A NOVEL APPROACH TO THE ESTIMATION OF THE LONG-RANGE DEPENDENCE PARAMETER

By

Mohammed El Houssain Ech-Cherif El Kettani

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

(Electrical Engineering)

at the

UNIVERSITY OF WISCONSIN – MADISON

2002

In the Name of Allah, the Most Gracious, the Most Merciful

"As for those who strive hard in Us, We will surely guide them to Our Paths.

And verily Allah is with the good doers"

Qur'an V.29:69

The will is infinite,

and the execution confined.

The desire is boundless,

and the act a slave to limit

Cesaro, 1905

In the memory of my father...

To my parents:

Professor Ali Kettani and Nuzha Kettani,

Without whom I would not have gone this far.

To my sister and brothers:

Doctor Hassna'a, Sheikh Hassan and Doctor Hamza,

In recognition of their continuous moral support...

To them all, I dedicate this work.

ACKNOWLEDGMENTS

I begin by saying *"All the praises and thanks be to Allah, Who has guided us to this, and never could we have found guidance, were it not that Allah had guided us"* Qur'an V.7:43. Indeed, all thanks go first and foremost to Allah for giving me the motivation, determination, patience and paving my path to achieve my goals.

The seed of this work started during my internship in the Summer of 2000 at Los Alamos National Laboratories (LANL), Los Alamos, New Mexico, under the supervision of Dr. Wu-chung Feng. The preliminary version of this work was presented in [21, 22]. This work was partially funded by Los Alamos National Laboratories under grant 25136-001-01-4T. I am grateful to Dr. Feng for his initial support in the theory and funding that gave this research a momentum.

I was blessed by having Professor John Gubner as my academic adviser. His availability, support, encouragement, and valuable comments made this thesis possible. In fact, I can say that he was the best academic adviser that I ever had. Thank you professor for making it a happy ending to my "student life."

I am grateful to various email contacts with Drs. J. Hosking and D. Veitch. Special thanks to Professors S. Robinson, O. Mangassarian, and B. Barmish for various discussions and advice.

I am grateful to the support and understanding expressed by Chancellor John Wiley and Madison community at large to myself and my community after the events of September 11th.

Special thanks to my friends at Associated Students of Madison (ASM), Muslim Students Association (MSA), Moroccan Students Association of Wisconsin (MSAW), and Madison Area Peace Coalition (MAPC) with whom I practiced my activism.

Special thanks to my Muslim friends at the Muslim House, Islamic Center, and Masjid Assuna for practicing my religion with them and making me feel this city like home for me.

I am grateful to Professor Edriss Makward and his wife Julie and to Sal and Judy Troia for their continuous support since I came to Madison. I am thankful to them for acting like my family here.

I gracefully acknowledge the help, support, and encouragements of Dr. Humayun Akhtar and his wife Yusriya since 1998.

The loss of my father in April 2001 had a profound impact on me. I wish he was around to share this happiness with me and my family. I pray Allah that He bestow His mercy on him. After all, I am proud that I fulfilled all my promises

to him.

Last but not least, I would like to express my sincere gratitude to my parents, professor Ali Kettani and Nuzha Kettani, my sister Dr. Hassna'a, my brothers Sheikh Hassan and Dr. Hamza, for their continuous encouragement and support throughout my life. I sincerely dedicate this work to them.

ABSTRACT

A new model-testing paradigm is introduced. This paradigm is illustrated through the two long-range dependent models: Second-order self-similar, and fractional ARIMA. We then consider the parameter-estimation problem when the process is known to follow a certain model. We illustrate this new estimation method on the two aformentioned long-range dependent models. The confidence intervals and biasedness are obtained for the estimates using the new method. This new method is then applied to pseudo-random data and to real traffic data. We compare the performance of the new method to that of the widely-used wavelet method, and demonstrate that the former is much faster and produces much smaller confidence intervals of the long-range dependence parameter estimate. We believe that the new method can be used as an on-line estimation tool for the long-range dependence parameter and thus be incorporated in the new TCP algorithms that exploit the known self-similar and long-range dependent nature of network traffic.

Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

Internet users are well-acquainted with congestion, sometimes known as the "world-wide wait" (WWW). Due to variable amounts of data sent over the Internet, buffers at routers can become quite full. Thus, you may experience delays in downloading files from the Internet because your data is waiting at a congested router until it can be forwarded to you. Sometimes, buffers at routers are full, and the router must drop packets. In this case, your wait may be even longer because the dropped packets must be retransmitted.

Internet routers were originally designed using statistical models of telephone traffic. However, recent findings have shown that telephone traffic models, e.g., Poisson models, do not hold for Local Area Network (LAN) traffic [24], Wide Area Network (WAN) traffic [34], and World Wide Web (WWW) traffic [6]. Instead, models exhibiting long-range dependence and self-similarity are more appropriate. Hence, there has been much recent work on modifying TCP algorithms, buffer sizes, algorithms for smart routers, etc., to deal better with the observed congestion and packet loss.

The most well-known models of long-range dependent processes are fractional Gaussian noise [28] (thus second-order self-similarity) and fractional ARIMA [9, 18]. Each of these models has a corresponding long-range dependence parameter

β ($\beta \in (0, 2)$). The smaller the value of β, the more long-range dependent the process is. Since the value of the parameter β indicates the intensity of this dependence structure, it is important to have a better tool to estimate it, so that the estimate is not biased and the confidence intervals are as small as possible. Moreover, if we would like to estimate β on-line, then the estimation tool should be as fast as possible.

Several methods for measuring the long-range dependence parameter β have been proposed (see Appendix A for details of each one). The most well-known methods are the R/S method [28, 23], variance-time analysis [23, 24], the periodogram method [23, 24], the Whittle estimator [43], and the wavelet method [1, 32]. All these methods except Whittle's are graphical. Only the last two give confidence intervals. The least computationally intensive method of these methods is the wavelet method.

The body of this thesis is composed of three chapters, Chapters 2, 3, and 4. Chapter 2 covers background definitions and theorems to be used later. Chapter 3 proposes a method for deciding whether or not a sample of a process has a given parametric long-range dependence correlation structure. If it does, we propose in Chapter 4 a method for estimating this parameter. When compared to the widely used wavelet method, the new method gives much smaller confidence intervals and proves to be much faster. We present concluding remarks and further research directions in Chapter 5.

CHAPTER 2

PRELIMINARIES

2.1. Introduction

Long-range dependence processes were observed as early as 1895 in astronomical data sets studied by the astronomer Newcomb. More concrete work on long-range dependence and self-similar processes was presented by Hurst and Mandelbrot in early 1950s. Since then, this phenomenon was empirically shown to exist in a number of fields, such as agronomy, astronomy, chemistry, economics, engineering, environmental sciences, geosciences, hydrology, mathematics, physics, and statistics. Since the early 1990s, evidence of long-range dependence and second-order self-similarity in aggregate network traffic has continued to accumulate [24, 34, 6]. As a result, several models of long-range dependent processes been introduced, including fractional Gaussian noise [28] (thus second-order self-similarity) and fractional ARIMA [9, 18].

In network traffic, long-range dependence corresponds to slowly decaying autocorrelation functions and heavy-tailedness. The former shows the existence of nontrivial correlation structure at large scales. This in turn leads to the "$1/f$ noise," which implies larger contributions of low frequency components. Heavy-tailedness on the other hand, indicates that large sample values have a nonnegligible probability. Thus, samples drawn from a heavy-tailed distribution result in a bulk of small values and another bulk of relatively very large values. Not sur-

prisingly, this corresponds to extreme variability and slows down the convergence rate of sample statistics. This in turn explains the burstiness observed in network traffic. Such burstiness forces packets to experience long delays and some packets are even dropped due to buffer overflow. This unpleasant behavior of network traffic introduces difficulty and complexity into traffic and resource management. Nevertheless, the long-range dependence structure helps predicting future sample values.

To study network traffic, we start by considering it as a random process Y_i, $i \in \mathbf{Z}$. It is of interest to develop a model of this process so that we can predict the future values with certain probability. This will allow us to develop better congestion control mechanisms, like Transmission Control Protocol (TCP), buffer sizes, packet spacing, routers, better algorithms for smart routers, etc. This in turn will reduce the loss of data being transmitted and increase the speed of transmission. In dealing with network traffic, we are most interested in the corresponding increment process. This process was shown to be long-range dependent in the work of Leland *et al.* [24]. Thus, to study network traffic, we need to invoke the theory of probability and random processes.

This chapter introduces some mathematical definitions and properties of second-order stationarity (in Section 2.2), second-order self-similarity (in Section 2.3), and fractional ARIMA(p, d, q) models (in Section 2.4), that are essential in understanding the content of this thesis. The relation of these models to long-range

dependence and the definition of the former are discussed in Section 2.5. Section 2.6 presents some sample statistics that we use to estimate the second-order statistics of the increment process of network traffic. The bias, covariance, and distribution of the autocorrelation estimates are also presented to characterize the reliability of the estimates.

2.2. Second-Order Stationarity

Consider a discrete-time stochastic process X_i, $i \in \mathbf{Z}$ [10, 32], where X_i is viewed as the increment process of network traffic, measured in packets, bytes, or bits. We say that X_i is *strongly stationary* if the families $(X_{i_1}, X_{i_2}, \ldots, X_{i_n})$ and $(X_{i_1+k}, X_{i_2+k}, \ldots, X_{i_n+k})$ have the same joint distribution for all $i_1, i_2, \ldots, i_n, k \in \mathbf{Z}$ and positive integers n. This form of stationarity turns out to be highly restrictive, and a weaker condition suffices for our purposes.

