

Содержание

Введение	3
1.Методы обнаружения спонтанных разладки наблюдаемых процессов	4
2.Алгоритм, основанный на теории случайных блужданий	7
3.Программная реализация и апробация.....	9
Выводы	16
Список литературы	17

Введение

Рассматривается задача наискорейшего и наиболее точного определения момента изменения свойств изучаемых объектов, характера протекающих в них процессов, часто именуемая задачей о разладке. Эта задача на практике встречается весьма часто: это и обнаружение аварийных ситуаций на производстве, при организации контроля за исправностью оборудования и качества готовой продукции, для обнаружения внезапно появляющихся объектов или сигналов (например, в сейсмологии), при построении адаптивных алгоритмов управления и т.д.

По сути разладка стохастических процессов – это спонтанное изменение характеристик наблюдаемого случайного процесса, происходящее в неизвестный, заранее непредсказуемый момент времени. Общепринятая математическая постановка задачи обнаружения в реальном времени разладки дискретного стохастического процесса $X(t)$ может быть описана следующим образом. Пусть наблюдается последовательность $x(t_1) = x_1 \quad x(t_2) = x_2, \dots$ Предполагается, что контролируемый объект может находиться в двух состояниях: «норма» (основное рабочее состояние), которому соответствует определенные вероятностные характеристики процесса $X(t)$, и «разладка», когда вероятностные свойства $X(t)$ отличаются от имевших место в состоянии «норма». Переход из состояния «норма» в состояние «разладка» происходит в некоторый заранее непрогнозируемый момент времени t_0 . Требуется по последовательно поступающим значениям x_1, x_2, \dots найти (обнаружить) t_0 , причем желательно с минимальным запаздыванием $\tau_{\text{зап}}$.

Ясно, что для оперативного обнаружения разладки необходимо использовать методы и алгоритмы, относящиеся к категории последовательных, пригодных для работы в реальном времени, в ритме с поступлением значений x_i .

Методы обнаружения спонтанных разладки наблюдаемых процессов

Все последовательные методы обнаружения основаны на вычислении по наблюдаемым дискретным значениям x_i ($i = 1, 2, \dots, n, \dots$) в темпе с их поступлением значений решающий функции $g_i = g_i(x_i, x_{i-1}, x_{i-2}, \dots)$. Эта функция вычисляются рекуррентно, исходя из значений статистики на предыдущем шаге и очередного измеренного значения x_i . Вычисленное значение g_i затем сопоставляется с определённым пороговым уровнем H , разделяющим область возможных значений статистики на две подобласти: Ω_0 и Ω_1 , где Ω_0 – подобласть продолжения наблюдений, Ω_1 – критическая подобласть. Если $g_i \in \Omega_0$ – то процесс контроля продолжается, т.е. считается, что разладка отсутствует; если же $g_i \in \Omega_1$, то подаётся сигнал о наличии разладки (сигнал тревоги). Контролирующая процедура при этом по желанию экспериментатора может запускаться заново.

Наиболее часто предполагается, что разладка проявляется не в виде изменения функции распределения вероятностей $f_{0x}(x)$, а в параметрической форме: в виде изменения математического ожидания процесса, его дисперсии, корреляционно-спектральных свойств и т.п.

Известны два принципиально различных подхода к решению указанной задачи: апостериорный и последовательный. При апостериорном подходе реализация $x(t)$ предварительно записывается, а затем в ходе обработки производится оценка неизвестного значения t_0 . В данном варианте задача может быть интерпретирована как обычная задача оценивания, а к соответствующим алгоритмам предъявляются стандартные требования получения оценок с наибольшей точностью, скажем, с минимальным смещением и дисперсией.

Однако больший практический интерес представляет второй вариант, когда обнаружение разладки происходит в реальном масштабе времени

одновременно с поступлением данных о значении случайного процесса $x(t)$. Требуется, производя текущую обработку данных, обнаружить разладку как можно быстрее после ее появления. В этом случае говорят о последовательных алгоритмах обнаружения разладки.

