PHYSA: 15379 Model 3Gsc pp. 1–21 (col. fig: NIL)

Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

Multifractal diffusion entropy analysis: Optimal bin width of probability histograms

Q1 Petr Jizba [a,b,*], Jan Korbel [a,c]

[a] *Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 11519, Prague, Czech Republic*

[b] *Institute of Theoretical Physics, Freie Universität in Berlin, Arnimallee 14, 14195 Berlin, Germany*

[c] *Max Planck Institute for the History of Science, Boltzmannstrasse 22, 14195 Berlin, Germany*

HIGHLIGHTS

- We propose a model of multiscale time series based on stable distributions.
- We analyze the series with multifractal diffusion entropy.
- We demonstrate the need for optimal bin-width in associated empirical histograms.
- We propose a method for optimal bin-width in general multifractal time series.
- Our proposal is illustrated with financial time series of S&P500 index.

ARTICLE INFO

Article history:
Received 12 March 2014
Received in revised form 20 May 2014
Available online xxxx

Keywords:
Multifractals
Rényi entropy
Stable distributions
Time series

ABSTRACT

In the framework of Multifractal Diffusion Entropy Analysis we propose a method for choosing an optimal bin-width in histograms generated from underlying probability distributions of interest. The method presented uses techniques of Rényi's entropy and the mean squared error analysis to discuss the conditions under which the error in the multifractal spectrum estimation is minimal. We illustrate the utility of our approach by focusing on a scaling behavior of financial time series. In particular, we analyze the S&P500 stock index as sampled at a daily rate in the time period 1950–2013. In order to demonstrate a strength of the method proposed we compare the multifractal δ-spectrum for various bin-widths and show the robustness of the method, especially for large values of q. For such values, other methods in use, e.g., those based on moment estimation, tend to fail for heavy-tailed data or data with long correlations. Connection between the δ-spectrum and Rényi's q parameter is also discussed and elucidated on a simple example of multiscale time series.

Q2

1. Introduction

The evolution of many complex systems in natural, economical, medical and biological sciences is usually presented in the form of time data-sequences. A global massification of computers together with their improved ability to collect and process large data-sets has brought about the need for novel analyzing methods. A considerable amount of literature has been recently devoted to developing and using new data-analyzing paradigms. These studies include such concepts as fractals and multifractals [1], fractional dynamics [2,3], complexity [4,5], entropy densities [5] or transfer entropies [6–8].

* Corresponding author at: Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 11519, Prague, Czech Republic. Tel.: +420 224358295; fax: +420 222320861.

E-mail addresses: p.jizba@fjfi.cvut.cz, jizba@zedat.fu-berlin.de (P. Jizba), korbeja2@fjfi.cvut.cz (J. Korbel).

http://dx.doi.org/10.1016/j.physa.2014.07.008
0378-4371/© 2014 Published by Elsevier B.V.

Particularly in the connection with financial time series there has been rapid development of techniques for measuring and managing the fractal and multifractal scaling behavior from empirical high-frequency data sequences. A non-trivial scaling behavior in a time data-set represents a typical signature of a multi-time scale cooperative behavior in much the same way as a non-trivial scaling behavior in second-order phase transitions reflects the underlying long-range (or multi-scale) cooperative interactions. The usefulness of the scaling approach is manifested, for instance, in quantifying critical or close-to-critical scaling which typically signalizes the onset of financial crises, including stock market crashes, currency crises or sovereign defaults [9]. A multifractal scaling, in particular, is instrumental in identifying the relevant scales that are involved in both temporal and inter-asset correlations [8]. In passing, one can mention that aside from *financial* data sequences, similar (multi)fractal scaling patterns are also observed (and analyzed) in time data-sets of heart rate dynamics [10,11], DNA sequences [12,13], long-time weather records [14] or in electrical power loads [15].

