|  |
| --- |
| **МИНОБРНАУКИ РОССИИ**  федеральное государственное бюджетное образовательное  учреждение высшего образования  **«Национальный исследовательский университет «МЭИ»** |

|  |  |
| --- | --- |
| **Институт** | **ИВТ** |
| **Кафедра** | **УИТ** |

**ЗАДАНИЕ**

**НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ (бакалаврскую работу)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Направление** | | | 27.03.04 Управление в технических системах | | | |
| (код и наименование) | | | | | | |
| **Направленность (профиль)** | | | | | | Управление и информатика в |
| технических системах | | | | | | |
|  | | | | | | |
| **Форма обучения** | | | | | **очная** | |
|  | (очная/очно-заочная/заочная) | | | | | |
| |  |  | | --- | --- | | **Тема:** | «Исследование влияния коррелированности значений решающей | | функции в алгоритмах обнаружения разладки стохастических сигналов | | | основанных на теории случайных блужданий» | | | | | | | | |
| **Студент** | | **А-01-19 Викторов Г.О.** | | | | |
| группа подпись фамилия и инициалы | | | | | | |
| **Научный руководитель** | | | | **к.т.н. доцент Рябов С.Н.** | | |
| уч. степень должность подпись фамилия и инициалы | | | | | | |
| **Консультант** | | | |  | | |
| уч. степень должность подпись фамилия и инициалы | | | | | | |
| **Консультант** | | | |  | | |
| уч. степень должность подпись фамилия и инициалы | | | | | | |
| **Зав. кафедрой** | | | | **д.т.н. доцент Бобряков А.В.** | | |
| уч. степень звание подпись фамилия и инициалы | | | | | | |
|  | | | | | | |
| **Место выполнения работы** | | | | | | НИУ «МЭИ» Кафедра Управления и |
| интеллектуальных технологий | | | | | | |

**СОДЕРЖАНИЕ РАЗДЕЛОВ ЗАДАНИЯ И ИСХОДНЫЕ ДАННЫЕ**

|  |
| --- |
| 1.Изучение литературы по проблематике обнаружения разладки |
| случайных процессов, известным методам обнаружения и их классификации. |
| Составление краткого обзора. |
| 2.Алгоритм, основанный на теории случайных блужданий : назначение, |
| структура, принцип действия и основные характеристики. |
| Программная реализация и апробация. |
| 3.Изучение и подбор методов задания корреляции для гауссовского процесса. |
| 4.Имитационное моделирование как инструмент исследования алгоритма. |
| Разработка программных средств имитационного моделирования . |
| Планирование имитационного эксперимента для |
| исследования варианта разладки коррелированного гауссовского процесса. |
| 5.Реализация имитационного эксперимента и обработка его результатов. |
| 6.Сопоставление исследуемого алгоритма с другими известными |
| алгоритмами обнаружения разладки по показателю эффективности. |
| 7.Анализ и обсуждение полученных результатов. |
| 8.Написание и оформление работы. |

**ПЕРЕЧЕНЬ ГРАФИЧЕСКОГО МАТЕРИАЛА**

|  |  |
| --- | --- |
| **Количество листов** | 12 |
|  |  |
| **Количество слайдов в презентации** | 12 |
|  |  |

**РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА**

|  |
| --- |
| **1. Феллер У.** Введение в теорию вероятностей и ее приложения. Том 1. |
| (An Introduction to Probability Theory and its Applications. Volume I), М.: Изд- |
| во «Мир», 1970. 499 с. |
| **2.** **Филаретов Г.Ф., Червова А.А., Бучаала З.** Последовательный |
| непараметрический алгоритм обнаружения разладки временнóго ряда // |
| Датчики и системы, 2020, №1, с.9 – 16. |
| **3.** **Филаретов Г.Ф., Червова А.А., Бучаала З.** Непараметрический |
| метод обнаружения разладки временнóго ряда c использованием механизма |
| случайных блужданий».// Известия ЮФУ. Технические науки. 2020, №4, с.107-117. |
| **4.** **Филаретов Г.Ф., Бучаала З.** Разработка алгоритмов обнаружения |
| разладки временных рядов на основе непараметрических критериев проверки |
| гипотез // Вестник МЭИ. 2021, №3. – с. 67 – 77. |
| **5. ФиларетовГ.Ф., Червова А.А.** Последовательный алгоритм |
| обнаружения момента изменения дисперсии временного ряда. // «Заводская |
| лаборатория. Диагностика материалов», Том 85, №3,2019, с. 75 -82. |

***Примечания:***

1. Задание брошюруется вместе с выпускной работой после титульного листа (страницы задания имеют номера 2, 3).
2. Отзыв руководителя, рецензия(и), отчет о проверке на объем заимствований и согласие студента на размещение работы в открытом доступе вкладываются в конверт (файловую папку) под обложкой работы.

# Введение

Исследуется задача, которая заключается в максимально быстром и точном определении момента, при котором происходит изменение свойств изучаемых объектов и характера протекающих в них процессов. Эта задача, которую часто называют задачей о разладке, имеет целью выявление и фиксацию существенных изменений в данных и системах.

Обнаружение разладки в стохастических сигналах является актуальной и важной проблемой во многих областях науки и техники. Стохастические сигналы широко применяются в таких областях, как финансовая аналитика, электроника, медицина, экология и многие другие. Разладка в сигналах может указывать на изменения в системе или среде, и ее обнаружение имеет большое значение для принятия решений и предотвращения потенциальных проблем.

Разладка стохастических процессов представляет собой неожиданное изменение характеристик наблюдаемого случайного процесса, которое происходит в неизвестный и непредсказуемый момент времени. Задача обнаружения разладки дискретного стохастического процесса X(t) в реальном времени может быть сформулирована следующим образом: предполагается, что имеется последовательность наблюдений x(t1) = x1, x(t2) = x2 и так далее. Возможны два состояния объекта: "норма" с определенными вероятностными характеристиками процесса X(t) и "разладка", когда вероятностные свойства X(t) отличаются от "нормальных". Переход из состояния "норма" в состояние "разладка" происходит в заранее неизвестный момент времени t0. Задача состоит в обнаружении этого момента t0 на основе последовательности значений x1, x2…, при этом желательно минимизировать задержку τзап между моментом разладки и его обнаружением.

Для оперативного обнаружения разладки необходимо использовать методы и алгоритмы, способные работать в режиме реального времени и обрабатывать последовательные значения по мере их поступления.

# Алгоритмы обнаружения спонтанной разладки наблюдаемых процессов

Методы определения разладки могут быть разделены на две категории: статистические методы и методы машинного обучения. Статистические методы основаны на анализе свойств данных, распределений и структуры сигнала, а также на применении статистических тестов и моделей. Методы машинного обучения, в свою очередь, используют обучающие данные для построения моделей, которые могут обнаруживать и анализировать разладку в данных.

Выбор метода определения разладки зависит от специфики задачи, доступных данных, требований к скорости обработки и точности обнаружения. Кроме того, эффективность методов может варьироваться в зависимости от типа разладки (внезапная или постепенная), структуры сигнала и уровня шума. Разработка эффективных алгоритмов определения разладки является активной областью исследования, и новые подходы и методы постоянно разрабатываются для улучшения точности и скорости обнаружения разладки.

Виды статистических методов разладки включают параметрические и непараметрические подходы:

1. Параметрические методы разладки: Эти методы предполагают известную модель данных и параметры, которые описывают статистические свойства процесса. Они основаны на статистических тестах, которые сравнивают ожидаемое поведение данных с наблюдаемыми значениями, чтобы обнаружить отклонения и разладки. Примерами параметрических методов являются контрольные карты Шухарта, экспоненциально взвешенное скользящее среднее (EWMA), а также методы, основанные на ARIMA-моделях и регрессии.

2. Непараметрические методы разладки: Эти методы не требуют предположения о распределении данных или модели процесса. Они основаны на непараметрических тестах и алгоритмах, которые исследуют структуру данных и обнаруживают аномалии, не полагаясь на конкретную модель. Примерами непараметрических методов являются методы, основанные на рангах (например, критерий Уилкоксона-Манна-Уитни), методы бутстрэпа и методы, использующие алгоритмы кластеризации и машинное обучение.

Выбор между параметрическими и непараметрическими методами зависит от характера данных, доступной информации о модели и статистическом свойстве процесса. Каждый подход имеет свои преимущества и ограничения, и выбор определенного метода требует анализа и понимания особенностей конкретной задачи обнаружения разладки.

Любой последовательный алгоритм обнаружения разладки должен иметь определенные оптимальные свойства. С одной стороны, желательно минимизировать среднее время обнаружения разладки (минимальную задержку в обнаружении разладки τзап), а с другой стороны, максимизировать среднее время между ложными тревогами (Tлт).

