МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МАТЕРИАЛЫ

X-й Международной молодёжной научной конференции

«МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННЫХ, ТЕХНИЧЕСКИХ И ЭКОНОМИЧЕСКИХ СИСТЕМ»

Томск, 26–29 мая 2023 г.

Под общей редакцией
кандидата технических наук И.С. Шмырина

Томск
Издательство Томского государственного университета
2023

ББК 22.17– 22.19
УДК 519.2, 519.7, 519.8
Т78

ЧЛЕНЫ КОЛЛЕГИИ, РУКОВОДИТЕЛИ НАУЧНЫХ РЕДАКЦИЙ
ПО НАПРАВЛЕНИЯМ:

д-р техн. наук, проф. **А.А. Глазунов** – научная редакция «Механика, математика»; д-р техн. наук, проф. **Э.Р. Шрагер** – научная редакция «Механика, математика»; д-р техн. наук, проф. **А.М. Горцев** – научная редакция «Информатика и кибернетика»; д-р техн. наук, проф. **С.П. Сущенко** – научная редакция «Информатика и кибернетика»; д-р физ.-мат. наук, проф. **В.Г. Багров** – научная редакция «Физика»; д-р физ.-мат. наук, проф. **А.И. Потекаев** – научная редакция «Физика»; д-р биол. наук, проф. **С.П. Кулижский** – научная редакция «Биология»; д-р геол.-минер. наук, проф. **В.П. Парначев** – научная редакция «Науки о Земле, химия»; канд. хим. наук, доц. **Ю.Г. Слижов** – научная редакция «Науки о Земле, химия»; д-р филол. наук, проф. **Т.А. Демешкина** – научная редакция «История, филология»; д-р ист. наук, проф. **В.П. Зиновьев** – научная редакция «История, филология»; д-р экон. наук, проф. **В.А. Уткин** – научная редакция «Юридические и экономические науки»; д-р ист. наук, проф. **Э.И. Черняк** – научная редакция «Философия, социология, психология, педагогика, искусствознание»; д-р психол. наук, проф. **Э.В. Галажинский** – научная редакция «Философия, социология, психология, педагогика, искусствознание»

НАУЧНАЯ РЕДАКЦИЯ ТОМА:

д-р техн. наук, проф. **А.М. Горцев**, д-р физ.-мат. наук, доц. **Ю.Г. Дмитриев**, д-р техн. наук, проф. **А.В. Замятин**, д-р физ.-мат. наук, доц. **А.Н. Моисеев**, д-р физ.-мат. наук, проф. **С.П. Моисеева**, д-р физ.-мат. наук, доц. **Л.А. Нежельская**, канд. техн. наук, доц. **С.А. Останин**, д-р техн. наук, проф. **С.П. Сущенко**, канд. техн. наук **И.С. Шмырин**.

Т78 Труды Томского государственного университета. – Т. 308. Серия физико-математическая: Математическое и программное обеспечение информационных, технических и экономических систем : материалы X-й Международной молодёжной научной конференции. Томск, 26–29 мая 2023 г. / под общ. ред. И.С. Шмырина. – Томск : Издательство Томского государственного университета, 2023. – 197 с.

ISBN 978-5-907572-24-9

Сборник содержит материалы X-й Международной молодёжной научной конференции «Математическое и программное обеспечение информационных, технических и экономических систем», проводившейся 26–29 мая 2023 г. на базе Института прикладной математики и компьютерных наук Томского государственного университета. Материалы сгруппированы в соответствии с работавшими на конференции секциями.

Для научных работников, преподавателей, аспирантов, магистрантов и студентов.

