

РАЗДЕЛЕНИЕ ТРЕНДОВОЙ И ХАОТИЧЕСКОЙ КОМПОНЕНТ ВРЕМЕННЫХ РЯДОВ И ОЦЕНКА ИХ ХАРАКТЕРИСТИК С ПОМОЩЬЮ ЛИНЕЙНЫХ СПЛАЙНОВ

*А. В. Крянев^{а, б, 1}, В. В. Иванов^{а, б}, А. О. Романова^а,
Л. А. Севастьянов^{б, в}, Д. К. Удудьян^{а, в, з}*

^а Национальный исследовательский ядерный университет «МИФИ», Москва

^б Объединенный институт ядерных исследований, Дубна

^в Российский университет дружбы народов, Москва

^з Университет Майами, Корал-Гейблс, Флорида, США

В статье рассматривается проблема разделения трендовой и хаотической компонент хаотических временных рядов в условиях отсутствия информации о характеристиках хаотической компоненты. Такого рода проблема возникает при решении задач в ядерной физике, биомедицине и многих других прикладных задач. Схема состоит из двух этапов. На первом этапе, используя сглаживающие линейные сплайны при различных значениях параметра сглаживания α , выделяют трендовую компоненту. На втором этапе с помощью метода наименьших квадратов находят неизвестную дисперсию шумовой компоненты σ^2 .

The paper deals with the problem of dividing the trend and the chaotic component of chaotic time series in the absence of information about the characteristics of the chaotic component. This kind of problem arises in solving problems in nuclear physics, biomedicine, and many other applied problems. The scheme consists of two stages. At the first stage, using smoothing linear splines for different values of the smoothing parameter, the “trend component” is highlighted. At the second stage, using the method of least squares, an unknown dispersion σ^2 of the noise component is found.

PACS: 02.50.Fz

ВВЕДЕНИЕ

Одной из основных проблем обработки данных во многих областях является проблема разделения трендовой и хаотической (шумовой) компонент хаотических временных рядов в условиях, когда характеристики компонент неизвестны. К настоящему времени разработано много различных методов и схем, которые решают различные частные задачи разделения трендовой и хаотической компонент временных рядов [1–6]. Для получения оценки дисперсии шумовой компоненты хаотического временного ряда

$$Y_1, Y_2, \dots, Y_n \quad (1)$$

¹E-mail: avkryanev@mephi.ru

необходимо выделить из исходного хаотического временного ряда (1) трендовую и шумовую компоненты:

$$Y_k = Y_{\text{trk}} + \varepsilon_k, \quad k = 1, \dots, n, \quad (2)$$

где Y_{trk} — значение трендовой компоненты; ε_k — значение шумовой компоненты.

Для разделения трендовой и шумовой компонент можно использовать ортогональные полиномы, линейные и кубические сплайны и другие базовые системы функций или такие схемы, как сингулярно-спектральный анализ, вейвлет-анализ, метрический анализ. Во всех этих методах и схемах присутствует параметр, определяющий уровень сглаживания при выделении тренда. Основной проблемой является выбор оптимального значения этого параметра в условиях отсутствия информации о характеристиках трендовой и шумовой компонент.

В представленной работе нами использовалась простая, но надежная схема разделения трендовой и хаотической компонент, основанная на сглаживающих линейных сплайнах [1].

1. СХЕМА РАЗДЕЛЕНИЯ ТРЕНДОВОЙ И ХАОТИЧЕСКОЙ КОМПОНЕНТ

В схеме разделения трендовой и хаотической компонент, представленной в данной работе, используются сглаживающие линейные сплайны.

Значения выделяемого тренда $Y_{\text{tr}}(\alpha, t_k)$ — сглаживающего линейного сплайна — являются решением следующей экстремальной задачи:

$$\alpha \sum_{k=1}^n (Y_k - Y_{\text{tr}}(\alpha, t_k))^2 + \sum_{k=2}^n (Y_{\text{tr}}(\alpha, t_k) - Y_{\text{tr}}(\alpha, t_{k-1}))^2, \quad (3)$$

где $\alpha > 0$ — параметр антисглаживания.

При $\alpha = 0$ значения трендовой компоненты $Y_{\text{tr}}(\alpha, t_k)$ являются константой, определяемой равенством $Y_{\text{tr}}(\alpha, t_k) = \frac{1}{n} \sum_{k=1}^n Y_k$, т.е. производится максимально возможное сглаживание. И, наоборот, при $\alpha \rightarrow \infty$ значения трендовой компоненты стремятся к значениям Y_k , т.е. отсутствует сглаживание (отсутствует выделение тренда) исходного временного ряда (1).