На практике, для оценки производительности контрольного алгоритма, часто используются две основные характеристики:

1. Среднее время между ложными тревогами (Tлт): Это среднее время между случаями, когда алгоритм срабатывает и сообщает о разладке, хотя разладки фактически нет. Меньшее значение Tлт указывает на более надежный алгоритм, который редко генерирует ложные тревоги.

2. Среднее время задержки обнаружения (τзап): Это среднее время, которое проходит от момента разладки до момента обнаружения разладки алгоритмом. Меньшее значение τзап означает, что алгоритм оперативно реагирует на разладку и быстро обнаруживает изменения в данных.

При настройке контрольного алгоритма необходимо учитывать информацию о характеристиках наблюдаемого процесса до и после разладки, а также желаемые свойства самого алгоритма. Настройка включает выбор номинального значения разладки и порогового значения. Для настройки и оценки производительности алгоритма необходима справочная информация, которая может быть представлена в графическом или табличном виде и содержать:

- Зависимость порогового значения от среднего времени между ложными тревогами для различных значений номинальной разладки.

- Зависимость времени задержки обнаружения от порогового значения или среднего времени между ложными тревогами для оценки скорости работы алгоритма при наличии номинальной разладки.

Использование такой справочной информации позволяет эффективно настроить контрольный алгоритм и оценить его производительность в соответствии с требованиями задачи.

Существуют два основных подхода к решению задачи обнаружения разладки: апостериорный и последовательный подходы.

В апостериорном подходе данные регистрируются и сохраняются, а затем, в процессе последующей обработки, выполняется оценка неизвестного момента t0. Этот подход интерпретируется как обычная задача оценивания, и для соответствующих алгоритмов применяются стандартные требования к получению оценок с наибольшей точностью, например, с минимальным смещением и дисперсией.

Однако больший интерес представляет последовательный подход, при котором обнаружение разладки происходит в реальном времени, одновременно с получением данных о значениях случайного процесса. В этом случае требуется обнаружить разладку как можно быстрее после ее возникновения при текущей обработке данных. Такие подходы называются последовательными алгоритмами обнаружения разладки.

Последовательные алгоритмы обнаружения разладки имеют практическую ценность, поскольку позволяют оперативно реагировать на изменения в данных и обнаруживать разладку практически в режиме реального времени. Эти алгоритмы основаны на анализе последовательности данных, с использованием статистических тестов, моделей или других методов, чтобы определить момент разладки и принять соответствующие меры.

Рассмотрим классификацию последовательных алгоритмов для обнаружения разладки в стохастических временных рядах. В данной классификации выделяются четыре основные группы: байесовские алгоритмы, эвристические алгоритмы, алгоритмы на основе подхода Неймана-Пирсона и алгоритмы, использующие идеи последовательного анализа.

Байесовские алгоритмы направлены на минимизацию среднего риска, который складывается из потерь от наличия нераспознанной разладки и ложных срабатываний. Они используют байесовский подход и учитывают априорную информацию о вероятности разладки и статистических свойствах процесса.

Эвристические алгоритмы основаны на интуитивных представлениях о наилучшей организации обнаружения разладки и удобны для практической реализации. Они не всегда имеют строгое математическое обоснование, но в некоторых случаях дают хорошие результаты. Однако в других случаях эти алгоритмы могут давать неверные результаты.

Алгоритмы на основе подхода Неймана-Пирсона используют классические критерии проверки гипотез. Они многократно проверяют гипотезы о виде функции распределения вероятностей или значениях параметров (например, среднего или дисперсии). Наиболее известным примером таких алгоритмов являются контрольные карты Шухарта, которые применяются для отслеживания изменчивости параметров и обеспечения стабильности технологических процессов.

Алгоритмы, основанные на идеях последовательного анализа, используют статистику отношения правдоподобия, но с особенностью - в них реализуется операция отражения от нижней решающей границы. Это позволяет эффективно обнаруживать разладку при анализе последовательности данных.

При применении последовательных алгоритмов на практике наиболее часто используют алгоритмы на основе подхода Неймана-Пирсона.

Алгоритмы, основанные на подходе Неймана-Пирсона, включают:

1. Алгоритм кумулятивных сумм (CUSUM-алгоритм): Он использует накопление сумм разностей между текущими значениями и их ожидаемыми значениями для обнаружения разладки.

2. Алгоритм скользящего среднего (Moving Average) или MA-алгоритм: Он вычисляет среднее значение последовательности данных и сравнивает его с пороговым значением для определения разладки.

3. Алгоритм экспоненциального сглаживания (EWMA-алгоритм): Он применяется для обнаружения разладки во временных рядах путем сглаживания последовательности и сравнения с пороговым значением.

4. Контрольные карты Шухарта: Это наиболее известный пример. Они используются для отслеживания изменчивости параметров и обеспечения стабильности технологических процессов.

Все эти алгоритмы основаны на статистических методах и используют критерии проверки гипотез, предложенные Нейманом и Пирсоном, для обнаружения разладки в данных.

Все последовательные методы обнаружения разладки основаны на вычислении решающих функций gi = gi (xi , xi-1 , xi-2,…)на основе наблюдаемых дискретных значений xi. Эти функции вычисляются рекуррентно, используя значения статистики на предыдущем шаге и текущее измеренное значение xi. Затем вычисленное значение gi сравнивается с пороговым уровнем H, который разделяет область возможных значений статистики на две подобласти: Ω0 и Ω1. Подобласть Ω0 соответствует отсутствию разладки, а подобласть Ω1- критической области. Если gi принадлежит к Ω0, процесс контроля продолжается, и считается, что разладки нет. Если gi принадлежит к Ω1, подается сигнал о наличии разладки (сигнал тревоги). По желанию экспериментатора контрольная процедура может быть запущена заново.

Чаще всего предполагается, что разладка проявляется в изменении параметров стохастического процесса, таких как математическое ожидание, дисперсия, корреляционно-спектральные свойства и т.д., а не в изменении функции распределения вероятностей .

Более подробно рассмотрим каждый из алгоритмов.

Алгоритм кумулятивных сумм (CUSUM-алгоритм)

Алгоритм кумулятивных сумм (CUSUM) - это последовательный метод обнаружения разладки, который используется для мониторинга изменений в стохастических процессах. Он основан на идее накопления суммарных отклонений от некоторого базового уровня.

Принцип работы CUSUM алгоритма заключается в следующем:

1. Определяется статистика C\_i на каждом шаге i, которая представляет накопленную сумму отклонений от некоторой базовой значения или среднего.

Ci = max(0, Ci-1 + xi - μ)

Где xi - текущее измерение ряда, μ - базовое значение или среднее.

2. Решающая функция gi вычисляется как разница между статистикой Ci и пороговым значением H.

gi = Ci - H

Если gi превышает нулевое значение, то это указывает на наличие разладки.

3. Если gi положительное, то считается, что разладка произошла, и сигнализируется об этом. Затем статистика Ci обнуляется и процесс начинается заново.

CUSUM алгоритм позволяет обнаруживать разладки с высокой чувствительностью, так как накапливает суммарные отклонения от базового значения. Он также способен обнаруживать как резкие, так и постепенные изменения в ряде данных.

CUSUM алгоритм широко используется в различных областях, включая производство, экономику, финансы, медицину и т.д., где необходимо раннее обнаружение и контроль изменений в процессах и системах.

Алгоритм скользящего среднего (Moving Average) или MA-алгоритм

Алгоритм скользящего среднего (Moving Average) может использоваться для обнаружения разладки во временных рядах. Он основан на следующем принципе: если точка данных во временном ряду значительно отличается от своего окружения, это может указывать на наличие разладки.

Для определения разладки с помощью алгоритма скользящего среднего используется следующий подход:

1. Вычисление скользящего среднего для каждой точки временного ряда. Размер окна выбирается в соответствии с характеристиками данных и желаемой чувствительностью алгоритма. Обычно используются окна размером от нескольких точек до нескольких временных интервалов.

2. Вычисление отклонения каждой точки данных от соответствующего скользящего среднего значения. Это можно сделать путем вычитания значения скользящего среднего из соответствующей точки временного ряда.

3. Оценка разладки. Если отклонение точки данных превышает определенный пороговый уровень, считается, что произошла разладка. Пороговый уровень выбирается на основе анализа исторических данных и требуемой чувствительности алгоритма.

4. Повторение алгоритма для последующих точек данных для обнаружения дополнительных разладок во временном ряду.

Алгоритм экспоненциального сглаживания (Exponential Smoothing) также может быть использован для обнаружения разладки во временных рядах. Он позволяет сгладить шумы и колебания в данных и выявить отклонения, которые могут свидетельствовать о разладке.