In order to identify fractal and multifractal scaling in time series generated by a complex system (of both deterministic and stochastic nature), several tools have been developed over the course of time. To the most popular ones belong the Detrended Fluctuation Analysis [12,16], Wavelets [17], or Generalized Hurst Exponents [18]. The purpose of the present paper is to discuss and advance yet another pertinent method, namely the Multifractal Diffusion Entropy Analysis (MF-DEA). In doing so we will stress the key rôle that Rényi's entropy (RE) plays in this context. To this end, we will employ two approaches for the estimation of the scaling exponents that can be directly phrased in terms of RE, namely, the monofractal approach of Scafetta et al. [19] and the multifractal approach of Huang et al. [20], with further comments of Morozov [21]. The most important upshot that will emerge from this study is the proposal for the optimal bin-width in empirical histograms. The latter ensures that the error in the RE (and hence also scaling exponents) evaluation, when the underlying probability density function (PDF) is replaced by its empirical histograms, is minimal in the sense of Rényi's information divergence and the associated L_2-distance. We further show that the ensuing optimal bin-width permits the characterization of the hierarchy of multifractal scaling exponents $\delta(q)$ and $D(q)$ in a fully quantitative fashion.

This paper is structured as follows: In Section 2 we briefly review foundations of the multifractal analysis that will be needed in following sections. In particular, we introduce such concepts as Lipschitz–Hölder's singularity exponent, multifractal spectral function and their Legendre conjugates. In Section 3 we state some fundamentals of the fluctuation collection algorithm and propose a simple instructive example of a heterogeneous multiscale time series. Within this framework we discuss the MF-DEA and highlight the rôle of Rényi's entropy as a multiscale quantifier. After this preparatory material we turn in Section 4 to the question of the optimal bin-width choice that should be employed in empirical histograms. In particular, we analyze the bin-width that is needed to minimize error in the multifractal spectrum evaluation. In Section 5, we demonstrate the usefulness and formal consistency of the proposed error estimate by analyzing time series from S&P500 market index sampled at a daily (end of trading day) rate basis in the period from January 1950 to March 2013 (roughly 16 000 data points). We apply the symbolic computations with the open source software *R* to illustrate the strength of the proposed optimal bin-width choice. In particular, we graphically compare the multifractal δ-spectrum for various bin-widths. Our numerical results imply that the proposed bin-widths are indeed optimal in comparison with other alternatives used. Implications for the $\delta(q)$-spectrum as a function of Rényi's q parameter are also discussed and graphically represented. Conclusions and further discussions are relegated to the concluding section. For the reader's convenience, we present in the Appendix the source code in the language *R* that can be directly employed for efficient estimation of the $\delta(q)$-spectrum (and ensuing generalized dimension $D(q)$) of long-term data sequences.

2. Multifractal analysis

Let us have a discrete time series $\{x_j\}_{j=1}^N \subset \mathbb{R}^D$, where x_j are obtained from measurements at times t_j with an equidistant time lag $\mathfrak{s}$. We divide the whole domain of the definition of x_j's into distinct regions K_i and define the probability of each region as

$$p_i \equiv \lim_{N\to\infty} \frac{N_i}{N} = \lim_{N\to\infty} \frac{\text{card}\{j \in \{1,\ldots,N\} \mid x_j \in K_i\}}{N}, \tag{1}$$

where "card" denotes the *cardinality*, i.e., the number of elements contained in a set. For every region, we consider that the probability scales as $p_i(\mathfrak{s}) \propto \mathfrak{s}^{\alpha_i}$, where α_i are scaling exponents also known as the Lipschitz–Hölder (or singularity) exponents. The key assumption in the multifractal analysis is that in the small-$\mathfrak{s}$ limit we can assume that the probability distribution depends *smoothly* on α and thus the probability that some arbitrary region has the scaling exponent in the interval $(\alpha, \alpha + \mathrm{d}\alpha)$ can be considered in the form

$$\mathrm{d}\rho(\mathfrak{s},\alpha) = \lim_{N\to\infty} \frac{\text{card}\{p_i \propto \mathfrak{s}^{\alpha'} \mid \alpha' \in (\alpha, \alpha+\mathrm{d}\alpha)\}}{N} = c(\alpha)\mathfrak{s}^{-f(\alpha)}\mathrm{d}\alpha. \tag{2}$$

The corresponding scaling exponent $f(\alpha)$ is known as the *multifractal spectrum* and by its very definition it represents the (box-counting) fractal dimension of the subset that carries PDF's with the scaling exponent α.

A convenient way how to keep track with various p_i's is to examine the scaling of the correspondent moments. To this end one can define a "partition function"

$$Z(q,\mathfrak{s}) = \sum_i p_i^q \propto \mathfrak{s}^{\tau(q)}. \tag{3}$$