Все алгоритмы обнаружения разладки можно разделить на две большие группы: методы параметрические и непараметрические.

Параметрические методы базируются на предположении, что наблюдаемый (контролируемый) процесс имеет априори известную функцию распределения вероятностей, входящую в некоторое параметрическое семейство распределений, причем разладка связана с изменением одного или нескольких параметров этого распределения.

Непараметрические методы не требуют какой-либо существенной информации о виде функций распределения контролируемого процесса. Поэтому их часто именуют процедурами «свободными от распределения». Их главная отличительная черта - они оперируют не с самими измеренными значениями, а с их знаковыми, ранговыми или порядковыми представлениями.

Любой такой алгоритм должен обладать некоторыми оптимальными свойствами, а именно желательно, с одной стороны, добиться минимального среднего времени в обнаружении разладки (минимальное среднее значение запаздывания в обнаружении разладки $\bar{t}_{\text{зап}}$, а с другой – максимизировать среднее значение интервала между ложными тревогами $\bar{T}_{\text{лт}}$.

Любой последовательный контролирующий алгоритм может быть охарактеризован с помощью определённых вероятностных характеристик. На практике наиболее часто используются следующие характеристики:

- 1) среднее значение интервала между ложными тревогами $T_{\text{лт}}$, т.е. среднее время между подачами сигналов о наличии разладки, когда в действительности её нет;

2) среднее время запаздывания в обнаружении номинальной разладки в ритме с поступлением данных (в динамике) $\tau_{\text{зап}}^*$;

На практике любой контролирующий алгоритм должен быть настроен с учётом имеющейся информации о характеристиках наблюдаемого процесса до и после разладки и желательных свойствах контролирующего алгоритма. В качестве настроечных параметров, как правило, выступают номинальное значение разладки и порог. Настройка состоит в определении решающего порога H при установленном значении номинальной разладки и выбранном, исходя из конкретных требований решаемой задачи, среднем интервале между ложными тревогами $T_{\text{лт}}$. Для реализации настройки и оценки потенциальных свойств контролирующего алгоритма необходима предварительно полученная справочная информация, которая – в графической или табличной форме – должна как минимум включать в себя:

- зависимости порога H от $T_{\text{лт}}$ для разных значений номинальной разладки;

- зависимости $\tau_{\text{зап}}$ от H или от $T_{\text{лт}}$ для оценки быстродействия алгоритма при номинальной разладке.

Алгоритм, основанный на теории случайных блужданий

Впервые идея использования механизма одномерных случайных блужданий как средства решения задачи обнаружения разладки была высказана в работе [20], опубликованной в 1975 г. В ней рассматривалась классическая схема одномерных случайных блужданий, когда на каждом шаге в некоторый дискретный момент времени n текущая координата точки z_n , отображающей процесс блуждания, с вероятностью p получает приращение $+1$ или с вероятностью $q=1-p$ – приращение -1 в соответствии с соотношением:

$$Z_n = \begin{cases} +1, & x_n \geq Me(x) \\ -1, & x_n \leq Me(x) \end{cases}, n = 1, 2, \dots$$

$Me(x)$ - значение медианы наблюдаемого временного ряда в состоянии «норма». Отметим, что в этом случае справедливо равенство $p = q = 0.5$; однако при наличии разладки эта симметрия нарушается. Можно, например, считать, что разладка приводит к сдвигу медианы в положительную сторону, следовательно, тогда будем иметь вероятность $p > 1/2$. Далее рассматривается одно из возможных направлений развития методов обнаружения разладки на базе применения механизма случайных блужданий, а именно с использованием теории серий «успехов».

Контролирующая процедура предлагаемого метода сводится на каждом n -ом такте работы к вводу очередного значения контролируемого процесса x_n , его преобразованию в значение приращения z_n в соответствии с указанным выше соотношением, подсчету k_n - текущей длины серии из следующих подряд значений $z_n = +1$: $k_n = k_{n-1} + 1$, если $z_n = +1$ и $k_n = 0$, если $z_n = -1$, фиксации номера такта n^* , на котором будет выполнено условие $k_n^* = k_n = k$, где k – заранее заданное пользователем значение, выбранное исходя из необходимости обеспечения желательного среднего времени между ложными тревогами $\bar{T}_{лт}$.

