Университет ИТМО

Лабораторная работа №1 «Обработка результатов измерений: статистический анализ числовой последовательности»

по дисциплине: Моделирование

вариант: 15

Выполнили: Соболев Иван, Верещагин Егор, P34312
Преподаватель: Тропченко Андрей Александрович

Санкт-Петербург
2024

Содержание

1. 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.

2. Обработка заданной числовой последовательности

Согласно выданному варианту, обработаем числовую последовательность и внесём результаты в Форму 1.

2.1. Математическое ожидание

Вычисление оценки математического ожидания происходит по формуле:

$$\tilde{m} = \frac{\sum ni=1\, X_i}{n}$$

В качестве значения n попеременно выбираем 10, 20, 50, 100, 200 и 300.

2.2. Дисперсия

Вычисление оценки дисперсии происходит по формуле:

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

В качестве значения n также попеременно выбираем 10, 20, 50, 100, 200 и 300.

2.3. Среднеквадратическое отклонение

Вычисление оценки среднеквадратического отклонения происходит по формуле:

$$\tilde{\sigma} = \sqrt{\tilde{D}}$$

2.4. Коэффициент вариации

Вычисление оценки коэффициента вариации происходит по формуле:

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

2.5. Доверительные интервалы

Вычисление длины полуинтервалов происходит по формуле:

$$\varepsilon_p = t_p * \tilde{\sigma}$$

Где в качестве t_p подставляются значения из таблицы для $p = 0{,}9;\ 0{,}95$ и $0{,}99$.

2.6. Относительное отклонение

Вычисление относительного отклонения для каждой из вычисленных выше значений происходит относительно эталонного значения, которым считается измерение каждой из величин для 300 чисел последовательности.

2.7. Характеристики заданной ЧП

Таблица 1. Характеристики заданной числовой последовательности

		10	20	50	100	200	300
Мат. Ож.	знач.	**253.19**	**242.10**	**229.50**	**216.66**	**206.13**	**206.06**
	%	22.87%	17.49%	11.38%	5.14%	0.03%	
Дисперсия	знач.	**5658.98**	**12927.59**	**14817.56**	**15179.97**	**14804.06**	**14476.84**
	%	60.91%	10.70%	2.35%	4.86%	2.26%	
С.к.о.	знач.	**75.23**	**113.70**	**121.73**	**123.21**	**121.67**	**120.32**
	%	37.48%	5.50%	1.17%	2.40%	1.12%	
К-т вариации	знач.	**0.30**	**0.47**	**0.53**	**0.57**	**0.59**	**0.58**
	%	49.12%	19.57%	9.16%	2.61%	1.09%	
Дов. инт. 0.9	знач.	**39.13**	**41.82**	**28.32**	**20.27**	**14.15**	**11.43**
	%	242.45%	265.99%	147.81%	77.36%	23.85%	
Дов. инт. 0.95	знач.	**46.62**	**49.83**	**33.74**	**24.15**	**16.86**	**13.62**
	%	242.45%	265.99%	147.81%	77.36%	23.85%	
Дов. инт. 0.99	знач.	**61.28**	**65.49**	**44.34**	**31.74**	**22.16**	**17.89**
	%	242.45%	265.99%	147.81%	77.36%	23.85%	

Согласно полученным данным, можно сказать, что наиболее близкой к эталонной последовательности (из 300 чисел) соответствует последовательность из 200 чисел, что вполне логично. Однако интересным наблюдением является то, что последовательность из первых 50 чисел также достаточно близка к эталону по признакам оценки математического ожидания, дисперсии, среднеквадратического отклонения и коэффициента вариации. Это можно объяснить тем, что на начальных 50 числах последовательность не имеет явных трендов роста или снижения, а в дальнейшей части числовой последовательности происходят и рост, и снижения, которые в итоге приводят общее математическое ожидание и другие признаки в то же состояние, что и для последовательности из первых 20 чисел.

2.8. График значений заданной числовой последовательности

Рисунок 1. График значений заданной числовой последовательности

По данному графику можно сделать вывод, что заданная числовая последовательность не является ни возрастающей, ни убывающей. Периодичной её также нельзя назвать.

