Chaos, Solitons and Fractals 116 (2018) 54–62

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Statistical test for fractional Brownian motion based on detrending moving average algorithm

Grzegorz Sikora

Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wroclaw University of Science and Technology, Janiszewskiego 14a, Wrocław 50-370, Poland

ARTICLE INFO

Article history:
Received 21 March 2018
Revised 28 August 2018
Accepted 31 August 2018

Keywords:
Detrending moving average algorithm
Statistical test
Fractional brownian motion

ABSTRACT

Motivated by contemporary and rich applications of anomalous diffusion processes we propose a new statistical test for fractional Brownian motion, which is one of the most popular models for anomalous diffusion systems. The test is based on detrending moving average statistic and its probability distribution. Using the theory of Gaussian quadratic forms we determined it as a generalized chi-squared distribution. The proposed test could be generalized for statistical testing of any centered non-degenerate Gaussian process. Finally, we examine the test via Monte Carlo simulations for two exemplary scenarios of anomalous diffusion: subdiffusive and superdiffusive dynamics as well as for classical diffusion.

1. Introduction

The theory of stochastic processes is currently an important and developed branch of mathematics [10,22,25]. The key issue from the point of view of the application of stochastic processes is statistical inference for such random objects [41,53,62,64]. This field consists of statistical methods for the reliable estimation, identification, and validation of stochastic models. Such a part of the theory of stochastic processes and the statistics developed for them are used to model phenomena studied by other fields such as physics [28,60,75], chemistry [28,66,75], biology [11,14,28,32,65,66], engineering [7,67], among others.

This work is motivated by growing interest and applications of the special class of stochastic processes, namely anomalous diffusion processes, which largely depart from the classical Brownian diffusion theory [50,63]. Such processes are characterized by a nonlinear power-law growth of the mean squared displacement (MSD) in the course of time. Their anomalous diffusion behavior manifested by nonlinear MSD is intimately connected with the breakdown of the central limit theorem, caused by either broad distributions or long-range correlations. More precisely, the MSD for anomalous diffusion processes behaves as a power law function t^{β}, where the exponent β is called the anomalous diffusion exponent. In the case of $\beta = 1$ we have classical diffusion, while $\beta \neq 1$ we deal with anomalous diffusion. When $\beta < 1$ the MSD increases slower than linearly which is the case of subdiffusion, while $\beta > 1$ the MSD increases faster than linearly and this is the superdiffusion scenario. Today, the list of systems displaying anomalous dynamics is quite extensive [26,31,35,44,56,59]. Therefore in recent years, there has been great progress in the understanding of the different mathematical models that can lead to anomalous diffusion [36,37,51]. One of the most popular of them is the fractional Brownian motion (FBM) [29,33,35,42,51,73,78]. Introduced by Kolmogorov [38] and studied by Mandelbrot in a series of papers [46,47], it is now a well-researched stochastic process. FBM is still constantly developed by mathematicians in different aspects [5,23,55,57,77].

The main subject considered in this work is the issue of rigorous and valid identification of the FBM model. The problem of FBM identification has been described in the mathematical literature for a long time [8,18]. However, most of the works mainly concern various methods of estimating the parameters of the FBM model. They are based, among others, on p-variation [45], discrete variation [19], sample quantiles [20] and other methods [9,12,21,27,43,52,74,81]. A certain gap in this theory is the lack of tools such as rigorous statistical tests to identify the FBM model in empirical data. Some approaches to FBM identification are known, e.g., application of empirical quantiles [13], distinguishing FBM from pure Brownian motion [40]. According to the author's current knowledge, the only statistical test for the FBM model is the test based on the distribution of the time average MSD [71]. Due to the lack of statistical tests specially designed for the FBM model, in this work, we propose such a statistical testing procedure.

