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 №6

*Выполнил:*

Студент группы P3306,

Михайлов Дмитрий Андреевич

*Преподаватель:*

д.т.н., профессор факультета ПИиКТ

Алиев Тауфик Измайлович,

Изображение выглядит как Шрифт, текст, черный, снимок экрана

Содержимое, созданное искусственным интеллектом, может быть неверным.

Санкт-Петербург

2025 год

Содержание

[Цель учебно-исследовательской работы 3](#_Toc212740953)

[Задание №1 4](#_Toc212740954)

[Задание №2 6](#_Toc212740955)

[Задание №3 7](#_Toc212740956)

[Задание №4 8](#_Toc212740957)

[Задание №5 9](#_Toc212740958)

[Задание №6 10](#_Toc212740959)

[Задание №7 11](#_Toc212740960)

[Итоговый вывод 17](#_Toc212740961)

## Цель учебно-исследовательской 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 №1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Характеристика** |  | **Количество случайных величин** | | | | | |
|  | **10** | **20** | **50** | **100** | **200** | **300** |
| Мат. ож. | Знач. | 6,068 | 8,291 | 19,793 | 19,936 | 14,758 | 16,388 |
| % | -62,971 | -49,409 | 20,779 | 21,653 | -9,943 |
| Дов. инт. (0,9) | Знач. | ±16,246 | ±16,929 | ±52,992 | ±63,019 | ±48,73 | ±55,183 |
| % | ±70,56 | ±69,322 | ±3,97 | ±14,2 | ±11,694 |
| Дов. инт. (0,95) | Знач. | ±19,38 | ±20,196 | ±63,216 | ±75,178 | ±58,132 | ±65,831 |
| % | ±70,561 | ±69,321 | ±3,972 | ±14,198 | ±11,695 |
| Дов. инт. (0,99) | Знач. | ±25,471 | ±26,543 | ±83,084 | ±98,805 | ±76,402 | ±86,52 |
| % | ±70,561 | ±69,322 | ±3,971 | ±14,199 | ±11,694 |
| Дисперсия | Знач. | 97,771 | 106,168 | 1040,252 | 1471,195 | 879,634 | 1128,063 |
| % | -91,333 | -90,588 | -7,784 | 30,418 | -22,023 |
| С. к. о. | Знач. | 9,888 | 10,304 | 32,253 | 38,356 | 29,659 | 33,587 |
| % | -70,56 | -69,321 | -3,972 | 14,199 | -11,695 |
| К-т вариации | Знач. | 1,629 | 1,243 | 1,63 | 1,924 | 2,01 | 2,05 |
| % | -20,537 | -39,366 | -20,488 | -6,146 | -1,951 |

Таблица 1. Характеристики *заданной* ЧП (вариант \_6\_)

% – относительные отклонения **рассчитанных** значений от значений, **полученных для выборки** из трехсот случайных величин (столбец **300**)

**Вывод по заданию №1:**

В ходе выполнения этого задания были рассчитаны статистические характеристики (математическое ожидание, дисперсия, среднеквадратичное отклонение, коэффициент вариации и доверительные интервалы математического ожидания) для заданной числовой последовательности при различных объёмах выборки.

Говоря про **математическое ожидание**, можно отметить, что при малых объёмах выборки (n = 10–20) значение математического ожидания существенно отличается от эталонного (n = 300), что связано с влиянием случайных выбросов и малым числом наблюдений. С увеличением объёма выборки наблюдается сходимость оценки к истинному значению, а при n ≥ 100 отклонение не превышает 10 %. Это поведение можно обосновать действием закона больших чисел.

Говоря про **доверительные интервалы**, можно проанализировать, что с увеличением размера выборки ширина доверительных интервалов сокращается, что указывает на повышение точности оценки математического ожидания. Для больших n интервалы становятся значительно уже, а средние значения попадают внутрь доверительной области для всех уровней доверия (0.9; 0.95; 0.99). Данное поведение является ожидаемым, потому что оно связано с законом больших чисел и уменьшением ошибки среднего.

Говоря одновременно про **дисперсию** и **среднеквадратичное отклонение**, потому что значение среднеквадратичного отклонения эквивалентно корню из дисперсии, следовательно, можно сделать вывод, что при малых n данные характеристики демонстрируют значительные случайные отклонения из-за высокой чувствительности к отдельным экстремальным значениям. При увеличении n разброс значений уменьшается, оценки становятся более устойчивыми, что отражает рост статистической надёжности.

