



МИНОБРНАУКИ РФ
Российский фонд
фундаментальных исследований
Национальный исследовательский
Томский государственный университет
НИИ прикладной математики и механики
Томского государственного университета
Физико-технический факультет
Совет молодых учёных ТГУ



**VIII Всероссийская молодежная научная конференция
«Актуальные проблемы современной механики
сплошных сред и небесной механики – 2018»
г. Томск, 26–28 ноября 2018 г.**

**VIII All-Russian Scientific Conference
«Current issues of
continuum mechanics and celestial mechanics – 2018»,
November, 26–28, 2018**

Томск-2019

ПОСЛЕДОВАТЕЛЬНОЕ ОБНАРУЖЕНИЕ РАЗЛАДКИ ВО ВРЕМЕННЫХ РЯДАХ

А.О. Шерстобитова, Т.В. Емельянова

Национальный исследовательский Томский государственный университет, Россия
annasherstobitova06@gmail.com

Ключевые слова: модели временных рядов, разладка, механизм генерации.

Аннотация. Рассматривается проблема разделения временных рядов произвольной природы на сегменты, порожденные одним механизмом, а также обнаружения момента смены одного механизма генерации другим.

SEQUENTIAL DETECTION OF CHANGEPOINTS IN TIME SERIES

A.O. Sherstobitova, T.V. Emelyanova

National Research Tomsk State University, Russian Federation
annasherstobitova06@gmail.com

Keywords: time series models, changepoint, generation mechanism.

Abstract. We consider the problem of separating time series of arbitrary nature into segments generated by one mechanism, as well as detecting the moment of changing one generation mechanism by another.

В анализе данных актуальной задачей является обработка неоднородных временных рядов, порожденных различными механизмами генерации. В конце XX в. была высказана идея о том, что задача обнаружения изменений в любой вероятностной характеристике случайного процесса может быть сведена к задаче обнаружения изменений математического ожидания в специальном диагностическом процессе, построенном на основе исходного ряда данных [1]. Эта идея позволяет обнаружить изменение любых вероятностных характеристик, а также осуществлять сегментацию временного ряда. В работе рассматривается проблема сегментации временных рядов произвольной природы на однородные сегменты. Кроме того, решается задача обнаружения точного момента смены одного механизма генерации другим на примере модели ARCH(2).

Пусть $X = \{x(t)\}_{t=1}^N$ – временной ряд с неизвестными моментами смены механизма генерации $t_i, i = 2, \dots, k$. Механизмы генерации ряда неизвестны.

Отрезки ряда $[t_i, t_{i+1}]$, $t_1 = 1, t_{k+1} = N$, которые генерируются одним и тем же механизмом, называем однородными. Б. С. Дарховским была сформулирована и доказана теорема [2], согласно которой связь ε -сложности функции

гельдерова класса, заданной конечным набором своих отсчетов на равномерной решетке с погрешностью аппроксимации ε характеризуется парой действительных чисел (A, B) , называемых коэффициентами сложности [2].

Ключевая идея предлагаемой методологии заключается в предположении о том, что на i -м сегменте $[t_i, t_{i+1}]$ временного ряда X_t для $t_i \leq t, t+n < t_{i+1}$ коэффициенты сложности удовлетворяют соотношению

$$R(j) = R_i + \xi^i(j), \quad (1)$$

где R_i – математическое ожидание последовательности $R(j)$ на отрезке $[t_i, t_{i+1}]$, а $\xi^i(j)$ – последовательность случайных величин.

На практике часто необходимо знать номер наблюдения, на котором происходит смена механизма генерации либо фиксируется аномальное поведение данных, не характерное для наблюдаемого ряда данных. В связи с этим предлагается модификация метода сегментации Б.С. Дарховского, которая позволяет получить конкретный номер наблюдения, с зафиксированной аномалией. В каждом эксперименте генерируется ряд данных, затем выполняется разделение полученного ряда на «окна» размера n . Каждое «окно» обрабатывается методом наименьших квадратов и определяются коэффициенты зависимости (A, B) . Далее с помощью скользящего окна выполняется обнаружение момента изменения механизма генерации («разладки») ряда. При помощи алгоритма, основанного на статистике

$$Y(s, \delta) = \left(\frac{(N_1 - s)s}{N_1^2} \right)^\delta \left(\frac{1}{s} \sum_{k=s+1}^{N_1} z(k) - \frac{1}{N_1 - s} \sum_{k=s+1}^{N_1} z(k) \right), \quad (2)$$

где $0 \leq \delta \leq 1, 1 \leq s \leq N_1 - 1, N_1 = \left[\frac{N}{n} \right], Z = \{z(k)\}_{k=1}^{N_1}$ – реализация компонент диагностической последовательности R [1], определяются моменты «разладки».

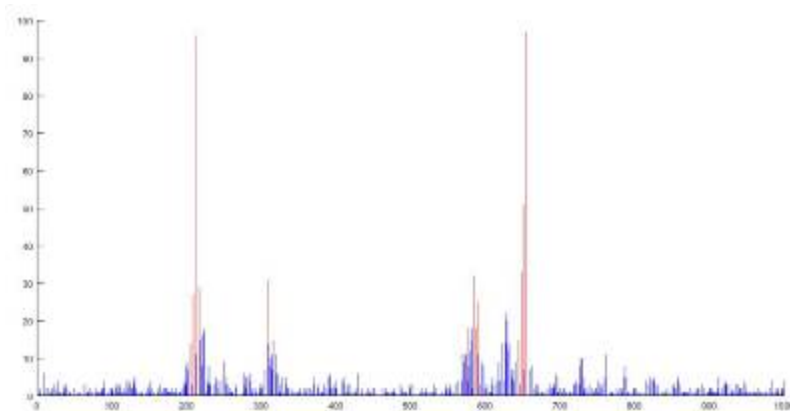


Рис. 1. Результаты численного моделирования: обнаружение точного момента смены механизма генерации для модели ARCH (2)

В работе проведено имитационное моделирование, результаты которого показали, что алгоритм сегментации Б.С. Дарховского позволяет не только выполнить сегментацию ряда данных на подмассивы, порожденные разными механизмами генерации, но и выполнить обнаружение аномалий в ряде данных и получить точное значение момента смены одного механизма генерации другим.

Литература

1. *Бродский Б.Е., Дарховский Б.С.* Непараметрическая статистическая диагностика: проблемы и методы. Dordrecht: Kluwer, 2000.
2. *Дарховский Б.С., Пирятинская А.* Новый подход к проблеме сегментации временных рядов произвольной природы // Труды института математики им. Стеклова, № 287. С. 54–67.

References

1. *Brodsky B.E., Darkhovsky B.S.* Non-parametric statistical diagnosis: Problems and methods. Dordrecht: Kluwer, 2000.
2. *Darkhovsky B.S., Piryatinska A.* New approach to the segmentation problem for time series of arbitrary nature: Proceedings of the Steklov Institute of Mathematics. 2014. Vol. 287. Pp. 54–67.