The proposed test has a test statistic which is the detrending moving average (DMA) statistic introduced in the paper [2]. For more than a decade, the DMA algorithm has become an important and promising tool for the analysis of stochastic signals. It is

E-mail address: grzegorz.sikora@pwr.edu.pl

https://doi.org/10.1016/j.chaos.2018.08.031
0960-0779/

constantly developed and improved [4,16,17,69], its multifractal version was created and used [15,30,34,80] and it is applied for different empirical datasets [39,58,61,68]. As one of the important method for fluctuation analysis, the DMA algorithm was often compared with other methods [6,79,82]

In Section 2 we show that the distribution of the DMA statistics follows the generalized chi-squared distribution. The main Section 3 demonstrates the statistical testing procedure based on computing the DMA statistic for empirical data. In Section 4 the results of Monte Carlo simulations of the proposed test are presented and discussed. Section 5 contains conclusions and final remarks. In the Appendix A, the Matlab code of the proposed test is presented.

2. Probability distribution of DMA statistic

The DMA algorithm was introduced in [2]. For a finite trajectory $\{X(1), X(2), \dots, X(N)\}$ of a stochastic process the DMA statistic has the following form

$$\sigma^2(n) = \frac{1}{N-n}\sum_{j=n}^{N}\left(X(j)-\tilde{X}_n(j)\right)^2, \quad n = 2, 3, \dots, N-1, \tag{1}$$

where $\tilde{X}_n(j)$ is a moving average of n observations $X(j), \dots, X(j-n+1)$, i.e.

$$\tilde{X}_n(j) = \frac{1}{n}\sum_{k=0}^{n-1} X(j-k).$$

The statistic $\sigma^2(n)$ is a random variable which computes the mean squared distance between the process $X(j)$ and its moving average $\tilde{X}_n(j)$ of the window size n. It has scaling law behavior $\sigma^2(n) \sim C_H n^{2H}$, i.e. $\lim_{n\to\infty}\frac{E[\sigma^2(n)]}{C_H n^{2H}} = 1$, where H is a self–similarity parameter of the signal [2,4]. The constant C_H has explicit expression computed in the case of fractional Brownian motion [4]. As a byproduct of this scaling law one can estimate the self–similarity parameter H from linear fitting on double logarithmic scale [6,17,70].

In this work, we leave the issue of DMA algorithm as an estimation method and concentrate on the probability characteristics of this random statistic. Throughout the paper, we assume that the stochastic process $X(j)$ is a centered Gaussian process. Therefore a finite trajectory $\mathbb{X} = \{X(1), X(2), \dots, X(N)\}$ is a centered Gaussian vector with covariance matrix $\Sigma = \{E[X(j)X(k)] : j, k = 1, 2, \dots, N\}$. Let introduce the process $Y(j) := X(j+n-1) - \tilde{X}_n(j+n-1)$, which is still a centered Gaussian process. We calculate the covariance matrix of the vector $\mathbb{Y} = \{Y(1), Y(2), \dots, Y(N-n+1)\}$

$$\begin{aligned} E[Y(j)Y(k)] &= E[X(j+n-1)X(k+n-1)] \\ &\quad - E\left[X(j+n-1)\tilde{X}_n(k+n-1)\right] \\ &\quad - E\left[\tilde{X}_n(j+n-1)X(k+n-1)\right] \\ &\quad + E\left[\tilde{X}_n(j+n-1)\tilde{X}_n(k+n-1)\right] \\ &= E[X(j+n-1)X(k+n-1)] \\ &\quad - \frac{1}{n}\sum_{m=k}^{k+n-1} E[X(j+n-1)X(m)] \\ &\quad - \frac{1}{n}\sum_{l=j}^{j+n-1} E[X(k+n-1)X(l)] \\ &\quad + \frac{1}{n^2}\sum_{j\le l\le j+n-1}\ \sum_{k\le m\le k+n-1} E[X(l)X(m)]. \end{aligned} \tag{2}$$