Для нахождения оптимального значения α_{i_0} , соответствующего искомым выделяемым трендовой и шумовой компонентам (2), в данной работе использована схема поиска значения α_{i_0} , соответствующего переходу от сильного к слабому изменению зависимости суммы квадратов невязок:

$$Y(\alpha) = \sigma^2(\alpha) = \sum_{k=1}^n (Y_k - Y_{\text{tr}}(\alpha, t_k))^2 \quad (4)$$

при различных значениях α . В этой схеме зависимость $Y(\alpha) = \sigma^2(\alpha)$ аппроксимировалась слева от α_{i_0} параболой, а справа — прямой.

Затем подсчитывалась сумма квадратов невязки между аппроксимационными значениями и значениями $\sigma^2(\alpha)$ в узлах сетки по α . Разделение трендовой и хаотической компонент производилось при значении $\alpha_{i_0} = \alpha_{\text{opt}}$, при котором сумма квадратов невязки между аппроксимационными значениями и значениями $\sigma^2(\alpha)$ достигала минимума.

Далее использовалось относительное отклонение σ_{model} от σ_{real} :

$$dS = \frac{|\sigma_{\text{real}} - \sigma_{\text{model}}|}{\sigma_{\text{real}}}. \quad (5)$$

Здесь σ_{model} — восстановленное среднеквадратическое отклонение; σ_{real} — среднеквадратическое отклонение шумовой компоненты, распределенной по нормальному закону.

Для проверки эффективности предложенной схемы оценки дисперсии шумовой компоненты проведено тестирование схемы на трех примерах с различными значениями дисперсии. В каждом из примеров на исходную функцию с помощью датчика случайных величин накладывалась (добавлялась) шумовая компонента с заданным значением дисперсии, а затем с помощью предложенной схемы выделялась трендовая компонента.

Исследовалась точность полученной оценки в зависимости от частоты сетки по $\alpha \in [0, 1]$ (количество точек равномерной сетки N).

Проведено сравнение точности полученных оценок дисперсии шумовой компоненты.

2. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Пример 1:

$$y = 2t + 1 + 5 \sin t + N(0, \sigma_{\text{real}}), \quad t = [0 : 0,05 : 20], \quad (6)$$

где $N(0, \sigma_{\text{real}})$ — нормально распределенная шумовая компонента с нулевым математическим ожиданием и среднеквадратическим отклонением σ_{real} ; 0,05 — временной шаг.

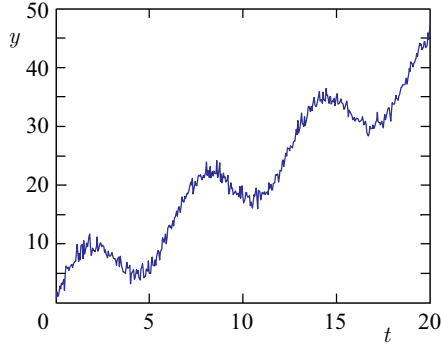


Рис. 1

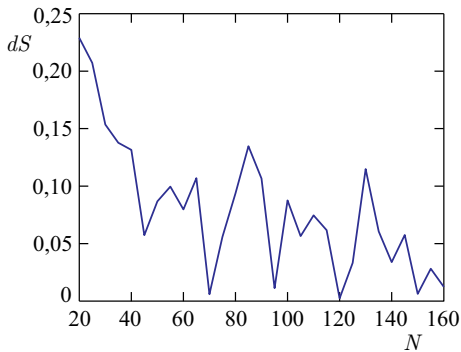


Рис. 2

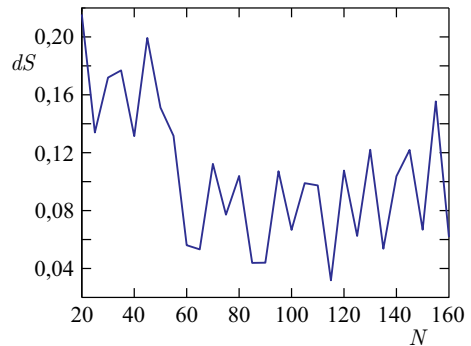


Рис. 3

График хаотического временного ряда (6) для $\sigma_{\text{real}} = 0,8$ приведен на рис. 1. На рис. 2, 3 представлена зависимость dS (ось ординат) от числа точек сетки N (ось абсцисс) по α от 20 до 150 для $\sigma_{\text{real}} = 0,8$ и $\sigma_{\text{real}} = 2,0$ соответственно.

