

Федеральное государственное автономное
образовательное учреждение высшего образования
«Научно-образовательная корпорация ИТМО»

Факультет программной инженерии и компьютерной техники
Направление подготовки 09.03.04 Программная инженерия

Отчёт по лабораторной работе № 1
По дисциплине «Моделирование» (семестр 5)

Студенты:

Дениченко Александр Р3312

Балин Артём Р3312

Кобелев Роман Р3312

Практик:

Мартынчук Илья Геннадьевич

Санкт-Петербург
2024 г.

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.

Порядок выполнения работы

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 считать 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.

1 Оценки и доверительные интервалы

Таблица 1: Расчетные характеристики

Характеристика	Формула
Оценка мат ожидания	$\tilde{m} = \frac{1}{n} \sum_{i=1}^n X_i$
Оценка дисперсии	$\tilde{D} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \tilde{m})^2$
Среднеквадратическое отклонение	$\sigma = \sqrt{\tilde{D}}$
Коэффициент вариации	$V = \frac{\sigma}{\tilde{m}} \times 100\%$
Оценка среднеквадратического отклонения мат ожидания	$\tilde{\sigma}_m = \sqrt{\frac{\tilde{D}}{n}}$
Отклонение	$\epsilon_p = t_p \cdot \tilde{\sigma}_m, t_p = \Phi'(\frac{1+p}{2})$
Доверительный интервал	$m \pm \epsilon_p$
Доверительный интервал (0.9)	$m \pm 1.643 \cdot \tilde{\sigma}_m$
Доверительный интервал (0.95)	$m \pm 1.960 \cdot \tilde{\sigma}_m$
Доверительный интервал (0.99)	$m \pm 2.576 \cdot \tilde{\sigma}_m$

Таблица 2: Форма 1. Характеристики заданной ЧП (вариант 2)

Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. ож.	Знач.	147.844	181.747	146.786	165.879	173.955	168.836
	%	-12.433	7.647	-13.06	-1.751	3.032	
Дов. инт. (0,9)	Знач.	± 30.879	± 46.27	± 21.414	± 16.096	± 13.063	± 11.039
	%	179.712	319.132	93.982	45.807	18.332	
Дов. инт. (0,95)	Знач.	± 36.836	± 55.197	± 25.546	± 19.202	± 15.584	± 13.169
	%	179.712	319.132	93.982	45.807	18.332	
Дов. инт. (0,99)	Знач.	± 48.413	± 72.545	± 33.575	± 25.237	± 20.481	± 17.308
	%	179.712	319.132	93.982	45.807	18.332	
Дисперсия	Знач.	3532.136	15861.648	8493.935	9597.771	12643.022	13543.708
	%	-73.92	17.115	-37.285	-29.135	-6.65	
С.к.о.	Знач.	59.432	125.943	92.163	97.968	112.441	116.377
	%	-48.932	8.219	-20.807	-15.819	-3.382	
К-т вариации	Знач.	0.402	0.693	0.628	0.591	0.646	0.689
	%	-41.681	0.532	-8.911	-14.318	-6.226	

% - относительное отклонение рассчитанных значений от значений, полученных для выборки из трехсот величин

Чем больше значений берется в выборке, тем точнее рассчитываются параметры. Значение коэффициента вариации приближено к 1, но все же меньше единицы.

2 График значений заданной ЧП

График отображает значения числовой последовательности из 300 элементов.

Промежуточный вывод:

- Значения колеблются вокруг некоторого среднего значения, не проявляя устойчивого роста или падения на протяжении всей последовательности.
- Хотя на графике наблюдаются некоторые повторяющиеся участки, нельзя сказать, что последовательность имеет четко выраженный период. Колебания выглядят довольно случайными.
- На графике видны несколько пиков, где значения значительно превышают средний уровень. Это говорит о наличии выбросов в данных.