УДК 539.3.004
ББК 22.17-22.19

ISBN 978-5-907572-24-9
© Томский государственный университет, 2023

IV. МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ ЦИФРОВОЙ ЭКОНОМИКИ

ОТСЛЕЖИВАНИЕ ТОЧЕК РАЗЛАДКИ АВТОРЕГРЕССИОННОГО СИГНАЛА

Пупков А.В.
Томский государственный университет
andrewpupkov@gmail.com

Введение

При создании моделей реальных процессов крайне важно учитывать возможность дестабилизации объекта и его переход к нетипичному поведению. В общем случае это означает изменение распределения процесса в некоторый неизвестный момент времени, именуемый разладкой. Есть два основных подхода к определению момента разладки. С точки зрения первого, момент разладки представляет из себя неизвестный, но детерминированный момент времени. Во втором случае предполагается, что момент отклонения есть случайная величина с некоторым распределением. В рамках данной работы будем считать, что момент отклонения детерминирован.

Для отслеживания разладки разработано множество процедур, таких как контрольные карты Шухарта, основанные на совокупностях выборки фиксированной длины, контрольные карты с экспоненциальным взвешенным скользящим средним (EWMA), процедуры кумулятивного типа, которые представляют из себя всевозможные модификации CUSUM, а также процедура Ширяева – Робертса [1].

В рамках работы представлена модификация процедуры кумулятивных сумм для отслеживания разладки авторегрессионного сигнала. В качестве составной части базовых статистик алгоритма выступают неасимптотические оценки параметров процесса авторегрессии [2]. Для работоспособности алгоритма, в качестве априорного предположения, вводится нижняя граница для нормы разницы параметров процесса до и после момента разладки. В зависимости от данного предположения получена достаточная нижняя граница для параметра, отвечающего за объем выборки неасимтотических оценок, гарантирующая отрицательность математического ожидания базовой статистики до разладки и положительность – после, что приводит к возможности использования базовых статистик в процедуре кумулятивных сумм.

Стоит отметить, что в [3] была предложена процедура отслеживания разладки, основанная на похожих базовых предположениях.

Работа включает следующие разделы. В разделе 1 приводится постановка задачи и вводится исследуемый процесс. В разделе 2 отражена процедура построения оценки дисперсии шума и приведены свойства построенной оценки. В разделе 3 приводится процедура взвешенного оценивания параметра авторегрессии, позволяющая построить оценку с неасимптотическим свойством ограниченности среднеквадратической ошибки. В разделе 4 вводятся базовые статистики процедуры и выводятся их свойства. В заключительном разделе 5 определяется модифицированная процедура кумулятивных сумм.

1. Постановка задачи

Пусть процесс AR(p) с разладкой в момент ν описывается схемой вида

$$x_k = X'_{k-1}\theta_1 + b_1\varepsilon_k,\ k = \overline{1, \nu - 1},$$

$$x_k = X'_{k-1}\theta_2 + b_2\varepsilon_k,\ k \geq \nu,$$

где $X_k = \left(x_k, \ldots, x_{k-p+1}\right)'$, $\theta_i = \left(\theta_i^{(1)}, \ldots, \theta_i^{(p)}\right)'$, $i = 1,2$ – векторы неизвестных параметров до и после разладки, при условии $\theta_1 \neq \theta_2$ (штрих (') означает транспонирование), b_i^2, $i = 1,2$ – неизвестные дисперсии шума процесса до и после разладки соответственно, $\{\varepsilon_k\}$ – последовательность независимых случайных величин с нулевым математическим ожиданием $\mathbb{E}\varepsilon_k = 0$ и единичной дисперсией $\mathbb{E}\varepsilon_k^2 = 1$. Процесс $\{\varepsilon_k\}$ не зависит от начальных значений $x_0, x_{-1}, \ldots, x_{1-p}$. Требуется оценить момент ν по наблюдениям процесса $\{x_k\}$ при условии, что θ_i и b_i, $i = 1,2$ неизвестны. Основная идея состоит в разделении процесса на интервалы случайной длины $\left(\tau_{i-1}(h), \tau_i(h)\right]$, $i > 0$, на каждом из которых происходит оценивание параметра $\theta \in \{\theta_1, \theta_2\}$ с гарантированным качеством. Затем вычисляется норма разницы между оценками, далее на норму накладываются дополнительные преобразования, и полученная статистика используется как базовая в алгоритме кумулятивных сумм.