We say that X_i is *second-order stationary* if its mean and autocovariance function, respectively, satisfy

$$E(X_i) = E(X_j), \tag{2.1}$$

and

$$\gamma(X_i, X_j) = \gamma(X_{i+k}, X_{j+k}), \tag{2.2}$$

for all i, j and $k \in \mathbf{Z}$. The notation γ denotes the autocovariance function defined

as

$$\gamma(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))]. \tag{2.3}$$

Thus, if X_i is a second-order stationary process, it has a constant mean, and its autocovariance function is a function of $k = i - j$ only, which allows us to write $\gamma(X_i, X_j) = \gamma(k)$. Note that $\gamma(k)$ is an even function for real processes, i.e., $\gamma(k) = \gamma(-k)$. We put $\sigma^2 = \gamma(0)$ and $\rho(k) = \frac{\gamma(k)}{\sigma^2}$ to denote the variance and autocorrelation function of the process X_i.

For traffic-modeling purposes, we would like X_i to be at least second-order stationary so that its behavior or structure is invariant with respect to shifts in time. Without this property, a model loses much of its usefulness as a compact description of the assumed tractable phenomena [32].

2.3. Second-Order Self-Similarity

The notion of self-similarity or scale-invariance arises in many fields. To get a deeper understanding of this notion we start by explaining the self-similarity phenomenon on geometric images. A geometric image is said to be self-similar if there exists a piece of this image (a self-similar piece) that if magnified properly will give exactly the same image. An example of such is given in Figure 2.1 which is known as the Sierpinski triangle. The dimension of such figures is a number D

defined as

$$D = \frac{\ln(\text{Number of self-similar pieces})}{\ln(\text{Magnification factor})}. \tag{2.4}$$

Geometric images with dimension D are called *fractals*. The notion of fractals was first introduced by Mandelbrot (see [28] for details). The triangle in Figure 2.1 has $D \approx 1.58$ since it has three self-similar pieces, each can be magnified by a factor of two to get the original triangle.

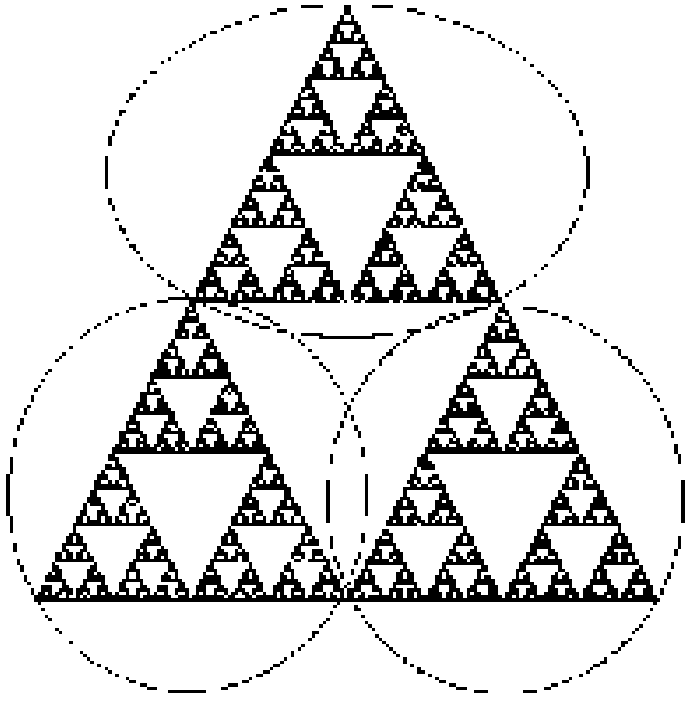

Figure 2.1. An example of a self-similar image.

Statisticians and probabilists on the other hand used this scale-invariant phenomenon to define self-similar processes, which are defined as follows [32]. Y_t is a self-similar process with self-similarity parameter H if for all $a > 0$ and $t \geq 0$,

$$Y_t =_d a^{-H} Y_{at},$$

where the notation $=_d$ denotes equality in distribution. This means that Y_t and its normalized (by a^{-H}) time scaled version Y_{at} have the same distribution.

In the network traffic modeling context, Y_t can be thought of as the cumulative process or the total traffic up to time t. Analogous to fractals, for $a > 1$ where time is dilated, a contraction factor a^{-H} is applied so that the magnitudes of Y_{at} and Y_t are comparable. Likewise, for $a < 1$ the opposite holds true.

The self-similarity parameter H is also known as the Hurst parameter (which explains the notation) named after the British hydrologist H. E. Hurst (1880–1978) who was studying the Nile river minima and published his observations in [19]. Negative values of H are prohibited since the corresponding Y_t is not a measurable process [40, 41]. The value $H = 0$ is not interesting since it implies that for all $t > 0$, $Y_t = Y_1$ with probability one.

We consider the case where Y_t has finite variance and stationary increments. We also take $Y_0 = 0$ with probability one. Thus, define the increment process $X_i = Y_i - Y_{i-1}$ $(i = 1, 2, \ldots)$. With this setup, it can be shown that X_i has zero

mean and a correlation given by

$$\rho(k) = \frac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}), \quad k \geq 1. \tag{2.5}$$

A process with the same second-order statistics as X_n is referred to as a *second-order self-similar process.*

Define the *aggregated process* $X_n^{(m)}$ of X_i at *aggregation level* m as

$$X_n^{(m)} = \frac{1}{m} \sum_{i=m(n-1)+1}^{mn} X_i. \tag{2.6}$$

That is, to create $X_n^{(m)}$, partition X_i into non-overlapping blocks of size m, average their values, and then use n to index these blocks. In network traffic, this can be viewed as "zooming out" in time, i.e., since many increments of X_i are juxtaposed in time, zooming out makes the juxtaposed increments appear as one increment, as shown in [23].

Let $\rho^{(m)}(k)$ denote the autocorrelation function of $X_n^{(m)}$. Then, X_i is said to be *asymptotically second-order self-similar* if for $k \geq 1$

$$\lim_{m\to\infty} \rho^{(m)}(k) = \frac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}). \tag{2.7}$$

Let $\rho(k)$ denote the *autocorrelation function* of X_i, i.e., $\rho(k) = \frac{\gamma(k)}{\sigma^2}$. A value $H > 1$ is prohibited since it contradicts the fact that $|\rho(k)| \leq 1$ for all k. The case $H = 1$ implies that $\rho(k) = 1$ for every k, which is of no practical importance. Note also that with $H = \frac{1}{2}$, the X_i's are uncorrelated. Hence, throughout this thesis, we consider the range $0 < H < 1$ only. The presence of self-similarity in

network traffic was observed by Leland *et al.* [23, 24]. Since then, second-order self-similarity has become a dominant framework for modeling network traffic.

An example of the self-similar process Y_t is fractional Brownian motion, first introduced by Mandelbrot [28]. The corresponding increment process X_i is fractional Gaussian noise. When $H = \frac{1}{2}$, fractional Brownian motion coincides with the ordinary Brownian motion.

2.4. Fractional ARIMA Models

We start this section by first defining autoregressive integrated moving-average processes or simply ARIMA(p, d, q). Let X_i be such process. We assume that $\mu = E(X_i) = 0$, otherwise X_i must be replaced by $X_i - \mu$ in all subsequent formulas. Let B denote the backward shift operator defined by $BX_i = X_{i-1}$. Hence, we can write $X_i - X_{i-1} = (1 - B)X_i$. Let p and q be non-negative integers.

Define the polynomials

$$\phi(x) = 1 - \sum_{i=1}^{p} \phi_i x^i,$$

and

$$\psi(x) = 1 - \sum_{i=1}^{q} \psi_i x^i,$$

with the notion that if p or q is zero then the sum is suppressed.

Assume that all solutions of $\phi(x) = 0$ and $\psi(x) = 0$ are outside the unit circle to guarantee stationarity. Furthermore, let ϵ_i be zero mean independent and identically distributed normal variables with variance σ_ϵ^2. An ARIMA(p, d, q) is defined to be the stationary solution of

$$\phi(B)(1 - B)^d X_i = \psi(B)\epsilon_i.$$

The fractional ARIMA(p, d, q) model proposed in [9, 17] is an extension of the ARIMA(p, d, q) in the sense that d is allowed to take any real value in the interval $(\frac{-1}{2}, \frac{1}{2})$. In this context, a binomial expansion in terms of the backward shift operator B is used to express the differencing operator $(1 - B)^d$, namely

$$(1 - B)^d = \sum_{i=0}^{\infty} \frac{\Gamma(i - d)}{\Gamma(i + 1)\Gamma(-d)} B^i,$$

where $\Gamma(\cdot)$ denotes the gamma function.

We remark that a relationship can be drawn between the general fractional ARIMA(p, d, q) process X_i and the *standard* fractional ARIMA$(0, d, 0)$ process X_i^*. More precisely, X_i is obtained by passing X_i^* through the linear filter

$$\xi(B) = \sum_{i=0}^{\infty} \xi_i B^i = \phi(B)\psi^{-1}(B).$$

The coefficients ξ_i can be calculated by matching the powers of the polynomial quotient $\phi(B)\psi^{-1}(B)$ with those of $\xi(B)$. For instance, if

$$\phi(B) = 1$$

and

$$\psi(B) = 1 - \psi_1 B,$$

then we obtain

$$\phi(B)\psi^{-1}(B) = 1 - \psi_1 B + \psi_1^2 B^2 - \cdots$$

Therefore,

$$\xi_i = (-1)^i \psi_1^i.$$

The next step, if we let $\gamma(k)$ and $\gamma^*(k)$ denote the autocovariances of the processes X_i and X_i^*, respectively, then the following relationship holds:

$$\gamma(k) = \sum_{i,j=0}^{\infty} \xi_i \xi_j \gamma^*(k+i-j).$$

In practice, we want both p and q to be small [17]. In fact, in the properties of an ARIMA(p,d,q) process at high lags or at low frequencies are similar to those of an ARIMA$(0,d,0)$ for the same d value.