Для определения разладки с использованием алгоритма экспоненциального сглаживания следуйте следующим шагам:

1. Инициализация начального значения сглаженного ряда (g0) и параметра сглаживания (α). Значение параметра α должно быть выбрано в диапазоне от 0 до 1, где более близкое к 1 значение означает большую чувствительность к последним значениям ряда.

2. Для каждой последующей точки данных Xt во временном ряду:

- Вычисление текущего сглаженного значения St с использованием формулы экспоненциального сглаживания:

gi = α \* Xi + (1 - α) \* gi-1

- Вычисление ошибки разладки (Et) путем нахождения разницы между текущим значением ряда и соответствующим сглаженным значением:

Ei = Xi - gi

3. Оценка разладки. Если ошибка разладки (Ei) превышает определенный пороговый уровень, то считается, что произошла разладка. Пороговый уровень выбирается на основе анализа исторических данных и требуемой чувствительности алгоритма.

4. Повторение алгоритма для последующих точек данных для обнаружения дополнительных разладок во временном ряду.

Контрольные карты Шухарта (Shewhart Control Charts) - это графический метод мониторинга и обнаружения разладки в процессах. Они были разработаны Уолтером Шухартом и используются для контроля статистических показателей процесса, таких как среднее значение, стандартное отклонение или диапазон.

Основная идея контрольных карт Шухарта заключается в следующем:

1. Измерение показателя процесса. Для каждой выборки из процесса измеряется соответствующий показатель, например, среднее значение, стандартное отклонение или диапазон.

2. Вычисление статистических показателей. Для каждой выборки вычисляются статистические показатели, такие как среднее значение и стандартное отклонение. Эти значения используются в дальнейшем для построения контрольных карт.

3. Построение контрольных карт. Для каждого статистического показателя строится контрольная карта. Контрольные карты имеют ось времени по горизонтали и ось значений показателя по вертикали. На графике отображается среднее значение показателя и контрольные пределы, которые определяются на основе статистических свойств процесса.

4. Обнаружение разладки. При мониторинге процесса, если точки показателя выходят за контрольные пределы или показывают другие необычные шаблоны или тренды, считается, что произошла разладка или нарушение процесса.

Контрольные карты Шухарта являются эффективным инструментом для наблюдения за процессами и своевременного обнаружения разладки. Они позволяют операторам и аналитикам проводить мониторинг качества процесса и принимать меры по устранению отклонений, что способствует повышению производительности и снижению доли брака.

# Алгоритм, основанный на теории случайных блужданий:назначение, структура, принцип действия и основные характеристики.

Впервые идея использования механизма одномерных случайных блужданий для обнаружения разладки была представлена в работе, опубликованной в 1975 году. В этой работе была рассмотрена классическая схема одномерных случайных блужданий, где на каждом шаге в дискретный момент времени n текущая координата точки zn, представляющей процесс блуждания, с вероятностью p получает приращение +1 или с вероятностью q=1–p приращение -1.

,

где Me(x) представляет собой значение медианы наблюдаемого временного ряда в состоянии "норма". Это соотношение описывает случай, когда p = q = 0.5, но при наличии разладки симметрия нарушается, и вероятность p становится больше 0.5, что соответствует сдвигу медианы в положительную сторону.

Затем рассматривается одно из возможных направлений развития методов обнаружения разладки, основанных на случайных блужданиях, а именно использование теории серий "успехов". Контрольная процедура предлагаемого метода сводится к вводу значения контролируемого процесса xn на каждом такте работы, его преобразованию в значение приращения zn в соответствии с вышеуказанным соотношением, подсчету текущей длины серии kn следующих значений zn = +1: kn = k­n-1 +1, если z­n = +1 и kn = 0, если zn = –1., также фиксируется номер такта n\* на котором будет выполнено условие kn\*= kn= k, где k представляет собой заранее заданное пользователем значение, выбранное с учетом желаемого среднего времени между ложными тревогами .

Общий алгоритм реализуется в виде следующей последовательности шагов:

1. Генерируется вариационный ряд длиной N с медианой *Me(X)* = 0.5 в состоянии «норма» и *Me(X)* = 0.8 в состоянии «разладка».
2. Образуется последовательность *kn*  из плюсов по следующему правилу:
3. Подсчитывается количество случаев, когда серия *kn* соответствует заданному значению *k* - число серий в совокупности .
4. Определяется такт, на котором впервые *kn* достигает заданного значения *k* после номинальной разладки.
5. Рассчитываются показатели и

# Программная реализация и апробация

В основной программе пользователем задаётся длина вариационного ряда – L. Помимо этого задаётся номер такта номинальной разладки – N и mx – медиана процесса с разладкой. Также пользователь может выбрать какую из 3 задач ему необходимо выполнить (вывод графика с результатами моделирования для заданного k, вывод таблицы зависимости , от k, определение оптимального k для заданного Tлт), для этого вводится параметр st. В функциях out\_table() и opt\_k\_for\_Tlt() вводится параметр m – число экспериментов для одного k.

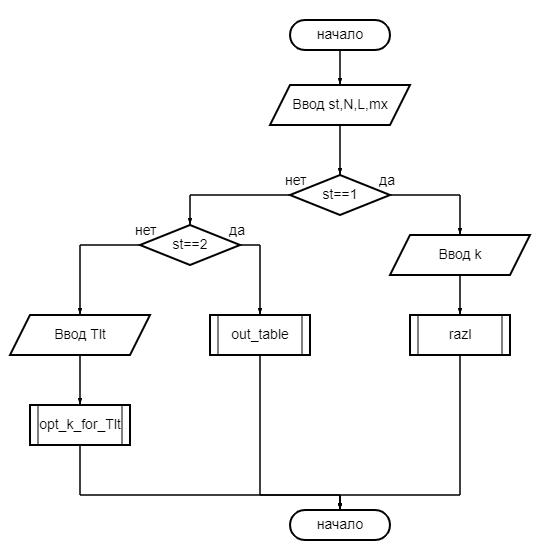


Рисунок 1 Структура основной программы

Рассмотрим функцию по определению разладки сигнала(opr\_razl()). В неё передаются 4 входных параметра: x-вариационный ряд длиной L, med-значение медианы вариационного ряда до разладки, k-длина серии и N-номер такта номинальной разладки.

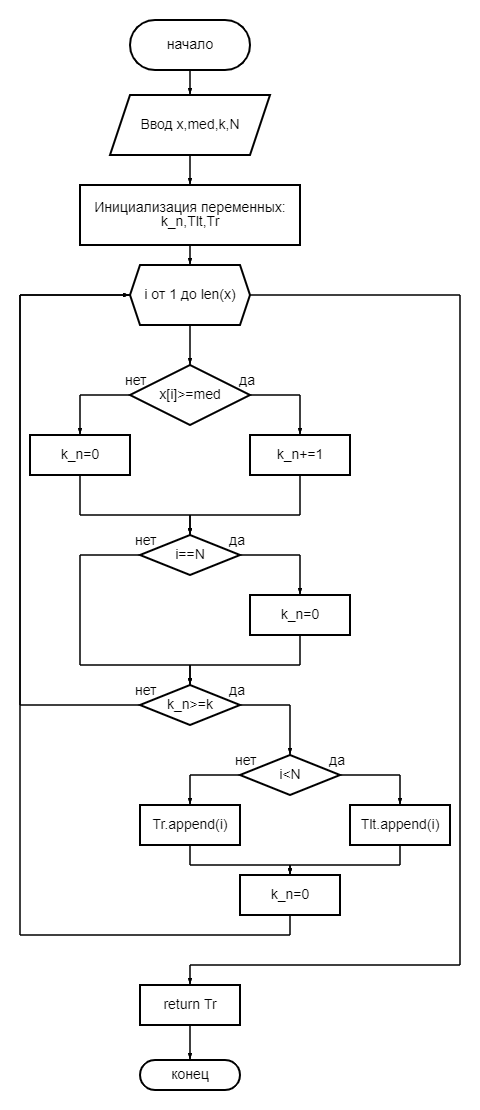


Рисунок 2 Структура функции по определению разладки

В функции opr\_razl() инициализируются следующие переменные:k\_n,Tlt,Tr, где k\_n – количество идущих подряд значений, больших значения медианы процесса до разладки. Tlt-моменты времени ложных тревог и Tr-моменты времени реальных тревог. Реализация метода происходит в цикле, где начиная с 0 элемента ряда, происходит сравнение его значения с медианой. Если значение ряда окажется больше медианы, то к переменной *kn* прибавляется единица *kn = kn+1,* а если меньше, то *kn*обнуляется. Если *kn* достигло заданного значения *k*, то проверяется, не зафиксировалось ли обнаружение разладки после наступления N, если это так, то значение текущего шага добавляется в список Tr, если это условие не выполнено, то значение шага добавляется в список Tlt. После успешного выполнения, функция возвращает значения моментов времени ложных и реальных тревог.

