Commun Nonlinear Sci Numer Simulat 46 (2017) 103–115

Contents lists available at ScienceDirect

Commun Nonlinear Sci Numer Simulat

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

Research paper

The ordinal Kolmogorov-Sinai entropy: A generalized approximation

J.S. Armand Eyebe Fouda[a,*], Wolfram Koepf[b], Sabir Jacquir[c]

[a] Department of Physics, Faculty of Science, University of Yaoundé I, P.O. Box 812, Yaoundé, Cameroon
[b] Institute of Mathematics, University of Kassel, Heinrich-Plett Str. 40, 34132 Kassel, Germany
[c] LE2I UMR 6306, CNRS, Arts et Métiers, University of Bourgogne Franche-Comté, F-21000 Dijon, France

ARTICLE INFO

Article history:
Received 20 April 2016
Revised 14 July 2016
Accepted 1 November 2016
Available online 2 November 2016

Keywords:
Complexity
Entropy
Ordinal pattern
Ordinal array

ABSTRACT

We introduce the multi-dimensional ordinal arrays complexity as a generalized approximation of the ordinal Komogorov-Sinai entropy. The ordinal arrays entropy (OAE) is defined as the Shannon entropy of a series of m-ordinal patterns encoded symbols, while the ordinal arrays complexity (OAC) is defined as the differential of the OAE with respect to m. We theoretically establish that the OAC provides a better estimate of the complexity measure for short length time series. Simulations were carried out using discrete maps, and confirm the efficiency of the OAC as complexity measure from a small data set even in a noisy environment.

1. Introduction

Measuring the complexity from series of observation is crucial for various fields of research. Existing theoretical complexity measures such as Lyapunov exponents, Kolmogorov-Sinai (KS) entropy, correlation dimension and many others are difficult to estimate from a finite data set [1–6], while empirical measures like the renormalized entropy or the approximate entropy often lack a theoretical foundation or are not well interpretable [7–9].

Considering the above limitations of the existing complexity measures, the permutation entropy (PE), an ordinal pattern analysis (OPA) based method, was introduced by Bandt and Pompe [10]. The PE is easily estimated from any type of data set and its theoretical foundation has been clearly established. Despite its well known performances, the PE does not well approximate the KS entropy for a finite data set. Therefore, the method cannot clearly state whether a dynamics is regular or not. Based on this observation, various modifications as well as other OPA based approaches have been developed [11–18]. Despite the improvements they brought, these methods have not yet clearly reached a zero complexity for regular dynamics as it should be. Indeed, although OPA based methods are very promising for applications of the dynamical system theory to real-world systems, they nevertheless present a fundamental limitation as their accuracy needs large embedding dimension and infinite time series length. In practice, such requirements are difficult to satisfy. Thus, it is judicious to try to improve the existing results for finite data sets. Recently the conditional entropy of ordinal patterns (CPE) was proposed where the average diversity of ordinal patterns succeeding a given one is characterized [19]. The CPE provides a much better practical estimation of the KS entropy than the permutation entropy, while having the same computational efficiency.

* Corresponding author.
E-mail addresses: efoudajsa@yahoo.fr (J.S.A. Eyebe Fouda), koepf@mathematik.uni-kassel.de (W. Koepf), sjacquir@u-bourgogne.fr (S. Jacquir).
URL: http://www.mathematik.uni-kassel.de/~koepf/ (W. Koepf)

http://dx.doi.org/10.1016/j.cnsns.2016.11.001
1007-5704/

However, similarly as the PE, the CPE is estimating the complexity from the frequency occurrence of ordinal patterns. Indeed, the idea behind OPA based methods is to consider order relations between values of time series instead of the values themselves. The original time series is firstly converted into a sequence of ordinal patterns of an order n, each of them describing order relations between n successive points of the time series. The entropy is then estimated using the frequencies of the various ordinal patterns. Using the ordinal pattern transformation undoubtedly reduces the number of patterns of the same order while considering the values of the time series. In fact two different embedding vectors derived from the time series may lead to the same ordinal pattern, hence directly calculating the entropy from the ordinal patterns frequencies leads to some information loss.