Hence, in the remainder of this thesis, we consider the fractional ARIMA$(0,d,0)$ process only. In this case, the autocovariances follow using a formula in [8]

$$\gamma(k) = \frac{\sigma_\epsilon^2 (-1)^k \Gamma(1-2d)}{\Gamma(1+k-d)\Gamma(1-k-d)}.$$

The correlation function then follows and can be further simplified using the extension formula of gamma functions

$$\begin{aligned} \rho(k) &= \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(1+k-d)} \\ &= \prod_{i=1}^{k} \frac{k-i+d}{k-i+1-d}. \end{aligned} \tag{2.8}$$

The asymptotic relation between second-order self-similar processes and fractional ARIMA$(0,d,0)$ processes, in addition to the long-range or short-range

dependence that they imply, depending on the value of their parameter is the topic of the following section.

2.5. Implications and Long-Range Dependence

Let X_i be a second-order self-similar or a fractional ARIMA$(0, d, 0)$ process. If the aggregated process $X^{(m)}$ is viewed as the sample mean of X, i.e.,

$$X^{(m)} = \frac{1}{m} \sum_{i=1}^{m} X_i,$$

then it is not hard to show that

$$(\sigma^{(m)})^2 \approx \sigma^2 m^{-\beta}, \tag{2.9}$$

as $m \to \infty$, where $(\sigma^{(m)})^2$ is the variance of $X^{(m)}$, $\beta = 2 - 2H$ in the second-order self-similar case and $\beta = 1 - 2d$ in the fractional ARIMA$(0, d, 0)$ case. In fact, for exact second-order self-similar processes, (2.9) holds with equality for all values of m.

It also follows from (2.5) and (2.8) that the autocorrelation functions decay hyperbolically rather than exponentially fast. More precisely,

$$\rho(k) \sim c_\rho k^{-\beta} \quad \text{as} \quad k \to \infty, \tag{2.10}$$

where β is as before, and the constant $c_\rho = H(2H - 1)$ in the second-order self-similar case, and $c_\rho = \Gamma(1 - d)/\Gamma(d)$ in the fractional ARIMA$(0, d, 0)$ case.

Let $f(\lambda)$ denote the spectral density function corresponding to the increment process X_i. By definition, the spectral density function $f(\lambda)$ is the discrete-time Fourier-transform of the autocorrelation function $\rho(k)$, i.e.,

$$f(\lambda) = \sum_{k=-\infty}^{\infty} e^{-j2\pi nk\lambda}\rho(k),$$

where $j^2 = -1$.

As noted in [5, p. 43], (2.10) is equivalent to[1]

$$f(\lambda) \sim c_f|\lambda|^{-\alpha} \quad \text{as} \quad \lambda \to 0, \tag{2.11}$$

where $\alpha = 1 - \beta$ and

$$c_f = \frac{\sigma^2 c_\rho}{\Gamma(1-\alpha)\sin(\frac{\alpha\pi}{2})}.$$

We say the process or time series X_i is *long-range dependent* if

$$\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty. \tag{2.12}$$

The process is *short-range dependent* otherwise.

Thus, in view of (2.10), for $\beta \in (0, 1]$ the process X_i is long-range dependent and it is short-range dependent for $\beta \geq 1$. From (2.11), when X_i is long-range

[1]The equivalence is in the sense that (2.10) and (2.11) each imply (2.12). Strictly speaking, from a purely mathematical point of view, this equivalence does not hold; see [12] for more details.

dependent, the spectral density $f(\lambda)$ diverges around the origin implying larger contributions of low frequency components. In this case we say that the spectral density obeys a power law near the origin and we deal with "$1/f$ noise" [28].

The presence of long-range dependence has both positive and negative effects. For such processes, the forecasting becomes easier in the sense that good short- and long-term predictions can be obtained when a long record of past values is available (see [5] for details). On the other hand, the classical time series approach to estimating the second-order statistics assumes short-range dependence. The presence of long-range dependence has been shown to significantly slow the speed of convergence of the estimates. The rate of convergence becomes slower as the long-range dependence increases (β decreases). The following section discusses this issue in more detail.

2.6. Statistical Sampling

Since we are dealing with measurements and no a priori probability density function, we use statistical sampling to estimate the second-order statistics of the process.

Let X_i be a second-order stationary process with mean, variance, and covariance μ, σ^2, and $\gamma(k)$, respectively. The *sample mean* and the *sample covariance*

are given by the following formulas:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{2.13}$$

and

$$\hat{\gamma}_n(k) = \frac{1}{n}\sum_{i=1}^{n-k}(X_i - \hat{\mu}_n)(X_{i+k} - \hat{\mu}_n), \tag{2.14}$$

where n is the number of samples to be used. The *sample variance* is given by

$$\hat{\sigma}_n^2 = \hat{\gamma}_n(0).$$

Likewise, the *sample autocorrelation* is given by

$$\hat{\rho}_n(k) = \frac{\hat{\gamma}_n(k)}{\hat{\sigma}_n^2}. \tag{2.15}$$

The relationship between these estimates and the estimated parameters (the asymptotic distribution of the difference, biasedness, etc.) was studied extensively for both the short-range and long-range dependence cases (see [3] for details on the former and [18] on the latter). Hosking's results [18] focused on processes with hyperbolically decaying autocorrelation functions as in (2.10). In this report, we will be interested in the relationship between $\hat{\rho}$ and ρ.

In [18], Hosking assumes that the process X_i has the representation

$$X_i = \mu + \sum_{j=0}^{\infty} \psi_j a_{i-j}, \tag{2.16}$$

where

$$\psi_j \sim \delta j^{-\frac{1}{2}(1+\beta)}, \quad \delta > 0, \quad \text{as} \quad j \to \infty, \tag{2.17}$$

and a_i is a white-noise process consisting of independent, identically distributed $N(0, \sigma^2)$ random variables. As noted by Hosking [18], both fractional Gaussian noise and fractional ARIMA processes have the representation (2.16) – (2.17). For such process, the following two theorems (see [18, Theorems 6 and 7]) provide the asymptotic bias, covariance, and limiting distribution of the sample autocovariances $\hat{\rho}_n(k), k \geq 1$.

Theorem 2.6.1. *Let X_i be a time series satisfying (2.16) and (2.17). Then as $n \to \infty$, the* asymptotic bias *and* covariance *of $\hat{\rho}(k)$, $k \geq 1$, are given by*

$$E(\hat{\rho}_n(k)) - \rho(k) \sim \frac{-2c_\rho(1-\rho(k))n^{-\beta}}{(1-\beta)(2-\beta)}, \tag{2.18}$$

$$\mathsf{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)) \sim 2c_\rho^2(1-\rho(k))(1-\rho(l))K_2 n^{-2\beta}, \quad \textit{if } 0 < \beta < \frac{1}{2}, \tag{2.19}$$

$$\mathsf{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)) \sim 4c_\rho^2(1-\rho(k))(1-\rho(l))n^{-1}\log(n), \quad \textit{if } \beta = \frac{1}{2}, \tag{2.20}$$

$$\begin{aligned}\mathsf{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)) \sim\ & n^{-1}\sum_{s=-\infty}^{\infty}\Big[\rho(s)\rho(s+k-l) + \rho(s)\rho(s+k+l) + 2\rho(k)\rho(l)\rho(s)^2 \\ & -2\rho(k)\rho(s)\rho(s+l) - 2\rho(l)\rho(s)\rho(s+k)\Big], \ \textit{if } \frac{1}{2} < \beta < 2,\end{aligned} \tag{2.21}$$

where K_2 is defined shortly.

Theorem 2.6.2. *Let X_i be a time series satisfying (2.16) and (2.17). Then as $n \to \infty$,*

1. *If* $0 < \beta < \frac{1}{2}$, *and* $R_k := n^{\beta}(\hat{\rho}_n(k) - \rho(k))/(1 - \rho(k))$, *then as* $n \to \infty$, *the common limiting distribution of the* R_k *has* r*th cumulant*

$$\kappa_r = c_\rho^r 2^{r-1}(r-1)! K_r, \tag{2.22}$$

where

$$K_1 = \frac{-2}{(1-\beta)(2-\beta)},$$

$$K_r = \int_0^1 \cdots \int_0^1 g(x_1, x_2) g(x_2, x_3) \cdots g(x_{r-1}, x_r) g(x_r, x_1) dx_1 dx_2 \cdots dx_r, \quad r \geq 2,$$

with

$$g(x,y) = |x-y|^{-\beta} - \frac{x^{1-\beta} + (1-x)^{1-\beta} + y^{1-\beta} + (1-y)^{1-\beta}}{1-\beta} + \frac{2}{(1-\beta)(2-\beta)}.$$

2. *If* $\beta = \frac{1}{2}$, *and* $R_k := (n/\log n)^{\frac{1}{2}}(\hat{\rho}_n(k) - \rho(k))/(1 - \rho(k))$, *then as* $n \to \infty$, *the common limiting distribution of the* R_k *is* $N(0, 4c_\rho^2)$.

3. *If* $\frac{1}{2} < \beta < 2$, *and* $R_k := n^{\frac{1}{2}}(\hat{\rho}_n(k) - \rho(k))$, *then as* $n \to \infty$, R_k *has a limiting distribution that is multivariate normal zero mean and covariances given by* n *times (2.21).*

We note here that this latter result for $\beta \in (\frac{1}{2}, 2)$ was found by Anderson [3].