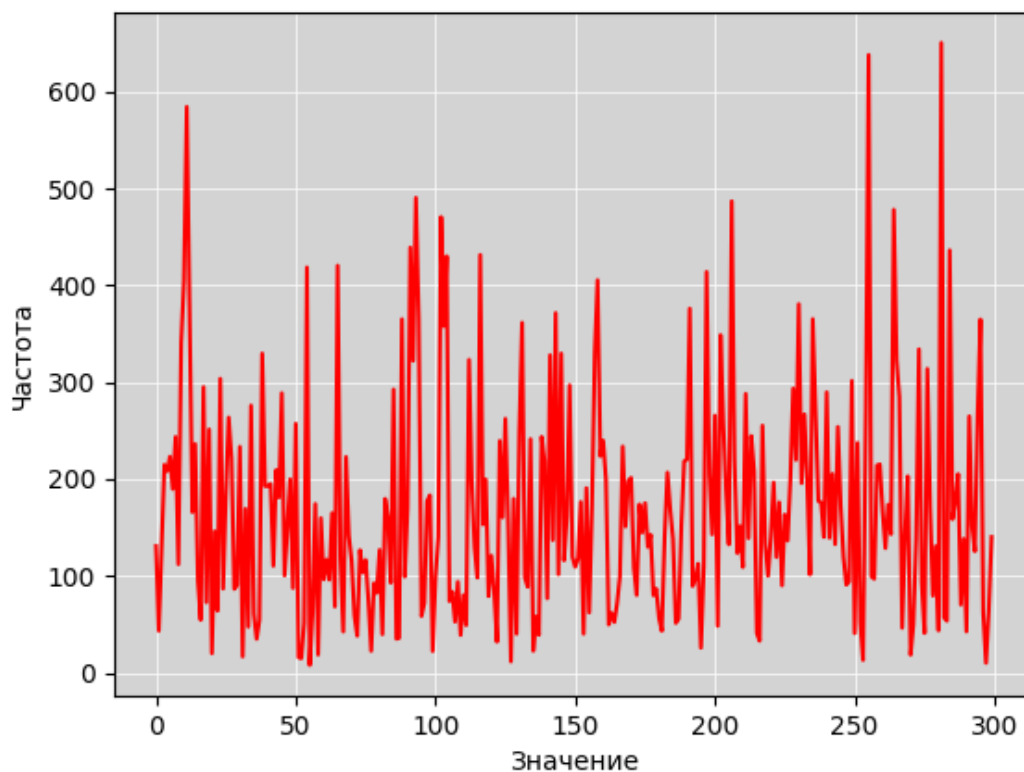


Рис. 1: График значений заданной ЧП

- Амплитуда колебаний достаточно велика, что свидетельствует о значительной изменчивости данных.

Стационарна (среднее значение и дисперсия постоянны во времени), но не периодическая.

3 Результаты автокорреляционного анализа

Расчёты коэффициентов автокорреляции происходили по следующей формуле:

$$r_{xk} \approx \frac{\sum_{i=1}^n (x_i - M[X])(x_{i+k} - M[X])}{\sum_{i=1}^n (x_i - M[X])^2}$$

Таблица 3: Форма 3. Коэффициенты автокорреляции

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. ЧП	0.0916	0.0318	0.0306	-0.0143	-0.0345	-0.0452	-0.0496	0.0065	0.0905	-0.0010
К-т АК для сгенерир. ЧП	0.0892	0.0879	-0.0259	0.0111	-0.0955	-0.0190	-0.0461	-0.0726	-0.0091	-0.1153
%	-2.62	176.4	-184.6	177.62	-176.81	57.96	7.06	-1216.92	-110.06	-11430

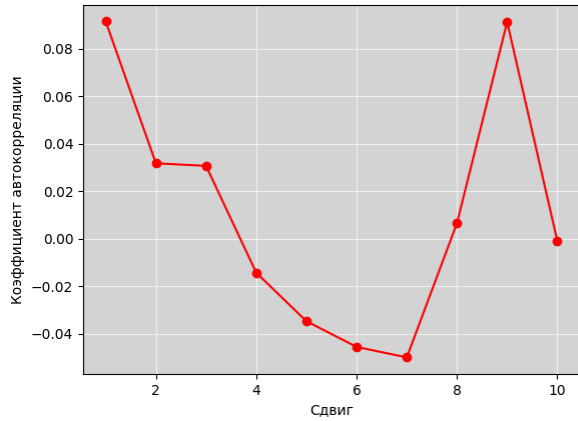


Рис. 2: График автокорреляции для лагов 1-10 для заданной числовой последовательности

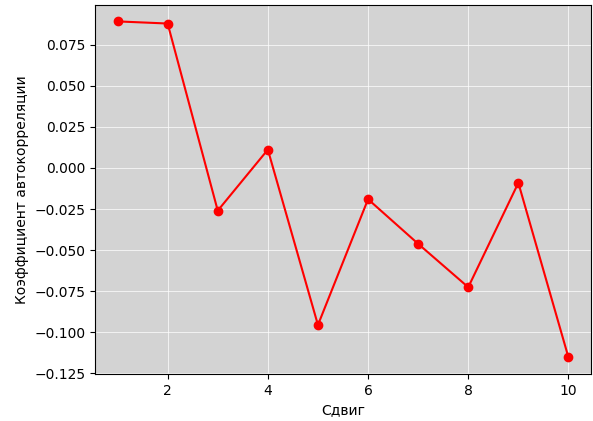


Рис. 3: График автокорреляции для лагов 1-10 для сгенерированной числовой последовательности

Промежуточный вывод: Коэффициент автокорреляции Сдвигов ЧП от 1 до 10 приближены к нулю, следовательно, можно сказать, что обе выборки случайны.