Предполагается известной следующая априорная информация о процессе:

1. Плотность распределения $f(x)$ случайных величин ε_k имеет симметричное и унимодальное распределение.
2. Известна константа $\|L\| > 0$: $\|\theta_2 - \theta_1\| \geq \|L\|$, где $\|x\|$ – евклидова норма вектора x.
3. Для гарантированного оценивания необходимо чтобы был известен порядок авторегрессии p.
4. Вводится положительная константа $C_{\text{var}} > 0$ такая, что $\mathbb{E}\left(\sum_{k=1}^{n_2} \varepsilon_k^2\right)^{-1} \leq C_{\text{var}}$, $n_2 > 0$.
5. Момент остановки $\tau_1(h)$, отвечающий за объем выборки первой неасимптотической оценки параметра авторегрессии, меньше точки разладки ν, т.е. $\tau_1(h) < \nu$ (последовательность моментов остановки $\{\tau_i(h)\}$ определим ниже).

В следующем разделе рассмотрим процедуру оценивания дисперсии шума процесса $b^2 \in \{b_1^2, b_2^2\}$.

2. Оценивание дисперсии шума процесса

Рассмотрим задачу построения неасимптотической оценки параметра $\theta \in \{\theta_1, \theta_2\}$ на некотором интервале $\left(\tau_{i-1}(h), \tau_i(h)\right]$, $i > 0$. Предполагаем, что данный интервал лежит полностью либо в доразладочной, либо в постразладочной части процесса. Для решения задачи сначала необходимо оценить дисперсию шума $b^2 \in \{b_1^2, b_2^2\}$. Для этой цели введем пилотную оценку параметра θ по фиксированному объему выборки

$$\tilde{\theta}_i(n_1) = \left(\sum_{k=\tau_{i-1}+1}^{\tau_{i-1}^*} X_{k-1} X_{k-1}'\right)^{-1} \sum_{k=\tau_{i-1}+1}^{\tau_{i-1}^*} X_{k-1} x_k,$$

где $\tau_{i-1}^* = \tau_{i-1} + n_1$, n_1 – объем выборки для оценивания параметра $\theta \in \{\theta_1, \theta_2\}$. Затем, используя оценку $\tilde{\theta}_i(n_1)$, построим оценку дисперсии шума по схеме $\Gamma_{n_1,n_2}^{(i)} = C_{\text{var}} S_{n_1,n_2}^{(i)}$, где n_2 – объем выборки для непосредственного оценивания дисперсии и

$$S_{n_1,n_2}^{(i)} = \sum_{k=\tau_{i-1}^*+1}^{\tau_{i-1}^{**}} \left(x_k - X_{k-1}' \tilde{\theta}_i(n_1)\right)^2, \quad C_{\text{var}} \geq \mathbb{E}\left(\sum_{k=n_1+1}^{n_1+n_2} \varepsilon_k^2\right)^{-1}, \quad \tau_{i-1}^{**} = \tau_{i-1}^*(h) + n_2.$$

По аналогии с результатом, представленным в [4], можем получить верхнюю границу верхнюю границу для величины $\mathbb{E}_j\left(1/\Gamma_{n_1,n_2}^{(i)}\right) \leq 1/b_j^2$, $j = 1,2$, где $\mathbb{E}_1$ соответствует

математическому ожиданию доразладочного распределения процесса, $\mathbb{E}_2$ – постразладочному распределению.

В следующем разделе приведем основные формулы для вычисления взвешенной оценки МНК параметра $\theta \in \{\theta_1, \theta_2\}$.