From (2.19) it is seen that the smaller β is (or equivalently, the higher the long-range dependence) the slower the decay of the bias of the sample autocorrelations as the sample size n increases.

Note that the covariance expression in (2.21) can be further simplified in the second-order self-similar and fractional ARIMA$(0, d, 0)$ cases. In the second-order self-similar case (2.21) is reduced to

$$\begin{aligned}\mathsf{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)) \sim n^{-1}\Big[&\sum_{s=-S}^{S}\Big(\rho(s)\rho(s+k-l) \\ &+\rho(s)\rho(s+k+l) + 2\rho(k)\rho(l)\rho(s)^2 - 2\rho(k)\rho(s)\rho(s+l) - 2\rho(l)\rho(s)\rho(s+k)\Big) \\ &+\{2H(2H-1)\}^2(1+\rho(k)\rho(l)-\rho(k)-\rho(l))\Big(\zeta(4-4H) - \sum_{s=-S}^{S} s^{4H-4}\Big)\Big], \\ &\text{if} \quad 0 < H < \frac{3}{4}, \end{aligned} \tag{2.23}$$

where S is a sufficiently large number for which an approximate equality holds in (2.10) and $\zeta(\cdot)$ is the *zeta function*, defined as

$$\zeta(x) = \sum_{k=1}^{\infty} k^{-x}.$$

In the fractional ARIMA$(0, d, 0)$ case, (2.21) is reduced to (this result follows from [18])

$$\begin{aligned}&\mathsf{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)) \sim \\ &\quad n^{-1}g(\rho^*(k-l) + \rho^*(k+l) + 2\rho(k)\rho(l) - 2\rho(k)\rho^*(l) - 2\rho^*(k)\rho(l)), \\ &\quad \text{if} \quad \frac{-1}{2} < d < \frac{1}{4}, \end{aligned} \tag{2.24}$$

where

$$g = \frac{\Gamma^4(1-d)\Gamma(1-4d)}{\Gamma^4(1-2d)},$$

$$\rho^*(k) = \frac{\Gamma(1-2d)\Gamma(k+2d)}{\Gamma(2d)\Gamma(1+k-2d)},$$

and $\rho(k)$ is given by (2.8).

2.7. Related Work

Several methods have been developed to estimate the long-range dependence parameter. The widely used method in the networking community is the wavelet method. Thus, in this section we consider the wavelet method, which we will use to compare the proposed method with. Other methods are described in Appendix A. The wavelet method is due to Abry and his co-workers (see [1, 2] and references therein), and is based on wavelets and the spectral relation of (2.11). A MATLAB program has been developed by Abry and Veitch that makes the use of this method even easier.

Let ψ_0 denote the mother wavelet. Construct other wavelets $\psi_{j,k}$ such that

$$\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k), k \in \mathbf{Z}\}.$$

Let $d_X(j,k)$ denote the projection of the data set X onto the wavelet $\psi_{j,k}$, namely

$$d_X(j,k) = \langle X, \psi_{j,k} \rangle, \tag{2.25}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. Then define the non-parametric unbiased random variable μ_j of the variance of the process $d_X(j,\cdot)$ as

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_X(j,k)|^2, \tag{2.26}$$

where n_j is the number of coefficients at octave j.

Let $y_j = \log_2(\mu_j)$, then it turns out that for a long-range dependent process X, y_j is asymptotically normally distributed:

$$y_j \sim N\Big(j\alpha + \log_2(c_f C), \frac{2(\log_2 e)^2}{n_j}\Big), \tag{2.27}$$

where α and c_f are defined in (2.11) and C is a constant.

This method gives a plot of y_j against the octave j together with confidence intervals about y_j (which is asymptotically normally distributed under some assumptions). This plot is called the *logscale diagram*. From (2.27), the logscale plot results in a straight line for large values of n_j. The slope of this straight line is the estimated long-range dependent parameter $\hat{\alpha}$.

2.8. Summary

Various notations and mathematical definitions and properties were introduced in this chapter which will be used throughout this thesis. Mainly, we started by defining second-order stationary processes, second-order self-similar processes, and fractional ARIMA processes. We then defined the long-range dependence phenomena, which is implied by the latter two processes. We than provided the distribution of the sample autocorrelation function. These ideas will be exploited to develop the new method.

CHAPTER 3
A NEW METHOD TO DECIDE WHETHER OR NOT A PROCESS HAS A GIVEN PARAMETRIC CORRELATION STRUCTURE

3.1. Introduction

We begin this chapter by proposing a method in Section 3.2 that uses the structure of the autocorrelation function of the model that we would like to fit the data to. We estimate the autocorrelation function of the data as described in Section 2.6, and then apply a curve-fitting criterion, which we call the optimization method. If the resultant error is high, then the given process fails the test and may not be considered to follow that particular model. Otherwise, the process is assumed to be long-range dependent following the assumed model with the parameter as estimated. The criteria that we followed to decide whether the error is large or not is the probability of false alarm. For a wide range of the parameter, we develop a relation between the the probability of false alarm and the cutoff to decide how large the error can be.

In Section 3.3 we perform an empirical study using artificial data to make a proper decision on the cutoff. The pseudo-random data was generated by MATLAB, simulating second-order self-similar, and fractional ARIMA$(0, d, 0)$ data. We used two sets of real data that are known to be second-order self-similar, the

yearly minimum water levels of the Nile River at the Roda Gauge, Egypt that was studied by H. E. Hurst [19], and Ethernet measurements for a local area network at Bellcore analyzed by Leland *et al.* We also applied the new testing method to artificial and real data that is known *not* to be second-order self-similar or fractional ARIMA.

Although we focus on processes that are long-range dependent, mainly second-order self-similar and fractional ARIMA$(0, d, 0)$ processes, the new method is readily generalizable to other processes. In other words, the new method can be used as a tool to check the validity of a model in characterizing certain process. We end the chapter by providing a summary in Section 3.4.

3.2. Optimization Method

Suppose we have a record of network traffic that records the total packets or bytes received in a period of time. Let X_i denote its increment process. There is always a vast interest in the process X_i by the networking community, since understanding the behavior of this process allows us to develop better congestion control mechanisms, like Transmission Control Protocol (TCP), buffer sizes, packet spacing, routers, better algorithms for smart routers, etc.

Define *the error function* $E_K(\beta)$ as

$$E_K(\beta) = \frac{1}{4K} \sum_{k=1}^{K} \{\rho(k) - \hat{\rho}_n(k)\}^2, \tag{3.1}$$

where $\rho(k)$ denotes the autocorrelation function of the model with parameter β that we would like to fit the data to, $\hat{\rho}_n(k)$ is the sample autocorrelation function of the data and K is the largest value of k for which $\hat{\rho}_n(k)$ is to be computed to reduce edge effects. Since we estimate the model parameter β based on optimizing (3.1), we call this estimation method the *optimization method*.

3.2.1. How Small Should $E_K(\beta)$ Be?

We expect $E_K(\beta)$ to be close to zero if X_i is close to the model. The estimated parameter $\hat{\beta}$ is chosen so that $E_K(\hat{\beta})$ is the minimum of the error function over the appropriate range of the parameter. It is easily seen that the highest $E_K(\hat{\beta})$ can be is 1. Thus, we consider that the prescribed model fits the process X_i if $E_K(\hat{\beta}) = e$, where e is "much smaller than 1".

Let U_K be the row vector with entries

$$u_k = \rho(k) - \hat{\rho}_n(k), \quad k = 1, 2, \ldots, K.$$

Thus, the error function can be rewritten as

$$E_K(\beta) = \frac{U_K U_K^T}{4K}. \tag{3.2}$$

Define *the probability of false alarm* as

$$P_{FA} = P(E_K(\hat{\beta}) \geq e | R = R_\beta), \tag{3.3}$$

where R is the covariance matrix of U_K and the condition $R = R_\beta$ denotes that the process follows the model with the corresponding covariance $R = R_\beta$. Thus, we would like to pick e so that $P_{FA} \leq 0.05$.

Next, note that

$$\begin{aligned} P_{FA} &= P(E_K(\hat{\beta}) \geq e | R = R_\beta) \\ &\leq P(E_K(\beta) \geq e | R = R_\beta) \\ &= P(\frac{U_K U_K^T}{4K} \geq e | R = R_\beta) \\ &= P(U_K U_K^T \geq 4Ke | R = R_\beta) \\ &= 1 - S_K(4Ke), \end{aligned} \tag{3.4}$$

where S_K is the cdf of $U_K U_K^T$.

Theorem 3.2.1. *For $\beta \in [\frac{1}{2}, 2)$, S_K is the cumulative distribution function of* Stacy's distribution *with the corresponding probability density function given by*

$$s_{K,\lambda}(t) = t^{\frac{K}{2}-1} \sum_{k=0}^{\infty} \frac{(-t)^k}{\Gamma(\frac{K}{2}+k)} \sum_{\sum_{l=1}^{K} m_l = k} \prod_{n=1}^{K} \frac{\Gamma(m_n + \frac{1}{2})(2\lambda_n)^{m_n + \frac{1}{2}}}{m_n! \sqrt{\pi}}, \tag{3.5}$$

where $\Gamma(\cdot)$ denotes the gamma function, and $\lambda = [\lambda_1, \ldots, \lambda_K]$, where λ_n, for $n = 1, 2, \ldots, K$, are the eigenvalues *of the covariance matrix R_β of U_K.*

Proof: From Theorem 2.6.2, for $\beta \in [\frac{1}{2}, 2)$, U_K is zero mean multivariate normally distributed with covariance R with the corresponding entries

$$r_{kl} = \mathsf{cov}(\hat{\rho}_n(k), \hat{\rho}_n(l)).$$

The distribution of $U_K U_K^T$ has the following characteristic function [37, p. 65]:

$$\begin{aligned} \Phi(\omega) &:= E[e^{j\omega U_K U_K^T}] \\ &= \prod_{n=1}^{K} (1 - j\omega 2\lambda_n)^{-\frac{1}{2}}. \end{aligned} \tag{3.6}$$

But this is the characteristic function of the sum of the independent random variables v_n, $n = 1, 2, \ldots, K$, each distributed as $\Gamma(\frac{1}{2}, 2\lambda_n)$, where $\Gamma(a, b)$ is the Gamma distribution with probability density function

$$\gamma_{a,b}(x) = \frac{x^{a-1} e^{-\frac{x}{b}}}{b^a \Gamma(a)}, \quad x > 0.$$

The distribution of such sums is discussed by Stacy [36], from which we get Stacy's distribution of the Theorem. This concludes the proof.