3. Автокорреляционный анализ

Для вычисления коэффициента автокорреляции воспользуемся формулой:

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

Наша числовая последовательность состоит из n=300 элементов, а максимальный сдвиг числовой последовательности составляет k=10. Поскольку n>>k, то можно использовать упрощённую формулу вычисления коэффициента автокорреляции:

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

Полученные коэффициенты автокорреляции внесены в таблицу ниже и в Форму 3:

Таблица 2. Коэффициенты автокорреляции для заданной ЧП

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. **ЧП**	0.08	0.05	0.05	0.02	-0.09	0.01	0.044	-0.135	0.11	0.01

А также представлены в виде графика:

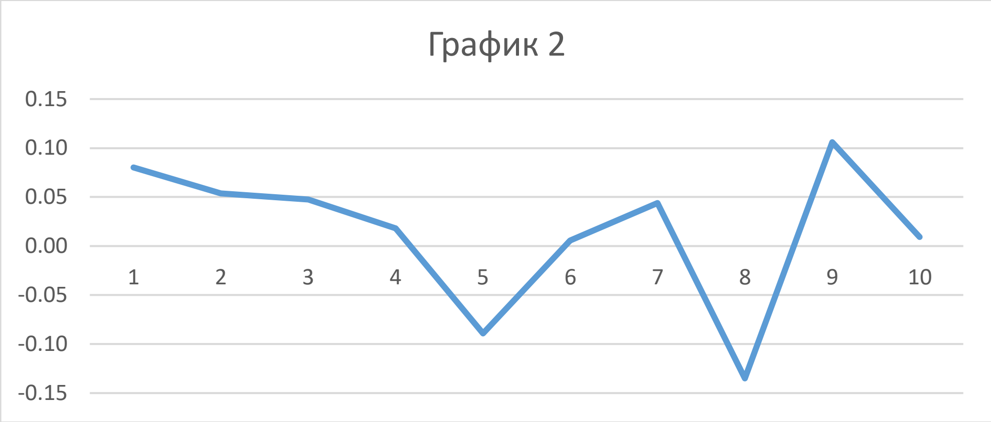

Рисунок 2. График изменения коэффициентов автокорреляции для заданной ЧП

Исходя из результатов автокорреляционного анализа можно сделать вывод, что заданная числовая последовательность является случайной, так как коэффициент автокорреляции составляет не более 0,135 в абсолютном измерении, что свидетельствует о случайности данной числовой последовательности.

4. Гистограмма распределения частот

Для заданной числовой последовательности построим гистограмму распределения частот с 10 интервалами, длина каждого из которых составляет 75:

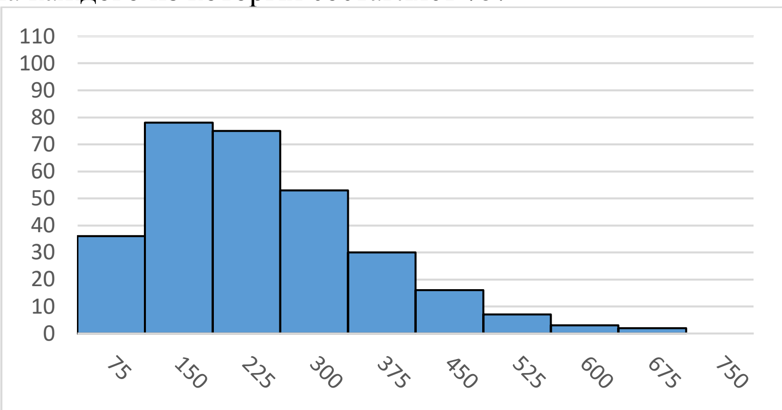

Рисунок 3. Гистограмма распределения частот заданной ЧП

По данной гистограмме видно, что большая часть значений заданной числовой последовательности находится в диапазоне от 0 до 225, в то время как значений в диапазоне 600–750 практически нет. Самое большое количество чисел находятся в диапазоне 75-150 – 78 значений.

5. Аппроксимация закона распределения

Для выполнения аппроксимации закона распределения заданной случайной последовательности необходимо выбрать один из законов распределения. Поскольку коэффициент вариации $\nu = 0{,}58 < 1$, то для аппроксимации был выбран закон гипоэкспоненциального распределения.

$$k \geq 1/0.58\hat{}2 \rightarrow \; k \geq 2{,}97; \; k \in Z \; \rightarrow \; k = 3$$

При использовании гипоэкспоненциального распределения у нас получается k_1 экспоненциальных случайных величин с параметром (матожиданием) M_1 и k_2 с параметром M_2, которые мы складываем для получения итогового значения.