4 Гистограмма распределения частот для заданной ЧП

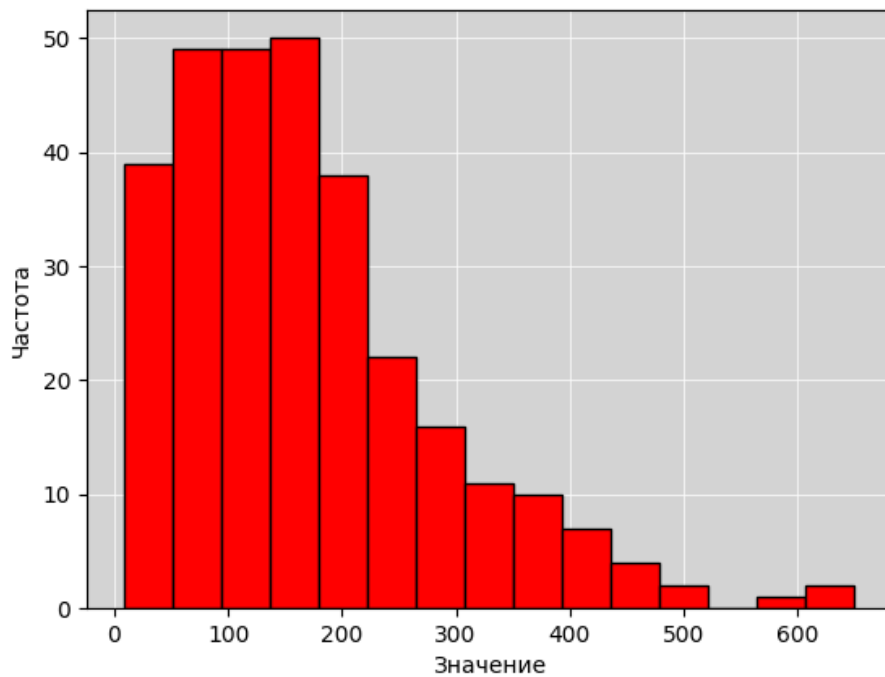


Рис. 4: Гистограмма заданной ЧП

5 Аппроксимирующий закон распределения

Подсчитанный коэффициент вариации $0.689 < 1$, при этом $0.689 > 0.577 \Rightarrow$ пробуем распределение Эрланга.

Просчитаем параметр формы:

$$k = \left(\frac{\tilde{m}}{\sigma} \right)^2 = \left(\frac{168.836}{116.377} \right)^2 \approx 2$$

Просчитаем параметр скорости:

$$\lambda = \left(\frac{k}{\tilde{m}} \right) = \left(\frac{2}{168.836} \right) = 0.0118$$

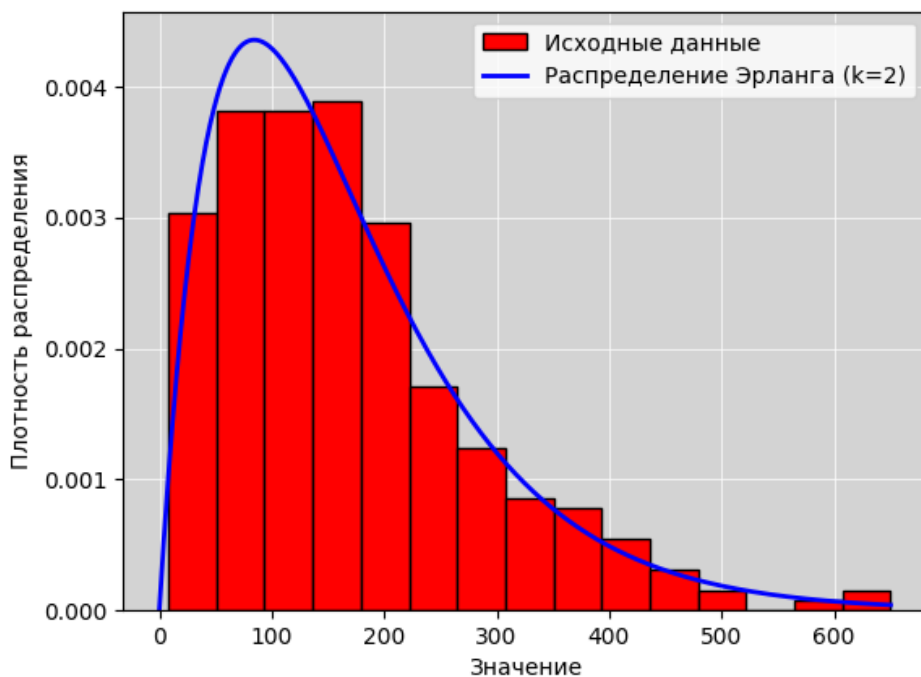


Рис. 5: Гистограмма значений заданной ЧП с аппроксимацией

В результате анализа коэффициента вариации, для аппроксимации выбран закон распределения Эрланга.