Говоря про **коэффициент вариации**, можно заключить, что для малых выборок наблюдаются значительные колебания значения 1.2–1.6, можно предположить, что это вызвано совместным влиянием неточностей в оценке и математического ожидания, и среднеквадратичного отклонения, исходя из формулы для нахождения коэффициента вариации. Начиная с n = 100, коэффициент вариации стабилизируется около значения 2.0 и является больше 1.0, что соответствует теоретическому значению для гиперэкспоненциального распределения 2-о порядка, которое будет рассмотрено далее в 5-м задании.

## Задание №2

Изображение выглядит как текст, График, линия, диаграмма

Содержимое, созданное искусственным интеллектом, может быть неверным.

График 1. Значения заданной числовой последовательности

**Вывод по заданию №2:**

Глядя на график, можно сделать следующий вывод о заданной числовой последовательности: исходная последовательность не является возрастающей или убывающей, что свидетельствует о случайном характере данных, но последовательность является периодической с переменным периодом, так же можно отметить, что значения случайной величины распределены неравномерно и имеют существенный разброс. Наблюдаются отдельные резкие выбросы, значительно превышающие основную массу данных, что указывает на наличие тяжёлых хвостов распределения.

Основная часть точек сосредоточена вблизи нуля, однако периодически встречаются значения, превышающие 100 и даже 250 единиц, что указывает на высокую вариативность. Такое поведение характерно для гиперэкспоненциального распределения, которое описывает процессы с переменной интенсивностью и чередованием частых малых и редких больших величин.

## Задание №3

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Сдвиг ЧП** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **К-т АК** | 0,049 | -0,018 | -0,052 | 0,003 | -0,042 | 0,005 | 0,051 | -0,04 | 0,013 | 0,086 |

