Перестановочная энтропия как инструмент анализа шумоподобных временных рядов

А. В. Лапшин, В.А. Панчелюга
victor.panchelyuga@gmail.com

Институт теоретической и экспериментальной биофизики РАН, Пущино, Россия

Предметом настоящей работы является исследование временных рядов флуктуаций или шумоподобных временных рядов. Очень часто такие временные ряды рассматриваются как эталон случайности. Так, например, временные ряды флуктуаций скорости альфа-распада обычно используются как очень высококачественный генератор случайных чисел. Считается, что локально такие временные ряды являются совершенно случайными и не несут в себе никакой информации. Поэтому, основные методы их исследования являются глобальными, например, функция распределения.

Ключевые слова: перестановочная энтропия, модель, шумоподобные временные ряды.

Permutation entropy as a tool for analyzing noise-like time series

A.V.Lapshin, V.A. Panchelyuga
victor.panchelyuga@gmail.com

Institute of Theoretical and Experimental Biophysics of RAS, Pushchino, Russia

The subject of this work is the study of time series of fluctuations or noise-like time series. Very often, such time series are considered as a standard of randomness. For example, time series of alpha decay rate fluctuations are commonly used as a very high quality random number generator. It is believed that locally such time series are completely random and do not carry any information. Therefore, the main methods of their research are global, for example, the distribution function.

Keywords: permutation entropy, model, noise-like time series.

1. Введение

Возможно, что одной из первых попыток локального анализа временных рядов являются исследования С.Э. Шноля [1, 2]. Используемый в его исследованиях гистограммный метод, основанный на попарном экспертном сравнении гистограмм [3], позволил обнаружить ряд периодов во временных рядах флуктуаций скорости альфа-распада, которые, как уже отмечалось, полагаются совершенно случайными.

Использование экспертного сравнения в процессе анализа шумоподобных временных рядов содержит ряд неустранимых трудностей: высокая трудоемкость, медленность анализа, что не позволяет обрабатывать большие массивы экспериментальных данных, субъективность – наличие систематических ошибок, связанных с особенностями работы человеческого сознания. Для преодоления недостатков гистограммного метода был разработан локальный фрактальный анализ методом всех сочетаний (МВС-анализ) [4-5], который обладает теми же свойствами, что и гистограммный метод [3]: результат экспертного сравнения пары гистограмм инвариантен относительно линейных преобразований (сдвиги, растяжения, зеркальные отражения) и относительно перестановок элементов отрезка временного ряда на основе которого строятся сравниваемые гистограммы. Результат МВС-анализа инвариантен относительно тех же преобразований.

Применение МВС-анализа к многолетнему массиву временных рядов флуктуаций скорости альфа-распада позволило обнаружить устойчивый спектр периодов [6-7]. Было показано, что большинство периодов из найденного спектра совпадает с периодами собственных колебаний Земли. Также был показан его универсальный характер: периоды, найденные во вре-

менных рядах флуктуаций процессов различной природы совпадают с соответствующей частью данного спектра.

Дальнейшие исследования позволили уточнить ряд периодов найденного спектра [8], а также выявить новые, в том числе, не принадлежащие спектру собственных колебаний Земли [9]. Наряду с совпадением найденного спектра со спектрами различных флуктуационных процессов «земного» происхождения, чрезвычайно интересным является его совпадение со спектрами периодов ряда астрофизических систем [10-14].

Также очень интересным и, в первую очередь с практической точки зрения, является обнаружение обсуждаемого спектра периодов в флуктуациях параметров ряда биологических систем [15-16]. Особенностью временных рядов биологического происхождения является, например, ограниченное временное разрешение, сложность в получении временных рядов большой длины. В этой связи, встает задача разработки и тестирования новых методов анализа шумоподобных временных рядов.

