Министерство образования и науки Российской Федерации
(МИНОБРНАУКИ РОССИИ)
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (ТГУ)
Механико-математический факультет
Кафедра математического анализа и теории функций

ДОПУСТИТЬ К ЗАЩИТЕ В ГЭК

Руководитель ООП
д-р физ.-мат. наук, профессор
_____ А.В. Старченко
«14» 06 2018г

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

ОБ ОБНАРУЖЕНИИ "РАЗЛАДКИ" В ПАРАМЕТРАХ ЛИНЕЙНОЙ МОДЕЛИ

по основной образовательной программе подготовки бакалавров
направление подготовки 02.03.01 - Математика и компьютерные науки

Клемешова Анна Ивановна

Руководитель ВКР

канд. физ.-мат. наук, доцент

_____ Т.В.Емельянова
подпись

«14» июня 2018г

Автор работы
студент группы № 04401
_____ А. И. Клемешова
подпись

Томск-2018

Аннотация

В работе рассмотрены два различных алгоритма обнаружения "разладки" в модели авторегрессии первого порядка и в моделях, описывающих появление сигнала на фоне независимых и зависимых шумов авторегрессионного типа. Симмуляционное моделирование показало, что при достаточной разнице ($|\varphi_0 - \varphi_1| \geq 0.3$ и $|\delta_0 - \delta_1| \geq 0.3$) в коэффициентах авторегрессионной модели лучшие результаты показывает алгоритм, основанный на отношении правдоподобия. Алгоритм CuSum применим для более широкого класса моделей. При незначительной разнице в параметрах ($|\varphi_0 - \varphi_1| < 0.3$ и $|\delta_0 - \delta_1| < 0.3$) он приводит к более качественному обнаружению $"$разладки$"$ при хорошем подборе коэффициентов α и β, где α - вероятность ложной тревоги, а β - вероятность ложного спокойствия.

Ключевые слова: "разладка", момент остановки, ложная тревога, ложное спокойствие.

СОДЕРЖАНИЕ

Введение 5

1 Доверительное оценивание для параметров биномиального распределения 9

1.1 Понятие доверительного интервала 9

1.2 Построение доверительного интервала с помощью центральной статистики 10

1.3 Построение доверительного интервала с использованием распределения точечной оценки параметра 11

1.4 Асимптотические доверительные интервалы. 13

1.5 Доверительный интервал для параметра бернуллиевской модели 14

1.6 Доверительное оценивание и его использование для обнаружения "разладки" 16

2 Обнаружение "разладки" в авторегрессионных моделях 26

2.1 Проверка гипотезы об обнаружении "разладки" в первом коэффициенте модели AR(1) 27

2.1.1 Построение критерия обнаружения "разладки" 27

2.1.2. Асимптотическое распределение тестовой статистики 29

2.2 Проверка гипотезы об обнаружении "разладки" во втором коэффициенте модели AR(1) 31

2.2.1 Построение критерия обнаружения "разладки" 31

2.2.2 Асимптотическое распределение тестовой статистики 33

2.3 Последовательный непараметрический метод обнаружения "разладок" случайных процессов рекуррентного типа 35

2.3.1 Постановка задачи 35

2.3.2 Построение решающей процедуры 36

2.3.3 Выбор параметров и исследование свойств решающей процедуры 39

3 Результаты численного моделирования 49

Вывод 58

Заключение 60

Список использованной литературы 62

Приложение А Первичная обработка данных 65

ВВЕДЕНИЕ

Описание и исследование временных рядов является одной из актуальных задач в статистике и ее приложениях. При моделировании временных рядов важное значение имеет сохранение вероятностных характеристик. Момент, когда свойства исследуемого процесса меняются можно считать моментом "разладки". Алгоритмы для решения такой задачи-задачи обнаружения "разладки" (change point) широко применяются в разных областях знаний, таких как контроль качества промышленности, финансовые рынки, медицинская диагностика.

История развития задач о "разладке" уходит в недалекое прошлое. Но, несмотря на столь небольшой промежуток времени, это направление развито достаточно хорошо и является актуальным по сей день. Учитывая, что уже опубликовано большое количество работ, сегодня есть над чем работать, что изучать и что развивать в данной области.

В 1960 году в городе Вильнюс на VIBсесоюзном совещании по теории вероятностей и математической статистике впервые была рассмотрена задача о "разладке" А.Н. Ширяевым. Результаты отражены в последующих его статьях [11],[12], [14]. Также Альберт Николаевич решал задачи о "разладке" для различных моделей: дифузионных и для временных рядов [8],[9],[10], [11].

Например, на производстве изготавливаются детали для каких-либо механизмов и нам важна точность размеров. В какой-то момент времени машина начинает изготавливать детали с небольшим отклонением. Нам требуется продиагностировать "разладку" вовремя и вызвать ремонтную службу иначе вся партия будет забракована и компания потерпит убытки.

Методы диагностики разладки можно использовать также при решении следующей задачи. В каждый час камерами магазина фиксируется количество посетителей. Необходимо исключить аномальное количество. Аномальные показатели могут возникать по разным причинам, одной из

таких является появление паутины (движение шариков при ветре), и камера может принять это за группу людей.

На сегодняшний день компьютерная безопасность занимает одну из важных ролей в жизни человека. Потому что не станем обращаться, например, в банк, где систему безопасности можно легко обойти. Пусть нам известно как система работает в нормальном режиме. В какой-то момент в систему входит "хакер". Требуется продиагностировать взлом системы, быстро отреагировать и улучшить безопасность. Поэтому алгоритмы обнаружения "разладки" могут применяться при решении такой задачи.

Немаловажную роль занимает здоровье как человека так и наших четвероногих друзей. С появлением новых лекарств вирусы стремятся выжить путем мутации и приспособления к новым условиям. Поэтому, ведя контроль заболеваний, важно определить не выходит ли за границу количество заболевших. Если да, то следует объявить карантин. "Разладкой" в этой задаче считается выход за верхнюю границу количества заболевших на 10000 человек. О задаче такого типа речь пойдет в первой главе.

Вторая глава посвящена методам диагностики "разладки" для авторегерессионной модели AR(p). В ней рассматривается алгоритм, основанный на отношении правдоподобия и правило CuSum и производится их сравнение.

Задача о "разладке" допускает различные формы постановки:

1. Общая постановка задачи различения двух гипотез:

На некотором измеримом пространстве $(\Omega, \mathcal{F})$ заданы две вероятностные меры P^0, P^1 и последовательность случайных величин $\xi_1, \xi_2, \ldots$, совместное распределение которых есть P^θ, где параметр θ принимает значение 0 и 1, и этот параметр неизвестен. Наша задача состоит в том, чтобы с минимальными "потерями" определить истинное значение неизвестного параметра по наблюдениям $\xi_1, \xi_2, \ldots$.

Далее, отталкиваясь от того, какие сделаны предположения о структуре

параметра θ, рассматриваются две постановки: байесовская и условно экстремальная.

I. Байесовская постановка задачи.

Пусть на измеримом пространстве $(\Omega, \mathcal{F})$ заданы случайные величины θ, $\xi_1, \xi_2, \ldots$ и вероятностная мера P^π, где π какое-то число из отрезка [0,1]. Предполагается, что случайная величина θ принимает значения 0 и 1 с вероятностями π и $1-\pi$ соответственно. Учитывая условие, что случайная величина θ равна j, $j = 0,1$, величины независимы и одинаково распределены (причем функция распределения не зависит от π).

Положим $M^\xi = \{\tau\}$ - совокупность марковских моментов $\tau = \tau(\omega)$, принимающих значения во множестве $N = \{0,1,2,\ldots\}$. Относительно задачи различия двух гипотез выбор момента τ означает выбор правила, по которому определяется момент, когда нужно прекратить наблюдения над величинами $\xi_1, \xi_2, \ldots$. Функция заключительного решения $d = d(\omega)$, которая измерима относительно $\mathcal{F}_\tau^\xi = \sigma\{\omega : \xi_1, \ldots, \xi_\tau\}$, показывает какую из гипотез нужно принять. Если $d(\omega) = 1$, то принимаем гипотезу $H_1 : \theta = 1$, если $d(\omega) = 0$, то гипотеза $H_0 : \theta = 0$.

II. Условно экстремальная постановка.

В этой постановке о неизвестном параметре θ никаких предположений не делается.

Положим, что на измеримом пространстве $(\Omega, \mathcal{F})$ заданы вероятностные меры P_0, P_1. Пусть также задана последовательность независимых случайных величин $\xi_1, \xi_2, \ldots$ (независимы по каждой из мер P_0, P_1).

Обозначим

$$\mathcal{F}_0^\xi = \sigma\{\varnothing, \Omega\},\ \mathcal{F}_n^\xi = \sigma\{\omega : \xi_1, \ldots, \xi_n\},\ n \geq 1,$$

$M^{\xi}=\{\tau\}$-класс моментов остановки $\tau=\tau(\omega)$ (относительно $F^{\xi}=\{\mathcal{F}_n^{\xi}\}$, $n\geq 1$), $D_{\tau}^{\xi}=\{d\}$- совокупность $\mathcal{F}_{\tau}^{\xi}$- измеримых функций $d=d(\omega)$, принимающих значение 0 и 1. Пусть еще $\Delta^{\xi}(\alpha,\beta)$-класс решающих правил $\delta=(\tau,d)$, таких, что $E_0\tau<\infty$ и $E_1\tau<\infty$. Вероятности ошибок

$$\alpha(\delta)=P_1\{d(\omega)=0\}\leq\alpha,\ \beta(\delta)=P_0\{d(\omega)=0\}\leq\beta,$$

где $\alpha,\beta>0$ и $\alpha+\beta<1$.

Относительно задачи различения двух гипотез постановка выглядит следующим образом.

Пусть заданы $\alpha,\beta>0$, такие что $\alpha+\beta<1$. Необходимо в классе $\Delta^{\xi}(\alpha,\beta)$ найти правило $\tilde{\delta}=(\tilde{\tau},\tilde{\delta})$ такое, чтобы одновременно выполнялось

$$E_0\tilde{\tau}<\infty \text{ и } E_1\tilde{\tau}<\infty$$

для всех $\delta=(\tau,d)\in\Delta^{\xi}(\alpha,\beta)$.

Работа содержит 70 страницы, 16 источников, 1 приложение.

1 ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ ДЛЯ ПАРАМЕТРОВ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Задача построения доверительных интервалов широко востребована при решении многих практических задач. Одной из таких задач является отыскание верхней границы количества людей находящихся в определенной зоне (например, в зоне накопления аэропорта, в вагоне электропоезда, в некоторой зоне магазина).

Доверительный интервал является показателем точности измерений. Это также показатель того, насколько стабильна полученная величина, то есть насколько близкая величина (к первоначальной величине) получится при повторении измерений (эксперимента). Главное преимущество интервальной оценки – возможность характеризовать уверенность в вычисленных параметрах распределения.

1.1 ПОНЯТИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА

В этой главе рассматривается доверительное оценивание скалярного параметра. Для этого находят две такие статистики $T_1 = T_1(X)$ и $T_2 = T_2(X)$, что $T_1 < T_2$, для которых при заданном $\gamma \in (0,1)$ выполнено условие:

$$P_\theta(T_1(X) < \theta < T_2(X)) \geq \gamma, \ \forall \theta \in \Theta. \qquad (1.1)$$

Полученный интервал $(T_1(X), T_2(X))$ называется $\gamma-$ *доверительным интервалом* (для параметра θ). Число γ – это *доверительный уровень* или *доверительная вероятность*.

Условие (1.1) означает, что проводится большое число не зависящих друг от друга экспериментов, в каждом из которых по n наблюдений над случайной величиной ξ. Далее оценивается параметр θ, используя следующее статистическое правило: если результаты эксперимента оцениваются выборкой X, то неизвестное значение параметра θ лежит

внутри интервала $(T_1(X), T_2(X))$. На практике часто пользуются значениями доверительного уровня γ из небольшого набора заранее выбранных, достаточно близких к 1, значений. Например, $\gamma = 0{,}9; 0{,}95; 0{,}99$ и т. д.

Иногда рассматриваются односторонние доверительные интервалы:

- Верхний вида $\theta < T_2(X)$;
- Нижний вида $T_1(X) < \theta$.

Они определяются условиями, которые аналогичны (1.1), где опускается соответствующая вторая граница.

1.2 ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА С ПОМОЩЬЮ ЦЕНТРАЛЬНОЙ СТАТИСТИКИ

Пусть модель $\mathcal{F}$ абсолютно непрерывна и существует случайная величина $G(X;\theta)$, зависящая от θ и такая, что: распределение $G(X;\theta)$ не зависит от θ и при каждом $x \in X$ функция $G(x;\theta)$ непрерывна и строго монотонна по θ - такая случайная величина называется *центральной статистикой* (для θ).

Пусть для модели $\mathcal{F}$ построена центральная статистика $G(X;\theta)$ и $f_G(g)$ - ее плотность распределения. Функция $f_G(g)$ от параметра θ не зависит, поэтому для любого $\gamma \in (0,1)$ можно выбрать величины $g_1 < g_2$ так, что бы

$$P_\theta\{g_1 < G(X;\theta) < g_2\} = \int_{g_1}^{g_2} f_G(g)dg = \gamma \tag{1.2}$$

Определяются два числа $T_i(x)$, $i = 1,2$, $x \in X$, где $T_1(x) < T_2(x)$ как решения относительно θ уравнений

$$G(x;\theta) = g_1, g_2 \tag{1.3}$$

(эти числа определяются однозначно, так как функция $G(x;\theta)$ непрерывна и строго монотонна). Тогда неравенства $g_1 < G(X;\theta) < g_2$ эквивалентны неравенствам $T_1(x) < G(X;\theta) < T_2(x)$ и (1.2) примет вид:

$$P_\theta\{T_1(X) < G(X;\theta) < T_2(X)\} = \gamma, \ \forall\theta.$$

Таким образом, построенный интервал $(T_1(X), T_2(X))$ в соответствии с (1.1) является γ-доверительным интервалом для θ.

На практике, чтобы устранить неопределенность в выборе чисел g_1 и g_2 в (1.2) рекомендуется выбирать их так, что бы выполнялись равенства

$$\int_{-\infty}^{g_1} f_G(g)dg = \int_{g_2}^{\infty} f_G(g)dg = \frac{1-\gamma}{2} \qquad (1.4)$$

В этом случае соответствующий доверительный интервал называется центральным.

1.3 ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНГО ИНТЕРВАЛА С ИСПОЛЬЗОВАНИЕМ РАСПРЕДЕЛЕНИЯ ТОЧЕЧНОЙ ОЦЕНКИ ПАРАМЕТРА

Из теории известно [1], что если имеется некоторая точечная оценка $T = T(X)$ для параметра θ и известна ее функция распределения $F_T(t;\theta)$, то доверительный интервал для θ можно построить, основываясь на этой функции. Предполагается, что распределение оценки T непрерывно и функция $F_T(t;\theta)$ непрерывна и монотонна по θ.

Пусть задан доверительный уровень γ. Нужно определить при каждом $\theta \in \Theta$ числа $t_i = t_i(\theta)$, $i = 1,2$, где $t_1 < t_2$ и

$$P_\theta\{t_1 < T(X) < t_2\} = F_T(t_2;\theta) - F_T(t_1;\theta) = \gamma \qquad (1.8)$$

Чтобы $t_i = t_i(\theta)$ определялись однозначно, надо выбирать эти числа так, чтобы выполнялись условия

$$F_T(t_1;\theta)=1-F_T(t_2;\theta)=\frac{1-\gamma}{2} \tag{1.9}$$

т.е. строится центральный доверительный интервал.

Пусть D_γ подмножество $\Theta\times\Theta$, $D_\gamma=\{(\theta,\theta'):t_1(\theta)<\theta'<t_2(\theta)\}$. Тогда $P_\theta\{(\theta,T(X)\in D_\gamma\}=\gamma,\forall\theta\in\Theta$. При фиксированном θ' определяется сечение $D_\gamma(\theta')$ множества D_γ

$$D_\gamma(\theta')=\{\theta:(\theta,\theta')\in D_\gamma\}$$

и рассматривается случайное множество $D_\gamma(T(X))\subset\Theta$. Событие $\theta\in D_\gamma(T(X))$ происходит тогда и только тогда, когда $T(X)\in(t_1(\theta),t_2(\theta))$, и следовательно, при каждом θ имеет вероятность γ. Таким образом, имеем случайное множество $D_\gamma(T(X))$, которое накрывает истинное значение параметра θ с вероятностью γ. Если это множество является интервалом, то построен (центральный) γ-доверительный интервал для θ. Из условия непрерывности и монотонности функции $F_T(t;\theta)$ по θ следует, что кривые $\theta'=t_i(\theta),\ i=1,2$ являются монотонными одного типа (т.е. одновременно либо возрастают, либо убывают).

Таким образом, при данных предположениях множество $D_\gamma(\theta')$ при каждом θ' представляет собой интервал; то есть определены его концы $\theta_1(\theta')<\theta_2(\theta')$, а, следовательно, и соответствующий γ-доверительный интервал $(T_1(x),T_2(x))$ для θ, где $T_i(X)=\theta_i(T(X)),\ i=1,2$.

Алгоритм построения центрального γ-доверительного интервала для θ, в случае, когда функция распределения $F_T(t;\theta)$ оценки $T=T(X)$ непрерывна и монотонна по θ:

Пусть $t=T(x)$-наблюдавшееся значение оценки. Разрешаются относительно θ уравнения

$$F_T(t;\theta)=\frac{1-\gamma}{2},\frac{1+\gamma}{2} \tag{1.10}$$

определяются два числа $\theta_1<\theta_2$. Пусть $\theta\in(\theta_1,\theta_2)$. Приведенная теория гарантирует, что при γ, близком к 1, вероятность ошибки в случае применения этого правила равна $1-\gamma$.

Аналогичные рассуждения проводятся и для случая дискретной модели. Единственное отличие состоит в том, что из-за ступенчатости функции распределения $F_T(t;\theta)$ в соотношении (1.8) можно обеспечить лишь выполнение неравенства

$$P_\theta\{t_1<T(X)<t_2\}=F_T(t_2-0;\theta)-F_T(t_1;\theta)\geq\gamma \tag{1.11}$$

поэтому вместо условий (1.9) вводятся условия

$$F_T(t_1;\theta)\leq\frac{1-\gamma}{2},1-F_T(t_2-0;\theta)\leq\frac{1-\gamma}{2}, \tag{1.12}$$

где t_1 -наибольшее, а t_2 - наименьшее значения T, удовлетворяющие этим неравенствам.

Алгоритм построения центрального γ-доверительного интервала для θ в дискретном случае в основном тот же, что и для непрерывного, только вместо уравнений (1.10) разрешаются относительно θ уравнения

$$\begin{gathered}F_T(t_1;\theta)\leq\frac{1-\gamma}{2},\\ 1-F_T(t_2-0;\theta)\leq\frac{1-\gamma}{2}\end{gathered} \tag{1.13}$$

t - наблюдавшееся значение оценки T.

1.4 АСИМПТОТИЧЕСКИЕ ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

Как известно[6], если имеется состоятельная асимптотически нормальная оценка $T_n=T_n(X),\ X=(X_1,X_2,...,X_n)$, для параметра θ, то строится асимптотический (при больших n) доверительный интервал.

В общем случае, пусть $n \to \infty$ и имеет место соотношение

$$L_\theta(\sqrt{n}(T_n - \theta)) \to N(0, \sigma^2(\theta)),\ \forall \theta \in \Theta$$

причем асимптотическая дисперсия $\sigma^2(\theta)$ непрерывна по θ. Тогда

$$L_\theta(\frac{\sqrt{n}(T_n - \theta)}{\sigma^2(T_n)}) \to N(0,1),\ \forall \theta \in \Theta$$

Отсюда следует, что при $n \to \infty$ и всех θ

$$P_\theta\left\{\frac{\sqrt{n}|T_n - \theta|}{\sigma(T_n)} < c_\gamma\right\} = 2\Phi(c_\gamma) - 1 = \gamma$$

если $c_\gamma = \Phi^{-1}((1+\gamma)/2)$. Переписав это соотношение в более удобном виде

$$P_\theta\left\{T_n - \frac{c_\gamma \sigma(Tn)}{\sqrt{n}} < \theta < T_n + \frac{c_\gamma \sigma(T_n)}{\sqrt{n}}\right\} \to \gamma$$

получается, что $\left(T_n \pm \frac{c_\gamma \sigma(T_n)}{\sqrt{n}}\right)$ - асимптотический γ-доверительный интервал для θ.

1.5 ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ ПАРАМЕТРА БЕРНУЛЛИЕВСКОЙ МОДЕЛИ

Требуется построить доверительный интервал для параметра θ модели $B(1,\theta)$. Если имеется соответствующая выборка $X = (X_1, X_2, ..., X_n)$, то оптимальной несмещенной оценкой для θ является выборочное среднее $T = \overline{X}$. Статистика T принимает значения вида k/n, $k = 0,1,...,n$ при этом

$$F_T\left(\frac{k}{n};\theta\right) = P_\theta\left\{T \le \frac{k}{n}\right\} = P_\theta\left\{\sum_{i=1}^{n} X_i \le k\right\} = \sum_{r=0}^{k} C_n^r \theta^r (1-\theta)^{n-r}$$

Здесь

$$\frac{d}{d\theta}F_T\left(\frac{k}{n};\theta\right)=-nC_{n-1}^{k}\theta^{k}(1-\theta)^{n-k-1}<0 \text{ при } k<n,$$

поэтому при $k<n$ функция $F_T\left(\frac{k}{n};\theta\right)$ монотонно убывает по θ.