We next consider the asymptotics of $S_K(t)$. To do this, we invoke the following form of the *Central Limit Theorem* (see [26, p. 287] for proof).

Theorem 3.2.2. *Let $w_1, w_2, \ldots$ be independent variables satisfying*

$$E(w_j) = 0, \quad \mathsf{var}(w_j) = \sigma_j^2, \quad E(|w_j^3|) < \infty$$

and such that

$$\frac{1}{\sigma(n)^3} \sum_{j=1}^{n} E(|w_j^3|) \to 0 \quad as \quad n \to \infty \tag{3.7}$$

where

$$\sigma(n)^2 = \sum_{j=1}^{n} \mathsf{var}(w_j).$$

Then as $n \to \infty$,

$$\frac{1}{\sigma(n)} \sum_{j=1}^{n} w_j \xrightarrow{D} N(0,1).$$

This gives us the following result.

Lemma 3.1. *For* $\beta \in [\frac{1}{2}, 1)$, *suppose (3.7) holds, then for large* K,

$$S_K \sim N(\sum_{j=1}^{K} \lambda_j, 2\sum_{j=1}^{K} \lambda_j^2).$$

Proof: The proof is obvious.

Note that in our analysis, (3.7) is equivalent to

$$\frac{\sum_{j=1}^{K} \lambda_j^3}{(\sum_{j=1}^{K} \lambda_j^2)^{\frac{3}{2}}} \to 0 \quad \text{as} \quad K \to \infty.$$

An empirical study shows that for $\beta \in [\frac{1}{2}, 1)$, as K increases the ratio decreases. Thus, in the remainder we apply Lemma 3.1 to find the right number e.

Figures 3.1 and 3.2 present the minimum e to get $P_{FA} < 0.05$ for different β values in $[\frac{1}{2}, 2)$ in the second-order self-similar case ($H = 1 - \frac{\beta}{2}$) and fractional ARIMA case ($d = \frac{1-\beta}{2}$), respectively. To obtain the corresponding e value, the Normal distribution of Lemma 3.1 was applied with $n = 4000$ and $K = 50$. From

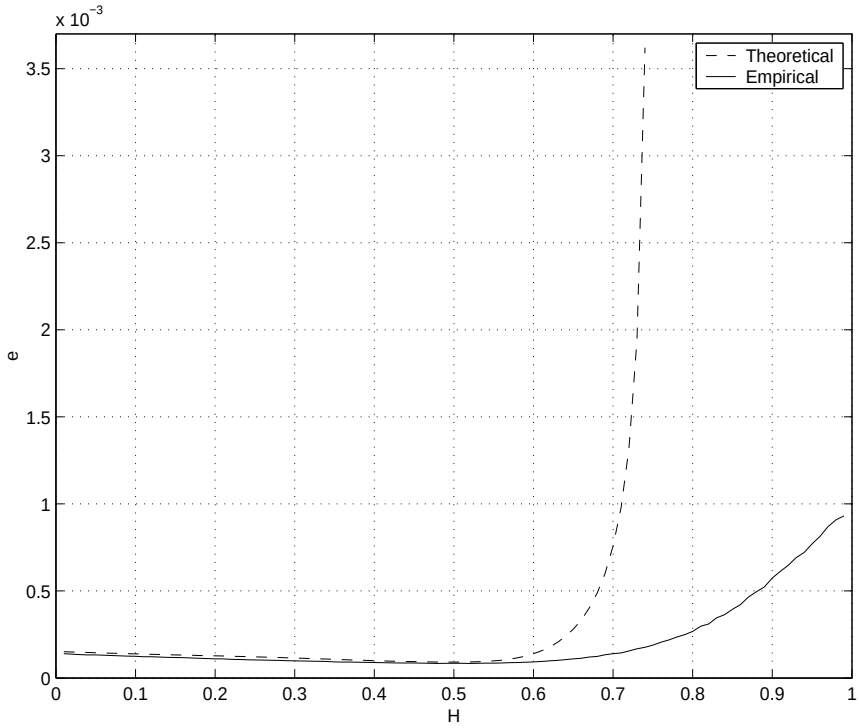

Figure 3.1. A plot of $argmin_e 1 - S_K(4Ke)$ in (3.4) and the empirical upper bound e vs. the Hurst parameter $H = 1 - \frac{\beta}{2}$ in the second order self-similar case. Here S_K is taken to be normal, based on Lemma 3.1.

Figure 3.1, we see that $e \approx 1.5 \times 10^{-4}$ for $H < 0.60$. As the long range dependence increases ($H \in (0.60, 0.75)$), the e increases up to about 3.5^{-3}. Similar comments apply to the fractional ARIMA$(0, d, 0)$ case (Figure 3.2). As can readily be seen from (3.4), this e is an upper bound on the actual e.

In the following section, an empirical study is performed to find the appropriate value e for the whole range of the parameter.

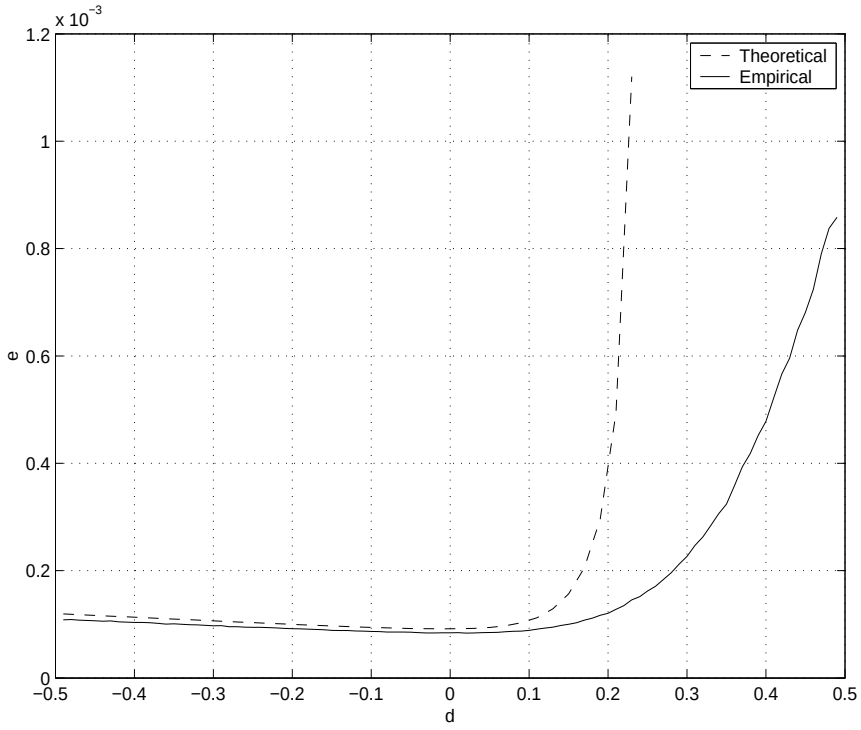

Figure 3.2. A plot of $argmin_e 1 - S_K(4Ke)$ in (3.4) and the empirical upper bound e vs. the difference parameter $d = \frac{1-\beta}{2}$ in the fractional ARIMA$(0, d, 0)$ case. Here S_K is taken to be normal, based on Lemma 3.1.

3.3. Empirical Study

In this section, we perform an empirical study to obtain the cutoff value e. For this purpose, we used second-order self-similar artificial data (Section 3.3.1), fractional ARIMA$(0, d, 0)$ artificial data (Section 3.3.2), and real data (Section 3.3.3).

3.3.1. Second-Order Self-Similar Artificial Data

For each $H = 0.01, 0.02, \ldots, 0.99$, we start by generating 10^4 realizations of a fractional Gaussian noise using MATLAB. The length of each realization is $n = 4000$ points. In our analyses we take $K = 50$ in (3.1) to reduce edge effects. Based on (3.3), for each H, e is chosen so that 95% of the obtained $E_K(\hat{H})$ are smaller than e. The plot of e versus H is given in Figure 3.1. For $H < 0.60$, the the upper bound on e (theoretical e in (3.4)) and the empirical value of e almost coincide. As the As the long-range dependence increases, the upper-bound becomes looser and looser. On the other hand, as the number of realizations increases, the empirical plot of e gets smoother and decreases slightly on the range $H \in (0.80, 1)$ (not shown).

Since the empirical plot of e vs. H is less than 10^{-3} for $H \in (0, 1)$, the cutoff that we take as our measuring criteria is the value $e = 10^{-3}$. Thus, if the error function is less than 10^{-3} (i.e., $E_K(H) < 10^{-3}$), then the process X_i is accepted as fitting the prescribed second-order self-similar model, otherwise it is not (which may be also due to the lack of sufficient data).