Настоящая работа посвящена применению одного из методов анализа таких временных рядов – метода перестановочной энтропии [17]. Данный метод в силу своей простоты и отмечаемой авторами эффективности в поиске закономерностей во временных рядах получил широкое распространение и развитие [18-21, 23-24]. В частности, к примеру, в [20] авторы модифицировали исходный алгоритм Бэндта-Помпа, введя такую величину, как взвешенная перестановочная энтропия и показав её лучшую эффективность по сравнению с собственно перестановочной энтропией. Рассмотрим подробнее, что она собой представляет, как её можно применять и нужно ли её модифицировать.

2. Перестановочная энтропия. Описание модели

Для данного временного ряда $\{x(i)\}_{i=1}^N$ выберем точечный отрезок $X(i)$ такой, что

$$X(i) = [x(i), x(i+1), \dots, x(i+(m-1)l)], \quad (1)$$

где m – размер вложенного во временной ряд отрезка, а l – т.н. *время запаздывания*, то есть интервал, разделяющий между собой значения временного ряда, падающие в $X(i)$.

Далее расположим элементы отрезка временного ряда $X(i)$ в порядке возрастания значений:

$$x[i+(j_1-1)l] \leq x[i+(j_2-1)l] \leq \dots \leq x[i+(j_m-1)l], \quad (2)$$

так, что j_t представляет собой порядковый номер элемента временного ряда x в отрезке $X(i)$. Таким образом, выбранный отрезок временного ряда кодируется так называемым порядковым паттерном

$$A(i) = [j_1, j_2, \dots, j_m], \quad (3)$$

m различных символов j_t в (3) даёт $m!$ различных символьных последовательностей $A(i)$.

Можно оценить вероятность появления каждой символьной последовательности, обозначив каждую из них $P_1, \dots, P_k, k \leq m!$:

$$P_k = \frac{s\{i | i \leq N-m, A(i) \in \pi\}}{N-m+1}. \quad (4)$$

Здесь $s\{\dots\}$ – мощность множества или количество перестановок $A(i)$, принадлежащих одному типу π .

Тогда перестановочная энтропия определяется как информационная энтропия по формуле Шеннона:

$$H(P) = -\sum_{v=1}^k P_v \log_2 P_v. \quad (5)$$

Стоит отметить, что максимальное значение $H(P)$ достигается для равномерного распределения всех перестановок, то есть в случае $H(P) = \log_2 m!$, когда $P_v = 1/m$ и $k = m!$, поэтому как правило, вводится нормирующий множитель:

$$h_p = H(P)/\log_2 m!, \quad (6)$$

в этом случае $0 \leq h_p \leq 1$.

Чем более регулярным (периодическим, квазипериодическим и т.п.) является сигнал, тем меньше значение h_p . И наоборот, чем более случайным является сигнал, тем ближе данное значение к единице. Таким образом, уменьшение значения h_p есть свидетельство проявления регулярной компоненты во временном ряду.

3. Параметры модели

Обсудим существенные параметры модели. Первый значимый параметр, N или длина временного ряда. По выбору исследователя можно оценить временной ряд как целое, рассчитав для него перестановочную энтропию, а можно разбить временной ряд на сумму пересекающихся или непересекающихся отрезков и оценить перестановочную энтропию каждого отрезка.

Следующим важнейшим параметром является длительность m отрезка $X(i)$, для которого строится порядковый паттерн $A(i)$. Бэндт и Помп [2] для практических целей рекомендуют выбирать значения $m \in [3,7]$, некоторые авторы [9] доходят до 8. А в [4] авторы отмечают, что зачастую более достоверные результаты достигаются для $m = 5,6,7$, тогда как $m =$

3,4 оказывается недостаточным. Однако при этом должно выполняться важное ограничение, связанное с предыдущим параметром, а именно:

$$m! \ll N. \quad (7)$$

Ещё одним значимым параметром является так называемое время запаздывания l . Этот параметр характеризует, как далеко друг от друга отстоят члены временного ряда, для участка которого строится порядковый паттерн $A(i)$. Бэндт и Помп [2] рекомендовали $l = 1$, то есть использовать для построения порядкового паттерна соседние значения ряда. Вместе с тем, некоторые авторы, подобно [4] используют большие значения времени задержки до $l = 10$. Точно так же, такой шаг требует, как минимум, на порядок усилить ограничение (7): $10m! \ll N$.