Рассмотрим функцию(razl()), данная функция выводит график процесса с результатами моделирования и генерирует случайный вариационный ряд с разладкой в момент времени N. В функцию передаются 4 входных параметра: N, L, mx, k. N-номер такта номинальной разладки, L-длина вариационного ряда, mx-медианна процесса с разладкой, k-число серий.

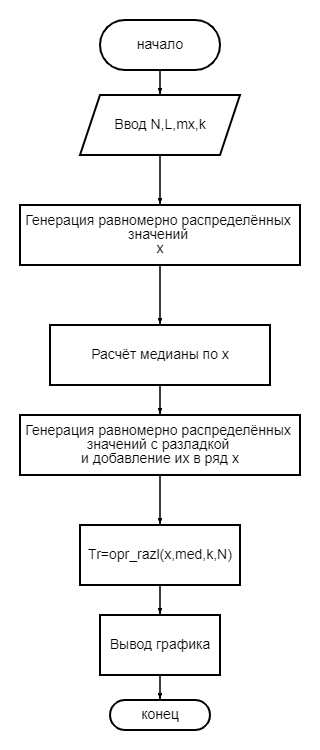


Рисунок 3

Проведём первоначальную апробацию программы. Выведем таблицу зависимости , от k и найдём по ней оптимальное значение k.

Исходные данные:

Медиана ряда с разладкой - 0.8

Количество повторений для одного k - 1000

Номер такта номинальной разладки - 5000

Длина сигнала - 10000

Таблица 1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Длина серии | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Среднее время между ложными тревогами | 62 | 125.9 | 255 | 489.8 | 950 | 1229 | - | - | - | - | - |
| Среднее время запаздывания | 10 | 14 | 19 | 24 | 32 | 40 | 54 | 67 | 84 | 109 | 130 |
| Номер такта обнаружения разладки | 5010 | 5014 | 5019 | 5024 | 5032 | 5040 | 5054 | 5067 | 5084 | 5109 | 5130 |

По результатам, представленным в таблице можно сделать вывод о том что с увеличением длины серии, увеличивается среднее время между ложными тревогами, при длине серий 11 и выше, они вообще пропадают, также увеличивается и среднее время запаздывания.

Здесь 10-оптимальное значение длины серий, так как отношение в этом случае будет наибольшим.

Выведем график для исходных данных и k=10.

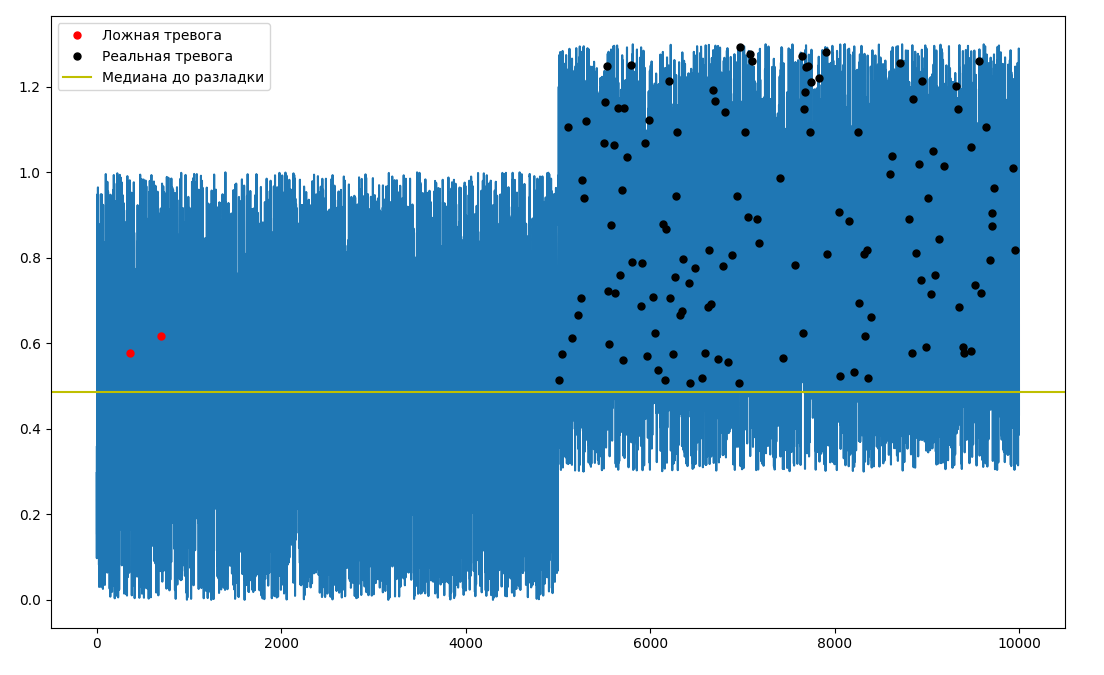


Рисунок 4 График процесса с данными о тревогах

Здесь красными точками обозначены ложные тревоги, в данном случае их всего две, что довольно хорошо. Черными точками обозначены моменты реальных тревог, видно, что первая реальная тревога наступает почти сразу после разладки, также алгоритм продолжает довольно часто сигнализировать пользователя о том, что произошла разладка.

Изменим медиану процесса с разладкой с 0.8 на 1 и сравним результаты моделирования.

Исходные данные:

Медиана ряда с разладкой - 1.0

Количество повторений для одного k - 1000

Номер такта номинальной разладки - 5000

Длина сигнала - 10000

Таблица 2

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Длина серии | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Среднее время между ложными тревогами | 62.2 | 125.9 | 256 | 505.7 | 924 | 1254 | - | - | - | - | - |
| Среднее время запаздывания | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Номер такта обнаружения разладки | 5005 | 5006 | 5007 | 5008 | 5009 | 5010 | 5011 | 5012 | 5013 | 5014 | 5015 |

По результатам, представленным в таблице видно, что в сравнении с предыдущими результатами, среднее время между ложными тревогами почти не поменялось, а среднее время запаздывания, наоборот, изменилось очень сильно и стало равным длине серии.

Выведем график для исходных данных и k=10.

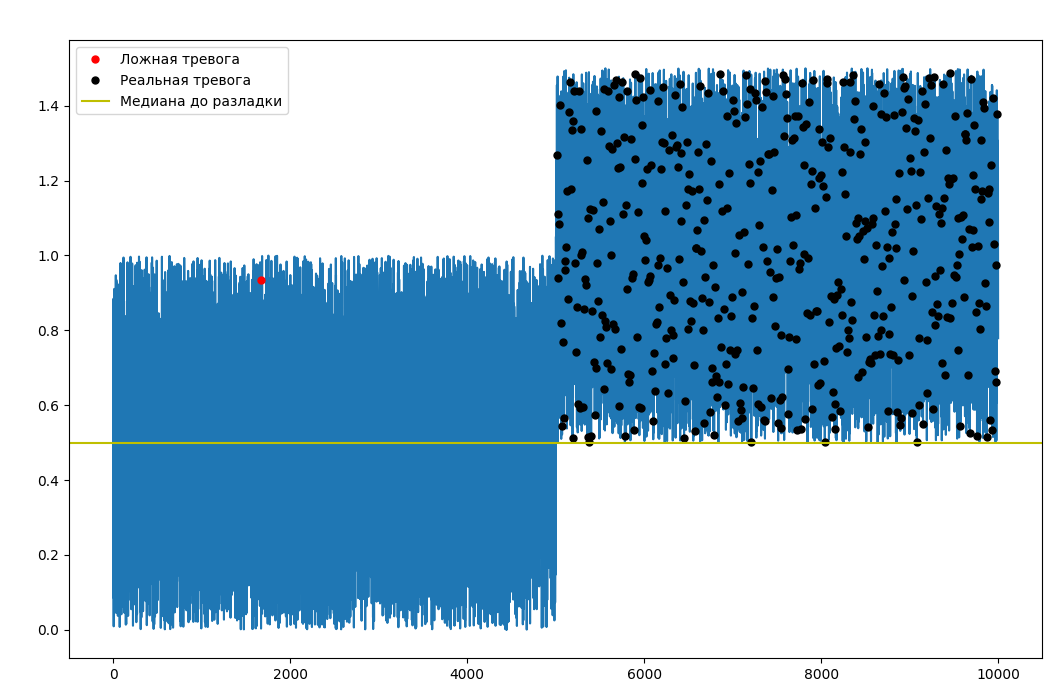


Рисунок 5 График процесса с данными о тревогах

Наблюдается всего одна ложная тревога. Видно, что сигналов о реальной тревоге в несколько раз больше, чем в прошлом случае.

Теперь выполним другую задачу: по заданному , определим наиболее оптимальное значение k.