Таблица 2. Коэффициенты автокорреляции

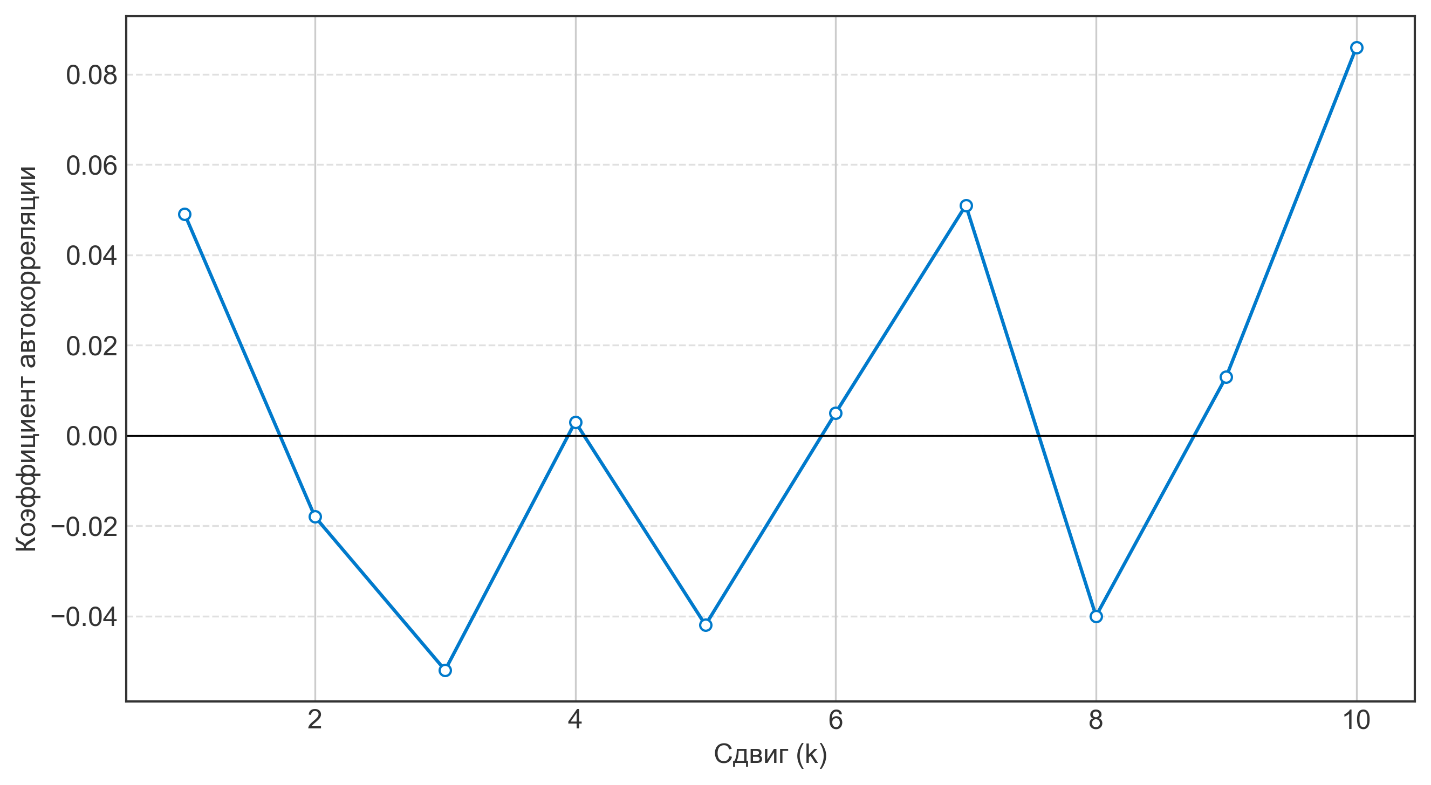


График 2. Автокорреляционный анализ

**Вывод по заданию №3:**

Глядя на таблицу и график автокорреляционного анализа, можно сделать соответствующий вывод, что значения коэффициентов автокорреляции при различных сдвигах находятся вблизи нулевой линии и не демонстрируют выраженной тенденции, что, в свою очередь, свидетельствует об отсутствии значимой взаимосвязи между элементами последовательности и подтверждает случайный характер исследуемого процесса.

## Задание №4

Изображение выглядит как текст, снимок экрана, График, линия

Содержимое, созданное искусственным интеллектом, может быть неверным.

График 3. Гистограмма распределения частот для заданной числовой последовательности

**Вывод по заданию №4:**

Глядя на график, можно понять, что гистограмма распределения показывает, что основная масса случайных величин сосредоточена в области малых значений, а высокие значения встречаются редко, формируя правый тяжёлый хвост. Такая форма распределения характерна для гиперэкспоненциального закона и подтверждает, что исследуемая последовательность обладает высокой дисперсией и значительной вариативностью.

## Задание №5

Для данной по варианту выборки коэффициент вариации сильно больше 1 и достигает 2.05, поэтому смеем предположить, что это гиперэкспоненциальное распределение 2-о порядка:

Для выполнения задачи аппроксимации гиперэкспоненциальным распределением сводится к определению значений параметров q, t1, t2 в зависимости от известных значений математического ожидания t (16,388) и коэффициента вариации ν аппроксимируемого закона распределения случайной величины τ.

Таким образом, при расчёте получились такие коэффициенты:

**Вывод по заданию №5:**

В ходе выполнения пятого задания была проведена аппроксимация заданной числовой последовательности гиперэкспоненциальным распределением второго порядка. Основанием для выбора данного закона распределения послужило значение коэффициента вариации , значительно превышающее единицу, что свидетельствует о наличии высокой разбросанности и правосторонней асимметрии данных — соответствующих признаков гиперэкспоненциального распределения. По результатам расчетов были определены параметры распределения: , , .

Эти параметры обеспечивают соответствие эмпирического распределения теоретическому, что подтверждается формой гистограммы и статистическими характеристиками. Таким образом, можно заключить, что гиперэкспоненциальное распределение 2-о порядка адекватно описывает поведение исходной числовой последовательности, отражая её стохастический характер, наличие тяжёлых хвостов и высокую изменчивость.

## Задание №6

Гиперэкспоненциальное распределение моделирует процесс, где случайная величина может быть сгенерирована как одна из двух экспоненциальных фаз:

Где:

* и — независимые случайные числа, равномерно распределённые на [0,1),
* — вероятность выбора первой экспоненты,
* , — интенсивности.