4. Модификации модели

Можно указать как минимум две существенные модификации модели, кроме того, можно использовать ещё один дополнительный параметр, в основу расчета которого положена перестановочная энтропия.

Первая из них так и называется авторами [3] – модифицированная перестановочная энтропия. Она решает одну существенную проблему, а именно проблему равных значений во временном ряду. На эту проблему указывали уже сами авторы, Бэндт и Помп [2], и предлагали как минимум два способа её решения. Во-первых, можно равным значениям временного ряда, то есть когда

$$x[i+(j_{k_1}-1)l] = x[i+(j_{k_2}-1)l], \quad (8)$$

присваивать разные значения индексов в соответствии с порядком следования этих значений во временном ряду. То есть когда $j_{k_1} < j_{k_2}$, в порядковый паттерн $A(i)$ j_{k_1} входит с меньшим на единицу значением, чем j_{k_2} . Другой способ состоит в том, что можно нарушить равенство значений добавкой малого случайного возмущения.

Оба этих способа существенно смещают результаты в случае большого количества равных значений во временном ряду. В этой ситуации авторы [3] фактически предлагают следующий выход – явно учитывать те порядковые паттерны, которые соответствуют отрезкам с равными значениями элементов временного ряда, присваивая таким значениям равные же значения индексов $j_{k_1} = j_{k_2}$. Вся модификация, таким образом сводится к учёту в знаменателе формулы (6) ещё и таких порядковых паттернов, их число становится больше, чем $m!$: для $m = 3$ их становится 13, для $m = 4 - 73$ и т.д. [3, Таб. I].

Следующая модификация алгоритма – т.н. взвешенная перестановочная энтропия [5]. Здесь в числитель и знаменатель формулы (4) вводятся весовые коэффициенты, получаемые как среднеквадратичное отклонение, рассчитанное для каждого отрезка временного ряда $X(i)$, по которому строится порядковый паттерн $A(i)$:

$$\omega_i = \frac{1}{m} \sum_{k=1}^m [x[i+(j-1)l] - \bar{x}_i^{m,l}]^2, \quad (9)$$

где $\bar{X}_i^{m,l}$ – среднее значение временного ряда на отрезке $X(i)$. Таким образом, каждый порядковый паттерн $A(i)$ входит со своим весом ω_i в общую формулу (4).

На основе перестановочной энтропии может быть рассчитана ещё одна величина – мера статистической сложности [7]:

$$C_{js} = Q_j[P, P_e] h_s[P], \quad (10)$$

где P рассчитывается по формуле (4), $P_e = 1/m!$ есть равномерное распределение, h_s – перестановочная энтропия (6) (можно использовать и другой такого рода параметр, но в данном случае использован этот), а $Q_j[P, P_e]$ рассчитывается по формуле:

$$Q_j[P, P_e] = \frac{H[(P+P_e)/2] - H[P]/2 - H[P_e]/2}{Q_{\max}}, \quad (11)$$

а $Q_{\max}$ – максимальное возможное значение $Q_j[P, P_e]$, достигаемое, когда одна из компонент P равна 1, а остальные исчезают, то есть,

$$Q_{\max} = -\frac{1}{2} \left[\frac{m+1}{m!} \log_2(m! + 1) - 2 \log_2(2m!) + \log_2 m! \right]. \quad (12)$$

Величина (10), таким образом, есть мера отклонения истинного распределения порядковых паттернов от равномерного распределения.

5. Методика обработки экспериментальных данных

Опираясь на материал предыдущего раздела, можно предложить следующий алгоритм обработки экспериментальных данных. Чтобы сравнить перестановочную энтропию с другими методиками обработки экспериментальных данных, будем поступать аналогично [1], то есть поделим временной ряд на 60-точечные участки, для каждого из которых будем считать перестановочную энтропию.