6 Описание программы для формирования новой ЧП

```
shape_param = 2
scale_param = d_best / m_best
erlang_300 = np.random.gamma(shape=shape_param, scale=scale_param, size=300).tolist()
```

7 Оценки и доверительные интервалы для новой сгенерированной ЧП

Таблица 4: Форма 2. Характеристики сгенерированной ЧП

Характеристика		Количество случайных величин					
		10	20	50	100	200	300
Мат. ож.	Знач.	83.078	207.15	167.297	167.826	160.307	156.587
	%	-43.807	13.977	13.974	1.174	7.846	
Дов. инт. (0,9)	Знач.	±31.385	±63.236	±22.575	±22.332	±15.128	±10.127
	%	1.638	36.668	5.423	38.743	15.806	
Дов. инт. (0,95)	Знач.	±37.44	±75.437	±26.931	±26.641	±18.047	±12.081
	%	1.64	36.669	5.422	38.74	15.802	
Дов. инт. (0,99)	Знач.	±49.207	±99.146	±35.395	±35.014	±23.718	±15.877
	%	1.64	36.669	5.421	38.74	15.806	
Дисперсия	Знач.	3648.885	29627.185	9439.845	18474.982	16955.314	11396.837
	%	-3.305	86.785	-11.136	92.492	34.108	
С.к.о.	Знач.	60.406	172.125	97.159	97.968	135.923	106.756
	%	-1.639	36.669	-5.421	-38.742	15.805	
К-т вариации	Знач.	0.727	0.831	0.581	0.81	0.812	0.682
	%	80.871	19.902	-7.523	37.039	25.739	

% - относительное отклонение рассчитанных значений от значений, полученных для выборки из трехсот величин

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

8 Результаты автокорреляционного анализа сгенерированной ЧП

Подсчитанный коэффициент вариации $0.682 < 1$, при этом $0.682 > 0.577 \Rightarrow$ пробуем распределение Эрланга.

Просчитаем параметр формы:

$$k = \left(\frac{\tilde{m}}{\sigma} \right)^2 = \left(\frac{156.587}{106.756} \right)^2 \approx 2$$

Просчитаем параметр скорости:

$$\lambda = \left(\frac{k}{\tilde{m}} \right) = \left(\frac{2}{156.587} \right) = 0.0128$$

В результате анализа коэффициента вариации, для аппроксимации выбран закон распределения Эрланга.

9 Сравнение гистограмм

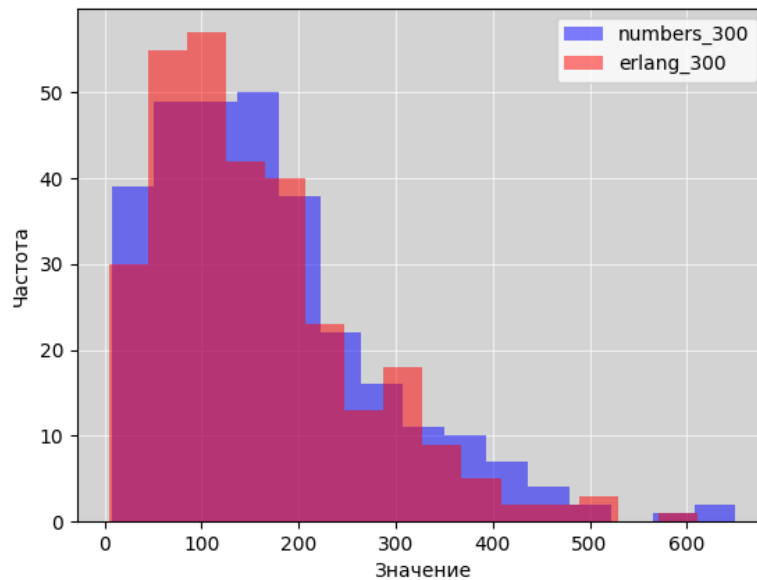


Рис. 6: Гистограммы исходной и сгенерированной ЧП

Сравнивая полученные гистограммы частот, можно сделать вывод, что сгенерированная нами последовательность практически идентична исходной (по варианту). Тем самым можно утверждать, что выбранная нами аппроксимация подходит.

Вывод

Проведенный анализ показал, что исходную числовую последовательность можно считать случайной. Закон распределения Эрланга с параметрами $k=2$ и $\lambda = 0.0118$ обеспечивает достаточно хорошую аппроксимацию исходных данных, что подтверждается сравнением числовых характеристик и гистограмм исходной и сгенерированной последовательностей. Небольшие расхождения в значениях характеристик могут быть связаны с естественной случайностью выборок и конечным размером исходной ЧП.