Следовательно, применима описанная методика в п. 1.2.2, согласно которой при $T=k/n$ центральным γ-доверительным интервалом для θ является интервал (θ_1,θ_2), где θ_1 и θ_2 определяются в соответствии с уравнениями (1.13), которые в данном случае принимают вид

$$\begin{gathered}1-F_T(\frac{k-1}{n};\theta_1)=\sum_{r=0}^{k}C_n^r\theta_1^r(1-\theta_1)^{n-r}=\frac{1-\gamma}{2},\\ F_T(\frac{k}{n};\theta_2)=\sum_{r=0}^{k}C_n^r\theta_2^r(1-\theta_2)^{n-r}=\frac{1-\gamma}{2}\end{gathered} \tag{1.14}$$

при этом, если наблюдалось $T=1$ (т.е. $k=n$), то принимается $\theta_2=1$, а при $T=0$ (т.е. $k=0$) полагается $\theta_1=0$.

Практически границы θ_1 и θ_2 вычисляются, пользуясь их связью с квантилями бета-распределения:

$$\sum_{r=0}^{k}C_n^r\theta_2^r(1-\theta_2)^{n-r}=B\left(\theta_1;k,n-k+1\right),$$

$$\sum_{r=0}^{k}C_n^r\theta_2^r(1-\theta_2)^{n-r}=1-\sum_{r=k+1}^{n}C_n^r\theta_2^r(1-\theta_2)^{n-r}=1-B\left(\theta_2;k+1,n-k\right).$$

Уравнения (1.14) принимают вид

$$B\left(\theta_1;k,n-k+1\right)=\frac{1-\gamma}{2},\ B\left(\theta_2;k+1,n-k\right)=\frac{1-\gamma}{2} \tag{1.15}$$

Если обозначить $z(p;a,b)$ p-квантиль бета-распределения $Be(a,b)$, то есть решение уравнения $B(p;a,b)=p$, $0<p<1$, то из (1.15) получается (заменяя k на nT):

$$\theta_1=z\left(\frac{1-\gamma}{2};nT,n-nT+1\right)\ \theta_2=z\left(\frac{1-\gamma}{2};nT+1,n-nT\right) \tag{1.16}$$

Доверительные интервалы (θ_1, θ_2) для широкого диапазона значений при $\gamma = 0{,}9;\ 0{,}95;\ 0.99$.

1.6 ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ И ЕГО ИСПОЛЬЗОВАНИЕ ДЛЯ ОБНАРУЖЕНИЯ РАЗЛАДКИ

Каждый час камерами видеонаблюдения фиксируется число людей в каждой зоне магазина. Всего четыре зоны. Требуется исключить аномальное количество посетителей в каждой зоне в каждый час в течении суток.

Из постановки задачи видно, что имеем дело со схемой Бернулли: проводится n опытов, в каждом из которых может произойти определенное событие ("успех") с вероятностью p, или не произойти -"неудача"-с вероятностью $1-p$). Применительно к этой задаче успех - попадание в зону, неудача- непопадание.

Для обработки данных предоставлены наблюдения почти за 3 месяца.

Проводя первичную обработку данных, было выявлено, что две зоны имеют нулевые показатели. Также построены гистограммы, в которых отображено количество людей в каждый час по двум зонам (приложение А).

Зона 1:

Анализируя построенные гистограммы, стоит отметить, что активность посетителей мала с1:00 до 7:00. Это связано с тем, что, например, люди спят или зона в это время закрыта. С 7:00 до 10:00 начинается поток покупателей увеличиваться. С 10:00 до 21:00 население активно. В оставшееся время поток уменьшается в интересующей зоне. Вечером с 21:00 до 1:00 население "засыпает".

Зона 2:

Проводя анализ гистограмм по этой зоне, можно сказать, что количество посетителей выше, чем в зоне 1. Так же стоит отметить, что с

5:00 до 6:00 утра активности нет. Возможно, в это время магазин закрыт. В остальное время наблюдается небольшая активность, как и в зоне 1. Можно предположить, что посещаемость этой зоны велика, потому что это зона касс, либо зона выхода из магазина.

Далее, данные по часам в течении дня усреднили. Отдельно будем рассматривать воскресенье, так как в этот день посещаемость очень маленькая по сравнению с остальными днями. Данные представлены в таблицах 1.1 и 1.2.

Таблица 1.1 - Среднее количество посетителей в будни и субботу.

№	Время	Средне кол-во в двух зонах (n)	Среднее кол-во в Зоне 1(m)	Среднее кол-во в Зоне 2 (m)
1	**0:00**	55	3	52
2	**1:00**	29	0	29
3	**2:00**	1	0	1
4	**3:00**	2	0	2
5	**4:00**	1	0	1
6	**5:00**	0	0	0
7	**6:00**	1	0	1
8	**7:00**	12	1	11
9	**8:00**	9	1	8
10	**9:00**	7	1	6
11	**10:00**	6	1	5
12	**11:00**	9	2	7
13	**12:00**	22	2	20
14	**13:00**	346	49	297
15	**14:00**	522	77	445
16	**15:00**	756	95	661
17	**16:00**	917	99	818
18	**17:00**	941	104	837
19	**18:00**	809	85	724
20	**19:00**	861	81	780
21	**20:00**	823	87	736

22	**21:00**	595	72	523
23	**22:00**	149	12	137
24	**23:00**	116	10	106

Таблица 1.2- Среднее количество посетителей в воскресенье.

№	**Время**	**Средне кол-во в двух зонах (n)**	**Среднее кол-во в Зоне 1(m)**	**Среднее кол-во в Зоне 2 (m)**
1	**0:00**	1	0	1
2	**1:00**	3	0	3
3	**2:00**	1	0	1
4	**3:00**	0	0	0
5	**4:00**	1	0	1
6	**5:00**	6	1	5
7	**6:00**	7	0	7
8	**7:00**	2	0	2
9	**8:00**	4	1	3
10	**9:00**	4	1	3
11	**10:00**	4	1	3
12	**11:00**	9	1	8
13	**12:00**	2	1	1
14	**13:00**	1	0	1
15	**14:00**	45	0	45
16	**15:00**	74	0	74
17	**16:00**	101	0	101
18	**17:00**	130	0	130
19	**18:00**	152	0	152
20	**19:00**	155	1	154
21	**20:00**	163	1	162
22	**21:00**	129	1	128
23	**22:00**	80	0	80
24	**23:00**	36	0	36

По полученным данным построили гистограммы, которые наглядно иллюстрируют посещение покупателями магазина в течении всего дня.

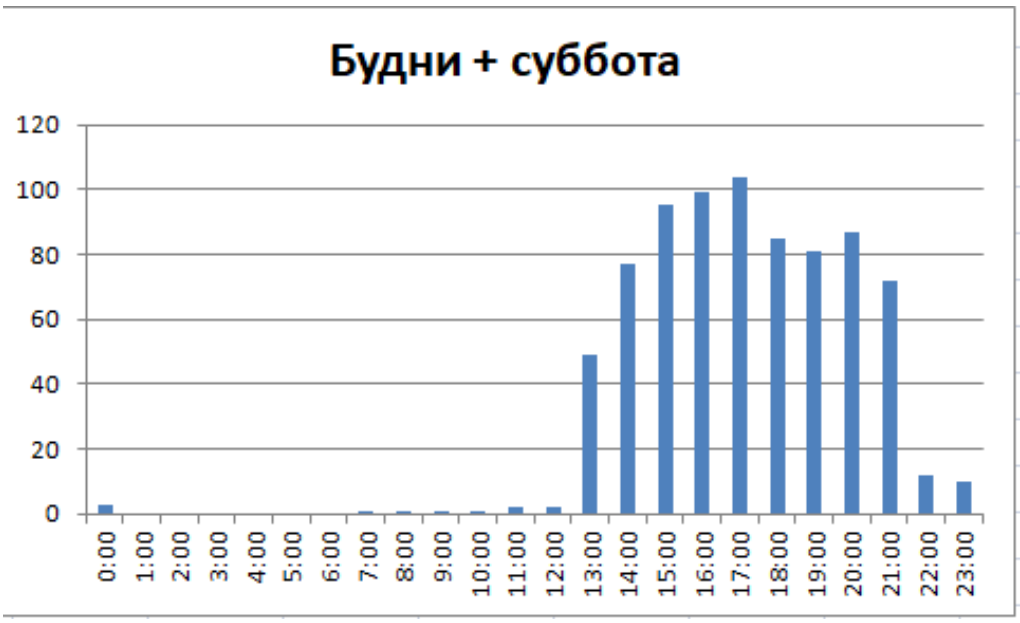


Рисунок 1.1 - Среднее количество покупателей в зоне 1 в течении дня (будни + суббота).

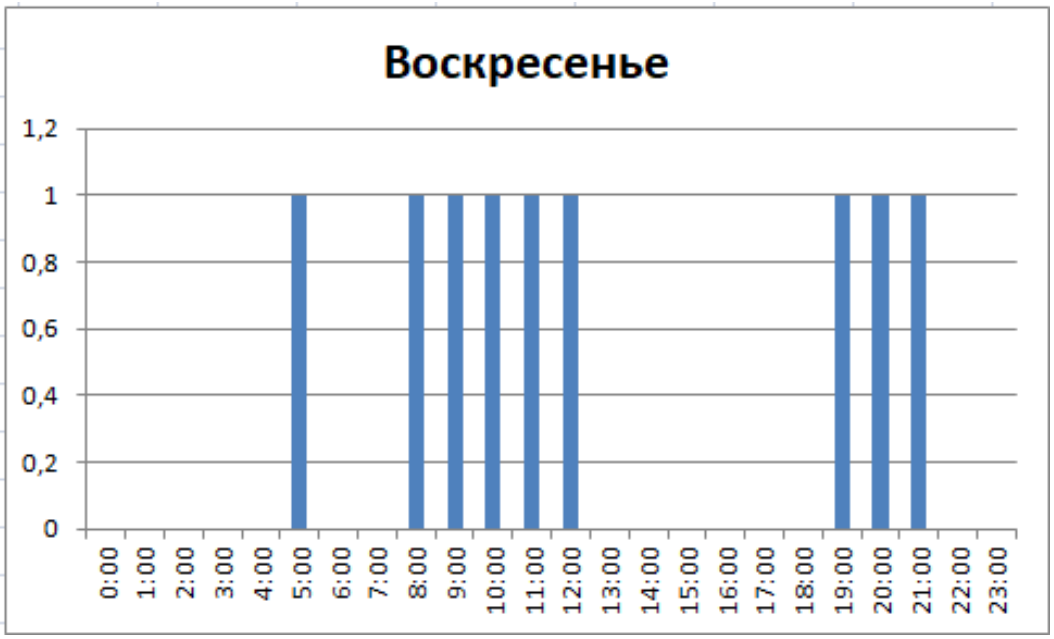


Рисунок 1.2- Среднее количество покупателей в зоне 1 в течении дня (воскресенье).



Рисунок 1.3 - Среднее количество покупателей в зоне 2 в течении дня (будни суббота).



Рисунок 1.4- Среднее количество покупателей в зоне 2 в течении дня (воскресенье).

Видно, что распределение посетителей близко к нормальному. Следовательно для построения доверительных интервалов можно применить интегральную теорему Муавра- Лапласа. В явном виде ее применить не

можем, так как вероятность p неизвестна. Доверительные интервалы будем строить для вероятности с уровнем доверия $\gamma = 0.95$.

Имеем:

$$P\left(\left|\frac{m-np}{\sqrt{np(1-p)}}\right| < \varepsilon\right) \approx \frac{1}{2\pi}\int_{-\varepsilon}^{\varepsilon} e^{\frac{x^2}{2}} dx$$

По таблицам распределения по заданному γ находим ε и все сводим к решению неравенства:

$$\left|\frac{m-np}{\sqrt{np(1-p)}}\right| < \varepsilon \qquad (*)$$

или равносильного ему неравенства $\frac{(m-np)^2}{np(1-p)} < \varepsilon^2$. Разрешаем последнее неравенство относительно p. Получаем:

$$\frac{m}{n} - \frac{\varepsilon^2 + \varepsilon\sqrt{4m + \varepsilon^2 - 4\frac{m^2}{n}} - \frac{2m\varepsilon^2}{n}}{2(n+\varepsilon^2)} < p < \frac{m}{n} + \frac{\varepsilon^2 + \varepsilon\sqrt{4m + \varepsilon^2 - 4\frac{m^2}{n}} - \frac{2m\varepsilon^2}{n}}{2(n+\varepsilon^2)}.$$

Если выражение $1-p$ заменить его максимальным значением 1/4, то полученный интервал $\left(\frac{m}{n} - \frac{\varepsilon}{2\sqrt{n}}, \frac{m}{n} + \frac{\varepsilon}{2\sqrt{n}}\right)$ шире и поэтому имеет больший коэффициент надежности.

Запрограммировав выведенные формулы, получили интервалы для вероятности.

По таблице стандартного нормального распределения $\varepsilon = 1.96$.

Таблица 1. 3- Границы для вероятности p (будни + суббота).

№	Время	Зона 1		Зона 2	
		нижняя	Верхняя	нижняя	верхняя
1	0:00	0	0,19	0,81	1,07
2	1:00	0	0,18	0,82	1,18
3	2:00	0	0,98	0,02	1,19
4	3:00	0	0,69	0,19	1,16
5	4:00	0	0,98	0	1,18
6	5:00	0	0	0	0
7	6:00	0	0,98	0	0,98
8	7:00	0	0,29	0,67	1,24
9	8:00	0	0,41	0,49	1,17
10	9:00	0	0,51	0,48	1,22
11	10:00	0	0,56	0,39	1,2
12	11:00	0	0,47	0,53	1,22
13	12:00	0	0,31	0,69	1,12
14	13:00	0,09	0,19	0,81	0,91
15	14:00	0,1	0,19	0,81	0,9
16	15:00	0,09	0,16	0,84	0,91
17	16:00	0,07	0,14	0,86	0,92
18	17:00	0,08	0,14	0,86	0,86
19	18:00	0,07	0,14	0,86	0,93
20	19:00	0,06	0,13	0,87	0,94
21	20:00	0,07	0,14	0,86	0,93
22	21:00	0,08	0,16	0,95	0,96
23	22:00	0	0,16	0,84	1
24	23:00	0	0,17	0,82	1

Таблица 1.3- Границы для вероятности p (воскресенье).

№	Время	Зона 1		Зона 2	
		нижняя	Верхняя	нижняя	верхняя
1	**0:00**	0	0,82	0,18	1,18
2	**1:00**	0	0,53	0,47	1,15
3	**2:00**	0	0,69	0	1,19
4	**3:00**	0	0	0	0
5	**4:00**	0	0,69	0	1,19
6	**5:00**	0	0,57	0,43	1,23
7	**6:00**	0	0,37	0,64	1,06
8	**7:00**	0	0,69	0,3	1,07
9	**8:00**	0	0,62	0,39	1,37
10	**9:00**	0	0,49	0,26	1,24
11	**10:00**	0	0,49	0,26	1,24
12	**11:00**	0	0,41	0,49	1,11
13	**12:00**	0	0,8	0	1,19
14	**13:00**	0	0,69	0	1,19
15	**14:00**	0	0,14	0,83	1,12
16	**15:00**	0	0,11	0,87	1,1
17	**16:00**	0	0,11	0,89	1,09
18	**17:00**	0	0,09	0,91	1,08
19	**18:00**	0	0,09	0,91	1,07
20	**19:00**	0	0,08	0,91	1,07
21	**20:00**	0	0,083	0,92	1,07
22	**21:00**	0	0,09	0,91	1,08
23	**22:00**	0	0,12	0,88	1,1
24	**23:00**	0	0,16	0,81	1,14

Теоретически получили интервалы для частоты попадания людей в зону. Чтобы построить интервалы для количества посетителей нужно полученные границы вероятности умножить на среднее количество людей в двух зонах. Результаты отображены в таблице 1.4 и таблице 1.5.

Таблица 1.4- Границы для количества людей (будни + суббота)

№	Время	Зона 1		Зона 2	
		нижняя	верхняя	нижняя	верхняя
1	**0:00**	0	11	44	60
2	**1:00**	0	6	24	35
3	**2:00**	0	2	0	2
4	**3:00**	0	2	0	2
5	**4:00**	0	1	0	2
6	**5:00**	0	0	0	0
7	**6:00**	0	1	0	1
8	**7:00**	0	4	7	15
9	**8:00**	0	4	4	10
10	**9:00**	0	4	3	10
11	**10:00**	0	4	2	7
12	**11:00**	0	4	4	11
13	**12:00**	0	7	15	25
14	**13:00**	30	68	279	316
15	**14:00**	53	99	422	467
16	**15:00**	68	122	633	687
17	**16:00**	68	128	787	847
18	**17:00**	73	134	807	808
19	**18:00**	57	113	695	752
20	**19:00**	51	109	750	809
21	**20:00**	58	116	707	763
22	**21:00**	48	96	567	568
23	**22:00**	0	24	125	149
24	**23:00**	0	20	94	116

Таблица 1.5- Границы для количества людей (воскресенье)

№	Время	Зона 1		Зона 2	
		Нижняя	Верхняя	нижняя	верхняя
1	**0:00**	0	1	0	1
2	**1:00**	0	2	1	4
3	**2:00**	0	0	0	0
4	**3:00**	0	0	0	0
5	**4:00**	0	0	0	0
6	**5:00**	0	4	3	8
7	**6:00**	0	3	4	8
8	**7:00**	0	2	0	3
9	**8:00**	0	3	1	6
10	**9:00**	0	2	1	5
11	**10:00**	0	2	1	5
12	**11:00**	0	4	5	11
13	**12:00**	0	1	0	2
14	**13:00**	0	1	0	1
15	**14:00**	0	7	38	51
16	**15:00**	0	9	65	82
17	**16:00**	0	11	91	110
18	**17:00**	0	12	118	141
19	**18:00**	0	13	139	163
20	**19:00**	0	13	141	166
21	**20:00**	0	14	149	175
22	**21:00**	0	12	117	139
23	**22:00**	0	10	70	88
24	**23:00**	0	6	29	41

Таким образом, получены верхняя и нижняя границы 95% доверительного интервала для количества людей в фиксированной зоне наблюдения. За момент "разладки" в этом случае естественно взять момент выхода за границу интервала. Учитывая уровень доверия к интервалу, легко увидеть, что одноразовый выход за границу, при работающем оборудовании, имеет вероятность $p = 0,05$. Вероятность повторного выхода (дважды подряд) не более 0,0025, а вероятность выхода трижды - не более 0,000125. Следовательно, при двукратном выходе за границу вероятность ложной тревоги не более 0,0025, а при выходе трижды - не более, чем 0,000125.

2 ОБНАРУЖЕНИЕ "РАЗЛАДКИ" В АВТОРЕГРЕССИОННЫХ МОДЕЛЯХ

Авторегрессионная модель (AR) - модель временных рядов, в которой каждое текущее значение временного ряда линейно зависит от предыдущего значения.

Модель AR(1) имеет вид:

$$X_t = \delta + \phi X_{t-1} + \omega_t,\ t = 1,2,3... \qquad (2.1)$$

Где X_t - процесс и δ, ϕ - константы, $|\phi| < 1$, ω_t - белый шум со средним 0 и дисперсией σ^2 .

Обычно параметр δ называют параметром сдвига, так как он определяет среднее значение процесса. А параметр ϕ представляет собой коэффициент автокорреляции между текущим и последующим значение процесса.

Временной ряд X_t является *стационарным*, если :

1) математическое ожидание постоянно и не зависит от времени t;
2) функция автоковариации зависит от s и t только от их разности $|s-t|$.

Пусть мы наблюдаем последовательность наблюдений $\{x_t; t = 1,2,...,n\}$, которые описываются стационарной моделью AR(1)

$$x_t = \delta + \phi X_{t-1} + \omega_t,\ t = 1,2,3.... .$$

Цель - построить критерии, которые могут проверить основную гипотезу о том, что процесс AR(1) не имеет "разладки" против альтернативной гипотезы о том, что последовательность имеет точку изменения. "Разладку" будем рассматривать как изменение коэффициента корреляции ϕ либо изменение сдвига δ. Поэтому будем проверять как изменяются два параметра: ϕ и δ. Если проверяем коэффициент δ, то

считаем, что ϕ фиксировано. Если проверяем коэффициент ϕ, δ считаем фиксированным.

2.1 ПРОВЕРКА ГИПОТЕЗЫ ОБ ОБНАРУЖЕНИИ "РАЗЛАДКИ" В ПЕРВОМ КОЭФФИЦИЕНТЕ МОДЕЛИ AR(1)

2.1.1 ПОСТРОЕНИЕ КРИТЕРИЯ ДЛЯ ОБНАРУЖЕНИЯ "РАЗЛАДКИ"

Построим критерий, для проверки основной гипотезы об отсутствии "разладки" против альтернативы о ее наличии :

$$\begin{aligned} &H_0: \qquad X_t = \delta + \phi X_{t-1} + \omega_t, \ t = 1,2,3,\ldots. \\ &H_1: \exists k \quad X_t = \begin{cases} \delta_0 + \phi X_{t-1} + \omega_t, & \mathrm{t} \le \mathrm{k}. \\ \delta_1 + \phi X_{t-1} + \omega_t & k < t \le n. \end{cases} \end{aligned} \tag{2.2}$$

где $\delta_0 \neq \delta_1$ и k - момент времени, когда происходит изменение.