Поскольку длина ряда, таким образом, будет ограничена $N = 60$ значениями, это накладывает жёсткие ограничения на параметры модели перестановочной энтропии, а именно $m = 3, l = 1$. То есть порядковый паттерн строится для 3-точечного отрезка, при этом сам отрезок не содержит пропусков, т.е. включаются в расчёт именно соседние значения.

Таким образом, экспериментальные данные обрабатываются в соответствии со следующим алгоритмом:

1. Берём временной ряд длительностью более 1000 точек.

2. Разбиваем временной ряд на непересекающиеся отрезки по 60 точек.

3. Берём первый 60-точечный отрезок.

4. Берём первых $m = 3$ значения временного ряда: $x_{i+(j_1-1)l=1}, x_{i+(j_2-1)l=2}, x_{i+(j_3-1)l=3}$.

5. Так как $i = 1, l = 1$, получаем, что $j_1 = 1, j_2 = 2, j_3 = 3$.

6. Проверяем значения ряда $x_{i+(j_1-1)l=1}, x_{i+(j_2-1)l=2}, x_{i+(j_3-1)l=3}$ на наличие равных значений.

А) Если таких нет, располагаем члены ряда в порядке возрастания значений, допустим $x_{i+(j_2-1)l=2} < x_{i+(j_1-1)l=1} < x_{i+(j_3-1)l=3}$, и получаем порядковый паттерн $A_{i=1} = [2, 1, 3] \in \pi_1$ первого типа.

Б) Если таковые находятся, к примеру, $x_{i+(j_2-1)l=2} < x_{i+(j_1-1)l=1} = x_{i+(j_3-1)l=3}$, то располагаем эти значения в соответствии с $j_1 < j_3$ (и в этом примере получаем паттерн того же типа $A_{i=1} = [2, 1, 3] \in \pi_1$).

7. Далее увеличиваем i на единицу ($i = 2$) и берём соответствующие три значения временного ряда $x_{i+(j_1-1)l=2}, x_{i+(j_2-1)l=3}, x_{i+(j_3-1)l=4}$.

8. Для них повторяем пункты 5. (получая также, что $j_1 = 1, j_2 = 2, j_3 = 3$, поскольку $i = 2$) и 6., получая следующий паттерн $A_{i=2} \in \pi_k$, где $k \in [1, 6]$, поскольку $3! = 6$.

9. Продолжаем процесс получения значений A_i до $A_{i=58}$.

10. Подсчитываем $s[\pi_k]$ – мощность множества значений каждого из 6 типов паттернов π_k .

11. Подставляем полученные величины в числитель формулы (4) при $N = 60, m = 3$ и получаем значения $P_k, k \in [1, 6]$.

12. Затем подставляем значения P_k в формулу (5), получая $H[P]$.

13. Нормируем значение $H[P]$ по формуле (6), получая h_p , т.е. значение перестановочной энтропии для первого отрезка временного ряда из 60 точек.

14. Берём следующий отрезок временного ряда из 60 точек, для которого повторяем изложенную выше последовательность действий 4. – 13.

15. Получаем построенный по 60-точечным отрезкам массив значений h_p для данного временного ряда, на основании которого можно говорить о степени регулярности временного ряда и характере её изменения.

Обращая внимание на проблему равных значений, представляется целесообразным дополнительно посчитать модифицированную перестановочную энтропию. Алгоритм практически тот же, меняются только следующие пункты:

6. Проверяем значения ряда $x_{i+(j_1-1)l=1}, x_{i+(j_2-1)l=2}, x_{i+(j_3-1)l=3}$ на наличие равных значений.