To illustrate the power of our test, consider the following short-range dependent process. Let X_i be an ARIMA$(0.90, 0, 0)$ = AR(0.90). Then applying the optimization method results in $E_K(\hat{H}) = 7.4 \times 10^{-3}$. Thus, the optimization method declares the process X_i not second-order self-similar in agreement with the truth.

3.3.2. Fractional ARIMA$(0, d, 0)$ Artificial Data

For each $d = -0.49, -0.48, \ldots, 0.49$, we start by generating 10^4 realizations of a fractional ARIMA$(0, d, 0)$ using MATLAB. The length of each realization is $n = 4000$ points. In our analyses we take $K = 50$ in (3.1) to reduce edge effects. Based on (3.3), for each d, e is chosen so that 95% of the obtained $E_K(\hat{d})$ are smaller than e. The plot of e versus d is given in Figure 3.2. The plot of e versus H is given in Figure 3.1. For $d < 0.10$, the the upper bound on e (theoretical e in (3.4)) and the empirical value of e almost coincide. As the long-range dependence increases, the upper-bound becomes looser and looser. On the other hand, as the number of realizations increases, the empirical plot of e gets smoother and decreases slightly on the range $H \in (0.30, 0.50)$ (not shown).

The empirical plot of e vs. d is less than 10^{-3} for $d \in (\frac{-1}{2}, \frac{1}{2})$. Therefore, the cutoff that we take as our measuring criteria is the value $e = 10^{-3}$. Thus, if the error function is less than 10^{-3} (i.e., $E_K(d) < 10^{-3}$), then the process X_i is accepted as fitting the prescribed fractional ARIMA$(0, d, 0)$ model, otherwise it is not (which may be also due to the lack of sufficient data).

To illustrate the power of our test, consider the following short-range dependent process. Let X_i be an ARIMA$(0.90, 0, 0)$ = AR(0.90). Then applying the optimization method results in $E_K(d) = 6.9 \times 10^{-3}$. Thus, the optimization

method declares the process X_i not fractional ARIMA$(0, d, 0)$ in agreement with the truth.

3.3.3. Real Data

In this section, we study standard measurements where second-order self-similarity was observed. Namely, the Nile river data (Figure 3.3), and the Bellcore data (Figure 3.4). Throughout this study, we take $K = 50$ in the error function formula (3.1). We then consider a process that is *not* second-order self-similar, namely the Variable Bit Rate data. A summary of the results is provided in Table 3.1

Figure 3.3 presents the yearly minimum water levels of the Nile river measured at the Roda Gauge near Cairo, Egypt, for the years $622 - 1281$ [39]. This data led Hurst to the observation of what was later called the Hurst effect, or self-similarity [19]. Applying the optimization method to this data results the error function $E = 2.15 \times 10^{-4}$. Thus, the optimization method acknowledges the fact that this process is second-order self-similar.

Figure 3.4 displays Ethernet measurements for a local area network traffic at Bellcore, Morristown, New Jersey [23, 24]. It was collected on August 29, 1989 and lasted for about fourteen minutes. Each observation represents the number of packets sent over the Ethernet per 100 ms. This data was considered by Leland *et*

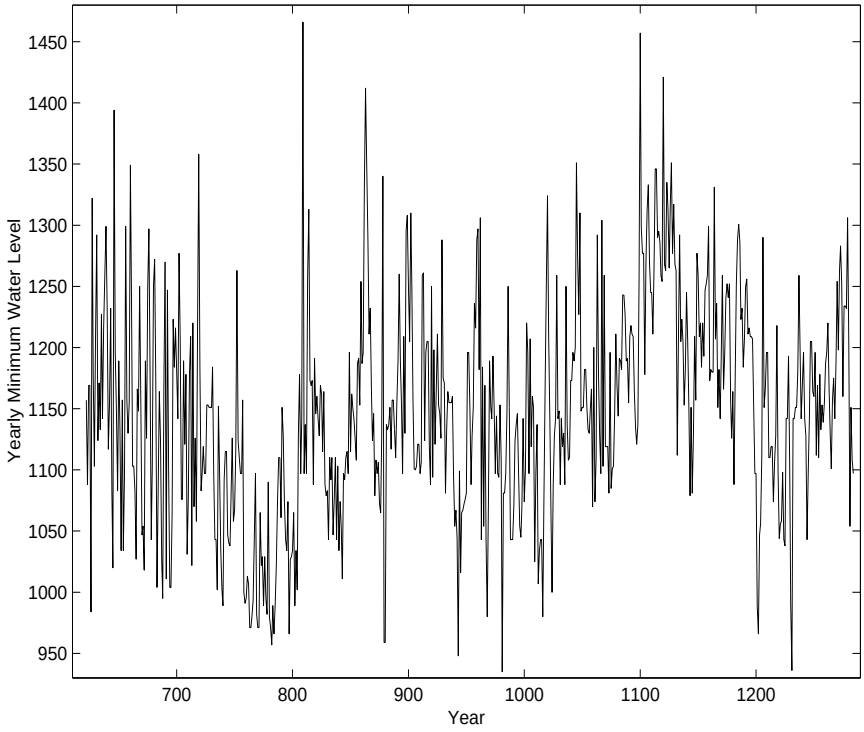

Figure 3.3. Yearly minimum water levels of the Nile River at the Roda Gauge near Cairo, Egypt (622-1281).

al. and used to show that network traffic is second-order self-similar. Passing this data through the optimization method gives the error function $E = 1.15 \times 10^{-4}$. Thus, the optimization method acknowledges again the fact that this process is second-order self-similar.

The last real data that we consider, is data that is known *not* to be second-order self-similar. Figure 3.5 displays the amount of codec information per frame for a certain video scene. The scene consists of a conversation between three people sitting at a table. No change in the background and no movement of

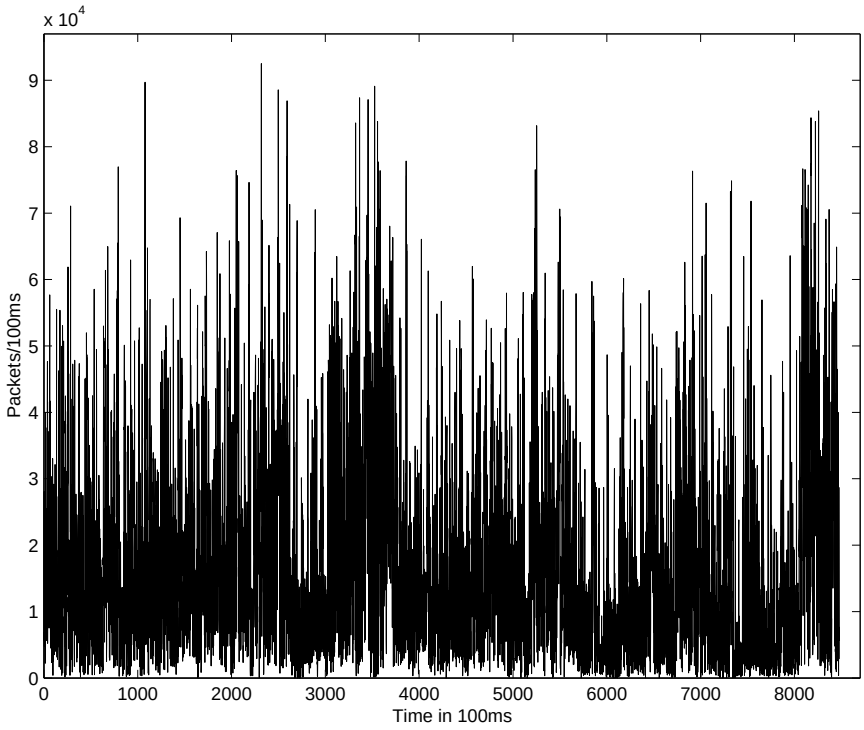

Figure 3.4. Ethernet measurements for a local area network traffic at Bellcore, Morristown, New Jersey (August 1989).

the camera exist. The codec that was used is called variable-bit-rate (VBR). This data was gathered in 1991 by engineers at Siemens, Munich, Germany (see [14, 15] and [5]). Application of the optimization method to this data results in $E = 6.74 \times 10^{-3}$. Since this value is greater than 10^{-3}, the optimization method declares that this data is not second-order self-similar, in agreement with what we already know.

Data	Nile	BC	VBR
$E_K(\hat{H})$	2.15×10^{-4}	1.15×10^{-4}	6.74×10^{-3}

Table 3.1. A summary of the results of the application of the optimization method to different sets of real data.

3.4. Summary

In this chapter, we have presented a new tool to test for long-range dependence parameter in network traffic. We applied the new method, which we called the optimization method, to pseudo random data and various real data that is known to be long-range dependent. The optimization method was shown to successfully answer the question of whether the process under study follows the given model or not.

In what follows, we present a summary of the optimization method for the second-order self-similar case, while noting that the fractional ARIMA case is similar.

