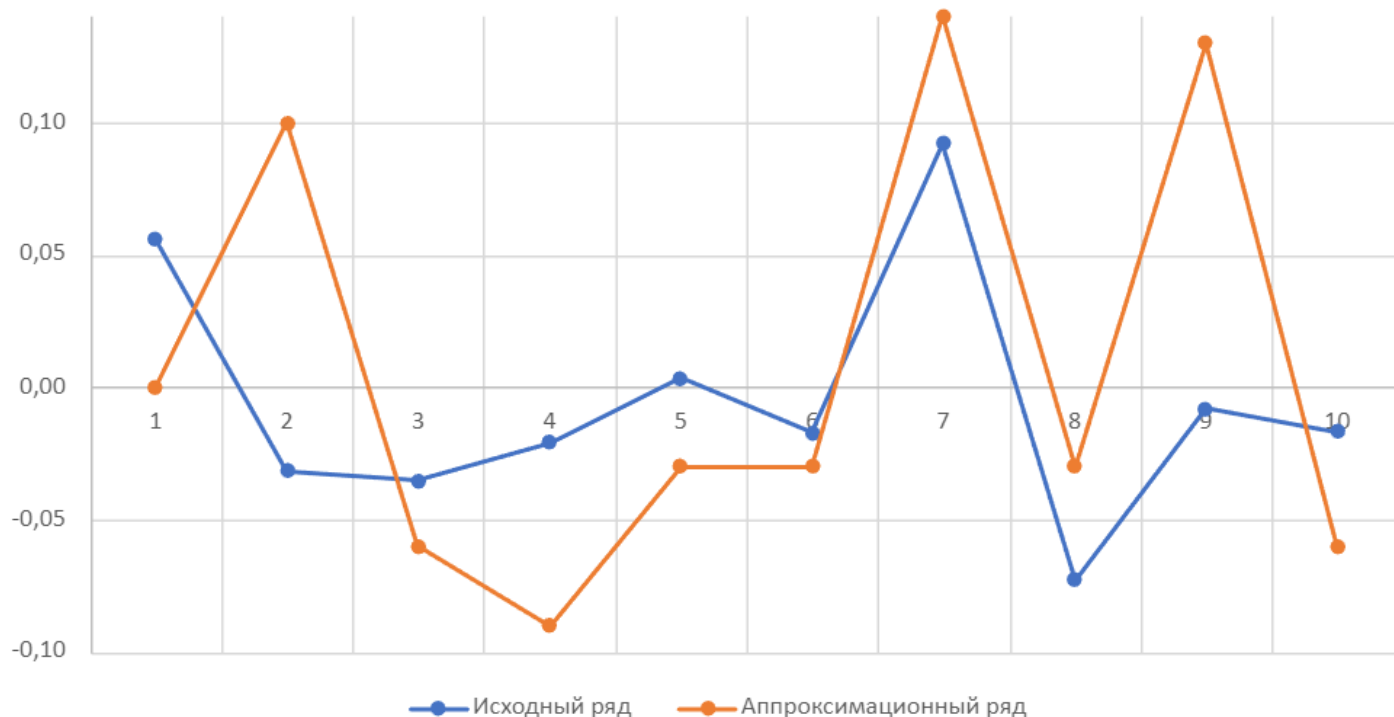
Как видно на графике полученный ряд, визуально, довольно хорошо аппроксимирует исходный.

Гистограмма распределения частот



Гистограмма распределения также схожа и исходной.

К-т АК относительно параметра сдвиг ЧП



К-т автокорреляции расходятся не сильно, а на некоторых участках ведут себя аналогично (5 - 8).

Характеристики сгенерированной случайной ЧП, для гиперэкспоненциальной функции

Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. Ож.	Знач	13,37	10,14	14,96	17,93	14,67	13,99
	%	22%	14%	15%	1%	3%	11%
Дов. Инт. (0.9)	Знач	8,61	4,41	6,72	5,09	2,75	2,03
	%	28,11%	28,59%	9,91%	7,02%	9,68%	25,51%
Дов. Инт. (0.95)	Знач	10,26	5,26	8,01	6,07	3,27	2,42
	%	28,11%	28,59%	9,91%	7,02%	9,68%	25,51%
Дов. Инт. (0.99)	Знач	13,48	6,91	10,53	7,97	4,30	3,18
	%	28,11%	28,59%	9,91%	7,02%	9,68%	25,51%
Дисперсия	Знач	273,98	144,03	834,98	957,80	557,39	456,81
	%	48%	49%	19%	14%	18%	45%
С.К.О.	Знач	16,55	12,00	28,90	30,95	23,61	21,37
	%	28,11%	28,59%	9,91%	7,02%	9,68%	25,51%
К-т вариации	Знач	1,24	1,18	1,93	1,73	1,61	1,53
	%	7,79%	16,80%	5,43%	6,03%	6,44%	16,42%

Из вычисленных данных видно, что финальная выборка на 300 элементов по своим числовым показателям примерно соответствует исходной последовательности.

Вычислим коэффициент корреляции двух рядов используя формулу:

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

Вычисленное значение коэффициента корреляции равно 0.77, что говорит о том, что полученная выборка довольно хорошо аппроксимирует исходную.

Выводы

В рамках лабораторной работы была дана числовая последовательность, для которой я определил математическое ожидание, дисперсию и другие параметры. Далее я проанализировал построенную гистограмму, по которой не было выявлено возрастания, убывания или периодичности последовательности. Исследуемую последовательность можно назвать случайной исходя из автокорреляционного анализа. Затем я вычислил параметры аппроксимирующего закона и по ним сгенерировал новую последовательность. Коэффициент вариации первой и второй последовательности приближен к двум, Коэффициент автокорреляции первой и второй последовательности варьируется около нуля, исходя из этого можно сказать то, что выборка случайна.

Математическое ожидание и дисперсия отличаются, но отличие не выходит за пределы доверительных интервалов.