Были получены следующие минимальные значения относительного отклонения σ_{model} от σ_{real} :

$dS_{\text{min}} = 0,0023$ — минимальное значение относительного отклонения (при $N_{\text{opt}} = 120$ — число точек сетки по α , при котором достигается минимум dS) (см. рис. 2);

$dS_{\text{min}} = 0,0318$ — значение относительного отклонения (при $N_{\text{opt}} = 115$) (см. рис. 3).

Пример 2:

$$y = t^2 + 5t + 7 + N(0, \sigma_{\text{real}}), \quad t = [-5 : 0,05 : 5]. \quad (7)$$

График хаотического временного ряда (7) для $\sigma_{\text{real}} = 0,8$ приведен на рис. 4. На рис. 5, 6 представлена зависимость dS (ось ординат) от числа точек сетки N (ось абсцисс) по α от 20 до 150 для $\sigma_{\text{real}} = 0,8$ и $\sigma_{\text{real}} = 2,0$ соответственно.

Были получены следующие минимальные значения относительного отклонения σ_{model} от σ_{real} :

$dS_{\text{min}} = 0,0073$, $N_{\text{opt}} = 85$ (см. рис. 5);

$dS_{\text{min}} = 0,0067$, $N_{\text{opt}} = 145$ (см. рис. 6).

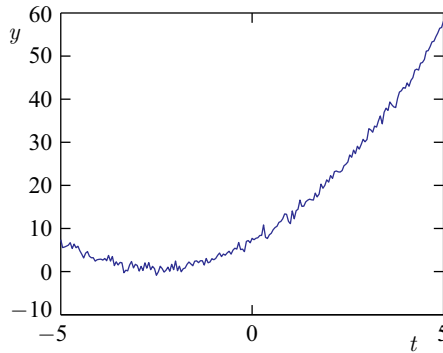


Рис. 4

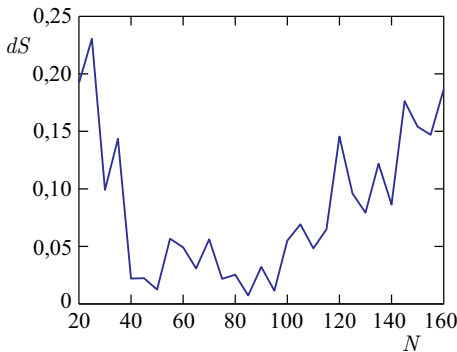


Рис. 5

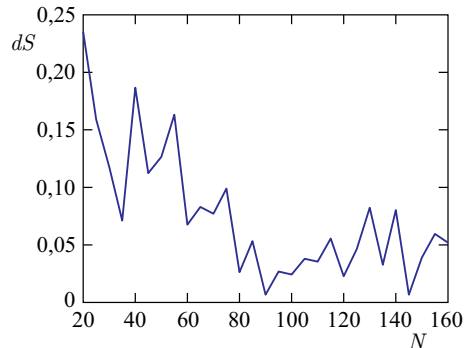


Рис. 6

Пример 3:

$$y = \sqrt{t} + N(0, \sigma_{\text{real}}), \quad t = [0 : 0,05 : 10]. \quad (8)$$

График хаотического временного ряда (8) для $\sigma_{\text{real}} = 0,3$ приведен на рис. 7. На рис. 8, 9 представлена зависимость dS (ось ординат) от числа точек сетки N (ось абсцисс) по α от 20 до 150 для $\sigma_{\text{real}} = 0,05$ и $\sigma_{\text{real}} = 0,3$ соответственно.