- Let $X_1, X_2, \ldots, X_n$ be a realization of a Gaussian second-order self-similar process,
- Compute $\hat{\rho}_n(k)$ as in (2.10),
- Compute the error function $E_K(H)$ as in (3.1),
- Let $E_K(\hat{H}) = e$, where $\hat{H}$ is the minimizer of the error function,

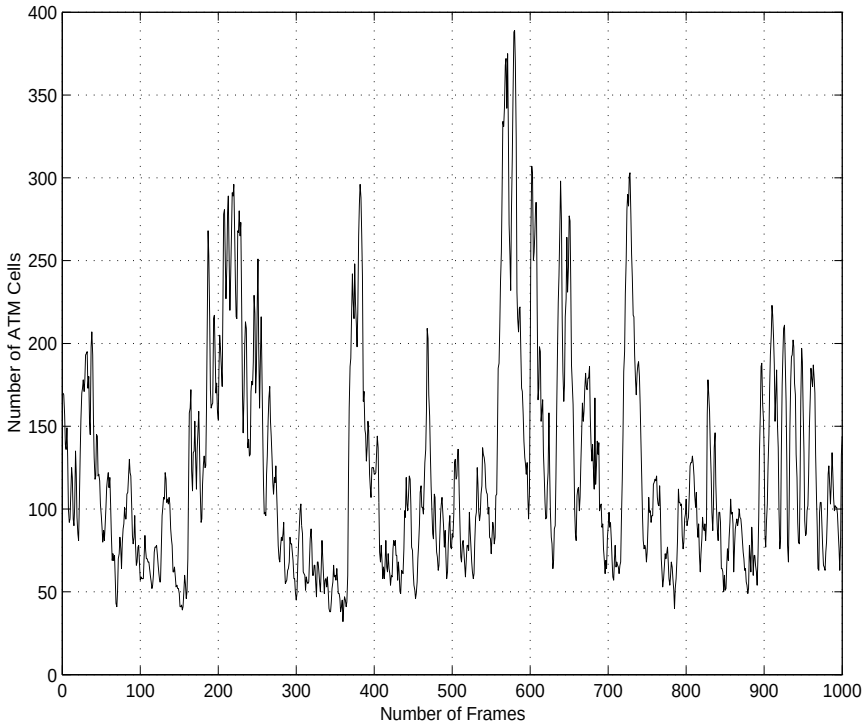

Figure 3.5. VBR data: the number of ATM cells per frame, gathered at Siemens, Munich, Germany.

- If $e < 10^{-3}$, then the process is second-order self-similar with self-similarity parameter $\hat{H}$, (or use the method of Chapter 4 for more accurate results). Otherwise, the process is not, or the data is not sufficient (n and K are not large enough) to make the right judgment.

In the following chapter, we consider the case when the long-range dependence structure is given, where the question is reduced to estimating the corresponding parameter.

CHAPTER 4
ESTIMATION OF THE LONG-RANGE DEPENDENCE PARAMETER

4.1. Introduction and Formulation

Many real processes possess the long-range dependent characteristics discussed in Chapter 2. In fact, some processes were shown to follow a particular model. In particular, we mention the results of Leland *et al.* [24], where it was shown that local area network (LAN) traffic is *exactly* second-order self-similar. Thus, knowledge of the Hurst parameter determines the second-order characteristics of such traffic. With this as motivation, we formulate the ideas of this chapter as follows.

Suppose that the process X_i is known to follow a particular model. Then the use of a one-lag autocorrelation function is justifiable, i.e., $K = 1$ in (3.1). In this case, the error function can be rewritten as

$$E_1(\beta) = [\rho(1) - \hat{\rho}_n(1)]^2. \tag{4.1}$$

This provides further simplifications of the optimization method. Mainly, it allows us to obtain theoretical confidence intervals of the estimated long-range dependence parameter $\hat{\beta}$ and makes the method much faster.

In the following two sections we discuss the two typical models for a long-range dependent process X_i. We discuss in Section 4.2 the case when X_i is second-order self-similar. The case when X_i is fractional ARIMA is discussed in Section 4.3. We then consider some examples to illustrate the use of the method on both artificial and real data and provide a comparison with the wavelet method in Section 4.4.

4.2. Second-Order Self-Similar Case

Suppose that X_i is known to be second-order self-similar. Then from (2.5)

$$\rho(1) = 2^{2H-1} - 1, \tag{4.2}$$

and $\hat{\rho}_n(1)$ is the one-lag sample autocorrelation function of the process as described in Section 2.6. Note that the global minimum of the error function $E_1(H)$ is attained and it is zero. The global minimizer is

$$\hat{H}_n = \frac{1}{2}\Big[1 + \log_2(1 + \hat{\rho}_n(1))\Big], \tag{4.3}$$

which is the estimated Hurst parameter of the process X_i.

This simple relationship between $\hat{H}$ and $\hat{\rho}_n(1)$ makes it sufficient to work with the latter to obtain statistical properties of the former. We can also obtain the asymptotic distribution of $\hat{H}$ assisted with what we know about that of $\hat{\rho}(1)$. Note that, in this case, if $\hat{\rho}_n(1)$ is $N(\mu_n, \sigma_n^2)$, then $\hat{H}$ has the following probability

density function:

$$f_{\hat{H}}(\hat{h}) = \frac{4^{\hat{h}} \log 2}{\sqrt{2\pi}\sigma_n} \exp\{\frac{(2^{2\hat{h}-1} - 1 - \mu_n)^2}{-2\sigma_n^2}\}. \tag{4.4}$$

In the case when $H \in (0, \frac{3}{4})$, from Theorem 2.6.1, (2.19), and (2.21), the one-lag sample autocorrelation function $\hat{\rho}_n(1)$ is asymptotically normally distributed with mean

$$\mu_n = \rho(1) - (1 - \rho(1))n^{2H-2}, \tag{4.5}$$

and variance given by (knowing that the autocorrelation function is an even function)

$$\begin{aligned}
\sigma_n^2 &= \frac{1}{n}\{(1 + 3\rho^2(1)) \\
&\quad +2\sum_{k=1}^{\infty}\Big[(1 + 2\rho^2(1))\rho^2(k) + \rho(k-1)\rho(k+1) - 4\rho(1)\rho(k-1)\rho(k)\Big]\} \\
&\approx \frac{1}{n}\{(1 + 3\rho^2(1)) \\
&\quad +2\sum_{k=1}^{K}\Big[(1 + 2\rho^2(1))\rho^2(k) + \rho(k-1)\rho(k+1) - 4\rho(1)\rho(k-1)\rho(k)\Big] \\
&\quad +\Big[2H(2H-1)(1-\rho(1))\Big]^2\Big[\zeta(4-4H) - \sum_{k=1}^{K} k^{4H-4}\Big]\},
\end{aligned} \tag{4.6}$$

where K is a sufficiently large number for which an approximate equality holds in (2.10) and $\zeta(\cdot)$ is the *zeta function*, defined as

$$\zeta(x) = \sum_{k=1}^{\infty} k^{-x}.$$

A plot of σ_n^2 as a function of the Hurst parameter H, with $K = 10^7$ in (4.6), is given in Figure 4.1. In the case when $H = \frac{3}{4}$, from (2.19) and (2.21), the one-lag sample autocorrelation function is asymptotically normally distributed with the same mean as in (4.5) and variance given by

$$\begin{aligned}\sigma_n^2 &= \frac{\log n}{n}[4H(2H-1)(1+\rho(1))]^2 \\ &= \frac{9}{2}\frac{\log n}{n} \quad \text{when } H = \tfrac{3}{4}. \end{aligned} \tag{4.7}$$

Thus, to construct the 95% confidence interval of $\hat{\rho}_n(1)$ we require

$$P(|\frac{\hat{\rho}_n(1)-\mu_n}{\sigma_n}| \leq 1.96) = 0.95,$$

i.e.,

$$\mu_n - 1.96\sigma_n \leq \hat{\rho}_n(1) \leq \mu_n + 1.96\sigma_n$$

holds with 95% probability. Using (4.3),

$$h_- \leq \hat{H}_n \leq h_+,$$

where

$$h_\pm = \frac{1}{2}[1 + \log_2(1+\rho(1) - (1-\rho(1))n^{2H-2} \pm 1.96\sigma_n)], \tag{4.8}$$

also holds with 95% probability.

In the case when $H \in (\frac{3}{4}, 1)$, The limiting distribution was found to be non-normal. The cumulants of which were given in (2.22). Recall that the first and second cumulants κ_1 and κ_2 correspond to the value of the mean and variance of the distribution. Thus, the variance of the limiting distribution of $\hat{\rho}_n(1)$ is

$$\sigma_n^2 = (1-\rho(1))^2 n^{2H-2}\kappa_2, \tag{4.9}$$

with κ_2 as in (2.22), which corresponds to the value of the variance of the limiting distribution. For instance, $\kappa_2 = 1.832, 0.518, 0.157$ and 0.003 for the values

$H = 0.80, 0.85, 0.90$, and 0.95, respectively [18].

Thus, to construct the 95% confidence interval of $\hat{\rho}_n(1)$ we assume normality and proceed as in the previous case of $H \in (0, \frac{3}{4})$.

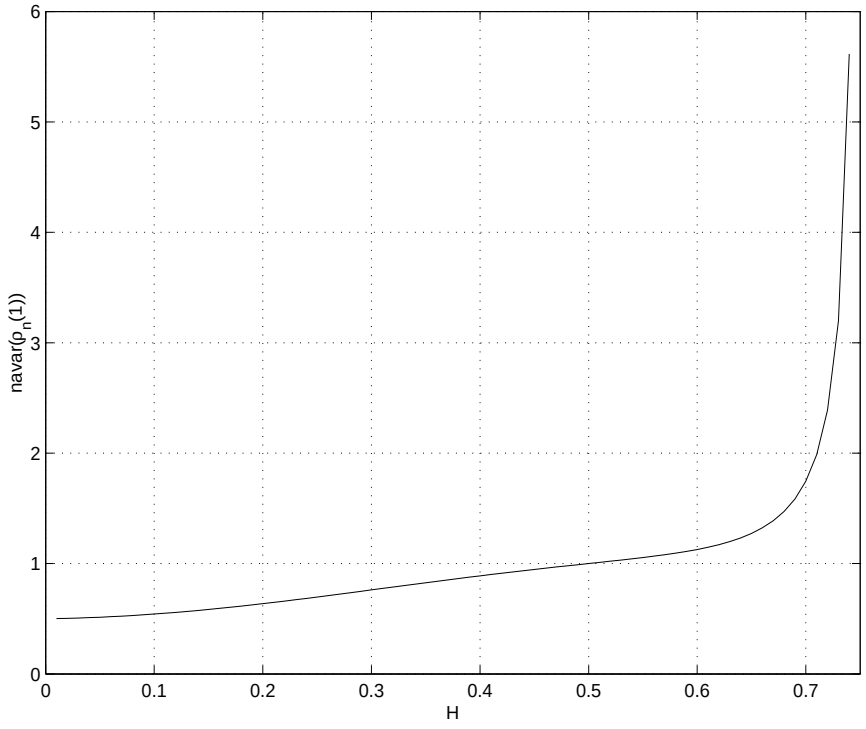

Figure 4.1. A plot of $n\sigma_n^2$ as a function of the Hurst parameter H, in the second-order self-similar case.