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 k_n из плюсов по следующему правилу:

$$z_i = +1, x_n > Me(X), n = 1, 2, \dots$$

3. Подсчитывается количество случаев, когда серия k_n соответствует заданному значению k - число серий в совокупности z_i .

4. Определяется такт, на котором впервые k_n достигает заданного значения k после номинальной разладки.

5. Рассчитываются показатели $\bar{\tau}_{зан}$ и $\bar{T}_{лт}$

Программная реализация и апробация

В основной программе пользователем задаётся длина вариационного ряда – L . Помимо этого задаётся номер такта номинальной разладки – N и m_x – 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 $\bar{T}_{лт}$, $\bar{\tau}_{зан}$ от k , определение оптимального k для заданного $T_{лт}$), для этого вводится параметр st . В функциях `out_table()` и `opt_k_for_Tlt()` вводится параметр m – число экспериментов для одного k .

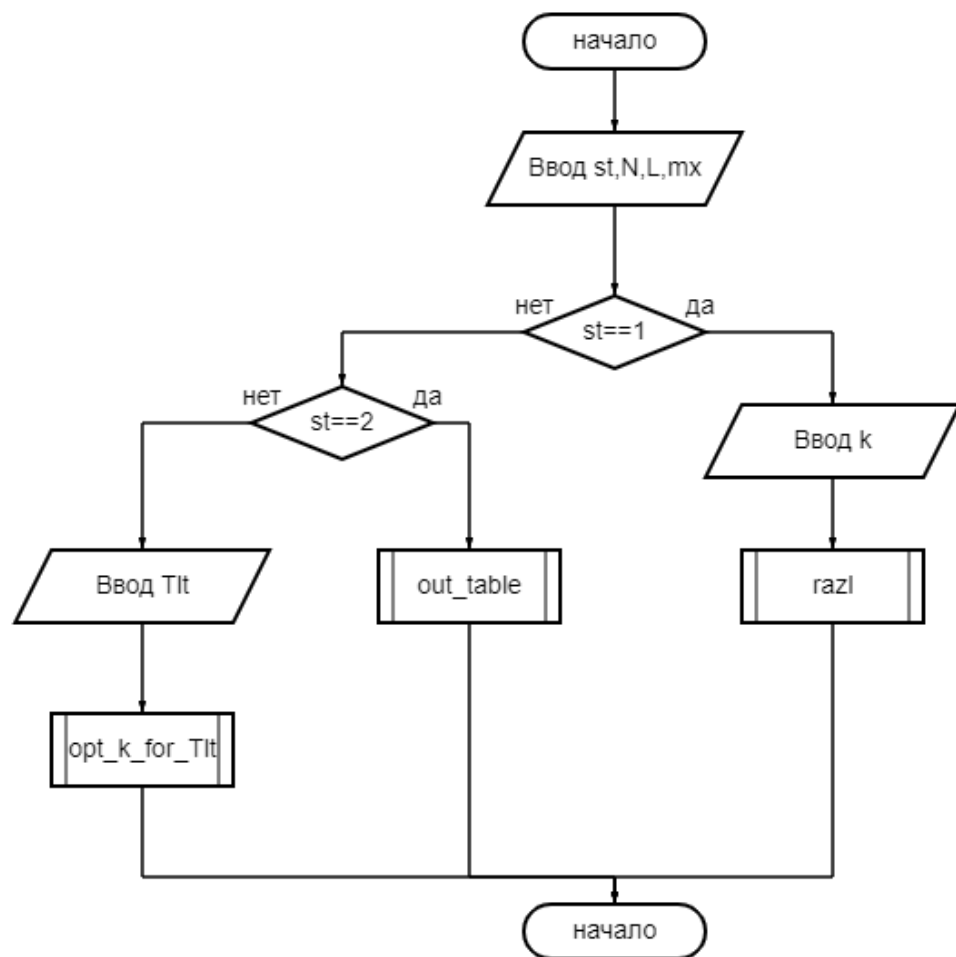


Рисунок 1 Структура основной программы

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.