Если $X_{t-1} = x_{t-1}$, то в (2.2) распределение X_t нормально:

$$f(X_t \mid x_{t-1}) \square N(\delta_0 + \phi x_{t-1}, \sigma^2). \tag{2.3}$$

Поэтому, в предположении, что гипотеза H_0 верна функция плотности для $\{X_2, X_3, ..., X_n\}$ имеет вид:

$$\begin{aligned} L_0 &= f(x_2, x_3, ..., x_n \mid x_1) = \\ &= \frac{f(x_1, x_2)}{f(x_1)} \frac{f(x_1, x_2, x_3)}{f(x_1, x_2)} \cdots \frac{f(x_1, x_2, ..., x_n)}{f(x_1, x_2, ..., x_{n-1})} = \\ &= f(x_2 \mid x_1) f(x_3 \mid x_2, x_1) \cdots f(x_n \mid x_{n-1}, x_{n-2}, ..., x_1) = \\ &= f(x_2 \mid x_1) f(x_3 \mid x_2) \cdots f(x_n \mid x_{n-1}) = \\ &= (2\pi\sigma^2)^{-\frac{n-1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \delta_0 - \phi x_{i-1})^2 \right\} \end{aligned} \tag{2.4}$$

В предположении, что верна альтернативная гипотеза H_1 функция плотности имеет вид:

$$L_1\left(\hat{\sigma}_k^2,\hat{\phi},\hat{\delta}_0,\hat{\delta}_1\right)=$$
$$=\left(2\pi\sigma^2\right)^{-\frac{n-1}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=2}^{k}\left(x_i-\delta_0-\phi x_{i-1}\right)^2+\sum_{i=k+1}^{n}\left(x_i-\delta_1-\phi x_{i-1}\right)^2\right]\right\} \quad (2.5)$$

Прологарифмировав L_0 и взяв производную функции логарифмического правдоподобия по δ и ϕ, получаем оценки максимального правдоподобия :

$$\hat{\phi}_n=\frac{\sum_{i=2}^{n}x_i\sum_{i=2}^{n}x_{i-1}-(n-1)\sum_{i=2}^{n}x_ix_{i-1}}{\left(\sum_{i=2}^{n}x_{i-1}\right)^2-(n-1)\sum_{i=2}^{n}x_{i-1}^2} \quad (2.6)$$

$$\hat{\delta}=\frac{\sum_{i=2}^{n}(x_i-\hat{\phi}x_{i-1})}{n-1} \quad (2.7)$$

Прологарифмировав L_1 и взяв производную, аналогичным образом получаем оценки максимального правдоподобия при альтернативной гипотезе:

$$\hat{\phi}_k=\frac{\sum_{i=2}^{k}x_ix_{i-1}+\sum_{i=k+1}^{n}x_ix_{i-1}-\frac{\sum_{i=2}^{k}x_i\sum_{i=2}^{k}x_{i-1}}{k-1}+\frac{\sum_{i=k+1}^{n}x_{i-1}\sum_{i=k+1}^{n}x_i}{n-k}}{\sum_{i=2}^{k}x_{i-1}^2+\sum_{i=k+1}^{n}x_{i-1}^2-\frac{\left(\sum_{i=2}^{k}x_{i-1}\right)^2}{k-1}-\frac{\left(\sum_{i=k+1}^{n}x_{i-1}\right)^2}{n-k}} \quad (2.8)$$

$$\hat{\delta}_0=\frac{\sum_{i=2}^{k}\left(x_i-\hat{\phi}_kx_{i-1}\right)}{k-1} \quad (2.9)$$

$$\hat{\delta}_1=\frac{\sum_{i=k+1}^{n}\left(x_i-\hat{\phi}_kx_{i-1}\right)}{n-k} \quad (2.10)$$

Оценка максимального правдоподобия для σ_n^2 и σ_k^2:

$$\hat{\sigma}_n^2=\frac{1}{n}\sum_{i=2}^{n}\left(x_i-\hat{\delta}-\hat{\phi}x_{i-1}\right)^2 \quad (2.11)$$

$$\hat{\sigma}_k^2=\frac{1}{n}\left(\sum_{i=2}^{k}\left(x_i-\hat{\delta}_0-\hat{\phi}x_{i-1}\right)^2+\sum_{i=k+1}^{n}\left(x_i-\hat{\delta}_1-\hat{\phi}x_{i-1}\right)^2\right) \quad (2.12)$$

Выпишем отношение правдоподобия:

$$\lambda_k = \frac{L_0\left(\hat{\sigma}_n^2, \hat{\phi}, \hat{\delta}\right)}{L_1\left(\hat{\sigma}_k^2, \hat{\phi}, \hat{\delta}_0, \hat{\delta}_1\right)} = \left(\frac{\hat{\sigma}_k^2}{\hat{\sigma}_n^2}\right)^{\frac{n-1}{2}} \tag{2.13}$$

Ее логарифм:

$$\Lambda_k = -2\log\lambda_k$$

Статистика критерия выражается формулой:

$$\Lambda_n = \max_{2\le k\le n-1}\left|-2\log\lambda_k\right| \tag{2.14}$$

Если нулевая гипотеза отвергается, значит происходит изменение параметров процесса в момент времени k ("разладка"). Тогда σ_n будет больше, чем σ_k. Значит λ_k мало , следовательно Λ_k большое. В результате, гипотеза H_0 отклоняется, если $|\Lambda_k|$ слишком велик. Момент времени k, когда Λ_k принимает наибольшее значение - момент "разладки" .

2.1.2 АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ТЕСТОВОЙ СТАТИСТИКИ

Дэвис, Хуан и Яо представили теорему в своей работе в 1995 году [15] , которая установила предельное распределение для Λ_n (тестовой статистики) для обнаружения "разладки" в модели AR (p).

Теорема 2.1 : Пусть $\{X_t\}$- процесс, который определяется равенством

$$\begin{aligned} X_t &= \phi_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t, \quad -\infty < t \le \tau \\ X_t &= \alpha_0 + \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \varepsilon_t, \quad t \ge \tau + 1 \end{aligned} \tag{2.15}$$

такой, что если верна гипотеза H_0 (нет "разладки"), ошибка $\{\varepsilon_t\}$ удовлетворяет условиям: $E(\varepsilon_t) = 0$ с $\sup_t E|\varepsilon_t|^{4+\delta} < \infty$ для некоторого $0 < \delta \le 1$ и

ковариационная функция этого процесса $\rho(n) << n^{-(1+\varepsilon)(1+4/\delta)}$ для некоторого $\varepsilon > 0$. Тогда при гипотезе H_0

$$P\left[\left(\sigma^{-2}\Lambda_n - b_n(p+1)\right)/a_n(p+1) \le x\right] \to \exp\left(-2e^{-x/2}\right) \qquad (2.16)$$

где $\Lambda_n = \max_{p<k\le n} \Lambda_n(k)$ является статистикой отношения правдоподобия,

$b_n(d) = \left(2\ln\ln n + (d/2)\ln\ln\ln n - \ln\Gamma(d/2)\right)^2 / (2\ln\ln n)$ и

$a_n(d) = \sqrt{b_n(d)/(2\ln\ln n)}$ нормировочные константы и $\Gamma(\cdot)$- гамма-функция.

AR(1) является частным случаем для AR(p)- модели, поэтому приведенная выше теорема верна для случая, исследуемого в работе. В исследуемой модели используется $b_n(2)$ и $a_n(2)$.

Теорема 2.2: Если верна гипотеза H_0 ("разладка" отсутствует), то предельно распределение статистики критерия

$$P\left[\left(\sigma^{-2}\Lambda_n - b_n(p+1)\right)/a_n(p+1) \le x\right] \to \exp\left(-2e^{-x/2}\right)$$

где $\Lambda_n = \max_{2<k\le n-1} |\Lambda_k|$ является статистикой отношения правдоподобия,

$$b_n(2) = \left(2\ln\ln n + \ln\ln\ln n\right)^2 / (2\ln\ln n)$$

$$a_n(2) = \left(2\ln\ln n + \ln\ln\ln n\right) / (2\ln\ln n)$$

-нормировочные константы и $\Gamma(\cdot)$- гамма-функция, $\Gamma(1/2) = \sqrt{\pi}$.

2.2 ПРОВЕРКА ГИПОТЕЗЫ ОБ ОБНАРУЖЕНИИ "РАЗЛАДКИ" ВО ВТОРОМ КОЭФФИЦИЕНТЕ МОДЕЛИ AR(1)

2.2.1 ПОСТРОЕНИЕ КРИТЕРИЯ ДЛЯ ОБНАРУЖЕНИЯ "РАЗЛАДКИ"

Построим критерий для проверки основной гипотезы об отсутствии "разладки" против альтернативы о ее наличии:

$$\begin{aligned} &H_0: \quad X_t = \phi X_{t-1} + \omega_t, \ \ t = 1,2,3,\ldots . \\ &H_1: \exists k \quad X_t = \begin{cases} \phi_0 X_{t-1} + \omega_t, & \mathrm{t} \le \mathrm{k}. \\ \phi_1 X_{t-1} + \omega_t & k < t \le n. \end{cases} \end{aligned} \tag{2.17}$$

Если $X_{t-1} = x_{t-1}$, то в (2.17) распределение X_t нормально:

$$f(X_t \mid x_{t-1}) \square N(\phi_0 x_{t-1}, \sigma^2) \tag{2.18}$$

Поэтому, в предположении, что гипотеза H_0 верна функция плотности для $\{X_2, X_3, ..., X_n\}$ имеет вид:

$$\begin{aligned} L_0 &= f(x_2, x_3, ..., x_n \mid x_1) = \\ &= \frac{f(x_1, x_2)}{f(x_1)} \frac{f(x_1, x_2, x_3)}{f(x_1, x_2)} \cdots \frac{f(x_1, x_2, ..., x_n)}{f(x_1, x_2, ..., x_{n-1})} = \\ &= f(x_2 \mid x_1) f(x_3 \mid x_2, x_1) \cdots f(x_n \mid x_{n-1}, x_{n-2}, ..., x_1) = \\ &= f(x_2 \mid x_1) f(x_3 \mid x_2) \cdots f(x_n \mid x_{n-1}) = \\ &= (2\pi\sigma^2)^{-\frac{n-1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \phi_0 x_{i-1})^2 \right\} \end{aligned} \tag{2.19}$$

При $\frac{\partial \log L_0}{\partial \phi_0} = 0$, получаем оценку максимального правдоподобия для ϕ_0:

$$\phi_0 = \frac{\sum_{i=2}^{n} x_i x_{i-1}}{\sum_{i=2}^{n} x_{i-1}^2}. \tag{2.20}$$

При $\frac{\partial \log L_0}{\partial \phi_0} = 0$, подставляем ϕ_0, получаем оценку максимального правдоподобия для σ_n^2

$$\sigma_n^2 = \frac{1}{n-1}\sum_{i=2}^{n}(x_i - \phi_0 x_{i-1})^2 \tag{2.21}$$

Если верна альтернативная гипотеза H_1, то функция плотности для $\{X_2, X_3, ..., X_n\}$ и оценки максимального правдоподобия для σ_k^2, ϕ_{0k} и ϕ_{1k} выражаются следующими формулами:

$$L_1 = (2\pi\sigma^2)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=2}^{k}(x_i - \phi_{0k}x_{i-1})^2 + \sum_{i=k+1}^{n}(x_i - \phi_{1k}x_{i-1})^2\right]\right\}; \tag{2.22}$$

$$\sigma_k^2 = \frac{1}{n-1}\left[\sum_{i=2}^{k}(x_i - \phi_{0k}x_{i-1})^2 + \sum_{i=k+1}^{n}(x_i - \phi_{1k}x_{i-1})^2\right]; \tag{2.23}$$

$$\phi_{0k} = \frac{\sum_{i=2}^{k} x_i x_{i-1}}{\sum_{i=2}^{k} x_{i-1}^2}; \qquad \phi_{1k} = \frac{\sum_{i=k+1}^{n} x_i x_{i-1}}{\sum_{i=k+1}^{n} x_{i-1}^2}. \tag{2.24}$$

Выпишем отношение правдоподобия:

$$\begin{aligned}
\lambda_k &= \frac{L_0(\sigma_n^2, \phi_0)}{L_1(\sigma_n^2, \phi_{0k}, \phi_{1k})} = \\
&= \frac{(2\pi\sigma_n^2)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2\sigma_n^2}\sum_{i=1}^{n}(x_i - \phi_0 x_{i-1})^2\right\}}{(2\pi\sigma_k^2)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2\sigma_k^2}\left[\sum_{i=2}^{k}(x_i - \phi_{0k}x_{i-1})^2 + \sum_{i=k+1}^{n}(x_i - \phi_{1k}x_{i-1})^2\right]\right\}} = \\
&= \frac{(2\pi\sigma_n^2)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2\sigma_n^2}\sigma_n^2(n-1)\right\}}{(2\pi\sigma_k^2)^{-\frac{n-1}{2}} \exp\left\{-\frac{1}{2\sigma_k^2}\sigma_k^2(n-1)\right\}} = \left(\frac{\sigma_k^2}{\sigma_n^2}\right)^{\frac{n-1}{2}}
\end{aligned} \tag{2.25}$$

Его логарифм выражается формулой:

$$\Lambda_k = -2\log\lambda_k = (n-1)(\log\sigma_n^2 - \log\sigma_k^2) \tag{2.26}$$

Если нулевая гипотеза отвергается, значит происходит изменение параметров в момент времени k ("разладка"). Тогда σ_n будет больше, чем σ_k. Значит λ_k

мало, а, следовательно, Λ_k большое. В итоге нулевая гипотеза H_0, если $|\Lambda_k|$ слишком велик. Момент времени k, когда Λ_k большое - момент "разладки" .

Таким образом, статистика критерия определяется формулой:

$$\Lambda_n = \max_{2\le k\le n-1}|\Lambda_k| = \max_{2\le k\le n-1}\left|(n-1)(\log\sigma_n^2 - \log\sigma_k^2)\right| \tag{2.27}$$

и оценка k:

$$k = \arg\left\{k; \Lambda_n = \max_{2\le k\le n-1}|\Lambda_k|\right\} \tag{2.28}$$

Примечание:

Применяемое здесь абсолютное значение не вносит никаких изменений с технической точки зрения $\lambda_k < 1$, а $\Lambda_k > 0$.

2.1.2. АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ТЕСТОВОЙ СТАТИСТИКИ

Теорема 2.1 : Пусть $\{X_t\}$ - процесс, который определяется

$$X_t = \phi_0 + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \varepsilon_t, \qquad -\infty < t \le \tau$$
$$X_t = \alpha_0 + \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \varepsilon_t, \qquad t \ge \tau + 1$$

такой, что если верна гипотеза H_0, ошибка $\{\varepsilon_t\}$ удовлетворяет условиям: $E(\varepsilon_t) = 0$ с $\sup_t E|\varepsilon_t|^{4+\delta} < \infty$ для некоторого $0 < \delta \le 1$ и ковариационная функция этого процесса $\rho(n) << n^{-(1+\varepsilon)(1+4/\delta)}$ для некоторого $\varepsilon > 0$. Тогда если гипотеза H_0 верна, то

$$P\left[\left(\sigma^{-2}\Lambda_n - b_n(p+1)\right)/a_n(p+1) \le x\right] \to \exp\left(-2e^{-x/2}\right)$$

где $\Lambda_n = \max_{p<k\le n} \Lambda_n(k)$ является статистикой отношения правдоподобия,

$b_n(d) = \left(2\ln\ln n + (d/2)\ln\ln\ln n - \ln\Gamma(d/2)\right)^2 / (2\ln\ln n)$ и

$a_n(d) = \sqrt{b_n(d)/(2\ln\ln n)}$ нормировочные константы и $\Gamma(\cdot)$- гамма-функция.

AR(1) является частным случаем для AR(p)- модели, поэтому приведенная выше теорема подходит для случая, исследуемого в работе. В работе рассматривается модель AR(1) без сдвига, поэтому используется $b_n(1)$ и $a_n(1)$.

Теорема 2.2: Если верна гипотеза H_0 (процесс без изменений), то предельное распределение статистики критерия

$$P\left[\left(\sigma^{-2}\Lambda_n - b_n(p+1)\right)/a_n(p+1) \le x\right] \to \exp\left(-2e^{-x/2}\right)$$

где $\Lambda_n = \max_{2<k\le n-1} |\Lambda_k|$ является статистикой отношения правдоподобия,

$$b_n(1) = \left(2\ln\ln n + (1/2)\ln\ln\ln n - \ln\Gamma(1/2)\right)^2 / (2\ln\ln n)$$

$$a_n(1) = \sqrt{b_n(1)/(2\ln\ln n)}$$

являются нормировочными константами и $\Gamma(\cdot)$- гамма-функция, $\Gamma(1/2) = \sqrt{\pi}$.

Последняя теорема дает асимптотическое распределение тестовой статистики, поэтому можно определить критическое значение тестовой статистики отношения правдоподобия.

2.3 ПОСЛЕДОВАТЕЛЬНЫЙ НЕПАРАМЕТРИЧЕСКИЙ МЕТОД ОБНАРУЖЕНИЯ "РАЗЛАДОК" СЛУЧАЙНЫХ ПРОЦЕССОВ РЕКУРЕНТНОГО ТИПА

Задача обнаружения "разладок" возникает во многих практических ситуациях при оценивании момента появления полезного сигнала, при диагностике технических систем, когда требуется скорейшим образом обнаружить нарушение нормального течения процесса и т. п. Теория обнаружения "разладок" продолжает интенсивно разрабатываться. Если на первых этапах развития этой теории преимущественно изучались задачи, относящиеся к схеме независимых наблюдений с полным статистическим описанием, то в последние годы все больший интерес вызывает непараметрический подход к задаче о "разладке", при котором распределение процессов до и после "разладки" точно известны, а предполагается только, что они принадлежат некоторым классам распределений. Наблюдается также тенденция к использованию для описания явления "разладки" зависимых случайных процессов.

2.3.1 ПОСТАНОВКА ЗАДАЧИ

Пусть наблюдается многомерный случайный процесс $x(t) = \{x_1(t), ..., x_n(t)\}^T$, который описывается системой стохастических разностных уравнений

$$x(t+1) = A_1(t, x) + B_1(t, x)\xi(t+1), \quad t = 0, 1, ..., \theta - 1, \tag{2.29}$$

до момента "разладки" θ и системой уравнений

$$x(t+1) = A_2(t, x) + B_2(t, x)\xi(t+1), \quad t = \theta, \theta+1, ..., \tag{2.30}$$

начиная с момента θ. Здесь $\{\xi(t)\}$ - не зависящая от $x(0)$ последовательность s-мерных одинаково распределенных случайных векторов с $E\xi(t) = 0$ и

$E\xi(t)\xi^T(t) = E$; E -единичная матрица. Распределение $\xi(t)$ неизвестно. Элементы матричных функций $A_i(t,x)$ и $B(t,x)$ (размера $n\times 1$ и $n\times s$) могут зависеть от реализации процесса x до момента t включительно. Момент "разладки" θ может быть детерминированным или случайным. В последнем случае предполагается , что случайная величина θ имеет неизвестное распределение и не зависит от процесса $\{\xi(t)\}$ и $x(0)$. Задача состоит в том, чтобы построить последовательную процедуру обнаружения "разладки" θ.

2.3.2 ПОСТРОЕНИЕ РЕШАЮЩЕЙ ПРОЦЕДУРЫ

Начнем рассмотрение со случая, когда $B_i(t,x) = B(t,x)$ в момент "разладки" θ не случаен. Учитывая, что распределение процесса неизвестно, для оценивания θ естественно использовать метод наименьших квадратов. Предположим, что имеется достаточно длинная реализация процесса (длины N), включающая момент "разладки". Тогда оценка θ определяется из условия минимума по m функционала

$$I(m) = \sum_{t=0}^{m-1}\left[x(t+1) - A_1(t,x)\right]^T \left(BB^T\right)^+(t,x)\left[x(t+1) - A_1(t,x)\right] + \\ + \sum_{t=m}^{N}\left[x(t+1) - A_2(t,x)\right]^T \left(BB^T\right)^+(t,x)\left[x(t+1) - A_2(t,x)\right]. \tag{2.31}$$

Этот функционал может быть преобразован в вид

$$I(m) = 2\sum_{t=0}^{m-1}\left(A_2(t,x) - A_1(t,x)\right)^T \left(BB^T\right)^+(t,x)\left[x(t+1) - \frac{A_1(t,x) + A_2(t,x)}{2}\right] + C_N,$$

где

$$C_N = \sum_{t=0}^{N}\left[x(t+1) - A_2(t,x)\right]^T (BB^T)(t,x)\left[x(t+1) - A_2(t,x)\right]$$

С помощью уравнений (2.29) и (2.30) получаем

$$I(m) = -\sum_{t=0}^{m\wedge\theta-1} r(t,x) + \sum_{t=m\wedge\theta-1}^{m-1} r(t,x) + C_N + 2\sum_{t=0}^{m-1}\eta(t+1), \tag{2.32}$$

где

$$r(t,x)=\left[A_2(t,x)-A_1(t,x)\right]^T\left(BB^T\right)^+(t,x)\left[A_2(t,x)-A_1(t,x)\right];$$

$$\eta(t+1)=\left[A_2(t,x)-A_1(t,x)\right]^T\left(BB^T\right)^+(t,x)B(t,x)\xi(t+1);$$

$\theta\wedge m$ означает минимальную из величин θ и m.