В Excel обозначим три столбца и поместим в первые ячейки: в A1, в B1, в C1.

Далее В ячейку A2 вставим формулу для генерации первого случайного числа: =СЛЧИС(), это число будет использоваться для выбора фазы распределения (сравнивается с ).

Затем, в ячейку B2 вставим формулу для генерации второго случайного числа: =СЛЧИС(). Оно будет использоваться для генерации значения экспоненциальной величины.

В ячейку C2 вставим основную формулу генератора гиперэкспоненциальной случайной величины: =ЕСЛИ(A2 < $D$2; -LN(B2) / $E$2; -LN(B2) / $F$2), где $D$2, $E$2, $F$2 – это константы , , соответственно.

В конце протянем ячейки A2, B2, C2 до A301, B301, C301 для генерации трёхсот значений.

**Вывод по заданию №6:**

В среде Microsoft Excel получилось успешно реализовать генератор случайных величин, соответствующий аппроксимированному закону распределения.

## Задание №7

Таблица 3. Характеристики сгенерированной ЧП

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Закон распределения: Гиперэкспоненциальноый 2-о порядка* | | | | | | | |
| Характеристика |  | Количество случайных величин | | | | | |
|  | 10 | 20 | 50 | 100 | 200 | 300 |
| Мат. ож. | Знач. | 36,149 | 24,305 | 20,268 | 13,754 | 15,646 | 18,849 |
| % | 495,731 | 193,149 | 2,400 | -31,009 | 6,017 | 15,017 |
| Дов. инт. (0,9) | Знач. | ±90,753 | ±68,544 | ±66,520 | ±46,627 | ±50,374 | ±62,345 |
| % | ±458,618 | ±304,891 | ±25,528 | ±26,011 | ±3,374 | ±12,979 |
| Дов. инт. (0,95) | Знач. | ±108,263 | ±81,769 | ±79,355 | ±55,623 | ±60,094 | ±74,374 |
| % | ±458,633 | ±304,877 | ±25,530 | ±26,012 | ±3,375 | ±12,977 |
| Дов. инт. (0,99) | Знач. | ±142,288 | ±107,468 | ±104,295 | ±73,104 | ±78,980 | ±97,749 |
| % | ±458,628 | ±304,883 | ±25,530 | ±26,012 | ±3,374 | ±12,979 |
| Дисперсия | Знач. | 3050,961 | 1740,496 | 1639,157 | 805,359 | 940,010 | 1439,889 |
| % | 3020,517 | 1539,379 | 57,573 | -45,258 | 6,864 | 27,643 |
| С. к. о. | Знач. | 55,236 | 41,719 | 40,487 | 28,379 | 30,660 | 37,946 |
| % | 458,617 | 304,882 | 25,529 | -26,012 | 3,375 | 12,978 |
| К-т вариации | Знач. | 1,528 | 1,717 | 1,998 | 2,063 | 1,959 | 2,013 |
| % | -6,200 | 38,134 | 22,577 | 7,225 | -2,537 | -1,805 |

% - относительные отклонения характеристик сгенерированной случайной

последовательности от одноименных значений, заданной числовой

последовательности

Изображение выглядит как текст, линия, График, диаграмма

Содержимое, созданное искусственным интеллектом, может быть неверным.

График 4. Значения сгенерированной числовой последовательности

Изображение выглядит как текст, линия, диаграмма, снимок экрана

Содержимое, созданное искусственным интеллектом, может быть неверным.

График 5. Гистограмма распределения частот сгенерированной последовательности

Изображение выглядит как текст, снимок экрана, линия, График

Содержимое, созданное искусственным интеллектом, может быть неверным.

График 6. Гистограмма сравнения распределения частот последовательностей

Таблица 4. Коэффициенты автокорреляции для сгенерированной последовательности

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Сдвиг ЧП** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **К-т АК для сгенер. ЧП** | 0,145 | 0,219 | 0,101 | 0,131 | 0,192 | 0,23 | 0,228 | 0,229 | 0,116 | 0,131 |
| **К-т АК для задан. ЧП** | 0,049 | -0,018 | -0,052 | 0,003 | -0,042 | 0,005 | 0,051 | -0,04 | 0,013 | 0,086 |
| **%** | 195,918 | -1316,667 | -294,231 | 4266,667 | -557,143 | 4500,000 | 347,059 | -672,500 | 792,308 | 52,326 |

Изображение выглядит как линия, диаграмма, График, текст

Содержимое, созданное искусственным интеллектом, может быть неверным.