Б) Если таковые находятся, к примеру, $x_{i+(j_2-1)l=2} < x_{i+(j_1-1)l=1} = x_{i+(j_3-1)l=3}$, считаем, что $j_1 = j_3 = 1$ и получаем паттерн типа $A_{i=1} = [2, 1, 1] \in \pi_1$

8. Для значений $x_{i+(j_1-1)l=2}, x_{i+(j_2-1)l=3}, x_{i+(j_3-1)l=4}$ повторяем пункты 5. (получая также, что $j_1 = 1, j_2 = 2, j_3 = 3$, поскольку $i = 2$) и 6. Получаем следующий паттерн $A_{i=2} \in \pi_k$, где $k \in [1, 13]$, поскольку число типов паттернов теперь составляет $k_3(3) = 3! + 3 \times 2! + 1! = 13$.

13. Нормируем значение $H[P]$ по формуле

$$h_p = H(P) / \log_2(k_m(m)) = H(P) / \log_2 13. \quad (13)$$

Таким образом, посчитав перестановочную энтропию h_p и модифицированную перестановочную энтропию $\hat{h}_p$, можно будет, во-первых, сравнить два подхода в рамках одного метода, во-вторых, посмотреть, какие закономерности во временном ряду можно выявить с помощью обоих типов параметризации.

Благодарность

Работа выполнена и опубликована при поддержке РФФИ, грант 19-07-00445.