Исходные данные:

Заданное среднее время между ложными тревогами - 500

Медиана ряда с разладкой - 0.8

Количество повторений для одного k - 1000

Номер такта номинальной разладки - 5000

Длина сигнала - 10000

Результаты моделирования:

Моменты времени ложных тревог - [463, 1291, 1384, 1842, 2224, 2910, 3174, 3767, 3782, 4464]

Среднее время между ложными тревогами - 444.55555555555554

Время запаздывания - 9

Оптимальное значение k - 8

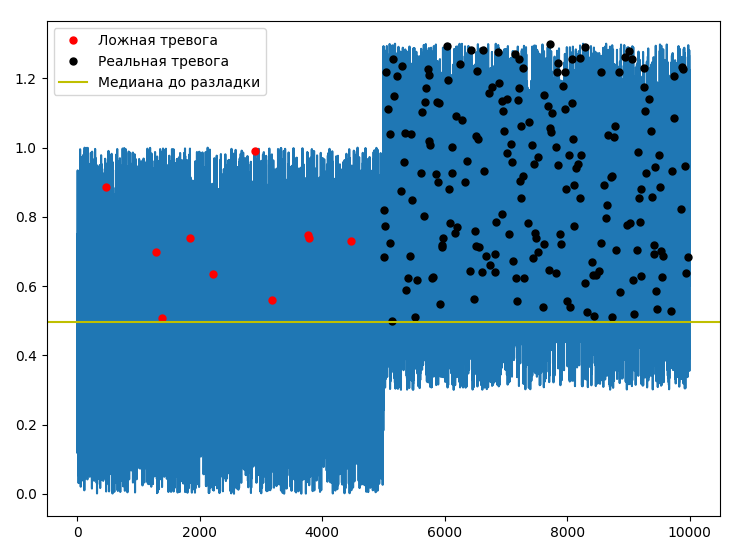


Рисунок 6 График процесса с данными о тревогах

На вход программы подано необходимое среднее время между ложными тревогами: . Алгоритм определил оптимальное значение длины серии k=8 и вывел график с результатами моделирования. Из результатов видно, что на практике среднее время между ложными тревогами получилось примерно равным 445, что довольно-таки хорошо, так как k может изменяться только в целом диапазоне чисел, поэтому сложно добиться высокой точности попадания в заданное.

# Выводы

В ходе данного исследования по проблематике обнаружения разладки временных рядов, реализована программа имитационного моделирования, основанная на теории случайных блужданий.

В ходе апробации получены зависимости и от длины серии k при различных значениях медианы разладки. По полученным данным выбрано оптимальное значение k. Выполнена задача определения оптимального значения длины серии k для заданного значения среднего времени между ложными тревогами.

# Изучение и подбор методов задания корреляции для гауссовского процесса.

Гауссовский процесс - это вероятностный процесс, в котором любая конечная коллекция случайных переменных имеет многомерное нормальное распределение. Он широко применяется в моделировании случайных функций и аппроксимации данных.

Корреляционная функция (или функция ковариации) играет ключевую роль в описании зависимости между значениями гауссовского процесса в разных точках. Она определяет степень связи или сходства между случайными переменными в разных местах. Выбор правильной корреляционной функции влияет на свойства и поведение гауссовского процесса.

Подбор методов задания корреляции для гауссовского процесса зависит от конкретной задачи и требований моделирования. Некоторые из наиболее распространенных методов включают следующие корреляционные функции:

1. Экспоненциальная корреляционная функция: Одна из наиболее простых и распространенных форм корреляционной функции. Она определяется параметром масштаба и описывает экспоненциальное затухание корреляции с увеличением расстояния между точками.

2. Гауссовская корреляционная функция: Используется в случаях, когда зависимость между точками может быть описана гауссовым распределением. Она имеет параметры масштаба и ширины и обладает гладкими свойствами.

3. Материал-корреляционная функция: Представляет собой корреляцию, которая уменьшается с расстоянием, но с медленной скоростью, что позволяет моделировать более длинные зависимости.

4. Квадратичная корреляционная функция: Используется для моделирования нерегулярных и неоднородных процессов. Она обладает большей гибкостью и позволяет учитывать нелинейные зависимости.

Выбор конкретного метода задания корреляции зависит от природы данных, предметной области и требований моделирования. Часто требуется проводить эксперименты и анализировать поведение гауссовского процесса с разными корреляционными функциями, чтобы найти наиболее подходящую для конкретной задачи.

Для задания корреляции гауссовского процесса выбрана модель авторегрессии второго порядка. Модель авторегрессии второго порядка представляет собой стохастическую модель, в которой каждое значение процесса зависит от двух предыдущих значений. Она является одной из моделей временных рядов, которая учитывает автокорреляцию и зависимость между значениями процесса на разных временных отрезках.

В модели авторегрессии второго порядка, текущее значение процесса y(t) выражается через два предыдущих значения y(t-1) и y(t-2) с определенными коэффициентами авторегрессии:

y(t) = ϕ₁ \* y(t-1) + ϕ₂ \* y(t-2) + ε(t),

где ϕ₁ и ϕ₂ - коэффициенты авторегрессии, определяющие веса предыдущих значений,

ε(t) - случайная ошибка или шум с нулевым средним и конечной дисперсией.

Модель авторегрессии второго порядка позволяет учесть зависимость между текущим значением процесса и двумя предыдущими значениями. Коэффициенты авторегрессии ϕ₁ и ϕ₂ определяют влияние предыдущих значений на текущее значение процесса. Они могут быть положительными или отрицательными, что указывает на положительную или отрицательную корреляцию между значениями процесса.

Основными характеристиками модели авторегрессии второго порядка являются способность учесть зависимость между предыдущими значениями, возможность предсказания будущих значений и анализ изменений в поведении процесса. Однако следует отметить, что выбор модели всегда зависит от конкретной задачи, данных и предполагаемой структуры временного ряда.

**Исследование работы метода случайных блужданий на бинарном коррелированном сигнале в отсутствии разладки:**

Разработана программа для получения коррелированного бинарного сигнала.

Настраиваемыми параметрами являются b1 и b2.

По корреляционным функциям определяются характеристики:

τ мк – абсцисса при вхождении в пятипроцентный диапазон.

τ к – сумма ординат графика.

**Сравнение работы метода для сигнала без корреляции и для коррелированного сигнала:**

Все исследования проводятся при числе серий k=5, длине сигнала 10000. Разладка не вводилась.

Исследования сигнала без корреляции.