Проанализируем поведение функционала $I(\cdot)$. В простейшем случае, когда шумы отсутствуют ($\eta(t)\equiv 0$), $I(\cdot)$ убывает при увеличении m от 0 до θ и возрастает после этого момента, причем приращение функционала

$$\Delta I(t)=I(t)-I(t-1)=\pm r(t,x)$$

зависит от различия моделей процесса в момент t. Далее стоит отметить, что для оценивания момента "разладки" по методу наименьших квадратов не требуется проводить все N наблюдений процесса x, необходимые для вычисления $I(m)$, а достаточно следить лишь за приращениями функционала $\Delta I(t)$, которые находятся по текущим наблюдениям $x(0),...,x(t)$. Присутствие шумов сказывается на поведении функционала $I(\cdot)$. Однако принципиальный вывод о том, что "разладку" с помощью метода наименьших квадратов можно обнаружить по приращениям $\Delta I(t)$ в текущем времени, остается верным и в этом случае. Естественно, что при наличии шумов выводы о "разладке" по отдельным значениям $\Delta I(t)$ являются ненадежным. Поэтому логично выделить некоторую возрастающую последовательность моментов $\{\tau_i, i\geq 0\}$ и принимать решения по приращениям $I(\tau_i)-I(\tau_{i-1})$. Выбором моментов τ_i можно регулировать как вероятность ложных тревог, так и величину запаздывания в обнаружении "разладки". Моменты τ_i, очевидно, должны выбираться с учетом различия конкурирующих моделей в отдельные моменты времени. В качестве меры близости моделей в момент t будем использовать величину $r(t,x)$, определяемую равенством (2.31).

Пусть H - положительное число. Введем последовательность моментов $\{\tau_i, i \geq 0\}$ по формулам:

$$\tau_0(H) = 0,\ \tau_i(H) = \inf\left\{k \geq \sigma_i(H):\ \sum_{t=\sigma_i(H)}^{k} r(t,x) \geq H\right\},\ i \geq 1,$$
$$\sigma_i(H) = \begin{cases} 0,\ i = 1 \\ \tau_{i-1}(H) + 1,\ i \geq 2 \end{cases} \tag{2.33}$$

Для конечности почти наверное величин τ_i ряд

$$\sum_{t=0}^{\infty} r(t,x) = +\infty \tag{2.34}$$

должен расходиться с вероятностью единица. С каждым интервалом $(\tau_{i-1}, \tau_i]$ (с i -м циклом наблюдений) свяжем статистику $Y_i(H)$ определяемую равенствами :

$$Y_i(H) = \frac{2}{H}\sum_{t=\sigma_i}^{\tau_i} \alpha_i(t)\left[A_2(t,x) - A_1(t,x)\right]\left(BB^T\right)^{+}(t,x)\left[x(t+1) - \frac{A_1(t,x) + A_2(t,x)}{2}\right],$$
$$i \geq 1;$$
$$\alpha_i(t) = \begin{cases} 1,\ \ \sigma_i \leq t \leq \tau_i; \\ \left[H - \sum_{s=\sigma_i}^{\tau_i - 1} r(s,x)\right] / r(\tau_i, x),\ \ t = \tau_i; \\ 0,\ \ t > \tau_i \ u\ t < \sigma_i. \end{cases} \tag{2.35}$$

Эта статистика совпадает с приращением функционала $I(\cdot)$ на интервале $(\tau_{i-1}, \tau_i]$ с точностью до постоянного множителя и последнего слагаемого, которое берется с весом $0 < \alpha_i(\tau_i) \leq 1.$

Используя (2.29) и (2.30) , получим

$$Y_i(H) = \begin{cases} -1 + \varepsilon_i(H), & \tau_i < \theta; \\ 1 + \varepsilon_i(H), & \tau_{i-1} \geq \theta; \end{cases} \tag{2.36}$$

(2.37)

$$\varepsilon_i(H) = \frac{2}{H} \sum_{t=\sigma_i}^{\tau_i} \alpha_i(t)\eta(t+1),$$

где $\eta(t+1)$ определяется формулой (2.32). В случае $\tau_{i-1} < \theta \le \tau_i$ статистика $Y_i(H)$ представляет собой сумму $\varepsilon_i(H)$ и некоторой случайной величины, не превосходящей по модулю единицы. Учитывая, что статистики $Y_i(H)$ претерпевают скачок после "разладки", процедуру обнаружения "разладки" естественно определить следующим образом. Зададим порог δ ($|\delta| < 1$) и при выполнении в i-м цикле неравенства

$$Y_i(H) > \delta \tag{2.38}$$

будем принимать решение о "разладке".

Решающая процедура обнаружения "разладки" построена.

2.3.3 ВЫБОР ПАРАМЕТРОВ И ИССЛЕДОВАНИЕ СВОЙСТВ РЕШАЮЩЕЙ ПРОЦЕДУРЫ

При решении вопроса о выборе параметров процедуры H и δ будем исходить из того, что заданы допустимые вероятности ложной тревоги α и ложного спокойствия β в каждом цикле. Поэтому ограничимся рассмотрением множества пар (H, δ), удовлетворяющих неравенствам:

$$P_0^i(H,\delta) \le \alpha,\ P_1^i(H,\delta) \le \beta,\ i \ge 1 \tag{2.39}$$

где $P_0^i(H,\delta)$ и $P_1^i(H,\delta)$ - соответственно вероятности ложной тревоги и ложного спокойствия в i-м цикле. В этом множестве оптимальной является пара (H,δ) с минимальным значением H.

Такая пара, обеспечивающая заданную надежность процедуры, приводит к минимальной длине цикла, а следовательно, и к минимальному запаздыванию в обнаружении момента "разладки". Будем предполагать, что

моменту появления "разладки" предшествует длительный процесс наблюдений. Тогда для большого числа циклов вероятность "захвата" момента "разладки" θ остается малой, т.е. $P\{\tau_i < \theta\}$ близка к единице. При этом условии верны формулы:

$$P_0^i(H,\delta) = \lim_{\theta\to\infty} P\{Y_i(H) > \delta \mid \tau_i < \theta\} = P\{\varepsilon_i(H) > 1+\delta\};$$
$$P_1^i(H,\delta) = P\{Y_i(H) > \delta \mid \tau_{i-1} \geq \theta\} = P\{\varepsilon_i(H) < -1+\delta \mid \tau_{i-1} \geq \theta\} \quad (2.40)$$

Поскольку точные значения самих вероятностей не могут быть найдены при сделанных предположениях относительно наблюдаемого процесса, получим верхние оценки вероятностей (2.40) и для них решим поставленную задачу параметрической процедуры. Полученная в результате процедура будет обладать требуемой надежностью (2.39).

ТЕОРЕМА 2.3: Пусть для процесса (2.29), (2.30) выполнено условие (2.34). Тогда:

1. Для любой пары (H,δ), $H>0$, $-1<\delta<1$ справедливы неравенства:

$$P_0^i(H,\delta) \leq \frac{4}{H(\delta+1)^2};\ P_1^i(H,\delta) \leq \frac{4}{H(\delta-1)^2} \quad (2.41)$$

2. Для любой пары (α,β), $0<\alpha$, $\beta<1$, последовательная процедура (2.35), (2.38) с параметрами:

$$H^* = \left(\frac{1}{\sqrt{\alpha}} + \frac{1}{\sqrt{\beta}}\right)^2;\ \delta^* = \frac{\sqrt{\beta}-\sqrt{\alpha}}{\sqrt{\alpha}+\sqrt{\beta}} \quad (2.42)$$

имеет вероятность ложной тревоги не больше, чем α, и вероятность ложного спокойствия не больше, чем β.

3. Для пары (H^*,δ^*) среднее число циклов между ложными тревогами N_0 и среднее число циклов запаздывания в обнаружении "разладки" N_1 удовлетворяет неравенствам:

$$N_0 \geq \frac{1}{\alpha};\ N_1 \leq \frac{1}{1-\beta} \tag{2.43}$$

Доказательство теоремы 2.3. Нам потребуется следующая

ЛЕММА 2.1: Пусть на вероятностном пространстве (Ω, F, P) заданы:

а) неубывающая система σ-подалгебр $\{F_n\}_{n\geq 0}$;

б) $\{\eta_n\}_{n\geq 0}$ и $\{\tau_n\}_{n\geq 0}$ - согласованные с $\{F_n\}$ последовательности случайных величин, такие, такие, что

$$E\left[\eta_{\sigma+k+1} \mid F_{\sigma+k}\right] = 0,\ E\left[\eta^2_{\sigma+k+1} \mid F_{\sigma+k}\right] \leq r_{\sigma+k},\ k \geq 0, \tag{2.44}$$

причем с вероятностью единица $\sum_{k\geq 0} r_k = +\infty$. Тогда для момента τ_σ, определяемого по формуле

$$\tau_\sigma = \inf\left\{k \geq \sigma : \sum_{j=\sigma}^{k} r_j \geq H\right\};\ \tau_\sigma = \inf\{\varnothing\} = +\infty,$$

при любом $H > 0$ имеет место соотношения:

$$E\left[\sum_{k=\sigma}^{\tau_\sigma} \alpha(k;\sigma,\tau_\sigma)\eta_{k+1} \mid F_\sigma\right] = 0;$$

$$E\left[\left(\sum_{k=\sigma}^{\tau_\sigma} \alpha(k;\sigma,\tau_\sigma)\eta_{k+1}\right)^2 \mid F_\sigma\right] \leq H, \tag{2.45}$$

где

$$\alpha(k;\sigma,\tau_\sigma) = \begin{cases} 1, & \sigma \leq k < \tau_\sigma; \\ \left[H - \sum_{j=\sigma}^{\tau_\sigma - 1} r_j\right] / r_{\tau_\sigma}, & \mathrm{k} = \tau_\sigma; \\ 0, & \mathrm{k} > \tau_\sigma. \end{cases}$$

F_σ - σ-алгебра событий, наблюдаемых до момента σ.

Доказательство леммы 2.1. Будем использовать свойства марковских моментов. Система σ-алгебр $\{F_\sigma, F_{\sigma+1}, \ldots\}$ является монотонно неубывающей, а

τ_σ-марковский момент относительно этой системы. Для проверки условия марковости момента τ_σ:

$$\left[\tau_\sigma \le k\right] \in F_{\sigma+k},\ k \ge 0,$$

требуется показать, что для любого k имеет место включение

$$\left[\tau_\sigma \le k\right] \bigcap \left[\sigma + k = l\right] \in F_l,\ l \ge k.$$

Это включение следует из равенства

$$\left[\tau_\sigma \le k\right] \bigcap \left[\sigma + k = l\right] = \left\{ \sum_{j=\sigma}^{k} r_j \ge H \right\} \bigcap \left[\sigma = l - k\right],$$

условия согласованности последовательности $\{r_n\}$ и марковости момента σ.

Введем усеченный марковский момент $\bar{\tau}_\sigma = \min(\tau_\sigma, N)$ и обозначим

$$T_N = \sum_{k=\sigma}^{\tau_\sigma} \alpha(k;\sigma,\tau_\sigma)\eta_{k+1}.$$

Докажем сначала, что

$$E\left[T_N \mid F_\sigma\right] = 0 \text{ и } E\left[(T_N)^2 \mid F_\sigma\right] \le H. \tag{2.46}$$

Имеем

$$\begin{aligned} E\left[T_N^2 \mid F_\sigma\right] &= E\left[\sum_{k=\sigma}^{\bar{\tau}_\sigma}\sum_{l=\sigma}^{\bar{\tau}_\sigma} \alpha(k;\sigma,\tau_\sigma)\alpha(l;\sigma,\tau_\sigma)\eta_{k+1}\eta_{l+1} \mid F_\sigma\right] = \\ &= E\left[\sum_{k=\sigma}^{\bar{\tau}_\sigma} \alpha(k;\sigma,\tau_\sigma)\eta_{k+1} \mid F_\sigma\right] + \\ &+ 2M\left[\sum_{\sigma \le k \le l \le \bar{\tau}_\sigma} \alpha(k;\sigma,\tau_\sigma)\alpha(l;\sigma,\tau_\sigma)\eta_{k+1}\eta_{l+1} \mid F_\sigma\right]. \end{aligned} \tag{2.47}$$

Рассмотрим первое слагаемое. Используя свойства условных математических ожиданий, условия (2.44) и учитывая, что $0 < \alpha \le 1$ получаем:

$$
\begin{aligned}
&E\left[\sum_{k=\sigma}^{\bar{\tau}_\sigma}(\cdot)\,|\,F_\sigma\right]=\\
&=E\left\{\sum_{k=0}^{N}\alpha^2(\sigma+j;\sigma,\tau_\sigma)\eta^2_{\sigma+j+1}\chi_{\left[\sigma+j\le\bar{\tau}_\sigma\right]}\,|\,F_\sigma\right\}=\\
&=\sum_{j=0}^{N}E\left\{\alpha^2(\sigma+j;\sigma,\tau_\sigma)\chi_{\left[\sigma+j\le\bar{\tau}_\sigma\right]}E\left(\eta^2_{\sigma+j+1}\,|\,F_\sigma\right)\right\}\le\\
&\le\sum_{j=0}^{N}E\left\{\alpha^2(\sigma+j;\sigma,\tau_\sigma)r_{\sigma+j}\chi_{\left[\sigma+j\le\bar{\tau}_\sigma\right]}\,|\,F_\sigma\right\}=\\
&=E\left[\sum_{k=\sigma}^{\bar{\tau}_\sigma}\alpha(k;\sigma,\tau_\sigma)r_k\,|\,F_\sigma\right]\le\\
&\le E\left[\sum_{k=\sigma}^{\tau_\sigma}\alpha(k;\sigma,\tau_\sigma)r_k\,|\,F_\sigma\right]=H
\end{aligned}
$$

Последнее равенство выполняется в силу выбора весовой функции $\alpha(k;\sigma,\tau_\sigma)$. Аналогично показывается, что слагаемое в (2.47) и $E\left[T_N\,|\,F_\sigma\right]$ равны нулю. Таким образом, установлены соотношения (2.46), из которых с помощью предельного перехода при $N\to\infty$ получается утверждение леммы 2.1.

Лемма 2.1. доказана.

1. Воспользуемся леммой 2.1. для доказательства утверждения 1. В качестве F_n возьмем σ-алгебру, порожденную случайными векторами $x(0),\xi(1),...,\xi(n)$, и проверим, что для последовательностей $\{r(n,x),n\ge0\}$ и $\{\eta(n),n\ge1\}$, определяемых формулами (2.32), выполнены условия леммы 2.1. Условие согласованности очевидно. Далее имеем

$$
\begin{aligned}
&E\left[\eta_{\sigma+k+1}\,|\,F_{\sigma+k}\right]=\\
&=\left[A_2(\sigma+k,x)-A_1(\sigma+k,x)\right]^T\cdot\\
&\cdot\left(B^T\right)^+(\sigma+k,x)E\left[\xi(\sigma+k+1)\,|\,F_{\sigma+k}\right];\\
&E\left[\eta^2_{\sigma+k+1}\,|\,F_{\sigma+k}\right]=\\
&=\left[A_2(\sigma+k,x)-A_1(\sigma+k,x)\right]^T\cdot\\
&\cdot\left(B^T\right)^+(\sigma+k,x)E\left[\xi(\sigma+k+1)\xi^T(\sigma+k+1)\,|\,F_{\sigma+k}\right]
\end{aligned}
\qquad (2.48)
$$

Для последовательности независимых одинаково распределенных векторов $\{\xi(n), n \geq 1\}$ выполняется строго марковское свойство [42]. Поэтому

$$E\left[\xi(\sigma+k+1)\,|\,F_{\sigma+k}\right] = E\left[\xi(1)\,|\,F_0\right] = E\xi(1) = 0, \quad F_0 = \sigma\{x(0)\};$$
$$E\left[\xi(\sigma+k+1)\xi^T(\sigma+k+1)\,|\,F_{\sigma+k}\right] = \\ = E\left[\xi(1)\xi^T(1)\,|\,F_0\right] = E\left[\xi(1)\xi^T(1)\right] = E. \tag{2.49}$$

Из (2.48), (2.49) следуют соотношения (2.44) и справедливость утверждений леммы 2.1. для последовательностей (2.32). Последовательность моментов $\{\tau_i, i \geq 0\}$ по определению является марковской относительно введенной системы σ-алгебр. Учитывая, что из соотношений (2.45) при $\sigma = \sigma_i$ для величины $\varepsilon_i(H)$, определяемой равенством (2.37), получаем

$$E\left[\varepsilon_i(H)\,|\,F_{\sigma_i}\right] = 0,\ E\left[\varepsilon_i^2(H)\,|\,F_{\sigma_i}\right] \leq \frac{4}{H}. \tag{2.50}$$

Отсюда и из (2.40) следует первое из неравенство в утверждении 1. Для доказательства второго неравенства (2.41) требуются дополнительные построения. Обозначим υ номер цикла, захватившего момент "разладки", т.е.

$$\upsilon = \inf\{j \geq 1 : \tau_j \geq \theta\}.$$

Случайная величина υ является марковским моментом относительно системы σ-алгебр $\{F_{\tau_k}, k \geq 0\}$. Для проверки условия марковости требуется показать, что для каждого j $(j \geq 0)$ имеет место включение

$$[\upsilon = j] \cap \left[\tau_j = l\right] \in F_l,\ l \geq 0$$

Это включение следует из равенства

$$[\upsilon = j] \cap \left[\tau_j = l\right] = \left[\tau_1 < \theta, ..., \tau_{j-1} < \theta, \tau_j \geq \theta\right] \cap \left[\tau_j = l\right]$$

и марковости момента τ_j относительно системы $\{F_n, n \geq 1\}$.

Оценим вероятность $P_1^i(H,\delta)$. Очевидно, что

$$P_1^i(H,\delta)=P\{\varepsilon_i(H)\le\delta-1|\,\tau_{i-1}\ge\theta\}\le P\{|\varepsilon_i(H)|\le\delta-1|\,\tau_{i-1}\ge\theta\}.$$

Так как

$$\{\tau_{i-1}\ge\theta\}=\{\upsilon\le i-1\}\in F_{\tau_{i-1}}\subseteq F_{\sigma_i},$$

то, используя определение условной вероятности, неравенство Чебышева и (2.50), получаем

$$P\{|\varepsilon_i(H)|>1-\delta,\tau_{i-1}\ge\theta\}=\int\limits_{\{\tau_{i-1}\ge\theta\}}P\{|\varepsilon_i(H)|>1-\delta\,|\,F_{\sigma_i}\}\le\frac{4}{H(1-\delta)^2}P\{\tau_{i-1}\ge\theta\}.$$

Отсюда и из (2.40) следует второе неравенство (2.41). Утверждение 1 доказано.

2. Для доказательства утверждения 2 достаточно убедиться, что во множестве пар (H,δ), удовлетворяющих неравенствам :

$$\frac{4}{H(\delta+1)^2}\le\alpha,\ \frac{4}{H(\delta-1)^2}\le\beta,$$

пара (H^*,δ^*) имеет минимальное значение H.