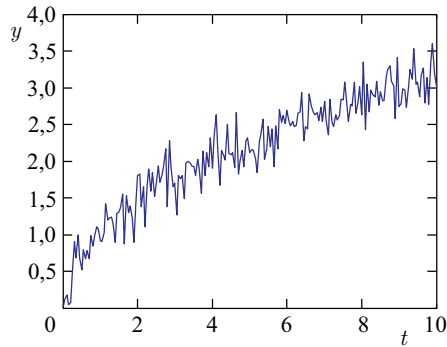


Рис. 7

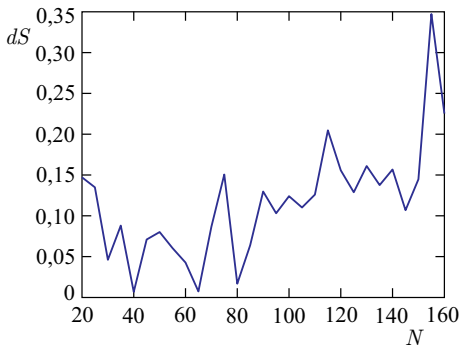


Рис. 8

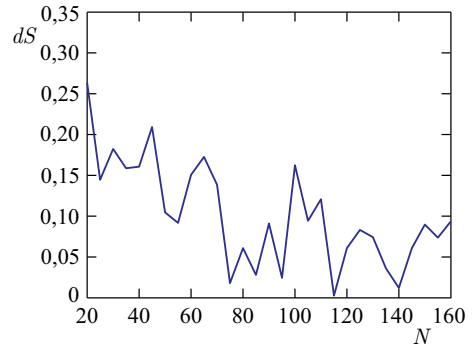


Рис. 9

Были получены следующие минимальные значения относительного отклонения σ_{model} от σ_{real} :

$$dS_{\text{min}} = 0,0062, \quad N_{\text{opt}} = 45 \text{ (см. рис. 8);}$$

$$dS_{\text{min}} = 0,0028, \quad N_{\text{opt}} = 115 \text{ (см. рис. 9).}$$

ЗАКЛЮЧЕНИЕ

В статье рассматривается проблема разделения трендовой и хаотической компонент хаотических временных рядов в условиях отсутствия информации об уровне вариации шумовой компоненты. Схема разделения состоит из двух этапов. На первом этапе, используя сглаживающие линейные сплайны при различных значениях параметра антисглаживания α , выделяют трендовую компоненту и подсчитывают сумму квадратов невязки между значениями исходного временного ряда и значениями сглаживающего линейного сплайна, получая зависимость суммы квадратов невязки $\sigma^2(\alpha)$. На втором этапе

с помощью метода наименьших квадратов аппроксимируют $\sigma^2(\alpha)$ для значений левее фиксированного значения α_0 полиномом второй степени, а правее α_0 — полиномом первой степени и подсчитывают сумму квадратов невязки между аппроксимационными значениями и значениями $\sigma^2(\alpha)$ в узлах сетки по α . Разделение трендовой и хаотической компонент производится при $\alpha_0 = \alpha_{\text{opt}}$, при котором сумма квадратов невязки между аппроксимационными значениями и значениями $\sigma^2(\alpha)$ достигает минимума. Представлены примеры разделения трендовой и хаотической компонент хаотических временных рядов, доказывающие эффективность рассматриваемой схемы. Результаты тестирования показали, что разработанная схема позволяет с высокой точностью устойчиво оценивать значения дисперсии шумовой компоненты исходного хаотического временного ряда. В представленных тестах относительная погрешность оценки искомой дисперсии находится в пределах 0,1–4 %. Предложенную схему можно обобщить на несколько случаев, встречающихся в практике реализации хаотических временных процессов:

- 1) при условии наличия аномальных выбросов (по отношению к трендовой компоненте);
- 2) при условии наличия корреляций между значениями шумовой (хаотической) компоненты;
- 3) при условии непостоянного значения дисперсии шумовой (хаотической) компоненты.

СПИСОК ЛИТЕРАТУРЫ

1. *Крянев А. В., Лукин Г. В.* Математические методы обработки неопределенных данных. М.: Физматлит, 2006. 216 с.
2. *Крянев А. В., Лукин Г. В., Удумыан Д. К.* Метрический анализ и обработка данных. М.: Физматлит, 2012. 308 с.
3. *Ivanov V. V., Kryanev A. V., Udumyan D. K., Lukin G. V.* // Appl. Math. Sci. 2014. V. 8, No. 22. P. 1053–1060.
4. *Амосов О. С., Муллер Н. В.* // Современные наукоемкие технологии. 2014. № 3. С. 122–124.
5. *Лукашин Ю. П.* Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Финансы и статистика, 2003. 203 с.
6. *Мишулина О. А.* Статистический анализ и обработка временных рядов. М.: Изд-во МИФИ, 2004. 180 с.

Получено 18 сентября 2017 г.