Chinese Journal of Physics 55 (2017) 2325–2335

Contents lists available at ScienceDirect

Chinese Journal of Physics

journal homepage: www.elsevier.com/locate/cjph

Scaling invariance embedded in very short time series: A factorial moment based diffusion entropy approach

Yue Yang[a], Lu Qiu[b], Tianguang Yang[c], Liying Hou[a,d], Changgui Gu[a], Huijie Yang[*,a]

[a] Business School, University of Shanghai for Science and Technology, Shanghai, 200093, China
[b] School of Finance and Business, Shanghai Normal University, Shanghai, 200234, China
[c] Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin, 300071, China
[d] School of Science, Shanghai Open University, Shanghai, 200433, China

A R T I C L E I N F O

Keywords:
Factorial moment based diffusion entropy
Scaling invariance
Very short time series

A B S T R A C T

How to evaluate scaling behaviors in very short time series is still an open problem, in which the mechanism-dependence and the bias of estimation of a statistical quantity become critical. We propose a new method called factorial moment based diffusion entropy (FMDE). A theoretical derivation and extensive calculations show that it can give us a high-confident and unbiased evaluation of scaling exponent from a time series with a length of $\sim 10^2$. It provides a reliable method to monitor evolutionary behaviors of complex systems. As an illustration, it is used to monitor the fractal gait rhythm for a volunteer in six stride trials. We find rich patterns in its physiological state.

1. Introduction

A stochastic process can be described with its probability distribution function (PDF) $p(x, t)$, namely, the occurring probability density in the interval of displacement $x \sim x + dx$ at time t. If the PDF obeys a specific form [1],

$$p(x, t) \sim \frac{1}{t^{\alpha}} F\left(\frac{x}{t^{\alpha}}\right), \tag{1}$$

the process behaves scale-invariant, where α is scaling exponent, and $F(\cdot)$ a function. Under a transformation of scale $x \to \frac{x}{t^{\alpha}}$, the PDF keeps unchanged.

To define scaling invariance for a single time series, let us map it to a diffusion process. From the series one can extract all the possible segments with a specified length. Regarding every segment as a trajectory of a particle starting from the original point, all the segments form an ensemble containing many realizations of a stochastic process. The displacements of the trajectories are simply the summations of the segments respectively. If the distribution of displacements obeys the form in Eq. (1), the time series behaves scale-invariant, where the segment length acts as time.

In literature there exist several variance based methods to evaluate scaling behaviors of time series such as the wavelet transformation modulus Maxima (WTMM) [2–9] and the de-trended fluctuation analysis (DFA) [10–17]. If a time series behaves scale-invariant, the standard deviation versus time obeys a power-law, whose scaling exponent is called Hurst exponent H. There are two serious limitations [18–20]. First, this kind of methods is dynamical mechanism dependent. For a fractional Brownian motion, the estimation is correct, namely, the power-law can be found and we have $H = \alpha$. For a Levy walk, the scaling behavior can be detected

* Corresponding author.
E-mail addresses: hjyang@ustc.edu.cn, hjyang@usst.edu.cn (H. Yang).

http://dx.doi.org/10.1016/j.cjph.2017.10.002
Received 10 September 2017; Received in revised form 5 October 2017; Accepted 5 October 2017
Available online 06 October 2017
0577-9073/

qualitatively, but the estimated scaling exponent does not equal to the real one, i.e., $\alpha = \frac{1}{3-2H}$. As for a Levy flight, the standard deviation does not obey a power-law at all. Second, the methods require an infinite length of time series. Extensive calculations show that to obtain a reliable estimation of scaling exponent lengths of 2^{13} [21] and 2^{15} [22] are required by the WTMM and DFA methods respectively. For the re-scaled fluctuation method (R/S) even a length of 2^{16} [23] can not guarantee a well-converged result.

To overcome the mechanism-dependent problem, a concept called diffusion entropy (DE) [20] is introduced in literature, in which instead of the variance the Shannon entropy is calculated from the PDF. A simple computation proves that for a scale-invariant time series, no matter by which dynamical mechanism it is produced, the entropy has a linear relation with the natural logarithm of time, the slope of which equals to the scaling exponent. A large amount of works (e.g., [24–41], and the references there-in) show its powerfulness, however, a concise estimation of scaling exponent requires still an infinite number of records.