3. Оценим сначала N_1. Обозначим n_1 число циклов запаздывания в обнаружении разладки. Воспользуемся формулой

$$N_1=En_1=\sum_{k\ge0}P\{n_1>k\}$$

Докажем вспомогательное равенство

$$P\{n>k\}\le\frac{4}{(\delta^*-1)^2H^*}P\{n_1>k-1\} \qquad (2.51)$$

По определению имеем

$$
\begin{aligned}
&P\{n_1 > k\} = P\{\varepsilon_{\upsilon+1} < \delta^* - 1, \ldots, \varepsilon_{\upsilon+k} < \delta^* - 1\} = \\
&= \sum_{j\geq 1} P\{\varepsilon_{\upsilon+1} < \delta^* - 1, \ldots, \varepsilon_{\upsilon+k} < \delta^* - 1, \upsilon = j\} = \\
&= \sum_{j\geq 1} P\{\varepsilon_{j+1} < \delta^* - 1, \ldots, \varepsilon_{j+k} < \delta^* - 1, \upsilon = j\}.
\end{aligned} \tag{2.52}
$$

Далее находим

$$
\begin{aligned}
&P\{\varepsilon_{j+1} < \delta^* - 1, \ldots, \varepsilon_{j+k} < \delta^* - 1, \upsilon = j\} = \\
&= \int\limits_{\{\varepsilon_{j+1} < \delta^* - 1, \ldots, \varepsilon_{j+k} < \delta^* - 1, \upsilon = j\}} P\{\varepsilon_{j+k} < \delta^* - 1 \mid F_{j+k}\} dP \leq \\
&\leq \frac{4}{(\delta^* - 1)^2 H^*} P\{\varepsilon_{j+1} < \delta^* - 1, \ldots, \varepsilon_{j+k} < \delta^* - 1, \upsilon = j\}.
\end{aligned}
$$

Отсюда и из (2.52) следует равенство (2.51). Применяя неравенство (2.51) повторно, получаем

$$
P\{n_1 > k\} \leq \left[\frac{4}{(\delta^* - 1)^2 H^*}\right]^k.
$$

После суммирования приходим к требуемой оценке

$$
N_1 \leq \frac{1}{1 - \dfrac{4}{(\delta^* - 1)^2 H^*}} = \frac{1}{1-\beta}.
$$

При выводе оценки (2.43) для среднего количества циклов между ложными тревогами предполагается, что "разладке" предшествует длительный процесс наблюдений. Пусть μ -номер последнего цикла, в котором произошла ложная тревога, а n_0-количество циклов, через которое ложная тревога повториться. Тогда

$$
N_0 = E_\mu n_0 = \sum_{k\geq 0} P_\mu\{n_0 > k\}.
$$

По определению n_0, используя (2.50) , получаем

$$P_\mu\{n_0 > k\} = P_\mu\{\varepsilon_{\mu+1} < \delta^* + 1, ..., \varepsilon_{\mu+k} < \delta^* + 1\} =$$
$$= \int_{\{\varepsilon_{\mu+1} < \delta^*+1, ..., \varepsilon_{\mu+k-1} < \delta^*+1\}} P_\mu\{\varepsilon_{\mu+k} < \delta^* + 1 | F_{\sigma_{\mu+k}}\} dP \geq$$
$$\geq \left[1 - \frac{4}{H^*(\delta^*+1)^2}\right] P_\mu\{n_0 > k-1\}$$

Отсюда $P_\mu\{n_0 > k\} \geq \left[1 - \frac{4}{H^*(\delta^*+1)^2}\right]$ и, следовательно, $N_0 \geq \frac{1}{\frac{4}{H^*(\delta^*+1)^2}} = \frac{1}{\alpha}$.

Теорема 2.1 доказана.

Замечание 2.1: Выше предполагалось, что заданы допустимые вероятности ложной тревоги и ложного спокойствия и при этих условиях производилась оптимизация параметров процедуры. Можно рассмотреть другой вариант задачи оптимизации. Введем функции:

$$Q(t_1, t_2) = \sum_{s=t_1}^{t_2} r(s, x);\ Q'(t_1, t_2) = \sum_{s=t_1}^{t_2}{}' r(s, x),$$

где $t_1 \leq t_2$, штрих у знака второй суммы означает, что величина $r(s, x)$ берется с весом $\alpha_i(\tau_i)$, если $s = \tau_i$, $i = 1, 2, \ldots$. Первая из этих сумм определяет суммарное различие между конкурирующими моделям (2.29), (2.30), накапливаемое на интервале наблюдений $[t_1, t_2]$; вторая совпадает с первой с точностью до отдельных слагаемых , взятых с весами. Пусть t' и t'' - моменты соседних ложных тревог, $t' < t''$, а t_3 -время запаздывания в обнаружении момента "разладки" θ. Тогда, используя теорему 2.3 и определение моментов τ_i, получаем неравенства:

$$EQ(t', t'') \geq N_0 H^* \geq \frac{H^*}{\alpha} = T_0(\alpha, \beta)$$

$$EQ'(\theta, \theta + t_3) \leq N_1 H^* \leq \frac{H^*}{1-\beta} = T_1(\alpha, \beta)$$

Зафиксируем величину $T_0(\alpha,\beta)$, нижнюю оценку среднего суммарного различия моделей, накапливаемого между ложными тревогами, и найдем параметры α и β из условия минимума $T_1(\alpha,\beta)$, верхней границы для среднего суммарного различия моделей, накапливаемого после "разладки". Разрешив уравнение

$$T_0 = \frac{1}{\alpha}\left(\frac{1}{\sqrt{\alpha}} + \frac{1}{\sqrt{\beta}}\right)^2$$

относительно β и подставляя β в T_1, получим

$$\left(\alpha\sqrt{T_0} - 1\right)^3 = 2\alpha^2\sqrt{T_0}\,. \tag{2.53}$$

Параметрическое решение этого уравнение имеет вид

$$\alpha = \frac{u}{\sqrt{T_0}};\ \sqrt{T_0} = \frac{2u^2}{(u-1)^3}\,. \tag{2.54}$$

Используя найденное α и соответствующее ему β в (2.42), получаем параметры процедуры, являющиеся решением поставленной задачи.

Отметим, что в частном случае, когда $r(s,x) = r_0$ до момента "разладки" θ и $r(s,x) = r_1$ для $s \geq \theta$, величина $EQ(t',t'')/r_0$ определяет среднее время между ложными тревогами, а величина $EQ'(\theta,\theta+t_3)/r_1$ - среднее время запаздывания в обнаружении "разладки".

3 РЕЗУЛЬТАТЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

В этой главе будут приведены результаты численного моделирования.

В пункте 2.1 настоящей работы описывался алгоритм обнаружения "разладки" для первого коэффициента - δ модели AR(1). В соответствии с теорией [16] коэффициент $\phi = 0,8 = const$. Варьируя значения δ_0 и δ_1 получили следующие результаты.

Моделировался процесс: $X_t = \begin{cases} -0,5 + 0,8X_{t-1} + \omega_t, & \text{t} \le 1500. \\ -0,6 + 0,8X_{t-1} + \omega_t & 1500 < t \le 2000. \end{cases}$

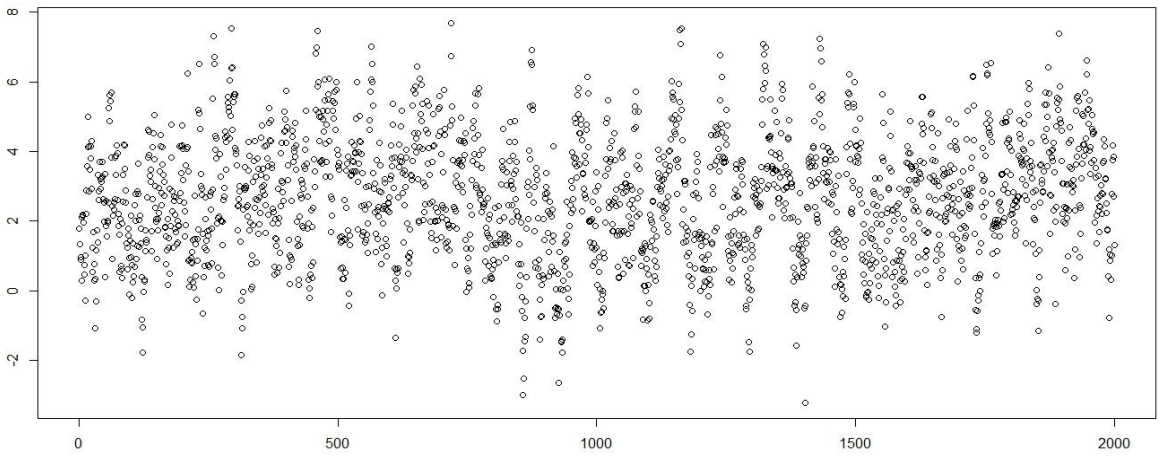


Рисунок 3.1- Диагностируемый момент разладки $t_k = 1741$;
статистика критерия: $\Lambda_k = 5.635$.

Моделировался процесс: $X_t = \begin{cases} -0,5 + 0,8X_{t-1} + \omega_t, & \text{t} \le 1500. \\ -0,6 + 0,8X_{t-1} + \omega_t & 1500 < t \le 2000. \end{cases}$

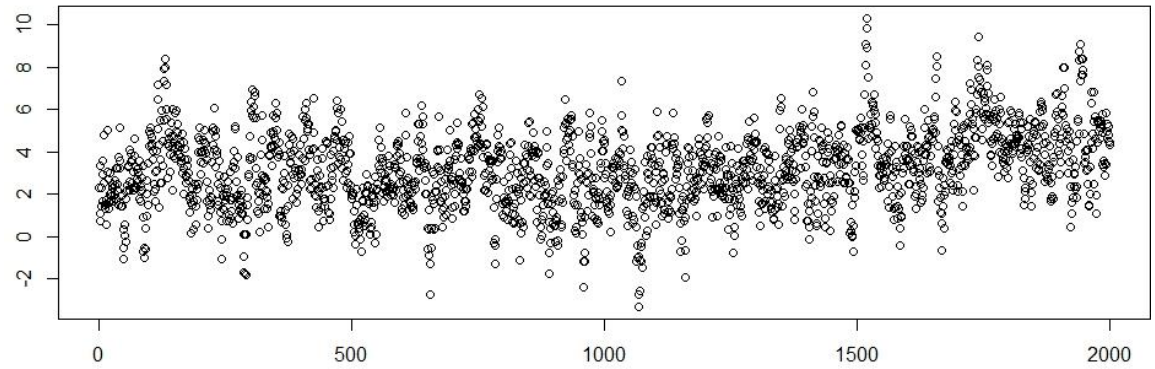


Рисунок 3.2- Диагностируемый момент "разладки" $t_k = 1504$;
тестовая статистика: $\Lambda_k = 29.692$.

Моделировался процесс: $X_t = \begin{cases} -0,4 + 0,8X_{t-1} + \omega_t, & \text{t} \le 1500. \\ -0,8 + 0,8X_{t-1} + \omega_t & 1500 < t \le 2000. \end{cases}$

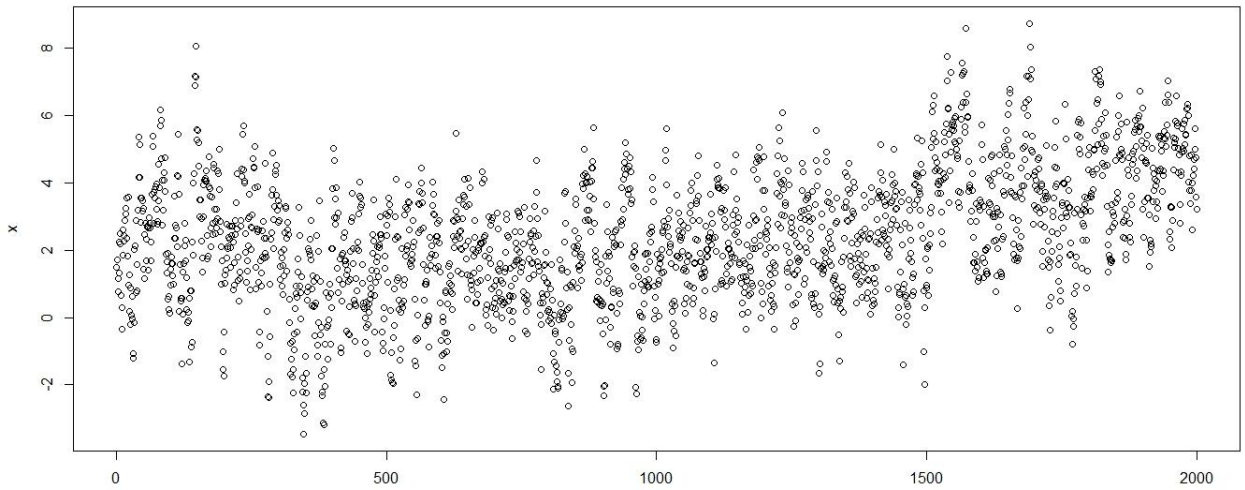


Рисунок 3.3- Диагностируемый момент "разладки" $t_k = 1504$;
тестовая статистика: $\Lambda_k = 57.143$.

Моделировался процесс: $X_t = \begin{cases} -0,3 + 0,8X_{t-1} + \omega_t, & \text{t} \le 1500. \\ -0,9 + 0,8X_{t-1} + \omega_t & 1500 < t \le 2000. \end{cases}$

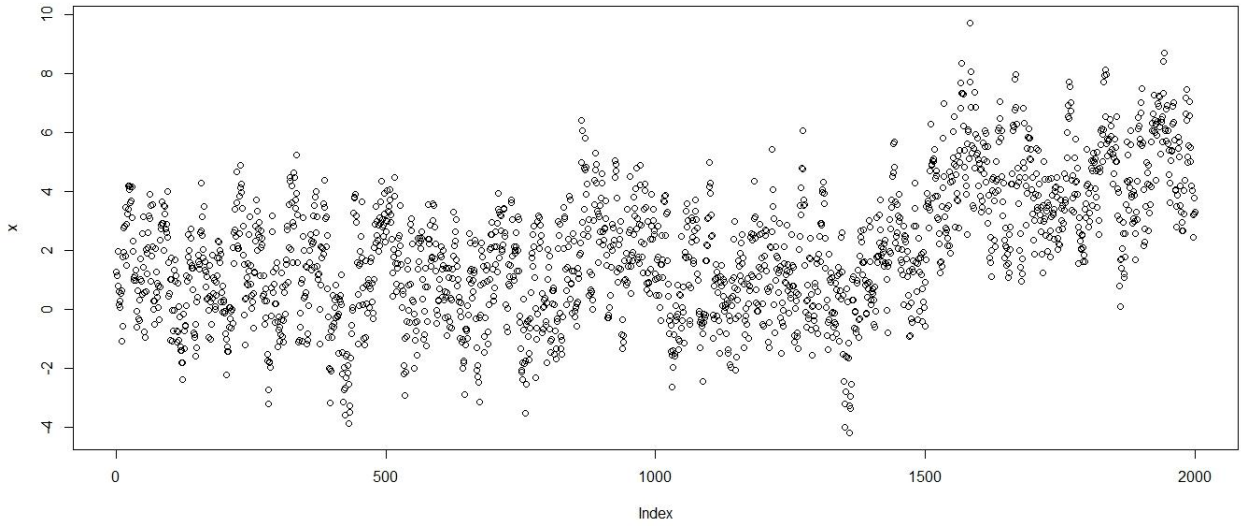


Рисунок 3.4- Диагностируемый момент "разладки" $t_k = 1499$;
тестовая статистика: $\Lambda_k = 88.0499$.

В пункте 2.2 настоящей работы описывался алгоритм обнаружения "разладки" для второго коэффициента- ϕ модели AR(1). Аналогично предыдущему, варьировались коэффициенты ϕ_0 и ϕ_1. При этом коэффициент $\delta = 0 = const$.

Моделировался процесс: $X_t = \begin{cases} 0,4X_{t-1} + \omega_t, & \text{t} \le 1000. \\ 0,5X_{t-1} + \omega_t & 1000 < t \le 1500. \end{cases}$

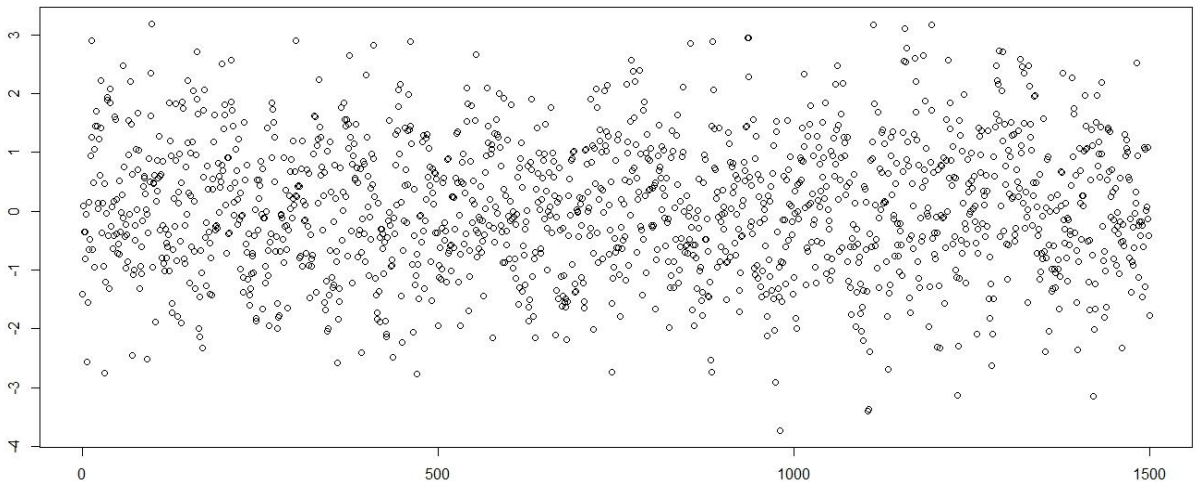


Рисунок 3.5- Диагностируемый момент "разладки" $t_k = 868$,
тестовая статистика: $\Lambda_k = 5.748$.

Моделировался процесс: $X_t = \begin{cases} 0,3X_{t-1} + \omega_t, & \text{t} \le 1000. \\ 0,6X_{t-1} + \omega_t & 1000 < t \le 1500. \end{cases}$

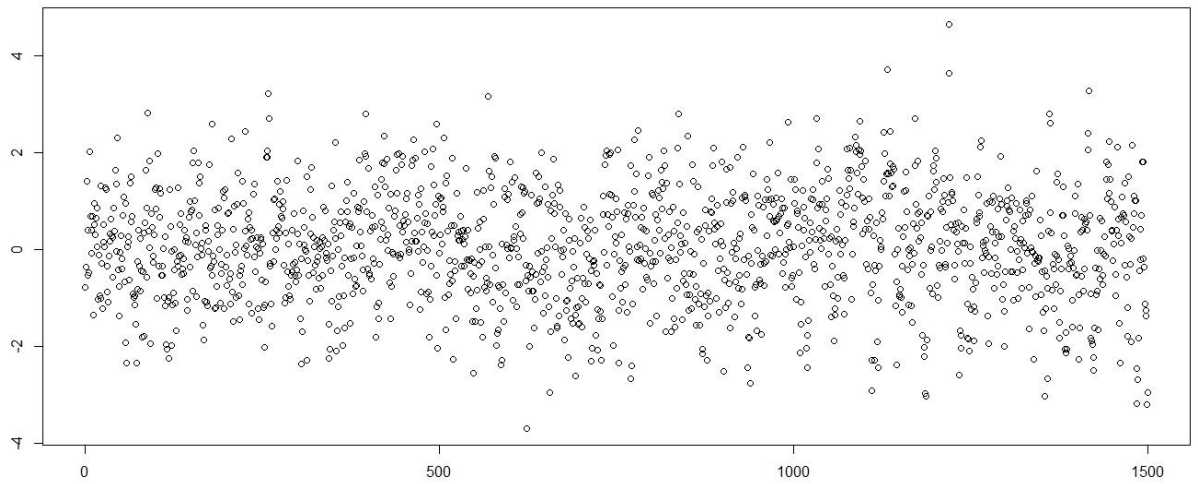


Рисунок 3.6- Диагностируемый момент "разладки" $t_k = 1005$,
тестовая статистика: $\Lambda_k = 47.616$.

Моделировался процесс: $X_t = \begin{cases} 0,4X_{t-1} + \omega_t, & \text{t} \le 1000. \\ 0,8X_{t-1} + \omega_t & 1000 < t \le 1500. \end{cases}$

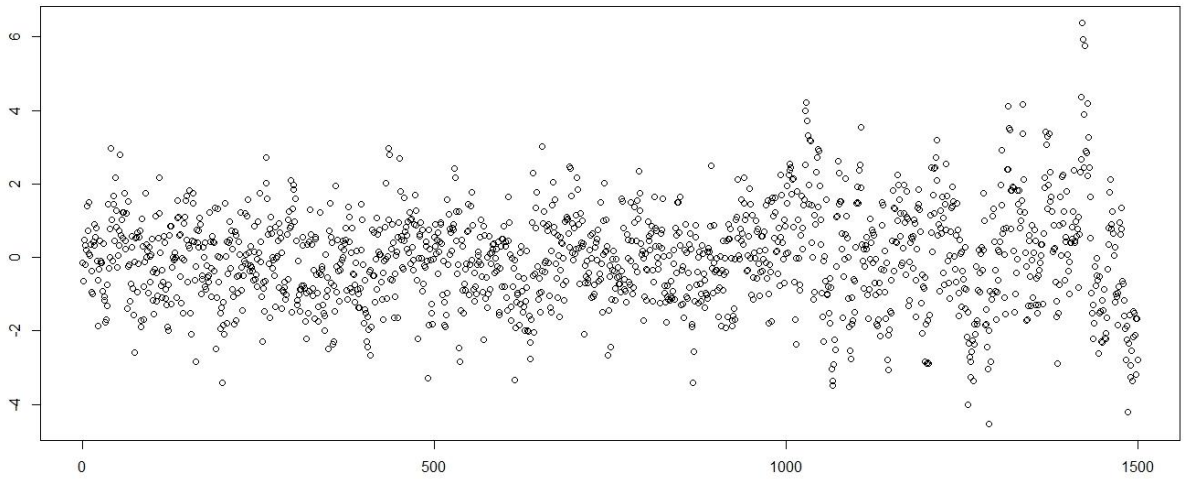


Рисунок 3.7- Диагностируемый момент "разладки" $t_k = 1003$,
тестовая статистика: $\Lambda_k = 97.033$.

