Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2007) 1047 – 1062
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Two approximation methods to synthesize the power spectrum of fractional Gaussian noise☆

Sergio Ledesma[a,∗], Derong Liu[b], Donato Hernández[a]

[a]FIMEE School Of Engineering, University of Guanajuato, 912 Tampico, Salamanca, Gto. 36730, México
[b]Department of Electrical and Computer Engineering, University of Illinois, 851 S. Morgan Street (MC 154), Chicago, IL 60607, USA

Available online 1 March 2007

Abstract

The simplest models with long-range dependence (LRD) are self-similar processes. Self-similar processes have been formally considered for modeling packet traffic in communication networks. The fractional Gaussian noise (FGN) is a proper example of exactly self-similar processes. Several numeric approximation methods are considered and reviewed, two methods are found that are able to provide a better accuracy and less running time than previous approximation methods for synthesizing the power spectrum of FGN. The first method is based on a second-order approximation. It is demonstrated that a parabolic curve can be indirectly used to approximate the power spectrum of FGN. The second method is based on cubic splines. Despite the fact that splines cannot be used directly to approximate the power spectrum of FGN, they can, however, considerably simplify the calculations while maintaining high accuracy. Both of the methods proposed can be used to estimate the Hurst parameter using Whittle's estimator. Additionally, they can be used on synthesis of LRD sequences.

Keywords: Self-similarity; Long-range dependence; Fractional Gaussian noise; Power spectrum synthesis; Hurst parameter

1. Introduction

A standard assumption of time series analysis is that observations separated by a large time span are roughly independent. However, for long-range dependence (LRD), these observations are not independent, see Beran (1986, 1992a, b, 1994) and Fox and Taqqu (1986). The Hurst parameter, denoted by H, is the measure of LRD. Larger values of H mean greater LRD. Two methods are proposed to approximate the power spectrum of fractional Gaussian noise (FGN) (a self-similar process). These methods can be used to synthesize self-similar series and estimate the value of H for a time series. Because both methods are faster and more accurate than previous methods, they are ideal when a quick value of H using Whittle's estimator is required.

Traffic measurements obtained from packet networks have convincingly established the existence of statistical features that are characteristic of fractal traffic processes (Adas and Mukherjee, 1995; Erramilli et al., 1996; Jeong, 2002), particularly in the sense that these features regularly span many time scales, see Erramilli et al. (1994) and Taqqu

☆ The synthetic LRD sequences generated using the Spline method are given in the supplementary material.

∗ Corresponding author. Department of Electrical Engineering and Computer Science, University of Illinois at Chicago, 851 S. Morgan Street (MC 154), Chicago, IL 60607, USA. Tel.: +1 312 359 4475.
E-mail address: selo@salamanca.ugto.mx (S. Ledesma).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2007.02.017

and Teverovsky (1997). It has been shown that the concept of self-similarity can be applied to WAN and LAN traffic, see Leland et al. (1994), Paxson and Floyd (1995) and Willinger et al. (1997).

This paper is organized as follows. In Section 2, we present background information on LRD and self-similarity. In Section 3, we present a brief description of existing methods for approximating the power spectrum of FGN. In Section 4, we develop two methods for reducing the computational complexity of the power spectrum of FGN. These two methods can be used to speedily generate self-similar time series and calculate Whittle's estimator. In Section 5, we perform a statistical analysis to compare existing methods with the two methods developed in this paper. Finally, Section 6 presents conclusions and direction for future work.

2. Background information

Consider a stochastic process $X = \{X_t; t = 0, 1, 2, \ldots\}$. For each $m = 1, 2, 3, \cdots$, let

$$X^{(m)} = \{X_k^{(m)}; k = 1, 2, 3, \ldots\},$$

denote a new time series obtained by averaging the original series X; that is,

$$X_k^{(m)} = \frac{X_{km-m} + \cdots + X_{km-1}}{m}.$$

The new sequence $X_k^{(m)}$ is frequently used to estimate the Hurst parameter, see Roughan et al. (1998). Generally, the Hurst parameter can be estimated using the:

- Analysis of the variances of the aggregate processes $X^{(m)}$.
- Time-domain analysis based on the re-scaled adjusted range (R/S-statistics).
- Frequency-domain analysis based on the periodogram or Whittle's estimator.
- Wavelet-based estimator, see Veitch and Abry (1998, 1999)

The most straightforward models with LRD are the self-similar processes, see Cox (1984). X is exactly or asymptotically second-order self-similar if the corresponding aggregate processes $X^{(m)}$ are the same as X or become indistinguishable from X, at least with respect to their autocorrelation functions, see Leland et al. (1994), Beran et al. (1995), Paxson and Floyd (1995) and Willinger et al. (1995). Self-similarity and second-order self-similarity are equivalent concepts for Gaussian process since their finite dimensional distributions are, by definition, Gaussian. Fractional Gaussian noise (FGN) is an example of exactly self-similar processes, and fractional autoregressive integrated moving average process is an example of asymptotically self-similar processes.

Paxson (1997) proposed an analytical method for synthesizing FGN using the results of Flandrin (1992). This ingenious method makes the assumption that if the power spectrum of FGN is given, then it is possible to easily create a frequency-domain sample path. However, long memory processes are characterized by a nonfinite spectrum at the zero frequency, and in practice, the frequency values must be chosen so that none of them takes the value of zero, i.e. $\pi/u, 2\pi/u, \ldots, \pi$ where $u > 1$ is an integer. Thus, using the fast Fourier transform (FFT) algorithm, it is feasible to utilize this frequency-domain sample path to synthesize a time-domain sample sequence for FGN.

Following the notation of Beran (1986) for an FGN process, and letting λ be the discrete frequency, the power spectrum for $0 < H < 1$ and $-\pi \leqslant \lambda \leqslant \pi$ is

$$f(\lambda; H) = (1 - \cos\lambda)\Upsilon(H)[|\lambda|^{-2H-1} + B(\lambda; H)], \tag{1}$$

where

$$\Upsilon(H) = 2\sin(\pi H)\Gamma(2H + 1)$$

and

$$B(\lambda; H) = \sum_{k=1}^{\infty}[(2\pi k + \lambda)^{-2H-1} + (2\pi k - \lambda)^{-2H-1}]. \tag{2}$$