$$k = k_1 + k_2, \text{ выберем } k_1 = 2;\ k_2 = 1$$

Значения параметров вычисляются по формулам:

$$M_1 = \frac{M[X]}{k}\left(1 + \sqrt{\frac{k_1}{k_2}(kv^2 - 1)}\right)$$

$$M_2 = \frac{M[X]}{k}\left(1 - \sqrt{\frac{k_2}{k_1}(kv^2 - 1)}\right)$$

Для полученных из заданной числовой последовательности значений:

$$M1 = 83.359067$$

$$M2 = 61.35189$$

Согласно этим значениям, сгенерируем новую числовую последовательность в Excel, воспользовавшись формулой:

$$= -M_1 * LN(\text{СЛЧИС}(\,)) - M_2 * LN(\text{СЛЧИС}(\,))$$

Поскольку функция СЛЧИС() в Excel возвращает случайное значение, полученное по закону нормального распределения, то для получения экспоненциального распределения мы вызовем функцию LN(x), возвращающую натуральный логарифм числа x. Затем преобразуем экспоненциальное распределение в гипоэкспоненциальное путём умножения каждого из случайных чисел в двух разных столбцах на одно из двух полученных ранее математических ожиданий, умноженных на -1, и сложим эти части для получения окончательной формулы числовой последовательности.
На основании новой числовой последовательности вычислим ей характеристики и заполним таблицу:

Характеристики сгенерированной случайной ЧП

Таблица 3. Характеристики сгенерированной случайной ЧП

		10	20	50	100	200	300
Мат. Ож.	знач.	**159.19**	**165.73**	**157.02**	**156.84**	**146.67**	**147.07**
	%	8.24%	12.69%	6.77%	6.64%	0.27%	
Дисперсия	знач.	**22264.94**	**14435.83**	**14462.98**	**11004.30**	**9732.77**	**9291.43**
	%	139.63%	55.37%	55.66%	18.43%	4.75%	
С.к.о.	знач.	**149.21**	**120.15**	**120.26**	**104.90**	**98.65**	**96.39**
	%	54.80%	24.65%	24.76%	8.83%	2.35%	
К-т вариации	знач.	**0.94**	**0.72**	**0.77**	**0.67**	**0.67**	**0.66**
	%	43.01%	10.61%	16.86%	2.05%	2.62%	
Дов. инт. 0.9	знач.	**77.61**	**44.19**	**27.98**	**17.25**	**11.47**	**9.15**
	%	747.87%	382.75%	205.61%	88.50%	25.35%	
Дов. инт. 0.95	знач.	**92.48**	**52.66**	**33.33**	**20.56**	**13.67**	**10.91**
	%	747.87%	382.75%	205.61%	88.50%	25.35%	
Дов. инт. 0.99	знач.	**121.54**	**69.20**	**43.81**	**27.02**	**17.97**	**14.34**

	%	747.87%	382.75%	205.61%	88.50%	25.35%	

Можно заметить, что отличия характеристик сгенерированной числовой последовательности от исходной не превышают 45%, а если не смотреть на дисперсию, которая имеет квадратные единицы измерения, то отличия составляют не более 30%, что указывает на успешное выполнение задачи выбора закона распределения заданной последовательности.