Моделировался процесс: $X_t = \begin{cases} 0,4X_{t-1} + \omega_t, & \text{t} \le 1000. \\ 0,9X_{t-1} + \omega_t & 1000 < t \le 1500. \end{cases}$

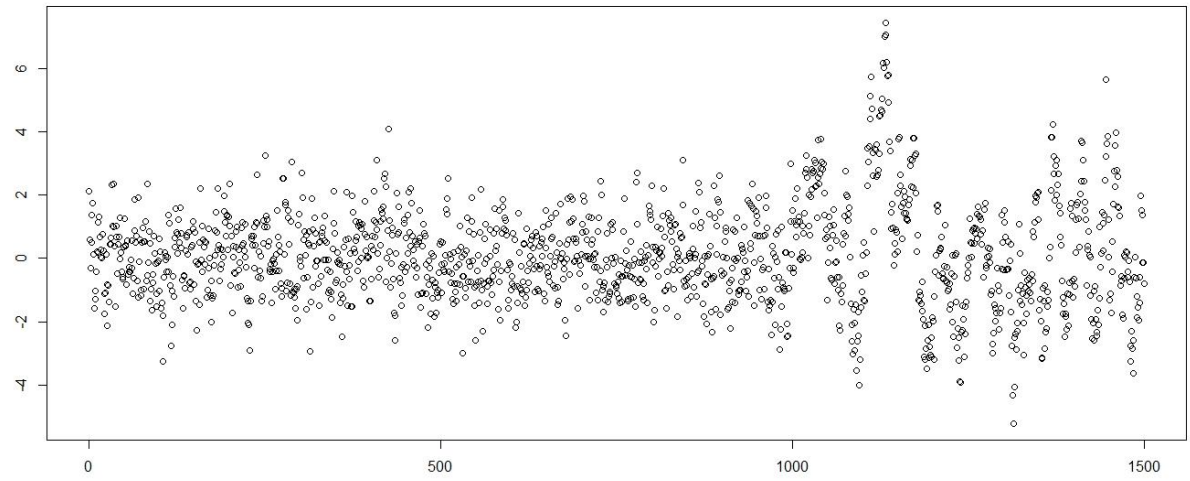


Рисунок 3.8- Диагностируемый момент "разладки" $t_k = 1003$,
тестовая статистика: $\Lambda_k = 178.55$.

Также в работе рассматривался второй непараметрический метод обнаружения "разладки" . Этот метод разработан не только для моделей вида AR(p), а для более широкого класса. Симуляционное моделирование было проведено на моделях двух видов.

Модель 1: появление детерминированного сигнала на фоне независимых шумов

$$x_t = \begin{cases} \xi_t, t < \theta \\ \sin(t) + \xi_t, t \geq \theta \end{cases}, \ \xi_t \sim N(0,1) \tag{3.1}$$

До момента разладки был шум, а после- появился сигнал.

Начнем построение процедуры с построения возрастающей последовательности моментов:

$$\tau_0 = 0; \ \tau_{k+1} = \inf\left\{ n \geq \tau_k : \sum_{t=\tau_k+1}^{n} \sin^2(t) \geq H \right\}; \tag{3.2}$$

Далее, строим статистику $Y_i(H)$, которая совпадает с приращением функционала $I(\cdot)$ на интервале $(\tau_{i-1}, \tau_i]$

$$Y_i = \frac{2}{H} \sum_{t=\tau_{i-1}+1}^{\tau_i} \alpha(t) \sin(t) \left[x_t - \frac{\sin(t)}{2} \right];$$
$$\alpha(t) = \begin{cases} 1, \ \tau_{i-1} + 1 \leq t \leq \tau_i \\ H - \sum_{t=\tau_{i-1}+1}^{\tau_i - 1} \sin^2(t) \\ \hline \sin^2(\tau_i) \end{cases}, \ t = \tau_i \tag{3.3}$$

Момент разладки определяем по правилу (3.4):

$$N_{CS} = \inf\left\{ n \geq 0 : \max_{0 \leq k \leq n} \sum_{i=k}^{n} Y_i \geq \delta \right\} \tag{3.4}$$

Моделировался процесс: $x_t = \begin{cases} \xi_t, t < 100 \\ \sin(t) + \xi_t, t \geq 100 \end{cases}$, $\xi_t \square N(0,1)$

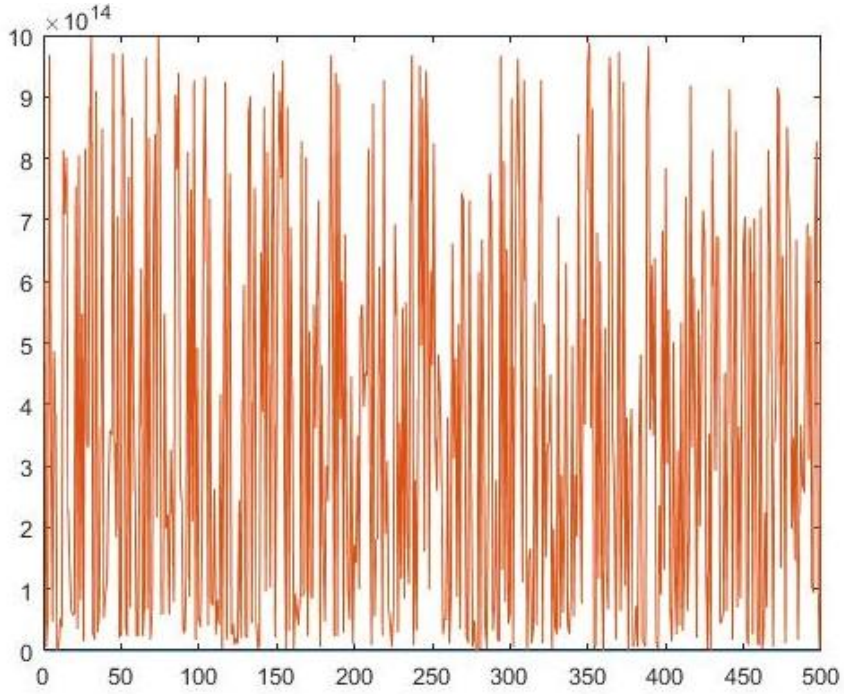


Рисунок 9- объем наблюдений: 500.

"Разладка" определяется близко к точному значению при наилучшем подборе коэффициентов α и β.

Таблица 3.1-коэффициенты α и β для $\theta = 100$

Teta =100		
alfa	**beta**	**момент разладки**
0,1	0,3	99
0,2	0,3	100
	0,6	99
0,3	0,6	102
0,4	0,9	102
0,5	0,7	102
	0,9	100
0,6	0,7	100
	0,8	99
0,8	0,9	102

Моделировался процесс: $x_t = \begin{cases} \xi_t, t < 250 \\ \sin(t) + \xi_t, t \geq 250 \end{cases}$, $\xi_t \square N(0,1)$

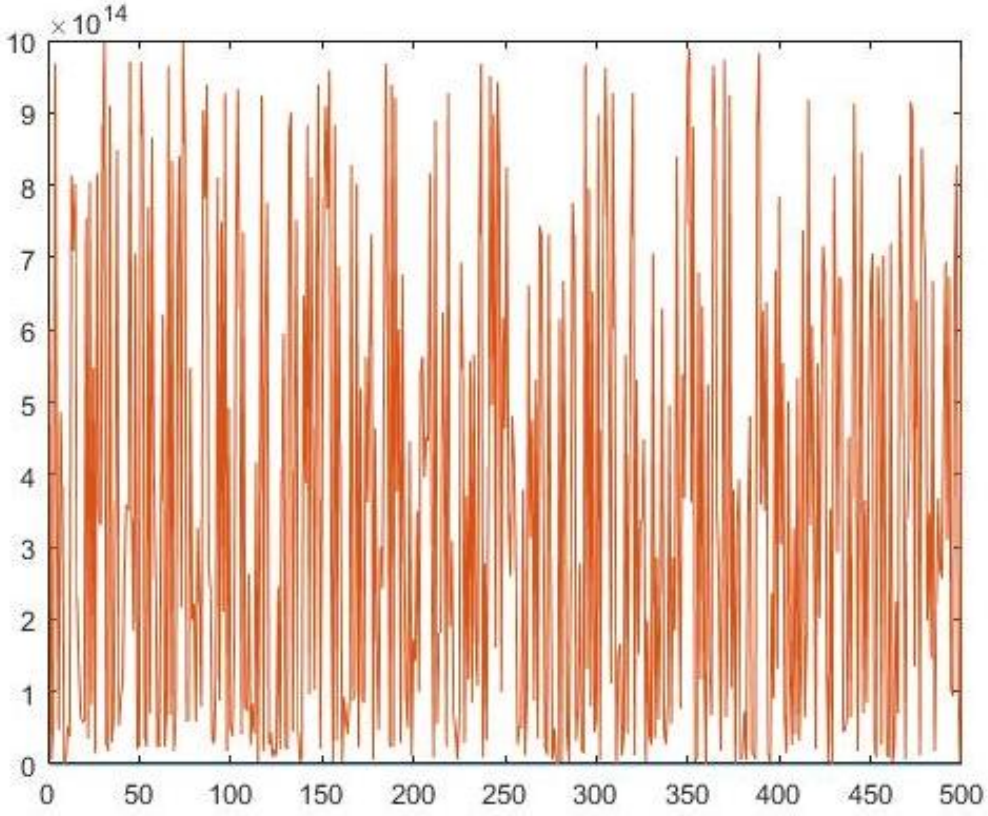


Рисунок 3.10- объем наблюдений: 500.

Таблица 3.2-коэффициенты α и β для $\theta = 250$

Teta 250		
alfa	**beta**	**момент разладки**
0,2	0,6	250
	0,8	257
0,3	0,6	244
	0,7	244
	0,8	256
0,4	0,5	244
	0,9	249
0,5	0,8	250
	0,9	250
0,6	0,9	257
0,7	0,9	253

Модель 2: появление детерминированного сигнала на фоне зависимых шумов

$$x_t = \begin{cases} \xi_t, t < \theta \\ \dfrac{1}{\ln(t+10)} + \xi_t, t \geq \theta \end{cases}, \ \xi_t = \delta\xi_{t-1} + \varepsilon_t, \xi_t \square\ N(0,1), \varepsilon_t \in N(0,1) \tag{3.5}$$

Начнем построение процедуры с построения последовательности моментов:

$$\tau_0 = 0; \ \tau_{k+1} = \inf\left\{ n \geq \tau_k : \sum_{t=\tau_k}^{n} r^2(t) \geq H \right\}; \tag{3.6}$$

Далее, строим статистику $Y_i(H)$ (формула (3.7)) , которая совпадает с приращением функционала $I(\cdot)$ на интервале $(\tau_{i-1}, \tau_i]$

$$Y_i = \frac{2}{H} \sum_{t=\tau_{i-1}}^{\tau_i} \alpha(t) r(t) \left[x_t - \frac{1}{2} r(t) \right]; \ r(t) = \frac{1}{\ln(t+10)};$$
$$\alpha(t) = \begin{cases} 1, \ \tau_{i-1} + 1 \leq t \leq \tau_i \\ \dfrac{H - \sum\limits_{t=\tau_{i-1}+1}^{\tau_i - 1} r^2(t)}{r^2(\tau_i)}, \ t = \tau_i \end{cases} \tag{3.7}$$

Момент разладки определяем по правилу (8):

$$N_{CS} = \inf\left\{ n \geq 0 : \max_{0 \leq k \leq n} \sum_{i=k}^{n} Y_i \geq \delta \right\} \tag{3.8}$$

Моделировался процесс: $x_t = \begin{cases} \xi_t, t < 100 \\ \dfrac{1}{\ln(t+10)} + \xi_t, t \geq 100 \end{cases}$,

$$\xi_t = \delta\xi_{t-1} + \varepsilon_t, \xi_t \square N(0,1), \varepsilon_t \in N(0,1)$$

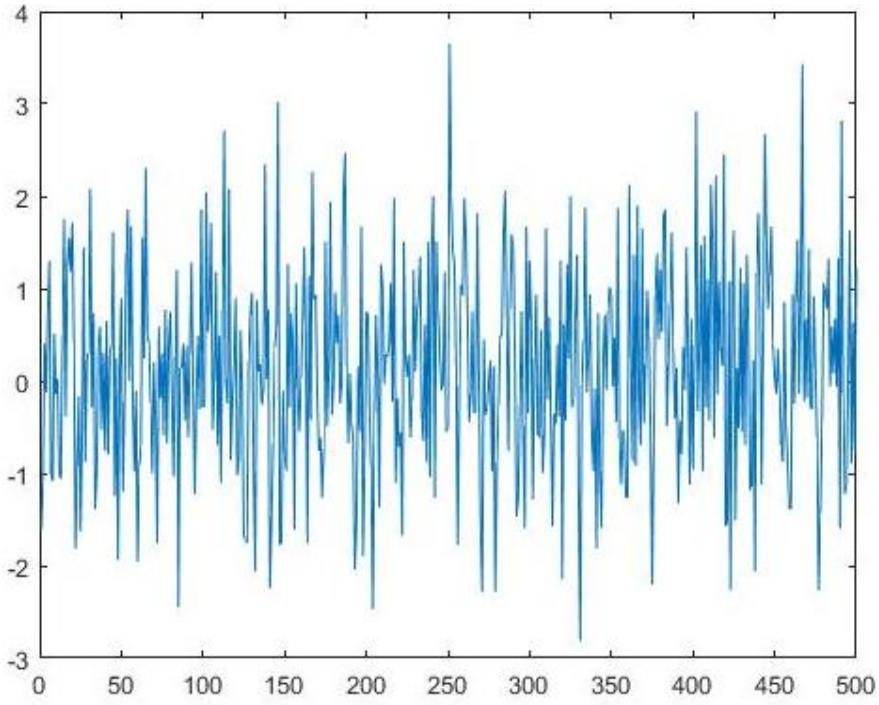


Рисунок 3.11- объем наблюдений: 500.

Таблица 3.3-коэффициенты α и β для $\theta = 100$.

Teta 100		
alfa	**beta**	**момент разладки**
0,4	0,7	117
	0,8	107
	0,9	100
0,5	0,7	97
	0,8	88
	0,9	82
0,6	0,7	83

Моделировался процесс: $x_t = \begin{cases} \xi_t, t < 250 \\ \dfrac{1}{\ln(t+10)} + \xi_t, t \geq 250 \end{cases}$,

$$\xi_t = \delta\xi_{t-1} + \varepsilon_t, \xi_t \square\ N(0,1), \varepsilon_t \in N(0,1)$$

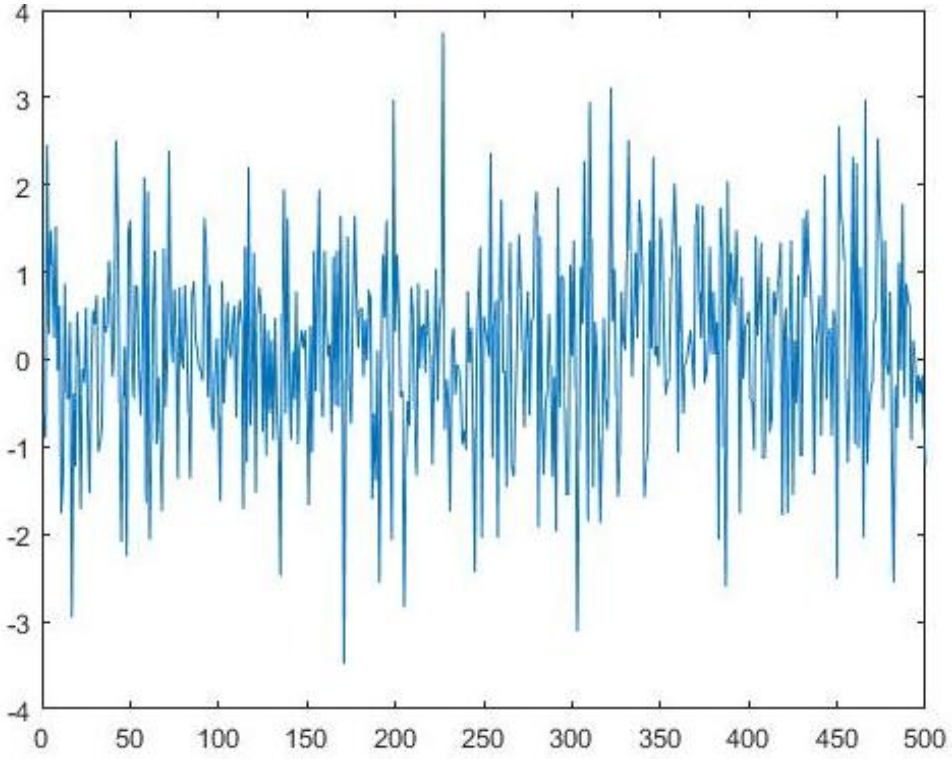


Рисунок 3.12- объем наблюдений: 500.

Таблица 3.4-коэффициенты α и β для $\theta = 250$.

Teta 250		
alfa	**beta**	**момент разладки**
0,2	0,8	221
0,3	0,4	221
0,5	0,7	285
0,6	0,7	244

ВЫВОД

В работе проведено сравнительное численное моделирование двух алгоритмов обнаружения "разладки".

1. При численном исследовании алгоритма, основанного на отношении правдоподобия, установлено, что в модели авторегрессии AR(1) при разнице в коэффициентах больше, чем 0,3 качество обнаружения "разладки" на модельных данных хорошее. А при разнице в коэффициентах меньше, чем 0,3 велика вероятность ложной тревоги (100%), время ложного спокойствия при этом нулевое.
   В работе моделировался временной ряд объемом 2000 и 1500 наблюдений. Точный момент разладки (t=1500 и t=1000 соответственно) определялся заранее. При разнице в первом и втором коэффициенте 0,3 ; 0,4 ($|\phi_0 - \phi_1| = 0.3; 0.4$; $|\delta_0 - \delta_1| = 0.3; 0.4$) "разладка" определялась с запаздыванием, то есть момент, когда произошло изменение в первом коэффициенте определялся как t=1504,1504 наблюдение; во втором коэффициенте: t=1505, 1003 наблюдение соответственно. Так же, моделировалась разница 0,5 для первого коэффициента ($|\delta_0 - \delta_1| = 0.5$) и 0,6 для второго коэффициента ($|\phi_0 - \phi_1| = 0.6$). Определяемый момент "разладки": t=1499 и t=1003 соответственно. Таким образом, можно сказать, что "разладка" при большой разнице в коэффициентах преимущественно определяется с запаздыванием (вероятность ложного спокойствия 0,99) , а при небольшой разнице происходит ложная тревога задолго до истинного момента, когда происходит изменение процесса (вероятность ложной тревоги 1).
2. При численном исследовании алгоритма CuSum моделирование проводилось на моделях двух типов:

1. Описывающих появление детерминированного сигнала на фоне независимых шумов;

2. Появление детерминированного сигнала на фоне зависимых шумов.

Численно установлено, что при хорошем подборе коэффициентов α и β (α - вероятность ложной тревоги, а β - вероятность ложного спокойствия), момент "разладки" определяется близко к истинному значению. Моделировалось два момента "разладки": $\theta = 100$ и $\theta = 250$. Для $\theta = 100$ при α и β (таблица 3.1) "разладка" определяется практически точно. Полученные результаты отклоняются от точного не более чем на два наблюдения. Ложная тревога наблюдалась в 30 % случаях, а ложное спокойствие в 20% случаях. При других же значениях α и β "разладка" определяется преимущественно много позже (то есть наблюдается ложное спокойствие в 99%). Для $\theta = 250$ (при α и β из таблицы 3.2) вероятность ложной тревоги равна 0,34, а вероятность ложного спокойствия 0,22. Построенная процедура определила "разладку" точно в 44 % случаях. Для второй модели для $\theta = 100$ при хорошем подборе α и β (таблица 3.3) вероятность ложной тревоги равна 0,37, а вероятность ложного спокойствия - 0,17. При других значениях α и β вероятность ложной тревоги равна 0,8, а вероятность ложного спокойствия - 0,2.Для $\theta = 250$ (при α и β из таблицы 3.4) Вероятность ложной тревоги 0,75, а вероятность ложного спокойствия - 0,2. При других значениях α и β вероятность ложного спокойствия равна 0,61, а ложной тревоги-0,39 .

Таким образом, при удачном подборе коэффициентов α и β момент "разладки" определяется точно, что обеспечивает своевременный ремонт оборудования и качественную работу всего механизма, будь это производство или другие сферы жизни современного человека.

ЗАКЛЮЧЕНИЕ

В настоящей работе были рассмотрены различные подходы к задаче об обнаружении "разладки".

В первой главе момент "разладке" определялся как момент выхода за границы 95% доверительного интервала, построенного для числа людей в зоне наблюдения. Вероятность однократного выхода за границы такого интервала равна 0,05. Вероятность двукратного и трехкратного выхода не более 0,0025 и 0,000125 соответственно. Таким образом, при двукратном выходе за границы доверительного интервала вероятность разладки фиксирующего оборудования 0,9975.