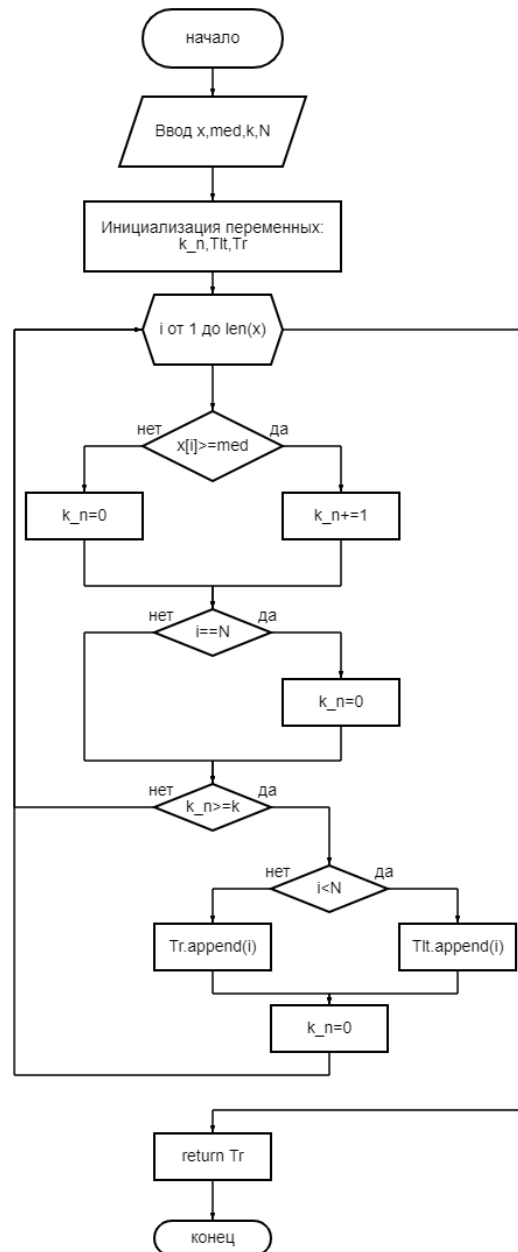


Рисунок 2 Структура функции по определению разладки

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 k_n прибавляется единица $k_n = k_n + 1$, а если меньше, то k_n обнуляется. Если k_n 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 , m_x , k . N -номер такта номинальной разладки, L -длина вариационного ряда, m_x -медианна процесса с разладкой, k -число серий.



Рисунок 3

Проведём первоначальную апробацию программы. Выведем таблицу зависимости $\bar{T}_{лт}$, $\bar{\tau}_{зап}$ 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 $\frac{\bar{T}_{лт}}{\bar{\tau}_{зап}}$ в этом случае будет наибольшим.

Выведем график для исходных данных и $k=10$.