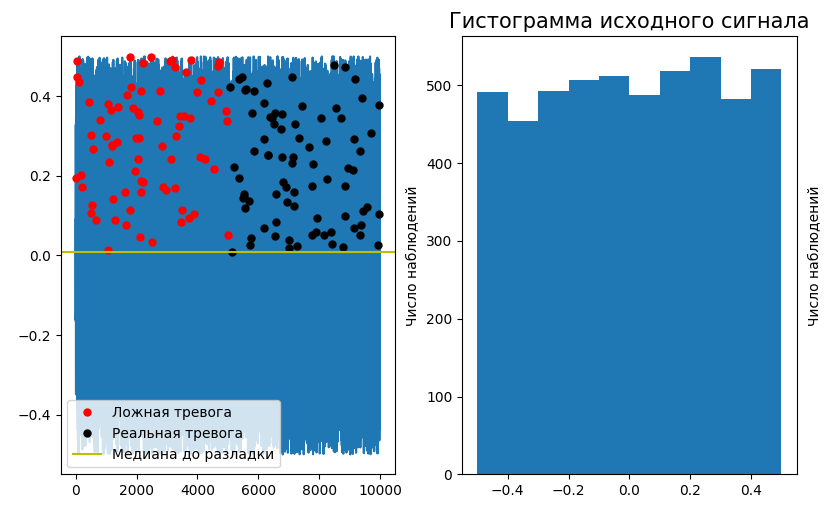


Рисунок 7 График процесса с данными о тревогах, гистограмма исходного сигнала

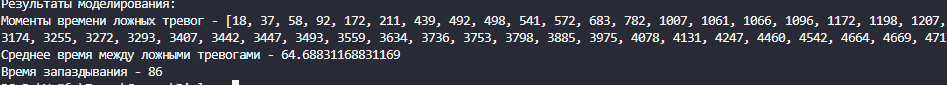


Рисунок 8 Результаты моделирования

Множество ложных тревог. Среднее время между ложными тревогами– 64

**Исследование коррелированного сигнала**

**Параметры возьмем следующие:**

b1 = 0.5, b2 = -0.75

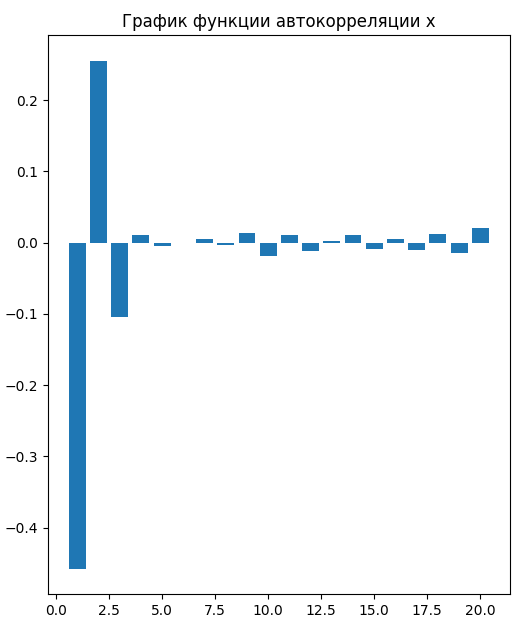
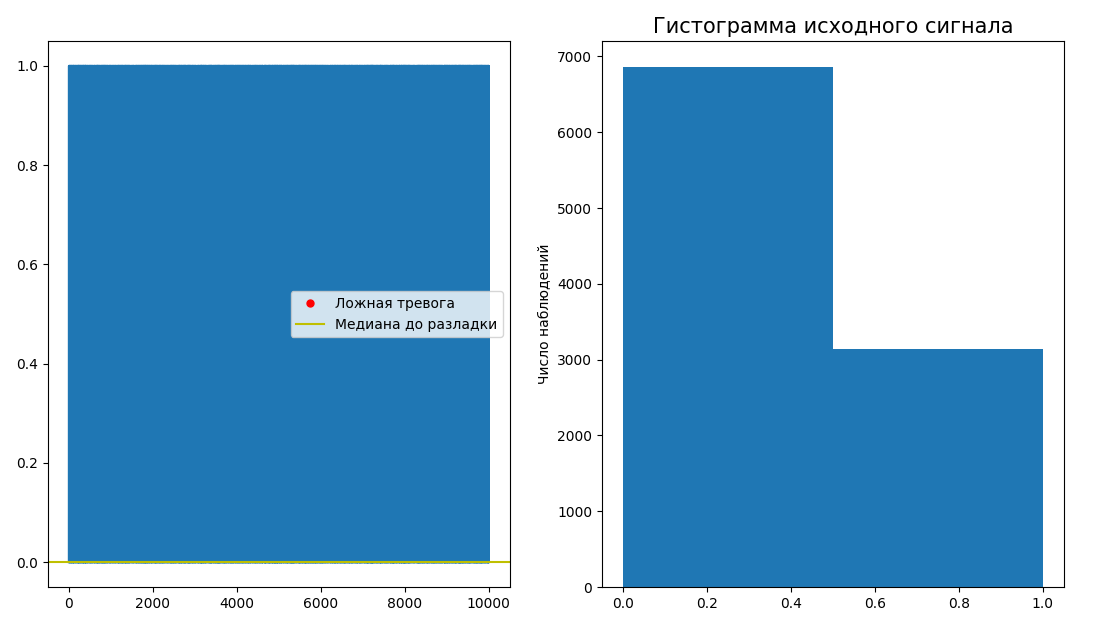


Рисунок 9 График процесса с данными о тревогах, гистограмма исходного сигнала, корреляционная функция

Результаты моделирования:

Ложные тревоги отсутствуют

Алгоритм не определил ни одной ложной тревоги, так как большая часть коррелированного ряда имеет значение, равное 0, что можно судить по гистограмме. Происходит это потому, что алгоритм считает серии только тех значений, которые лежат выше медианы.

**Возьмем другие параметры:**

b1 = 0.25, b2 = -0.25

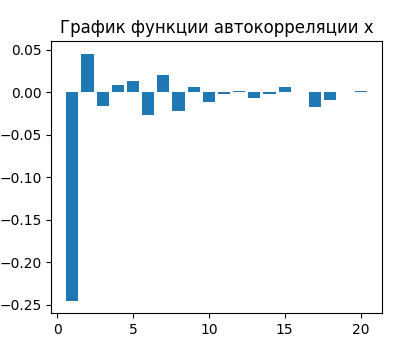
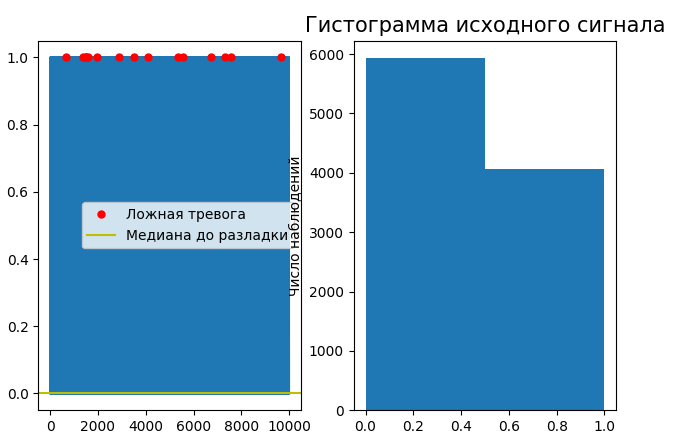


Рисунок 10 График процесса с данными о тревогах, гистограмма исходного сигнала, корреляционная функция



Рисунок 11 Результаты моделирования

В данном случае алгоритм определил множество ложных тревог. Это происходит, потому что в данном случае из 10000 значений ряда, примерно 4000 значений равны 1 , следовательно может возникнуть больше серий подряд идущих значений, равных 1, нежели в предыдущем эксперименте, и алгоритм фиксирует данные серии, если их длина больше или равна 5.

**Возьмём параметры b1 и b2 равными друг другу:**

b1 = 0.25, b2 = 0.25

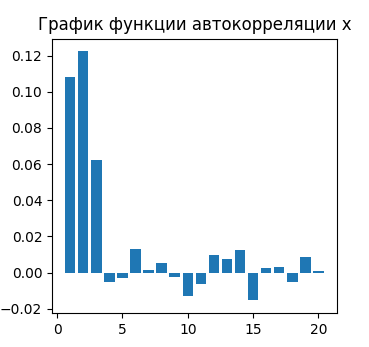
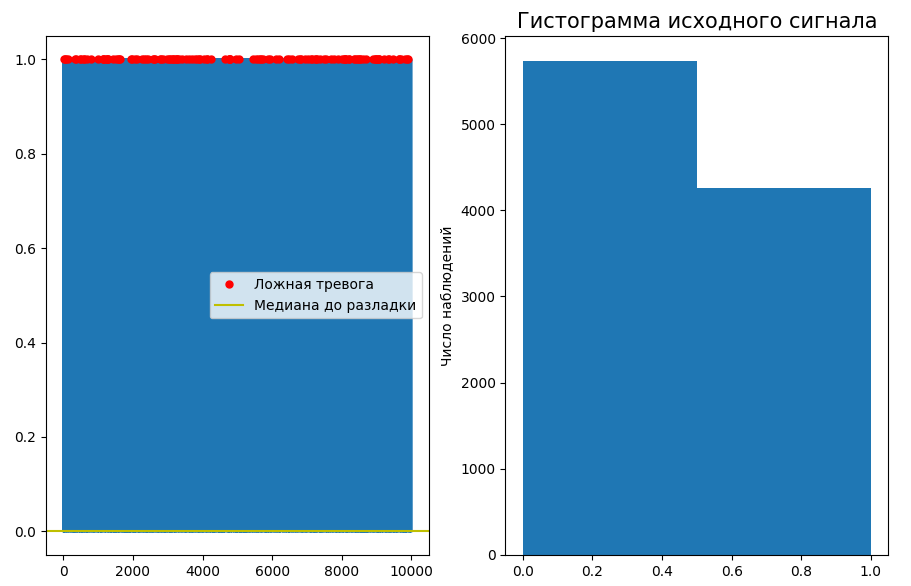
****

Рисунок 12 График процесса с данными о тревогах, гистограмма исходного сигнала, корреляционная функция

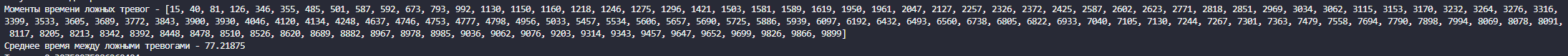
****

Рисунок 13 Результаты моделирования

Здесь тревог в разы больше, нежели в прошлом эксперименте, хотя число значений, равных 1, не сильно отличается значений в предыдущем эксперименте.

**Исследование работы алгоритма при коррелированном бинарном сигнале**

Исследуем другие пары значений и построим таблицу с получившимися характеристиками:

Таблица 3

|  |  |  |
| --- | --- | --- |
|  | τ мк | τ к |
| b1 = 0.5, b2 = -0.75 | 4 | 0.28 |
| b1 = 1.5, b2 = -0.75 | 3 | 0.08 |
| b1 = 0.25, b2 = -0.25 | 2 | 0.21 |
| b1 = 0.25, b2 = -0.5 | 2 | 0.23 |
| b1 = 1, b2 = -0.5 | 3 | 0.19 |

Посмотрим, как меняется среднее время между ложными тревогами при различных значениях коэффициентов в зависимости от числа серий.