3. Взвешенная оценка наименьших квадратов

Взвешенная оценка МНК вычисляется на основе выборки случайного объема, величина которого находится в соответствии с правилом $\tau_i(h) = \inf\{n > \tau_{i-1}^{**} : \lambda_n^{\min}(i) \geq h\}$, где $h > 0$ – параметр, задаваемый исследователем и отвечающий за объем выборки, $\lambda_n^{\min}(i)$ – минимальное собственное значение матрицы $A_n = \sum_{k=\tau_{i-1}^{**}(h)+1}^{n} \nu_{k-1} X_{k-1} X'_{k-1}$; здесь $\{\nu_{k-1}\}$ – система весовых коэффициентов, начальные значения которых рассчитываются по схеме

$$\nu_k = \begin{cases} \dfrac{1}{\sqrt{\Gamma_{n_1,n_2}^{(i)} X'_k X_k}}, \text{ если } X_k \text{ линейно не зависит от } X_{\tau_{i-1}^{**}}, \ldots, X_{k-1}; \\ 0, \text{ иначе,} \end{cases}$$

до первого момента n_3, определяемого из условия $n_3 = \inf\{n \geq \tau_{i-1}^{**} + 1 : \det A_n \neq 0\}$. После момента n_3, коэффициент ν_{n-1} отыскивается как решение уравнения

$$\frac{\lambda_n^{\min}(i)}{\Gamma_{n_1,n_2}^{(i)}} = \sum_{k=n_3+1}^{n} \nu_{k-1}^2 X'_{k-1} X_{k-1}, \; n > n_3.$$

Отыскать значение ν_{n-1} можно, используя метод деления отрезка пополам, поскольку величина $\lambda_n^{\min}(i)$ также зависит от коэффициента ν_{n-1}. Стоит отметить, что коэффициент $\nu_{n-1} \in \left(0, 1/\Gamma_{n_1,n_2}^{(i)}\right)$. На последнем шаге, в момент остановки $\tau_i(h)$, весовой коэффициент $\nu_{\tau_i(h)-1}$ уменьшается до тех пор, пока не будет выполняться условие $|\lambda_{\tau_i(h)}^{\min}(i) - h| < \delta$, где $\delta > 0$ – точность вычисления.

Взвешенная оценка МНК вычисляется по формуле

$$\hat{\theta}_i(h) = \left(A_{\tau(h)}^{(i)}\right)^{-1} \sum_{k=\tau_{i-1}^{**}+1}^{\tau_i(h)} \nu_{k-1} X_{k-1} x_k, \; A_{\tau(h)}^{(i)} = \sum_{k=\tau_{i-1}^{**}+1}^{\tau_i(h)} \nu_{k-1} X_{k-1} X'_{k-1}.$$

В [2] получен следующий результат: $\mathbb{E}_j \left\| \hat{\theta}_i(h) - \theta_j \right\|^2 \leq \dfrac{h+p-1}{h^2}$, $j = 1,2$, который является основой для алгоритма отслеживания разладки.

4. Свойства базовых статистик

В качестве базовых статистик процедуры кумулятивных сумм будем рассматривать следующую последовательность

$$Z_k = \left\| \hat{\theta}_k(h) - \hat{\theta}_1(h) \right\| - 2h^{-1}\sqrt{h+p-1}, \; k > 1, \tag{1}$$

при условии, что параметр h, отвечающий за объем выборки, будет ограничен снизу величиной

$$h > 4\|L\|^{-2}\left(2 + \sqrt{4 + \|L\|^2 (p-1)}\right), \tag{2}$$

где $\hat{\theta}_i(h)$ взвешенная оценка МНК параметра θ на интервале $\left(\tau_{i-1}(h),\tau_i(h)\right]$. Введем обозначение ξ_i – уклонение оценки на интервале $\left(\tau_{i-1},\tau_i\right]$, определяемое выражением

$$\xi_i=\hat{\theta}_i(h)-\theta_j=A^{-1}_{\tau_i(h)}\sum_{s=\tau_{i-1}(h)+1}^{\tau_i(h)}\nu_{s-1}X_{s-1}\varepsilon_s b_j,\ j=1,2.$$