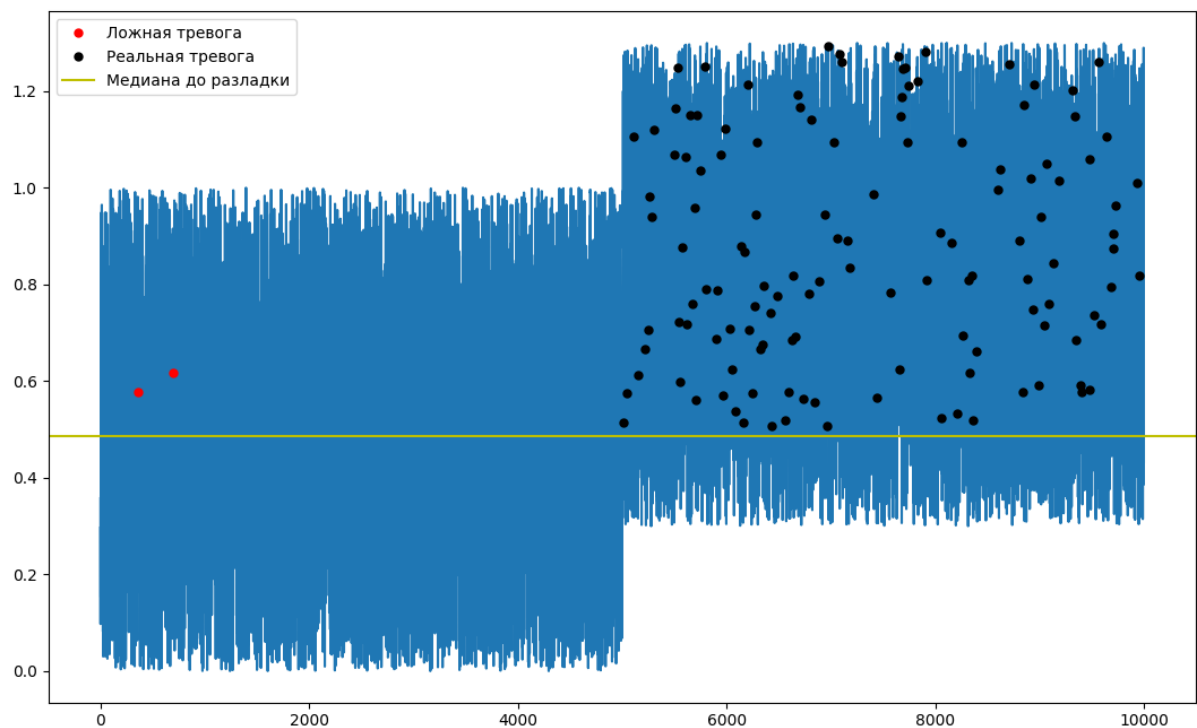


Рисунок 4 График процесса с данными о тревогах

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 $k=10$.