In this paper, we suggest to apply the OPA to the values of the series of ordinal patterns derived from the time series. As the effective number of symbols can be evaluated for ordinal patterns of order n, the entropy of the new source of symbols can be well estimated. This approach allows to increase the number of patterns which has been reduced by the conversion into ordinal patterns, therefore to better approximate the entropy of the time series. For instance, for a given real-world time series with Λ distinct symbols or values, the number of distinct ordinal patterns of order n is $n!$, while the number of distinct embedding vectors of the same order derived from the time series values is Λ^n. Now considering the series of ordinal patterns of order n, the number of possible distinct embedding vectors of order n which can be derived from the time series values increases as $(n!)^n \gg n!$.

The paper presents first a brief recall of the permutation entropy in Section 2. The ordinal arrays entropy (OAE) is introduced in Section 3 followed by the definition of the ordinal arrays complexity (OAC) in Section 4. In Section 5 some simulation results are given, followed by concluding remarks in Section 6.

2. Brief recall of the permutation entropy

Let $\{x_t\}_{t=0,1,\ldots,T-1}$ be a time series of length T where t is the time index. The PE of order n is defined as a measure of the probabilities of permutations of order n [10]. Permutations of order n are obtained from the comparison of neighboring values (increasing order) in embedding vectors $\mathbf{x}_k = \left(x_{k\tau_0}, x_{k\tau_0+\tau} \ldots, x_{k\tau_0+l\tau}, \ldots, x_{k\tau_0+(n-1)\tau}\right)$, where $k \in \mathbb{N}$, n is the embedding dimension (number of values in $\mathbf{x}_k$), $\tau_0 \in \mathbb{N}_{\geqq 1}$ is the delay time of the embedding vectors, $\tau \in \mathbb{N}_{\geqq 1}$ is the delay time of samples and $l+1$ the index of $x_{k\tau_0+l\tau}$ in $\mathbf{x}_k$, $l \in \mathbb{N}$. Let P_k be the permutation derived from $\mathbf{x}_k$, with $\tau_0 = 1$. $P_k = \left(\begin{smallmatrix}1,2,3,\ldots,n\\5,n,1,\ldots,3\end{smallmatrix}\right)$ for example is obtained by sorting the values of $\mathbf{x}_k$ in ascending order, with $x_{k+4\tau} < x_{k+(n-1)\tau} < x_k < \ldots < x_{k+2\tau}$. Identical values are sorted by ascending order of their time index. The permutation entropy of order n is thus given by

$$H_p(n) = -\sum p(\theta) \cdot \ln(p(\theta)), \tag{1}$$

where

$$p(\theta) = \frac{\#\{k \mid k \leq T - n\tau, P_k = \theta\}}{T - n\tau + 1}, \tag{2}$$

is the probability of the permutation θ and # denotes the cardinality [10].

Definition 1 [20]**.** A time series $\{x_t\}$ is called period-L cycle or simply L-periodic, if there exists a basic pattern of length q samples containing L distinct values ($L \leq q$) periodically repeated, independently of the time origin. q is known as the time space period and L as the phase space period.

Choosing $\tau > 1$ can lead to some misinterpretations in complexity values in the case of regular dynamics. Indeed, we showed in [20] that the number of distinct permutations in the case of L-periodic dynamics is fluctuating between $\delta = \frac{L}{\gcd(L,\tau)}$ and L, depending on the ordering of the time series. It therefore results that the PE is such that $\ln(\delta) \leq H_p(n) \leq \ln(L)$. Considering this dependence on the ordering of the time series, the bias between the PE and the largest Lyapunov exponent in the case of regular dynamics cannot be determined rigourously.

3. Ordinal arrays entropy

3.1. Designing of M-dimensional ordinal arrays

Measuring the Shannon entropy of an information source requires that the number of symbols is known and that the set of symbols (often called as the alphabet $\mathcal{A}$) is of finite size; otherwise, the Shannon entropy equals infinity. It is common to define the information entropy in bits. In the case of continuous random variables such as the speech signal for example, a uniform quantizer can be used to convert the corresponding continuous alphabet into a discrete one. Symbols of the discrete alphabet thus obtained are encoded with a finite number of bits. Similarly, the ordinal pattern analysis approach which considers the order relations between values of the time series instead of the values themselves can be seen as a quantizer. It then converts the continuous alphabet of the source into an ordinal pattern alphabet. Each ordinal pattern is encoded using positive integers similarly as bits are commonly used. The set of ordinal patterns can then be considered as a pseudo-source of information. If n is the length of ordinal patterns, then the cardinality of the ordinal pattern alphabet is equal to $n!$, while defining patterns of the same length using the values of the time series would have led to an Λ^n-length alphabet, Λ being the number of distinct values of the time series. It appears that considering the order relation