Среднее базовых статистик до разладки ограничено сверху нулем, т.е. $\mathbb{E}_1 Z_k<0$. Покажем это, используя неравенство Йенсена и свойства норм:

$$\mathbb{E}_1\left\|\hat{\theta}_k(h)-\hat{\theta}_1(h)\right\|=\mathbb{E}\|\xi_k-\xi_1\|\le\mathbb{E}\|\xi_k\|+\mathbb{E}\|\xi_1\|\le\sqrt{\mathbb{E}\|\xi_k\|^2}+\sqrt{\mathbb{E}\|\xi_1\|^2}\le 2\frac{\sqrt{h+p-1}}{h}.$$

Далее, вычитая из последнего неравенства величину $2h^{-1}\sqrt{h+p-1}$, приходим к доказываемому утверждению.

В случае, когда интервал $\left(\tau_{k-1}(h),\tau_k(h)\right]$ полностью лежит в постразладочной области, т.е. $\nu\le\tau_{k-1}(h)$, имеем

$$\mathbb{E}_2\left\|\hat{\theta}_k(h)-\hat{\theta}_1(h)\right\|=\mathbb{E}\|\theta_2-\theta_1+\xi_k(h)-\xi_1(h)\|\ge\|\theta_2-\theta_1\|-\mathbb{E}\|\xi_k-\xi_1\|\ge$$

$$\ge\|L\|-\left(\mathbb{E}\|\xi_k\|+\mathbb{E}\|\xi_1\|\right)\ge\|L\|-\left(\sqrt{\mathbb{E}\|\xi_k\|^2}+\sqrt{\mathbb{E}\|\xi_1\|^2}\right)\ge\|L\|-2h^{-1}\sqrt{h+p-1}.$$

Вычитая величину $2h^{-1}\sqrt{h+p-1}$, получаем нижнюю границу для математического ожидания базовой статистики после разладки в виде $\mathbb{E}_2 Z_k\ge\|L\|-4h^{-1}\sqrt{h+p-1}$. Далее, накладывая на полученную нижнюю границу условия неотрицательности и разрешая полученное квадратное неравенство относительно h, получаем условие (2).

В следующем разделе приведем модифицированную процедуру кумулятивных сумм.

5. Модифицированная процедура CUSUM

Возьмем за основу классическую процедуру кумулятивных сумм, но заменим базовые статистики, в качестве которых обычно используют отношение логарифмического правдоподобия, на статистики (1). В результате процедура отслеживания разладки примет вид $N^*(h)=\inf\{n>\tau_1(h):T_n(h)>C\}$, где $T_n(h)=\max\left(0,T_{n-1}(h)+Z_n(h)\right)$, а $C>0$ – некоторый порог, задаваемый исследователем.

Заключение

В работе представлена модификация процедуры кумулятивных сумм для отслеживания момента изменения авторегрессионного параметра θ. Использование оценок параметров для отслеживания разладки является перспективным направлением, поскольку требует минимальный объем информации о процессе, но, с другой стороны, нуждается в выборках большого объема. Дальнейшее развитие построенной процедуры может подразумевать расширение класса моделей, в которых необходимо отслеживать разладку.

ЛИТЕРАТУРА

1. *Tartakovsky A., Nikiforov I., Basseville M.* Sequential Analysis: Hypothesis Testing and Changepoint Detection. – Chapman and Hall, 2015. – 579 p.

2. *Meder N., Vorobejchikov S.* On guaranteed estimation of parameters of random processes by the weighted least squares method // IFAC Proceedings Volumes. – 2002. – V. 35. – N. 1. – P. 409–412.

3. *Vorobeychikov S., Burkatovskaya Yu.* Change point detection of autoregressive process with unknown parameters // IFAC Proceedings Volumes. – 2011. – V. 44. – N. 1. – P. 13215–13220.

4. *Vorobeychikov S., Burkatovskaya Yu.* Non-asymptotic Confidence Estimation of the Autoregressive Parameter in AR(1) Process with an Unknown Noise Variance // Austrian Journal of Statistics. – 2020. — V. 49. – N. 4. – P. 19–26.