Библиографический список

- [1] Шноль С.Э. Космофизические факторы в случайных процессах. Стокгольм, Шведский физический архив, 2009 – 338с.
- [2] Шноль С.Э., Коломбет В.А., Пожарский Э.В., Зенченко Т.А., Зверева И.М., Конрадов А.А. О реализации дискретных состояний в ходе флуктуаций в макроскопических процессах // УФН, 1998, 168(10) с. 1129-1140.
- [3] С.Э. Шноль, В.А. Панчелюга Феномен макроскопических флуктуаций. Методика измерений и обработки экспериментальных данных. // Мир измерений, 2007, №6, с. 49-55.
- [4] Панчелюга В.А., Панчелюга М.С. Фрактальная размерность и гистограммный метод: методика и некоторые предварительные результаты анализа шумоподобных временных рядов // Биофизика, 2013, том. 58, вып. 2, с. 377-384.
- [5] Панчелюга В.А., Панчелюга М.С. Локальный фрактальный анализ шумоподобных временных рядов методом всех сочетаний // Гиперкомплексные числа в геометрии и физике, 2014, том. 11, вып. 1, №21, с.107-133.
- [6] Панчелюга В.А., Панчелюга М.С. Локальный фрактальный анализ шумоподобных временных рядов методом всех сочетаний в диапазоне периодов 1-115 мин // Биофизика, 2015, том. 60, вып. 2, с. 395-410.
- [7] Панчелюга В.А., Панчелюга М.С. Некоторые предварительные результаты локального фрактального анализа шумоподобных временных рядов методом всех сочетаний // Гиперкомплексные числа в геометрии и физике, 2014, том. 11, вып. 1, №21, с. 134-156.
- [8] В.А. Панчелюга, М.С. Панчелюга, О.Ю. Серая Предварительные результаты исследования внутрисуточных периодов во временных рядах флуктуаций скорости альфа-распада // Гиперкомплексные числа в геометрии и физике, 2016, т. 13, вып. 2, № 25, с. 211-216.
- [9] Панчелюга В.А., Владимирский Б.М., Панчелюга М.С., Серая О.Ю., Панихин В.А. Выраженность периодов 50, 80 и 160 мин во временных рядах флуктуаций скорости альфа-распада // Сборник трудов XXI Всероссийской конференции «Солнечная и солнечно-земная физика – 2017», Санкт-Петербург, Пулково, 10-14 октября 2017 г., с. 261-264.
- [10] S.Siparov, V.Samodurov, and G.Laptev Origin of observed periodic components in astrophysical masers' spectra // Monthly Notices of the Royal Astronomical Society, 2017, 467, pp. 2813–2819.
- [11] Панчелюга В.А., Панчелюга М.С. О связи между спектром периодов во временных рядах флуктуаций скорости альфа-распада и периодическими компонентами в спектрах астрофизических мазеров // Материалы XIII Международной конференции "Финслеровы обобщения теории относительности" (FERT-2017) / Ред.: Павлов Д.Г., Панчелюга В.А. - Москва, 11-й формат, 2017 - с.69-70.
- [12] Панчелюга В.А., Владимирский Б.М., Панчелюга М.С. О совпадении спектра периодов во временных рядах флуктуаций скорости альфа-распада с периодическими компонентами в спектрах астрофизических мазеров // Система "Планета Земля": XXV лет семинару «Система «Планета Земля» (1994-2019) М., ЛЕНАНД, 2019 – с.115-118.
- [13] Панчелюга В.А., Владимирский Б.М., Панчелюга М.С. О некоторых возможных механизмах связи между спектром периодов во временных рядах флуктуаций скорости альфа-распада и периодическими компонентами в спектрах астрофизических мазеров // Материалы XIV Международной конференции «Финслеровы обобщения теории относительности» (FERT-2018) / Ред.: Павлов Д.Г., Панчелюга В.А., М., 11-й формат, 2018, с.89-91.
- [14] В.А. Панчелюга, М.С. Панчелюга О совпадении спектра периодов в флуктуациях скорости альфа-распада со спектром вращательных периодов астероидов // Материалы XV Международной конференции «Финслеровы обобщения теории относительности» (FERT-2019) / Ред.: Павлов Д.Г., Панчелюга В.А. — Москва, 11-й формат, 2019 - с.27-29.
- [15] М.Е. Диатроптов, В.А. Панчелюга, М.С. Панчелюга О совпадении спектра периодов во временных рядах флуктуаций температуры скворцов и крыс со спектром периодов флуктуаций скорости альфа-распада // Материалы XV Международной конференции «Финслеровы обобщения теории относительности» (FERT-2019) / Ред.: Павлов Д.Г., Панчелюга В.А. — Москва, 11-й формат, 2019 - с.30-35.
- [16] Панчелюга В.А., Тирас Х.П., Новиков К.Н., Панчелюга М.С., Нефёдова С.Е. О совпадении спектра периодов во временных рядах хемолюминесценции планарий со спектром периодов, найденным во временных рядах флуктуаций скорости альфа-распада // Материалы XV Международной конференции «Финслеровы обобщения теории относительности» (FERT-2019) / Ред.: Павлов Д.Г., Панчелюга В.А. — Москва, 11-й формат, 2019 - с.36-41.
- [17] C. Bandt, B. Pompe Permutation Entropy: A Natural Complexity Measure for Time Series // Phys. Rev. Lett., 2002, vol.88, No.17, 174102.
- [18] C. Bian, C. Qin, Q. D. Y. Ma, Q. Shen Modified permutation-entropy analysis of heartbeat dynamics // Phys. Rev., 2012, E 85, 021906.
- [19] Y. Cao, W.-W. Tung, J. B. Gao, V. A. Protopopescu, L. M. Hively Detecting dynamical changes in time series using the permutation entropy // Phys. Rev., 2004, E70, 046217.
- [20] B. Fadlallah, B. Chen, A. Keil, and J. Principe Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information // Phys. Rev., 2013, E 87, 022911.
- [21] Q. Li, Z. Fu Permutation entropy and statistical complexity quantifier of nonstationarity effect in the vertical velocity records // Phys. Rev., 2014, E 89, 012905.
- [22] R. López-Ruiz, H. L. Mancini, X. Calbet // Phys. Lett., 1995, A 209, 321.
- [23] P. J. Week, D. A. Schaffner, M. R. Brown Permutation entropy and statistical complexity analysis of turbulence in laboratory plasmas and the solar wind // Phys. Rev., 2015, E 91, 023101.
- [24] L. Zunino, M.C. Soriano, I. Fischer, O.A. Rosso, C.R. Mirasso Permutation-information-theory approach to unveil delay dynamics from time-series analysis // Phys. Rev., 2010, E82, 046212.