Длина сигнала 5000, число усреднений 1000

Таблица 4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 5 | 6 | 7 | 8 | 9 | 10 |
| b1 = 0.5, b2 = -0.75 | - | - | - | - | - | - |
| b1 = 1.5, b2 = -0.75 | - | - | - | - | - | - |
| b1 = 0.25, b2 = -0.25 | 802.47 | - | - | - | - | - |
| b1 = 0.25, b2 = -0.5 | - | - | - | - | - | - |
| b1 = 1, b2 = -0.5 | - | - | - | - | - | - |

Как видно из таблицы, алгоритм почти не выдаёт сообщения о тревогах.

# Выводы

В ходе исследования работы метода случайных блужданий на бинарном коррелированном сигнале в отсутствии разладки, выяснено, что эффективность работы метода случайных блужданий сильно зависит от параметров b1 и b2. Так как данные параметры определяют веса предыдущих значений, от них будет зависеть вид гистограммы анализируемого процесса. Следовательно будет менятся эффективность работы алгоритма.

**Исследование работы метода случайных блужданий на “гауссовском” коррелированном сигнале:**

Разработана программа для получения коррелированного гауссовского сигнала.

Настраиваемыми параметрами являются b1 и b2.

По корреляционным функциям определяются характеристики:

τ мк – абсцисса при вхождении в пятипроцентный диапазон.

τ к – сумма ординат графика.

**Сравнение работы метода для сигнала без корреляции и для коррелированного сигнала:**

Все исследования проводятся при числе серий k=10, длине сигнала 10000. Разладка начинается с 5000 такта.

Медиана до разладки – 0.5 , Медиана после разладки – 0.7.

*Исследования для сигнала без корреляции*

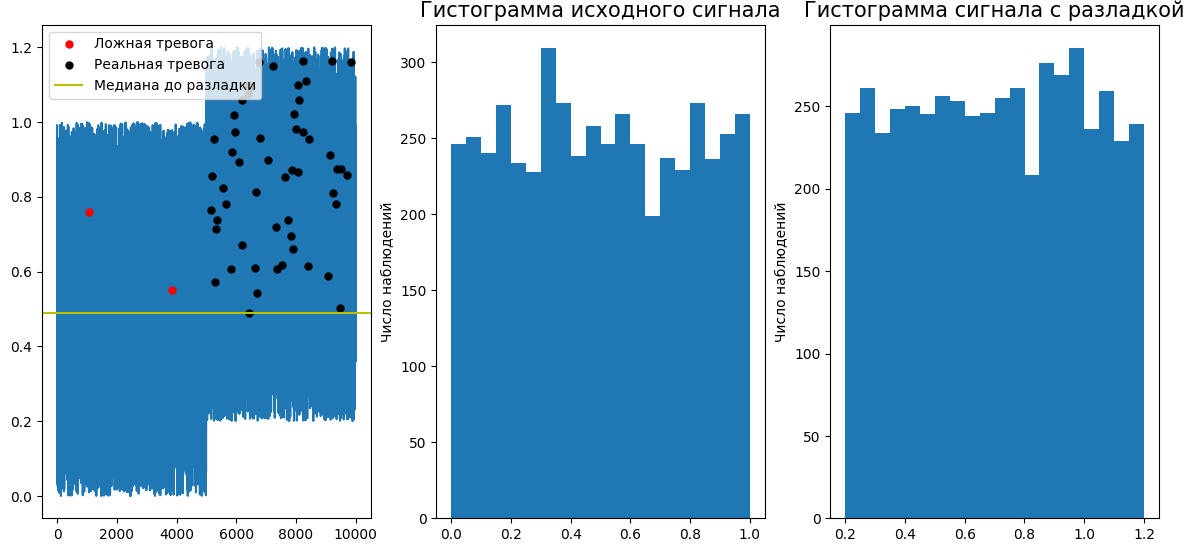


Рисунок 14 График процесса с данными о тревогах, гистограмма исходного сигнала и сигнала с разладкой

Результаты моделирования:

Моменты времени ложных тревог - [1080, 3863]

Среднее время между ложными тревогами – 2783

Время запаздывания - 94

Наблюдаем всего две ложные тревоги и множество реальных тревог. Время запаздывания довольно мало, что хорошо.

***Исследование для сигнала с корреляцией***

**Параметры возьмем следующие:**

b1 = 0.5, b2 = -0.75

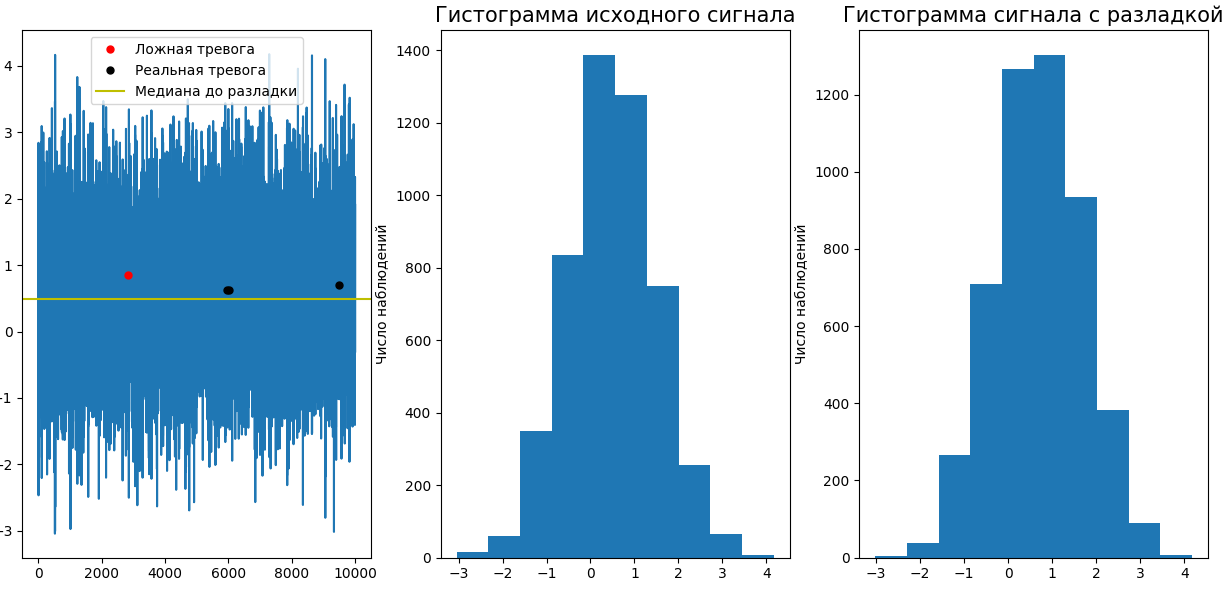
**

Рисунок 15 График процесса с данными о тревогах, гистограмма исходного сигнала и сигнала с разладкой

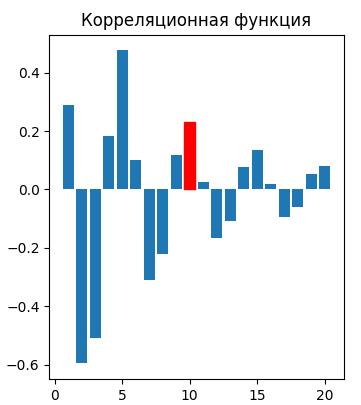
**

Рисунок 16 Корреляционная функция

Результаты моделирования:

Моменты времени ложных тревог - [2831]

Среднее время между ложными тревогами - -

Время запаздывания - 949

τ мк = 10

τ к = 0.236

Наблюдаем одну ложную тревогу и всего две реальные тревоги, что не хорошо. Также время запаздывания имеет довольно большое значение.