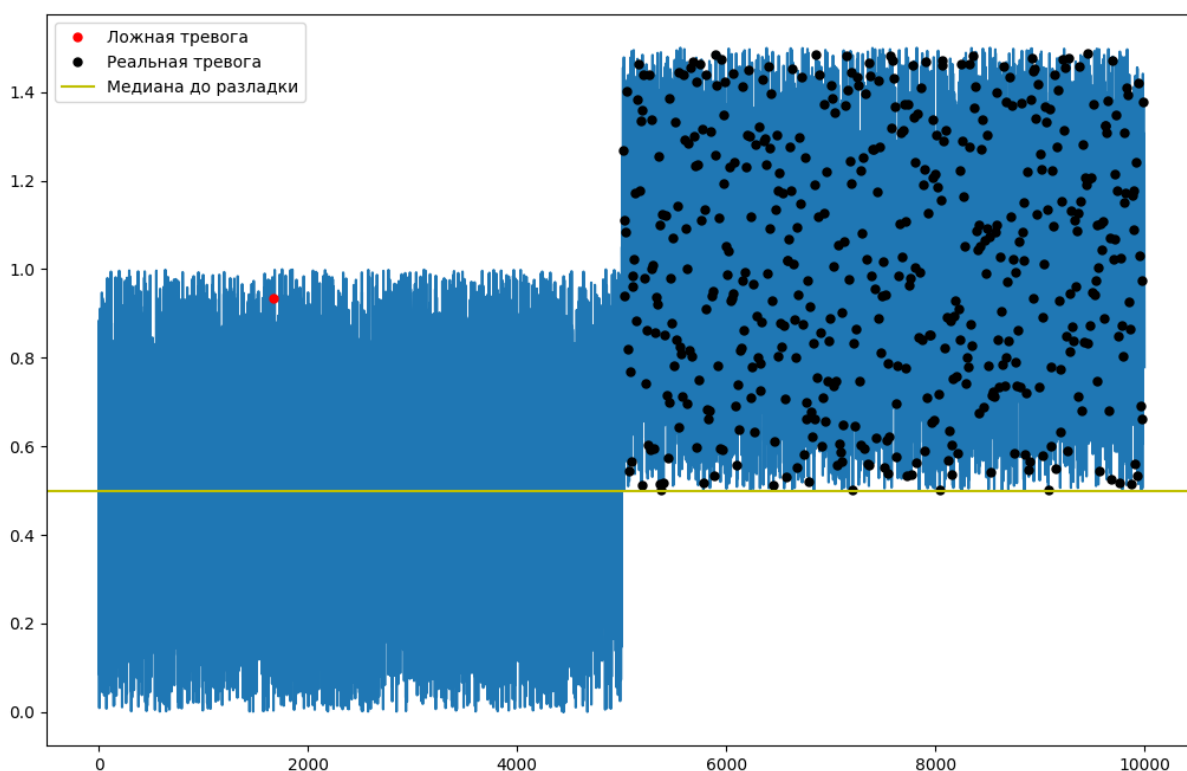


Рисунок 5 График процесса с данными о тревогах

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 $\bar{T}_{\text{лт}}$, 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

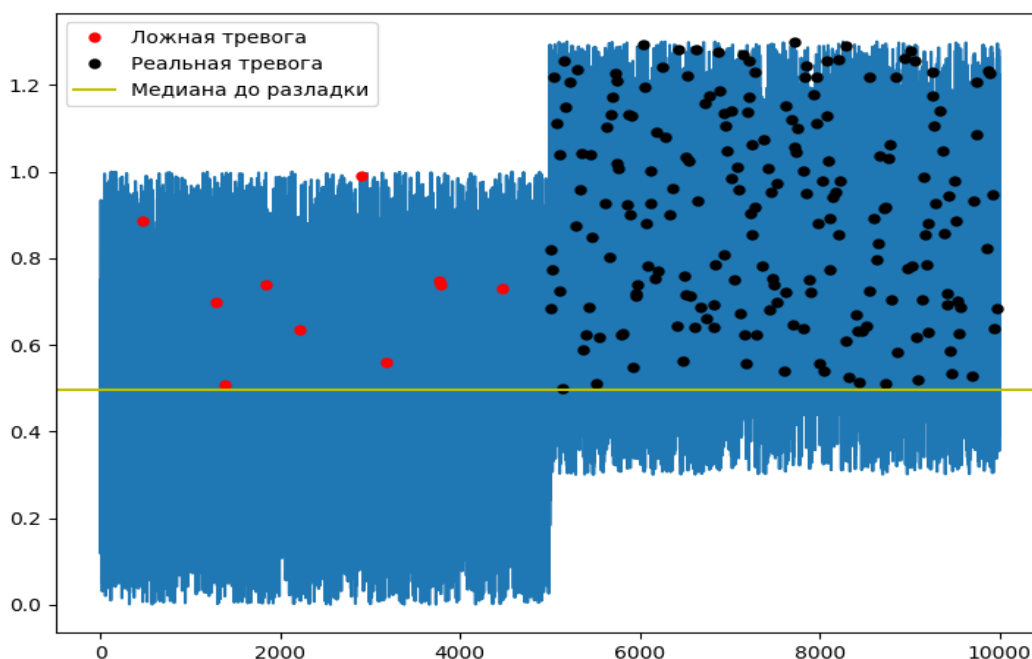


Рисунок 6 График процесса с данными о тревогах

На вход программы подано необходимое среднее время между ложными тревогами: $\bar{T}_{лт} = 500$. 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 $\bar{T}_{лт}$.

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 $\bar{T}_{лт}$ и $\bar{t}_{зап}$ 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.

Список литературы

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. Непараметрический

метод обнаружения разладки временного ряда с использованием

Механизма случайных блужданий» »// Известия ЮФУ. Технические

науки. 2020, №4, с.107-117.