Вторая глава посвящена сравнительному численному исследованию двух алгоритмов обнаружения "разладки". Первый алгоритм, который был рассмотрен, основан на отношении правдоподобия. Проводя симуляционное моделирование, было выявлено, что при разнице в коэффициентах авторегрессионной модели AR(1) $|\phi_0 - \phi_1| > 0.3$ и $|\delta_0 - \delta_1| > 0.3$ алгоритм хорошо находит "разладку". Но, например, в промышленности, при производстве деталей ошибка даже на 0,001 может оказать значительное влияние на работу всего механизма. Поэтому этот алгоритм хорошо работает только в том случае, когда разница в коэффициентах больше 0,3. Второй алгоритм - алгоритм CuSum. Этот алгоритм разработан для более широкого класса моделей, включающих в том числе и модель авторегрессии AR(1), поэтому симуляционное моделирование было проведено на моделях , описывающих появление детерминированного сигнала на фоне независимых шумов и появление детерминированного сигнала на фоне зависимых шумов. То есть, были рассмотрены модели, в которых до момента "разладки" был шум, а после - появлялся какой-то сигнал. Этот алгоритм практически точно определяет, когда появляется сигнал при хорошем подборе коэффициентов α и β, где

α - вероятность ложной тревоги, а β - вероятность ложного спокойствия. Соответственно, лучше определяет "разладку" второй алгоритм, чем алгоритм, основанный на отношении правдоподобия.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Ивченко Г.И., Медведев Ю.И. Математическая статистика -Москва: Высшая школа, 1984. С.81-84, С.89-92.
2. Исаева Н.А., Кривякова Э.Н. Оценивание параметров распределения: метод. пособие. -Томск-1990. С.9-10.
3. Конев В. В. Экспериментальное исследование непараметрической процедуры CUSUM для обнаружения разладки в авторегрессионных моделях с неизвестным конечным параметром / В. В. Конев, Б. Н. Назаренко // Материалы IV Международной молодежной научной конференции "Математическое и программное обеспечение информационных, технических и экономических систем", Томск, 20-21 мая 2016 г. Томск, 2016. С.66-67.
4. Конев В.В. Последовательные оценки параметров стохастических динамических систем - Томск: Томский университет, 1985. С. 134-151.
5. Конев В.В., Назаренко Б.Н. Экспериментальное исследование непараметрической процедуры CUSUM в неасимптотической постановке //Труды Томского государственного университета. Серия физико-математическая. Томск: Издательский Дом ТГУ, 2015. Т. 297. Математическое и программное обеспечение информационных, технических и экономических систем : материалы III Всероссийской молодежной научной конференции. Томск, 22–23 мая 2015 г.С.176-178.
6. Крамер Г. Математические методы статистики- Москва: Мир, 1975. С. 533 п.3.
7. Липцер И.Ш., Ширяев А.Н. Статистика случайных процессов - Москва: Наука, 1974.
8. Ширяев А.Н. Задача скорейшего обнаружения стационарного режима, ДАН СССР 138, 5 (1961), 1039-1042.
9. Ширяев А.Н. Некоторые точные формулы в задаче о "разладке", Теория вероятностей и ее применение 10, 2 (1965), 380-385.
10. Ширяев А.Н. О двух задачах последовательного анализа, Кибернетика 2, 1967, 79-80.
11. Ширяев А.Н. Об оптимальных методах в задачах скорейшего обнаружения, Теория вероятностей и ее применение 8, 1, 1963, 26-51.
12. Ширяев А.Н. Обнаружение спонтанно возникающих эффектов, ДАН СССР 138, 4 (1961), 794-801.
13. Ширяев А.Н. Статистический последовательный анализ - Моксва: Наука, 1976, С. 206.

14. Ширяев А.Н., О марковских достаточных статистиках в неаддитивных баессовских задачах последовательного анализа, Теория вероятностей и ее применение 9,4 (1964), 670-686.
15. Davis R., Huang D., Yao Y. Testing for a Change in the Parameter Values and R. Davis, D. Huang, and Y. Yao. Testing for a Change in the Parameter Values and Order of an Autoregressive Model. The Annals of Statistics, 1995
16. Fang Yuan. Tests for Change-Point of the AR(1) Model/ Fang Yuan , Kang Ling James - The UNIVERSITY OF MINNESOTA, 2013, p 5-21.

ПРИЛОЖЕНИЕ А

Данные представлены в течении недели по времени суток каждые 3 часа.

Зона 1

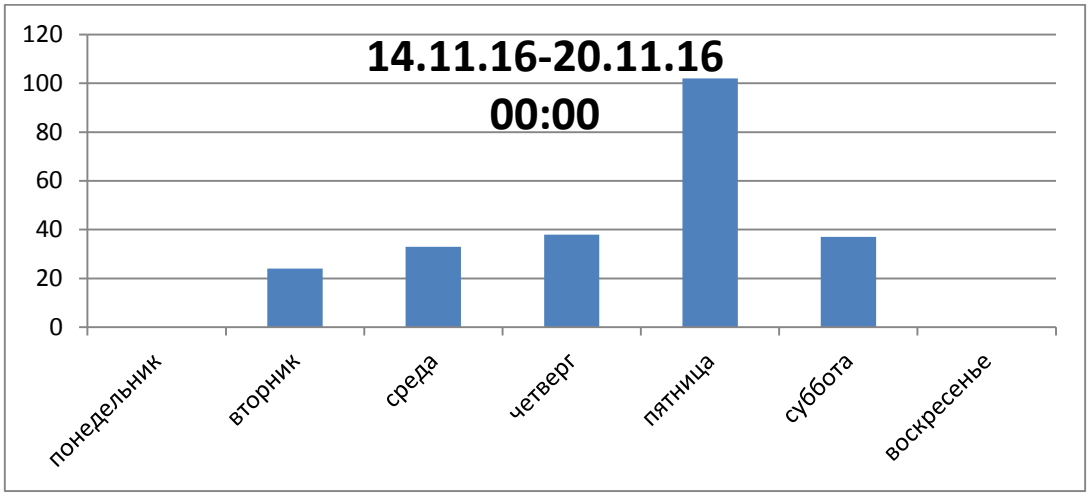


Рисунок 1 - Количество человек. 14.11.16.-20.11.16. Время 00:00.

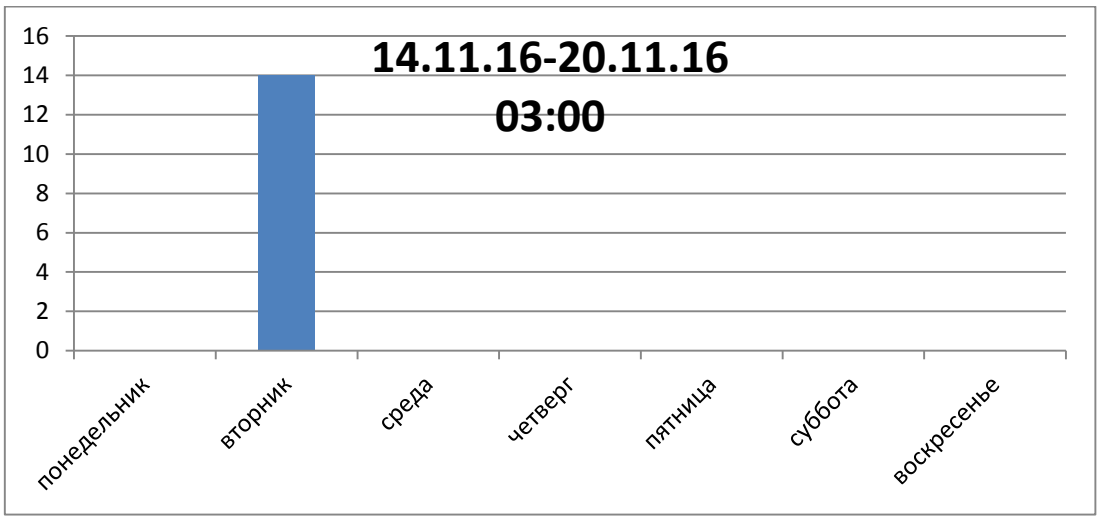


Рисунок 2 - Количество человек. 14.11.16.-20.11.16. Время 03:00.

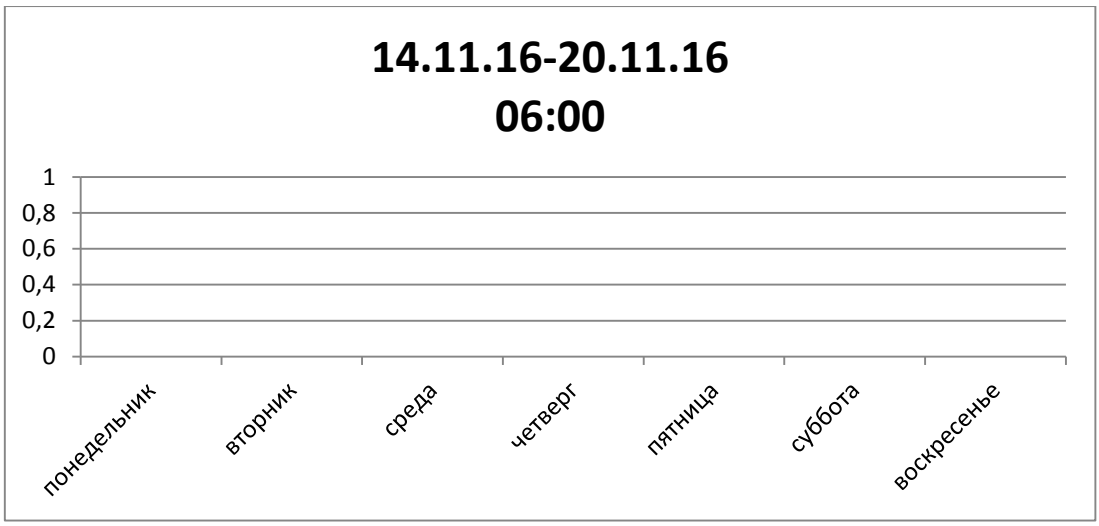


Рисунок 3 - Количество человек. 14.11.16.-20.11.16. Время 06:00.

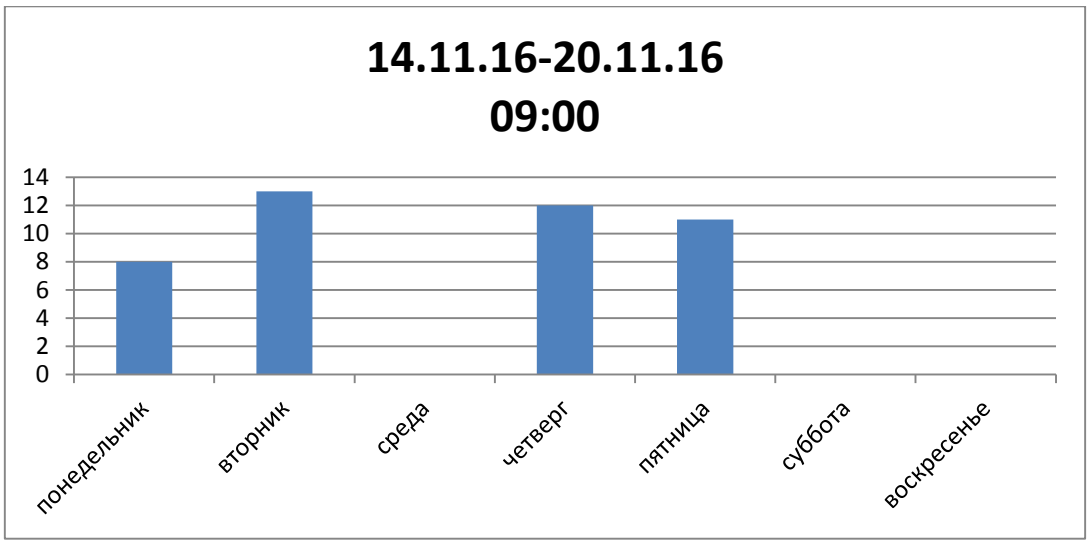


Рисунок 4 - Количество человек. 14.11.16.-20.11.16. Время 09:00.

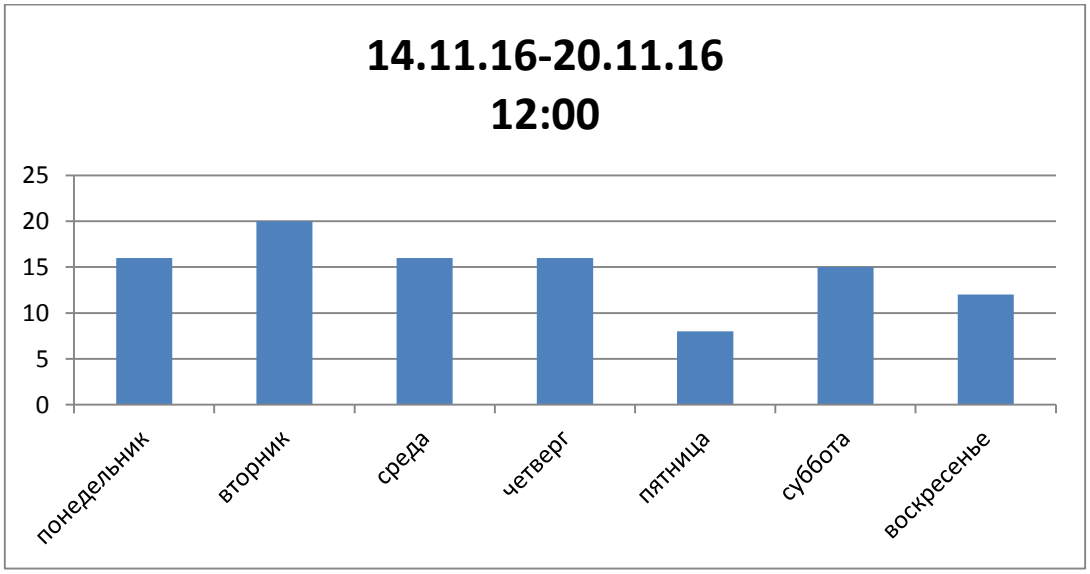


Рисунок 5 - Количество человек. 14.11.16.-20.11.16. Время 12:00.

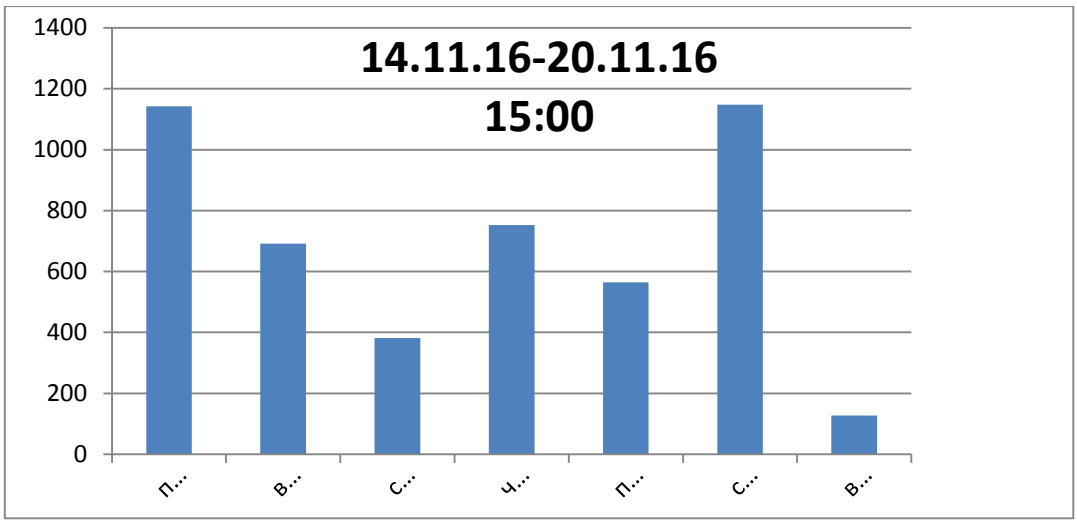


Рисунок 6 - Количество человек. 14.11.16.-20.11.16. Время 15:00.

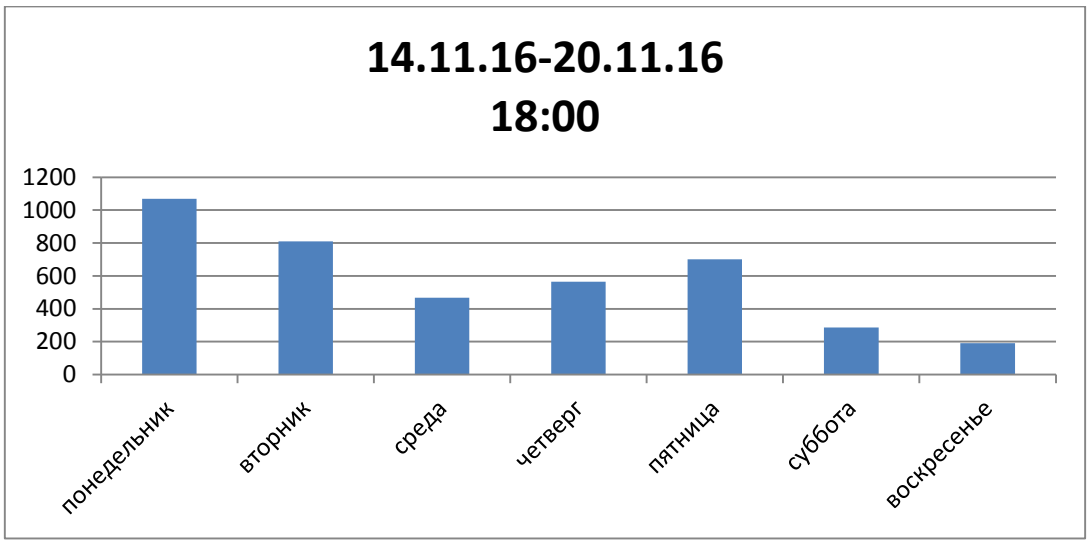


Рисунок 7 - Количество человек. 14.11.16.-20.11.16. Время 18:00.

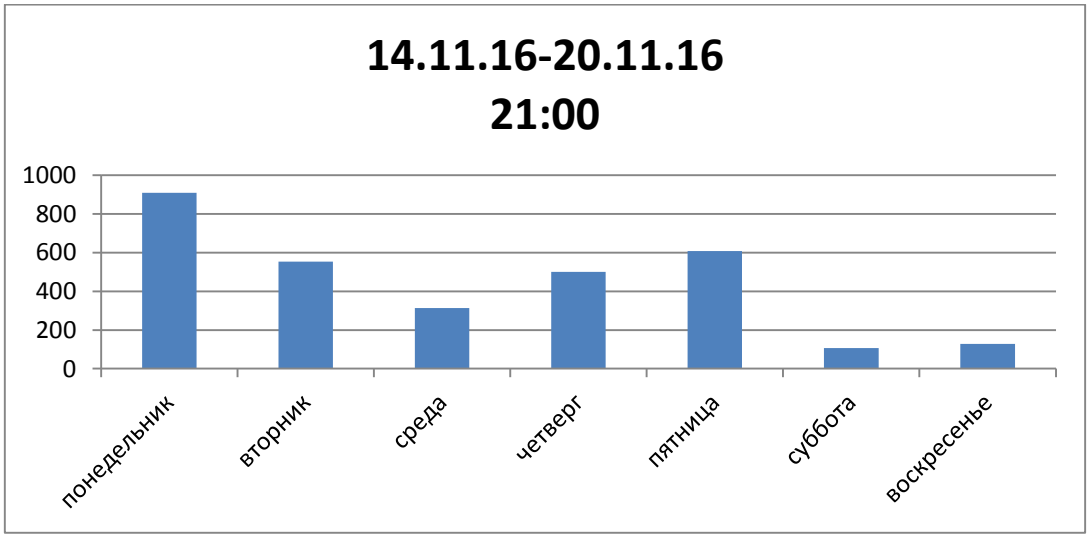


Рисунок 8 - Количество человек. 14.11.16.-20.11.16. Время 21:00.

Зона 2.

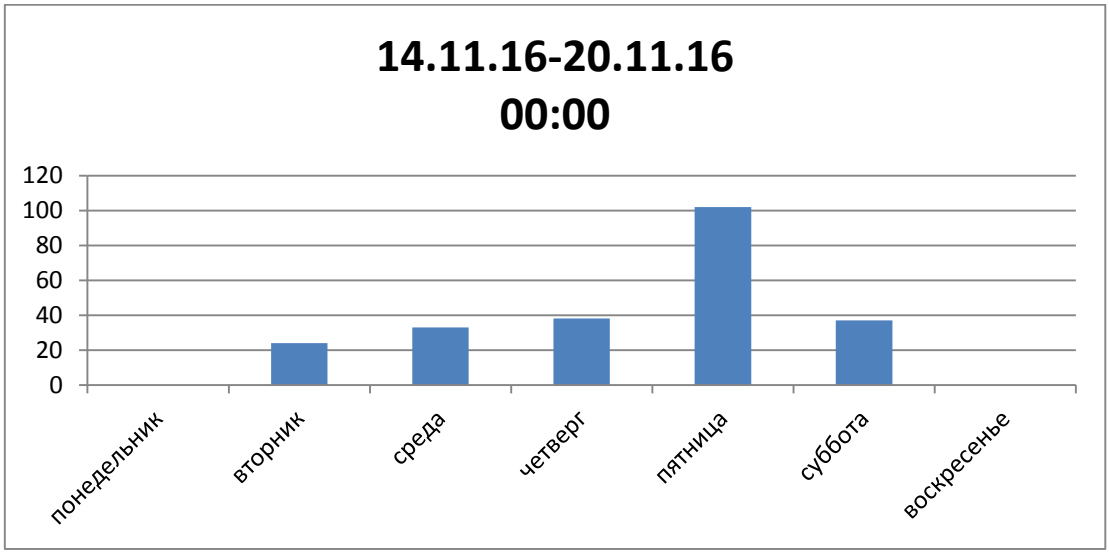


Рисунок 9 - Количество человек. 14.11.16.-20.11.16. Время 00:00.