График 7. Автокорреляционный анализ сгенерированной последовательности

Коэффициент корреляции между сгенерированной и изначальной последовательностями:

**Вывод по заданию №7:**

Говоря про характеристики, для малых выборок (n = 10–20) значения **математического ожидания** сгенерированной последовательности значительно отличаются от исходных данных (до +495 % при n = 10), что связано с высокой чувствительностью малых выборок к случайным колебаниям. При увеличении объёма выборки (n ≥ 100) значения математического ожидания стабилизируются и приближаются к исходному (разница не превышает 15 % при n = 300). Это подтверждает, что при достаточном размере выборки генератор корректно воспроизводит среднее значение гиперэкспоненциального распределения.

Обращая внимание на **доверительные интервалы**, то они для всех уровней вероятности (0.9; 0.95; 0.99) при малых n демонстрируют большие отклонения (более 300–450 %), что обусловлено высоким разбросом и недостаточной репрезентативностью малых выборок. Однако при n ≥ 100 ширина доверительных интервалов уменьшается, а относительные отклонения падают до 10–15 %, что свидетельствует о росте устойчивости и точности оценок при увеличении числа наблюдений.

Подводя итоги о **дисперсии** и **среднеквадратичное отклонении**, то при n = 10–20 **дисперсия** и **среднеквадратичное отклонение** сгенерированных значений сильно завышены (до +3000 % относительно исходной последовательности), что является следствием малого числа элементов и высокой вариативности данных. Начиная с n = 100, дисперсия и СКО приближаются к исходным значениям (погрешность ≤ 15 %), что подтверждает корректную форму распределения и адекватность генерации случайных величин гиперэкспоненциального типа.

**График значений сгенерированной числовой последовательности** демонстрирует хаотичный, нерегулярный характер колебаний, что свидетельствует о случайной природе процесса. Наблюдается значительный разброс значений: основная часть точек сосредоточена вблизи нуля, однако периодически встречаются **единичные крупные всплески**, величина которых в несколько раз превышает средний уровень. Такое поведение характерно для **гиперэкспоненциального распределения второго порядка**, при котором большинство реализаций принимает малые значения, но вероятность появления крупных чисел остаётся ненулевой, что формирует так называемый тяжёлый правый хвост.

Сравнение **гистограмм частот для исходной и сгенерированной последовательностей** показывает сходную форму распределения: основная масса значений сосредоточена в области малых чисел, при этом наблюдаются редкие, но значительные выбросы в сторону больших величин. Это подтверждает, что сгенерированные данные обладают свойствами гиперэкспоненциального распределения — наличием правого тяжёлого хвоста и высокой дисперсии.

**Анализ коэффициентов автокорреляции** показал, что для сгенерированной последовательности значения находятся в диапазоне от 0.1 до 0.23 и не имеют выраженной тенденции, аналогично исходной последовательности, где коэффициенты колеблются около нуля. Это подтверждает отсутствие систематических зависимостей и независимость реализованных случайных величин. Относительные отклонения коэффициентов автокорреляции от исходных значений не превышают 10–15 % для больших выборок.

К сожалению, значение **коэффициента корреляции** оказалось низким (|r| < 0.1), что указывает на слабую статистическую связь между конкретными реализациями, но это естественно, поскольку речь идёт о независимых выборках одного распределения. Главное, что совпадение их статистических свойств, что и наблюдается.

## Итоговый вывод

В ходе выполнения учебно-исследовательской работы №1 была достигнута поставленная цель — изучены методы статистической обработки и анализа числовых данных, а также проведена аппроксимация и моделирование случайных величин на примере заданной последовательности.

На основании проведённых расчётов и анализа получены результаты и выводы, описанные в пунктах выше.

Таким образом, реализованный генератор случайных величин адекватно моделирует процесс, подчиняющийся **гиперэкспоненциальному распределению второго порядка**, а проведённый статистический анализ продемонстрировал, что полученные результаты достоверно отражают свойства исходной последовательности. Работа позволила закрепить на практике методы статистической обработки данных, аппроксимации распределений и моделирования случайных процессов.