**Изменим значения параметров:**

b1 = 0.25, b2 = -0.25

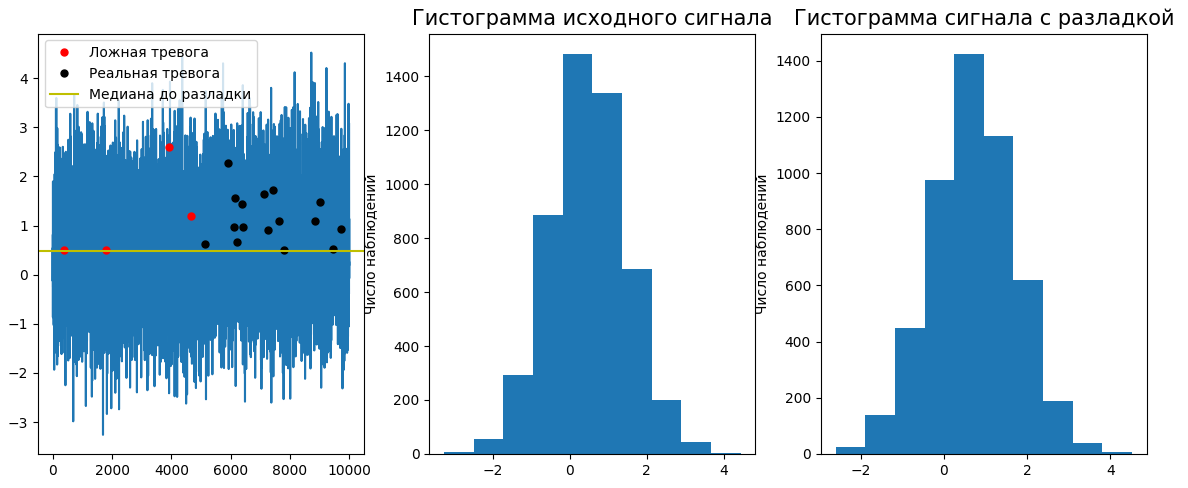
****

Рисунок 17 График процесса с данными о тревогах, гистограмма исходного сигнала и сигнала с разладкой

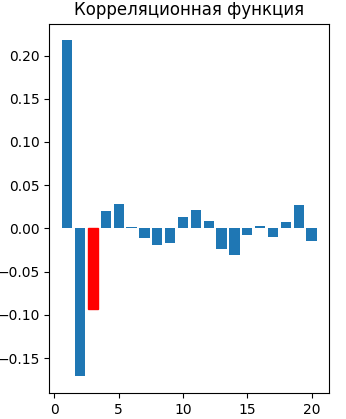


Рисунок 18 Корреляционная функция

Результаты моделирования:

Моменты времени ложных тревог - [385, 1812, 3905, 4674]

Среднее время между ложными тревогами - 1429.6666666666667

Время запаздывания - 120

τ мк = 3

τ к = 0.047

Наблюдаем 4 ложных тревоги и множество реальных тревог, что лучше, чем в прошлом эксперименте. Время запазывания, относительно прошлого эксперимента уменьшилось, что хорошо

# Выводы

В ходе исследования работы метода случайных блужданий на “гауссовском” коррелированном сигнале в присутствии разладки, выяснено, что эффективность работы метода случайных блужданий также сильно зависит от параметров b1 и b2.

В сравнении с некоррелированным сигналом, алгоритм показывает себя хуже.

Теперь выполним другую задачу: по заданному , определим наиболее оптимальное значение k.

Исходные данные:

Заданное среднее время между ложными тревогами - 500

Медиана ряда с разладкой - 0.7

Количество повторений для одного k - 1000

Номер такта номинальной разладки - 5000

Длина сигнала – 10000

***Результаты моделирования для некоррелированного сигнала:***

Оптимальное значение k - 8

Моменты времени ложных тревог - [420, 1684, 1890, 2691, 3768, 4489, 4720]

Среднее время между ложными тревогами - 716.6666666666666

Время запаздывания – 86

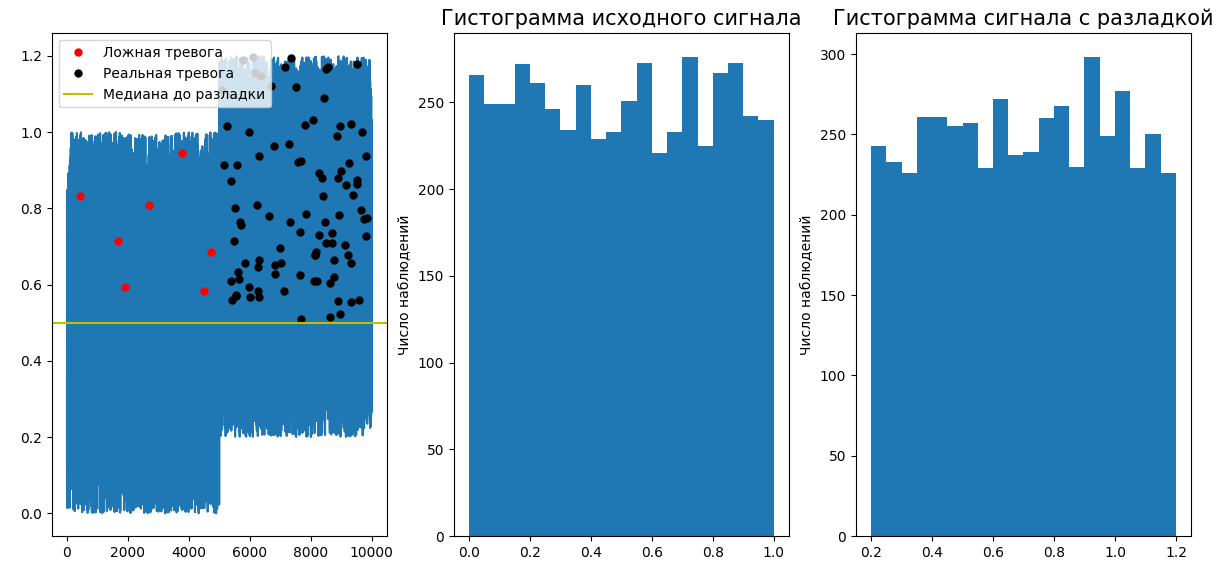


Рисунок 19 График процесса с данными о тревогах, гистограмма исходного сигнала и сигнала с разладкой

На вход программы подано необходимое среднее время между ложными тревогами: . Алгоритм определил оптимальное значение длины серии k=8 и вывел график с результатами моделирования. Из результатов видно, что на практике среднее время между ложными тревогами получилось примерно равным 716, что довольно-таки хорошо, так как k может изменяться только в целом диапазоне чисел, поэтому сложно добиться высокой точности попадания в заданное.

***Результаты моделирования для коррелированного сигнала:***

Оптимальное значение k – 6

Моменты времени ложных тревог - [45, 462, 647, 1268, 1407, 1604, 1903, 1938, 2296, 2450, 2746, 2917, 3186, 4178]

Среднее время между ложными тревогами - 317.9230769230769

Время запаздывания - 109

τ мк = 10

τ к = 0.26

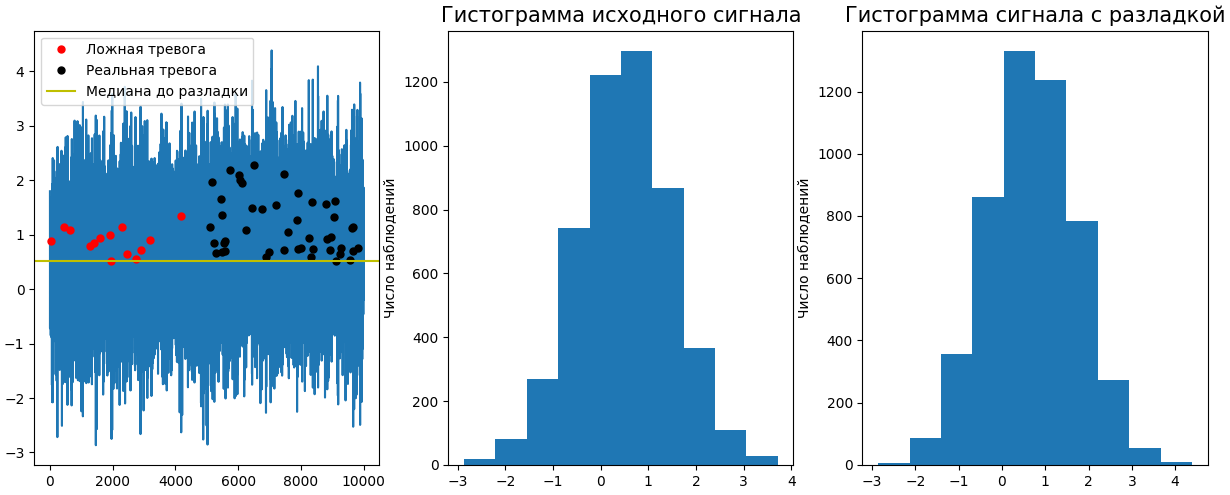


Рисунок 20 График процесса с данными о тревогах, гистограмма исходного сигнала и сигнала с разладкой

На вход программы подано необходимое среднее время между ложными тревогами: . Алгоритм определил оптимальное значение длины серии k=6 и вывел график с результатами моделирования. Из результатов видно, что на практике среднее время между ложными тревогами получилось примерно равным 317, что довольно-таки хорошо, так как k может изменяться только в целом диапазоне чисел, поэтому сложно добиться высокой точности попадания в заданное.