That matrix we denote by $\tilde{\Sigma} = \{E[Y(j)Y(k)] : j, k = 1, 2, \dots, N-n+1\}$. We see that the dependence structure of the process $Y(i)$ is fully determined by the covariance of the process $X(i)$. Moreover the covariance $E[X(l)X(m)]$ in formula (2) has a prefactor

$$\begin{cases} \left(1-\frac{1}{n}\right)^2, & \text{for} \quad l = j+n-1 \wedge m = k+n-1, \\ \frac{1}{n^2} - \frac{1}{n}, & \text{for} \quad (l = j+n-1 \wedge m \neq k+n-1) \\ & \quad \vee (l \neq j+n-1 \wedge m = k+n-1), \\ \frac{1}{n^2}, & \text{for} \quad l \neq j+n-1 \wedge m \neq k+n-1. \end{cases}$$

Therefore we can rewrite the formula (2) in the equivalent form

$$\begin{aligned} E[Y(j)Y(k)] &= \left(1-\frac{1}{n}\right)^2 E[X(j+n-1)X(k+n-1)] \\ &\quad + \left(\frac{1}{n^2}-\frac{1}{n}\right)\left[\sum_{m=k}^{k+n-2} E[X(j+n-1)X(m)] \right. \\ &\quad \left. + \sum_{l=j}^{j+n-2} E[X(l)X(k+n-1)]\right] \\ &\quad + \frac{1}{n^2}\sum_{j\le l\le j+n-2}\ \sum_{k\le m\le k+n-2} E[X(l)X(m)]. \end{aligned} \tag{3}$$

The average value of random variable $\sigma^2(n)$ we can now express based on (2) and (3) by covariance structure of the process $X(j)$

$$\begin{aligned} E\left[\sigma^2(n)\right] &= \frac{1}{N-n}\sum_{j=n}^{N} E\left[\left(X(j)-\tilde{X}_n(j)\right)^2\right] \\ &= \frac{1}{N-n}\sum_{j=n}^{N} E\left[Y^2(j-n+1)\right] \\ &= \frac{1}{N-n}\sum_{j=n}^{N}\left\{\left(1-\frac{1}{n}\right)^2 E\left[X^2(j)\right] + 2\left(\frac{1}{n^2}-\frac{1}{n}\right)\sum_{m=j-n+1}^{j-1} \right. \\ &\quad E[X(j)X(m)] + \frac{1}{n^2}\sum_{m=j-n+1}^{j-1} E\left[X^2(m)\right] \\ &\quad \left. + \frac{2}{n^2}\sum_{j-n+1\le k<m\le j-1} E[X(m)X(l)]\right\}. \end{aligned} \tag{4}$$

We can also express the variance of the random variable $\sigma^2(n)$

$$\begin{aligned} Var\left[\sigma^2(n)\right] &= \frac{1}{(N-n)^2} Var\left[\sum_{j=n}^{N} Y^2(j-n+1)\right] \\ &= \frac{1}{(N-n)^2}\sum_{l,m=n}^{N} Cov\left[Y^2(l-n+1), Y^2(m-n+1)\right] \\ &= \frac{1}{(N-n)^2}\sum_{l,m=n}^{N} E\left[Y^2(l-n+1)Y^2(m-n+1)\right] \\ &\quad - E\left[Y^2(l-n+1)\right]E\left[Y^2(m-n+1)\right]. \end{aligned} \tag{5}$$

The terms $E\left[Y^2(l-n+1)\right]$ and $E\left[Y^2(m-n+1)\right]$ one can compute from covariance of the process $Y(j)$ according to (3). The 4th–order moment $E\left[Y^2(l-n+1)Y^2(m-n+1)\right]$ can be expressed by covariance structure of process $Y(j)$ according to Isserlis' theorem [72], which states that for zero-mean multivariate normal vector $(X_1, X_2, \dots, X_{2n})$ the higher moment $E[X_1, X_2, \dots, X_{2n}] = \sum\prod E\left[X_i X_j\right]$, where the notation $\sum\prod$ means summing over all distinct ways of partitioning $X_1, X_2, \dots, X_{2n}$ into pairs X_i, X_j and each summand is the product of the n pairs. Therefore we have