Very recently, a solution is proposed to estimate entropy from a finite set of records [42], the key idea of which is to minimize simultaneously the statistical fluctuation and bias, called balanced estimator of entropy. Replacing the naive estimation of entropy in the DE with this estimator, we develop a new version of DE, called correlation-dependent balanced estimator of diffusion entropy (cBEDE) [43–49], which can obtain reliable scaling behaviors from very short time series with a length of $\sim 10^2$ [45].

In the research branch of high-energy collision, a basic task is to extract information of quantum chromo-dynamics from a limited number of experimental records. An important concept called factorial moment (FM) [50–53] is proposed to detect intermittency behaviors (multi-fractal) from several tens of experiments. Dividing the phase space of a particle into many bins, the PDF can be represented with the occurring frequencies in the bins. The partition function, namely, q-order probability moment of the PDF is proportional to the summation of the qth power of the frequencies. But this naive estimation has a non-ignorable bias when the number of records is finite, which is corrected in FM by replacing qth power of every frequency with a factorial moment. FM provides us an unbiased estimation of the partition function. Several tens of experiments can guarantee a reliable evaluation of scaling behaviors. However, the FM stands only when q is an integer, and an extension of the idea to a real number of q turns out to be a non-trivial task [54,55].

In the present paper the concept of FM is used to provide an unbiased estimation of Shannon entropy, based upon which we propose a new version of DE, called factorial moment based estimator of diffusion entropy (FMDE). We approximate every term of the initial definition of entropy with a polynomial function of probability, and then replace each term of which with a factorial moment to present an unbiased estimation of the polynomial expansion. A theoretical derivation and extensive calculations on fractional Brownian motions show that the FMDE can give a reliable evaluation of scaling behavior of a very short time series. It is then used to monitor evolutionary behaviors embedded in gait time series, as a typical illustration. From six stride trials of a volunteer we find rich patterns in its fractal gait rhythm.

2. Method

Herein we propose a new method to evaluate scaling behaviors in very short time series, called factorial moment based estimator of diffusion entropy (FMDE).

2.1. Diffusion entropy

Let us start from a stationary time series, $Y = \{y_1, y_2, ..., y_N\}$. All the possible segments with a predefined length s read, $Y^m = \{y_m, y_{m+1}, ..., y_{m+s-1}\}$, $m = 1, 2, ..., N-s+1$. Regarding each segment as a trajectory of a particle starting from origin, all the segments form a bundle of realizations of a stochastic process, whose displacements read,

$$Z(s) = \left\{ z_m(s) = \sum_{k=1}^{s} y_{m+k-1},\ m = 1, 2, ..., N-s+1 \right\}. \tag{2}$$

Selecting a certain fraction of standard deviation of the initial series Y as width of a bin, denoted with $\frac{std(Y)}{\epsilon}$, one can separate the distribution interval of displacements into a total of $M(s) = ceil\left[\epsilon \cdot \frac{max(Z) - min(Z)}{std(Y)}\right]$ bins. Here $ceil(\cdot)$ rounds a real number to the nearest integer greater than or equal to it. Reckoning the occurrence frequency of the elements of Z in every bin, $n(i, s)$, $i = 1, 2, ..., M(s)$, a simple estimation of PDF reads,

$$\hat{p}(s) \equiv \left\{ \hat{p}_i(s) = \frac{n(i, s)}{N-s+1},\ i = 1, 2, ..., M(s) \right\}. \tag{3}$$

The corresponding Shannon entropy is,

$$\hat{S}_{DE}(s) = -\sum_{i=1}^{M(s)} \hat{p}_i(s) \ln[\hat{p}_i(s)]. \tag{4}$$

If the original series Y behaves scale-invariant, namely, $\hat{p}(s)$ obeys the form in Eq. (1), a simple computation leads to,

$$\hat{S}_{DE}(s) = A + \alpha \ln(s). \tag{5}$$

Hence, the slope of the curve $\hat{S}_{DE}(s)$ versus $\ln(s)$ can be used to evaluate the scaling exponent. This procedure is called diffusion entropy method (DE) [20]. The advantage of this method over the variance based methods is that it is dynamical mechanism