5.1. График значений сгенерированной числовой последовательности

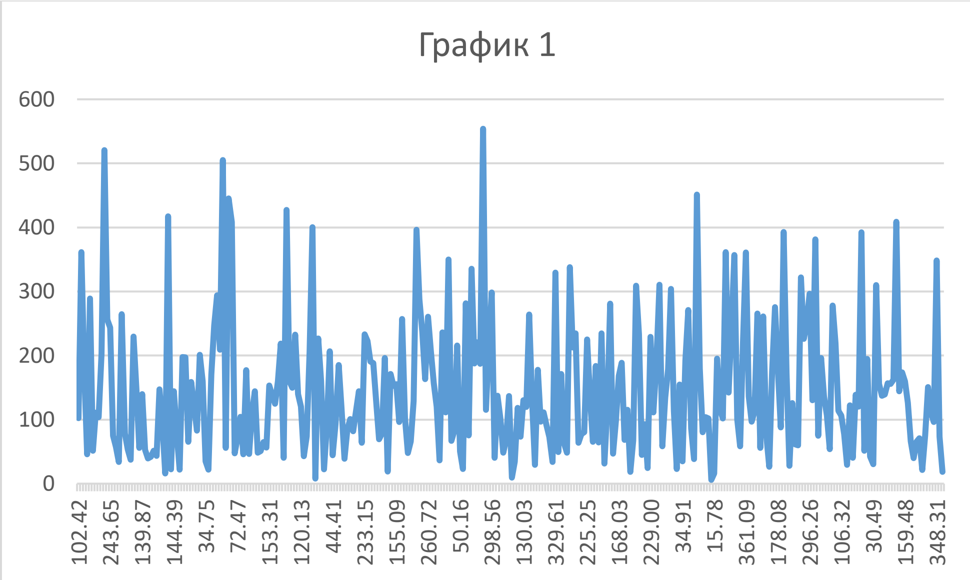

Рисунок 4. График значений сгенерированной ЧП

По данному графику можно увидеть, что сгенерированная числовая последовательность более случайна, чем заданная по варианту. В исходной на аналогичном графике можно было увидеть отрезки, имеющие тренд на рост и на спад, здесь же все колебания происходят относительно значения математического ожидания без трендов на рост или спад.

5.2. Результаты автокорреляционного анализа

Таблица 4. Коэффициенты автокорреляции

Сдвиг ЧП	1	2	3	4	5	6	7	8	9	10
К-т АК для задан. **ЧП**	0.08	0.05	0.05	0.02	-0.09	0.01	0.044	-0.135	0.11	0.01
К-т АК для сгенер. **ЧП**	-0.03	-0.01	0.04	-0.02	-0.07	0.06	-0.003	-0.131	-0.01	0.02

Как можно заметить, коэффициенты автокорреляции отличаются. В сгенерированной последовательности они уменьшились и теперь составляют не более 0,07 по абсолютному показателю кроме одного значения, в то время как у заданной по варианту числовой последовательности есть целых 4 значения, которые больше 0,07 по абсолютному показателю. Это значит, что сгенерированная величина более случайная, чем заданная. Для более наглядного различия приведены графики автокорреляционного анализа для заданной и сгенерированной числовых последовательностей.

Рисунок 5. Графики автокорреляционного анализа исходной и сгенерированной ЧП

5.3. Гистограмма распределения частот

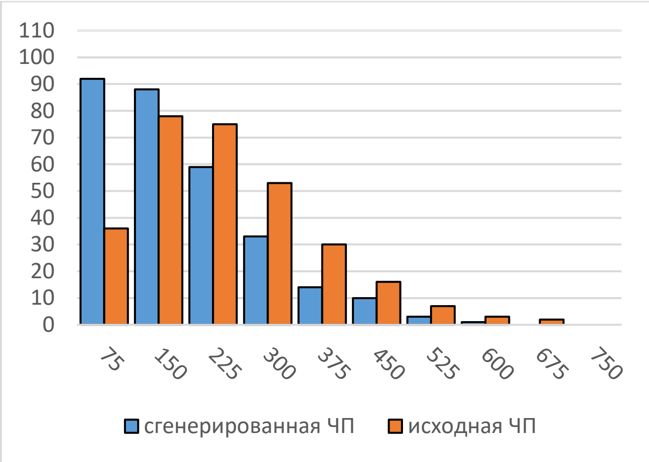

Рисунок 6. Гистограммы распределения частот исходной и сгенерированной ЧП

Гистограмма распределения частот для сгенерированной последовательности имеет схожие черты с гистограммой для заданной по варианту, что также свидетельствует о верности выбора закона аппроксимации.

5.4. Оценка корреляции между заданной и сгенерированной числовыми последовательностями

Для вычисления коэффициента корреляции между двумя числовыми последовательностями воспользуемся формулой:

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

Тогда для наших последовательностей:

$$r_{XY} = -0.04159$$

Корреляция между исходной и сгенерированной числовой последовательностью достаточно слабая и носит характер обратной зависимости.

6. Вывод

В процессе выполнения данной лабораторной работы был проведён анализ заданной по варианту и сгенерированной числовых последовательностей, были оценены такие характеристики случайной величины, как математическое ожидание, доверительные интервалы математического ожидания, дисперсия, среднеквадратическое отклонение и коэффициент вариации.

Поскольку исходная числовая последовательность имела гипоэкспоненциальное распределение, сгенерированная последовательность также использовала этот закон. При сравнении характеристик двух числовых последовательностей было выявлено, что вторая (сгенерированная) последовательность имеет более случайный характер.