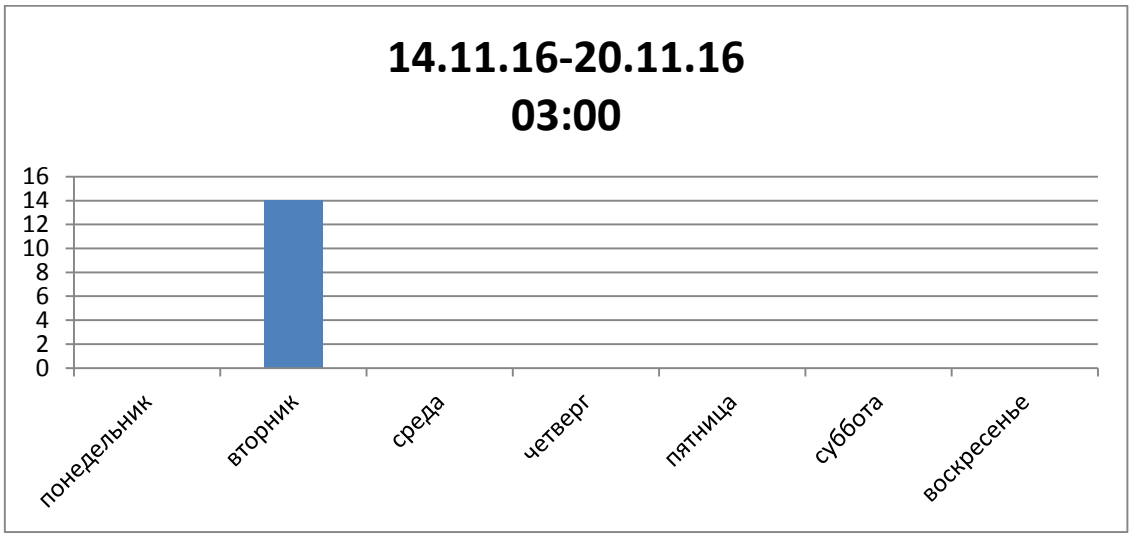


Рисунок 10 - Количество человек. 14.11.16.-20.11.16. Время 03:00.

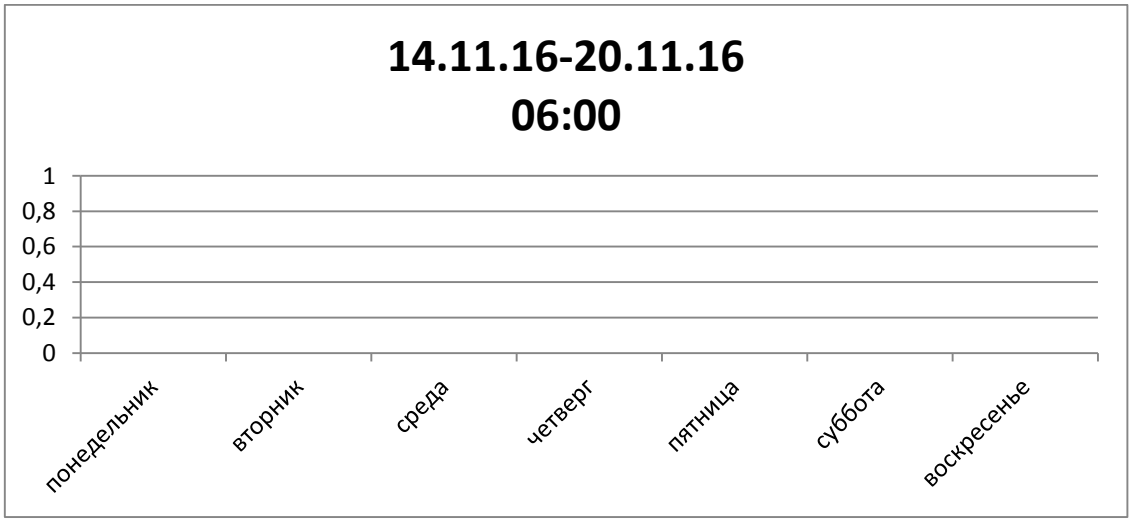


Рисунок 11 - Количество человек. 14.11.16.-20.11.16. Время 06:00.

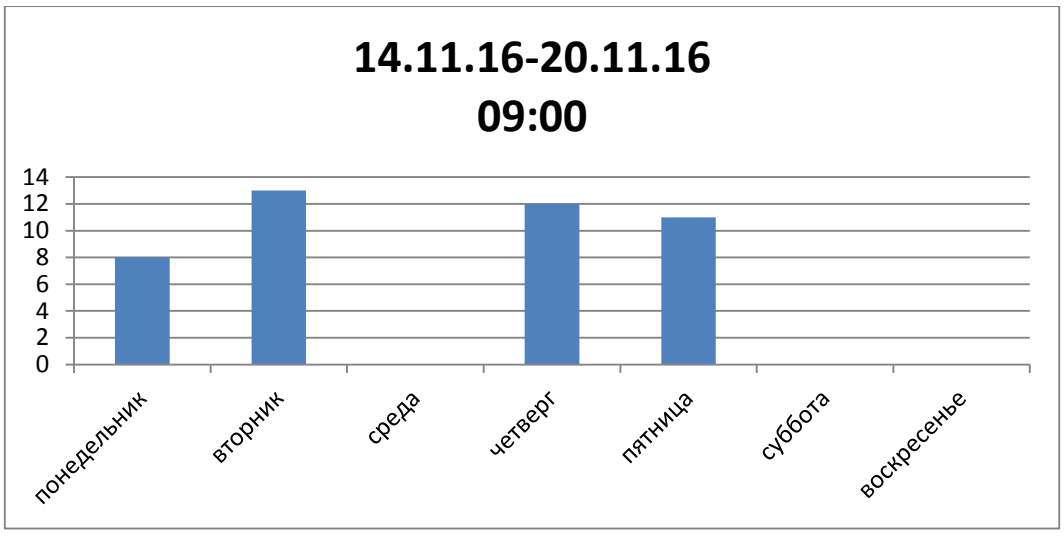


Рисунок 12 - Количество человек. 14.11.16.-20.11.16. Время 09:00.

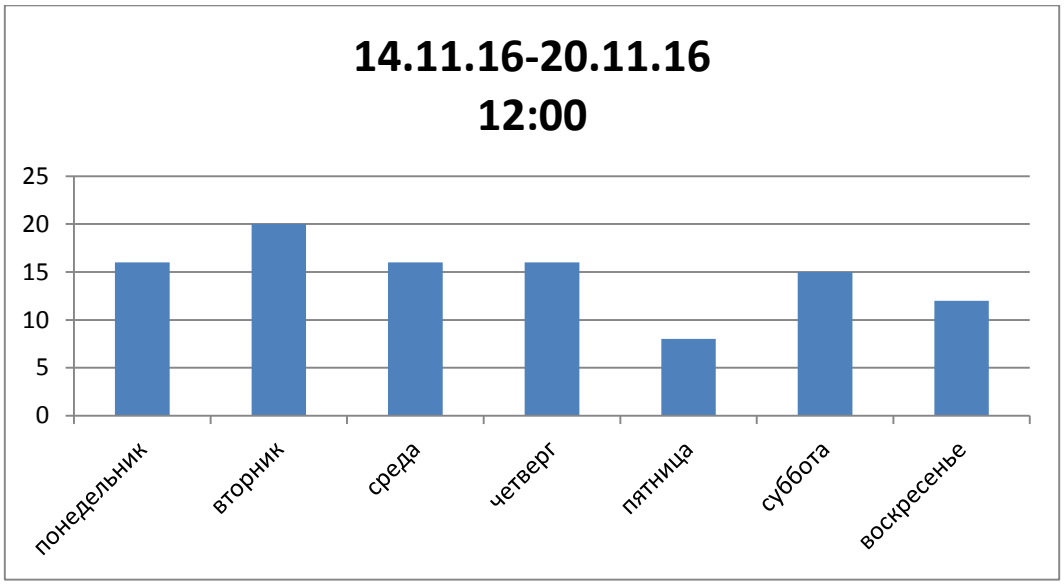


Рисунок 13 - Количество человек. 14.11.16.-20.11.16. Время 12:00.

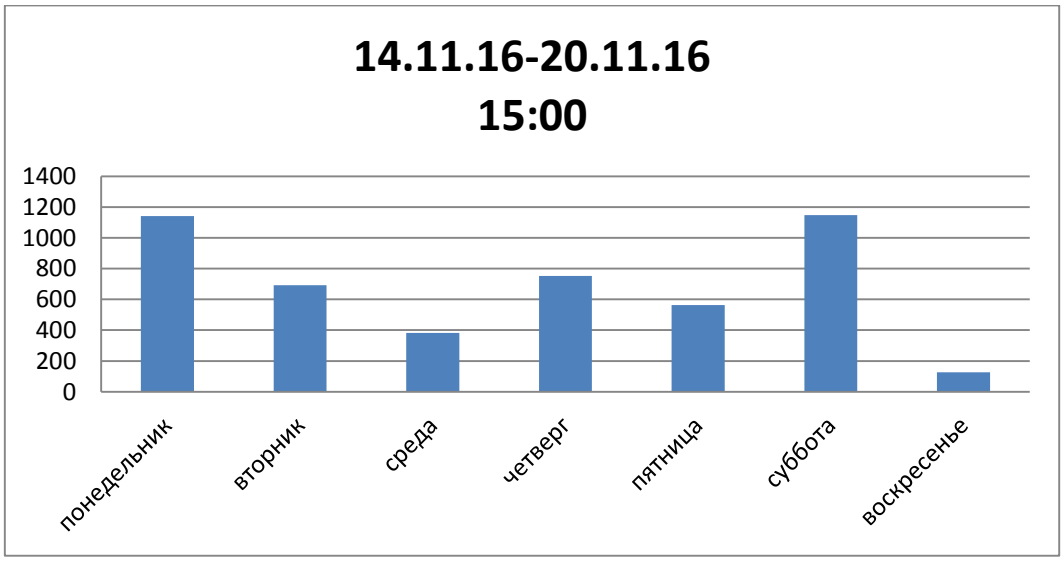


Рисунок 14 - Количество человек. 14.11.16.-20.11.16. Время 15:00.

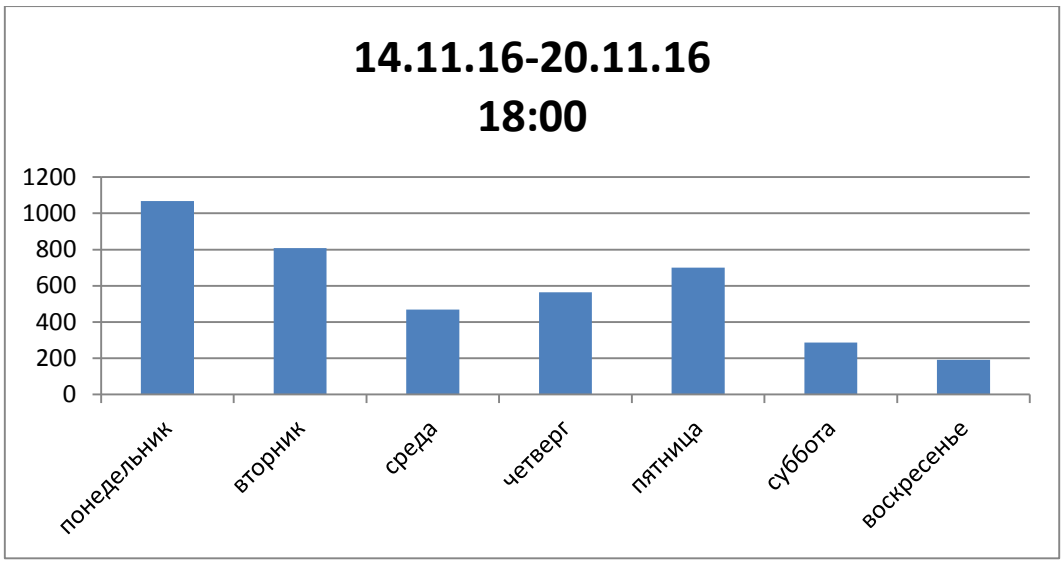


Рисунок 15 - Количество человек. 14.11.16.-20.11.16. Время 18:00.

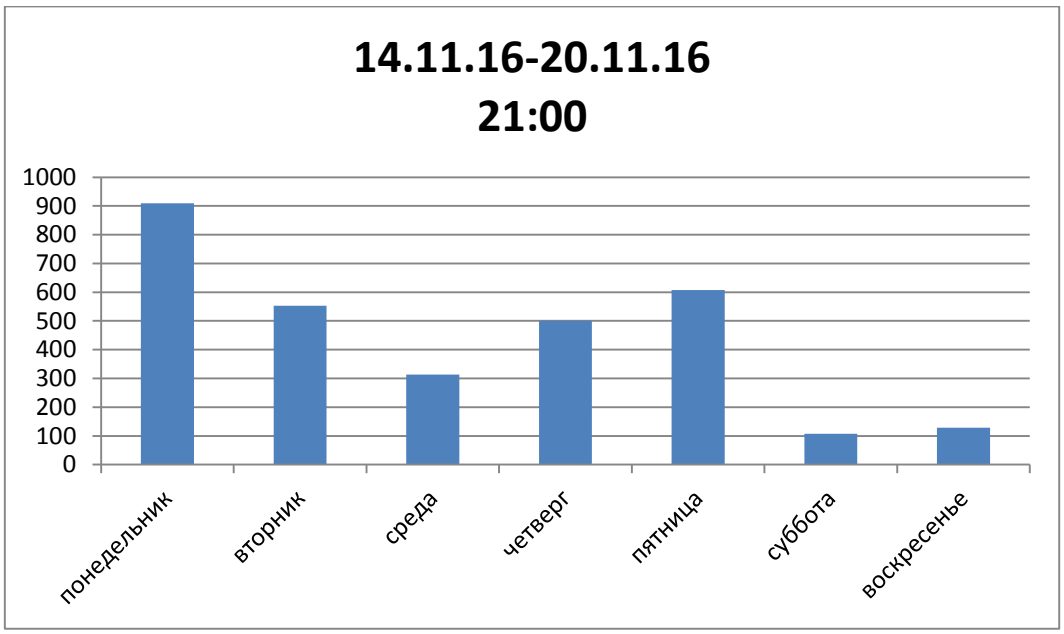


Рисунок 16 - Количество человек. 14.11.16.-20.11.16. Время 21:00.

(/index.php/ru/) (/index.php/en/)

Exactus Like (/index.php)

Поиск заимствований в научных текстах$^{\beta}$

Введите текст:

...или загрузите файл:

Файл не выбран... **Выбрать файл...**

Укажите год публикации: 2018

Выберите коллекции

- **Все**
- **Рефераты** **Википедия** **Российские журналы**
- **Авторефераты** **Российские конференции** **Энциклопедии**
- **Иностранные конференции** **Иностранные журналы** **Англоязычная википедия**
- **PubMed**

Анализировать

Проверить по расширенному списку коллекций системы Руконтекст (http://text.rucont.ru/like)

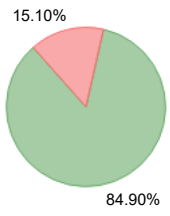
Обработан файл:
BKP.pdf.

Год публикации: 2018.

Оценка оригинальности документа - 84.9%

Процент условно корректных заимствований - 0.0%

Процент некорректных заимствований - 15.1%



Просмотр заимствований в документе

Время выполнения: 21 с.

Документы из базы

Источники заимствования	В списке литературы	Источники Заимствования
1. Последовательный метод обнаружения разладок случайных процессов рекуррентного типа (http://www.mathnet.ru/links/a7e25f80147a10c5098c8af4cc7ed302/at4706.pdf) Авторы: С. Э. Воробейчиков , В. В. Конев. Год публикации: 1984. Тип публикации: статья научного журнала. http://www.mathnet.ru/links/a7e25f80147a10c5098c8af4cc7ed302/at4706.pdf (http://www.mathnet.ru/links/a7e25f80147a10c5098c8af4cc7ed302/at4706.pdf) Показать заимствования (30)		11.3%

2. Проблемы и методы вероятностной диагностики (http://www.mathnet.ru/links/f3177a92f42b3a7db41b25e04660396b/at131.pdf)

Авторы: Б. Е. Бродский a , Б. С. Дарховский b.
Год публикации: 1999. Тип публикации: статья научного журнала.
http://www.mathnet.ru/links/f3177a92f42b3a7db41b25e04660396b/at131.pdf (http://www.mathnet.ru/links/f3177a92f42b3a7db41b25e04660396b/at131.pdf)
Показать заимствования (6)

1.62%

3. Об обнаружении разладок в динамических системах (http://www.mathnet.ru/links/e1a622ac3e1b41da3b6068dc034da645/at5324.pdf)

Авторы: С. Э. Воробейчиков , В. В. Конев.
Год публикации: 1990. Тип публикации: статья научного журнала.
http://www.mathnet.ru/links/e1a622ac3e1b41da3b6068dc034da645/at5324.pdf (http://www.mathnet.ru/links/e1a622ac3e1b41da3b6068dc034da645/at5324.pdf)
Показать заимствования (5)

1.35%

4. Об оптимальных методах в задачах скорейшего обнаружения (http://www.mathnet.ru/links/f096f47542333b4611b2b883334efa0e/tvp4645.pdf)

Авторы: А. Н. Ширяев.
Год публикации: 1963. Тип публикации: статья научного журнала.
http://www.mathnet.ru/links/f096f47542333b4611b2b883334efa0e/tvp4645.pdf (http://www.mathnet.ru/links/f096f47542333b4611b2b883334efa0e/tvp4645.pdf)
Показать заимствования (3)

0.97%

5. Обнаружение разладок и оценивание параметров авторегрессионных процессов по зашумленным наблюдениям (http://dlib.rsl.ru/loader/view/01000280091?get=pdf)

Авторы: Буркатовская, Юлия Борисовна.
Год публикации: 2000. Тип публикации: автореферат диссертации.
http://dlib.rsl.ru/loader/view/01000280091?get=pdf (http://dlib.rsl.ru/loader/view/01000280091?get=pdf)
Показать заимствования (5)

0.97%

6. Гарантированная оценка параметров и обнаружение момента разладки GARCH(1,1)-процесса (http://cyberleninka.ru/article/n/garantirovannaya-otsenka-parametrov-i-obnaruzhenie-momenta-razladki-garch-1-1-protsessa)

Авторы: Сергеева Екатерина Евгеньевна, Воробейчиков Сергей Эрикович.
Год публикации: 2011. Тип публикации: статья научного журнала.
http://cyberleninka.ru/article/n/garantirovannaya-otsenka-parametrov-i-obnaruzhenie-momenta-razladki-garch-1-1-protsessa (http://cyberleninka.ru/article/n/garantirovannaya-otsenka-parametrov-i-obnaruzhenie-momenta-razladki-garch-1-1-protsessa)
Показать заимствования (4)

0.82%

7. Непараметрическая сегментация электрических сигналов мозга (http://www.mathnet.ru/links/11d8f07b7644ec970186f552bce7c445/at2320.pdf)

Авторы: Б. Е. Бродский a , Б. С. Дарховский a , А. Я. Каплан b , С. Л. Шишкин b.
Год публикации: 1998. Тип публикации: статья научного журнала.
http://www.mathnet.ru/links/11d8f07b7644ec970186f552bce7c445/at2320.pdf (http://www.mathnet.ru/links/11d8f07b7644ec970186f552bce7c445/at2320.pdf)
Показать заимствования (3)

0.81%

8. Последовательные процедуры обнаружения момента разладки случайных процессов авторегрессионного типа с условной неоднородностью (http://dlib.rsl.ru/loader/view/01005018856?get=pdf)

Авторы: Сергеева, Екатерина Евгеньевна.
Год публикации: 2012. Тип публикации: автореферат диссертации.
http://dlib.rsl.ru/loader/view/01005018856?get=pdf (http://dlib.rsl.ru/loader/view/01005018856?get=pdf)
Показать заимствования (3)

0.79%

9. Обнаружение разладки процесса авторегрессии, наблюдаемого с помехами (http://www.mathnet.ru/links/665708114eb5b04a34ac057c6bbdce11/at250.pdf)

Авторы: Ю. Б. Буркатовская , С. Э. Воробейчиков.
Год публикации: 2000. Тип публикации: статья научного журнала.
http://www.mathnet.ru/links/665708114eb5b04a34ac057c6bbdce11/at250.pdf (http://www.mathnet.ru/links/665708114eb5b04a34ac057c6bbdce11/at250.pdf)
Показать заимствования (3)

0.57%

Дополнительно

Значимые оригинальные фрагменты

Библиографические ссылки

Искать в Интернете

© 2015 2018 Институт системного анализа Российской академии наук (http://www.isa.ru/index.php?lang=ru)