4.2.1. Comments on the Confidence Intervals

For known H, the 95% confidence interval of the estimate $\hat{H}_n$ is $[h_-, h_+]$, with h_- and h_+ as in (4.8), with σ_n as in Figure 4.1 if $H \in (0, \frac{3}{4})$, as in (4.7) if $H = \frac{3}{4}$

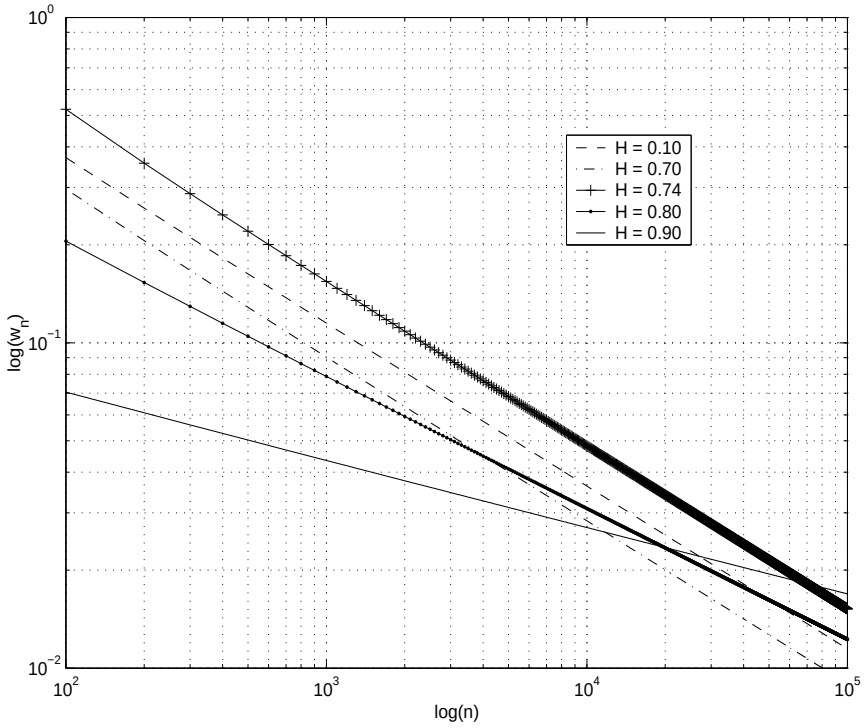

Figure 4.2. A plot of the width of the 95% confidence intervals for different H values

and as in (4.9) if $H \in (\frac{3}{4}, 1)$. Let w_n denote the width of such intervals, i.e.,

$$\begin{aligned} w_n &= h_+ - h_- \\ &= \log_4 \frac{1 + \rho(1) - (1 - \rho(1))n^{2H-2} + 1.96\sigma_n}{1 + \rho(1) - (1 - \rho(1))n^{2H-2} - 1.96\sigma_n} \end{aligned}$$

A log-log plot of w_n versus the number of samples n is given in Figure 4.2 for different values of H. It is remarkable to see the plot resembling a straight line for each value of H. Thus, the width w_n can be written as

$$w_n \approx an^{-b}, \tag{4.10}$$

where a and b are constants for fixed H. The values of these constants are given

in Table 4.1.

It is interesting to note that the width w_n is upper bounded by the w_n at $H = 0.74$. Hence in the case when H is not known (which is the typical case with real data), we choose the confidence interval centered around $\hat{H}_n$ with width

$$w_n = \frac{5}{\sqrt{n}}. \tag{4.11}$$

4.2.2. Summary of the Algorithm

In what follows, we present a summary of the new method:

- Let $X_1, X_2, \ldots, X_n$ be a realization of a Gaussian second-order self-similar process,
- Compute $\hat{\rho}_n(1)$ as in (2.15),
- Compute $\hat{H}_n$ as in (4.3), which is the estimated Hurst parameter,
- The 95% confidence interval of H is centered around the estimate $\hat{H}_n$ with width as in (4.11).

4.3. Fractional ARIMA(0,d,0) Case

Suppose on the other hand, that X_i is known to be a fractional ARIMA$(0, d, 0)$ process. Then from (2.8)

$$\rho(1) = \frac{d}{1-d},$$

H	a	b
0.10	3.65	0.50
0.20	3.44	0.50
0.30	3.28	0.50
0.40	3.08	0.50
0.50	2.85	0.50
0.60	2.65	0.50
0.70	2.92	0.50
0.74	5.00	0.50
0.75	1.63	0.45
0.80	1.28	0.40
0.90	0.18	0.21

Table 4.1. The values of the constants a and b in (4.10) for each value of H.

and $\hat{\rho}_n(1)$ is the one-lag sample autocorrelation function of the process as described in Section 2.6. Note that, as in the second-order self-similar case, the global minimum of the error function $E_1(d)$ is attained and it is zero. The global minimizer is

$$\hat{d} = \frac{\hat{\rho}_n(1)}{1 + \hat{\rho}_n(1)}, \tag{4.12}$$

which is the estimated parameter of the process X_i.

As in the second-order self-similar case, this simple relationship between $\hat{d}$ and $\hat{\rho}_n(1)$ makes it sufficient to work with the latter to obtain statistical properties of the former. We can also obtain the asymptotic distribution of $\hat{H}$ assisted with what we know about that of $\hat{\rho}(1)$. Note that, in this case, if $\hat{\rho}_n(1)$ is $N(\mu_n, \sigma_n^2)$, then $\hat{H}$ has the following probability density function:

$$f_{\hat{d}}(\hat{d}) = \frac{1}{\sqrt{2\pi}\sigma_n(1-\hat{d})^2} \exp\{\frac{((1+\mu_n)\hat{d} - \mu_n)^2}{-2\sigma_n^2(1-\hat{d})^2}\}. \tag{4.13}$$

In the case when $d \in (\frac{-1}{2}, \frac{1}{4})$, from (2.19) and (2.21), the one-lag sample autocorrelation function is asymptotically normally distributed with mean

$$\mu_n = \rho(1) - (1 - \rho(1))n^{2d-1}, \tag{4.14}$$

and variance given by (This result follows from that obtained by Hosking [18] where he applied Dougall's formula to (2.21) together with properties of gamma function.)

$$\sigma_n^2 = \frac{\Gamma(1-4d)\Gamma^4(1-d)(1-3d)}{n\Gamma^4(1-2d)(1-2d)(1-d)^2}. \tag{4.15}$$

A plot of $n\sigma_n^2$ as a function of the difference parameter d, is given in Figure 4.3. In the case when $d = \frac{1}{4}$, from (2.19) and (2.21), the one-lag sample autocorrelation

function is asymptotically normally distributed with the same mean as in (4.5) and variance given by

$$\begin{aligned} \sigma_n^2 &= \frac{\log n}{n}\Big[\frac{2(1-2d)\Gamma(1-d)}{(1-d)\Gamma(d)}\Big]^2 \\ &\approx 0.305\frac{\log n}{n} \quad \text{when } d = \tfrac{1}{4}. \end{aligned} \tag{4.16}$$

Thus, to construct the 95% confidence interval of $\hat{\rho}_n(1)$ we require

$$P(|\frac{\hat{\rho}_n(1)-\mu_n}{\sigma_n}| \leq 1.96) = 0.95,$$

i.e.,

$$\mu_n - 1.96\sigma_n \leq \hat{\rho}_n(1) \leq \mu_n + 1.96\sigma_n$$

holds with 95% probability. Using (4.12),

$$d_- \leq \hat{d}_n \leq d_+,$$

where

$$d_\pm = \frac{\rho(1) - (1-\rho(1))n^{2d-1} \pm 1.96\sigma_n}{1+\rho(1) - (1-\rho(1))n^{2d-1} \pm 1.96\sigma_n}, \tag{4.17}$$

also holds with 95% probability.

In the case when $d \in (\frac{1}{4}, \frac{1}{2})$, the limiting distribution was found to be non-normal. The cumulants of which were given in (2.22). Recall that the first and second cumulants κ_1 and κ_2 correspond to the mean and variance of the distribution. Thus, the variance of the limiting distribution of $\hat{\rho}_n(1)$ is

$$\sigma_n^2 = (1-\rho(1))^2 n^{2d-1}\kappa_2, \tag{4.18}$$

where κ_2 as in (2.22), which corresponds to the value of the variance of the limiting distribution. For instance, $\kappa_2 = 1.497, 0.433, 0.136$ and 0.027 for the values $d = 0.30, 0.35, 0.40$ and 0.45, respectively [18].

Thus, to construct the 95% confidence interval of $\hat{\rho}_n(1)$ we assume normality and proceed as in the previous case of $d \in (\frac{-1}{2}, \frac{1}{4})$.

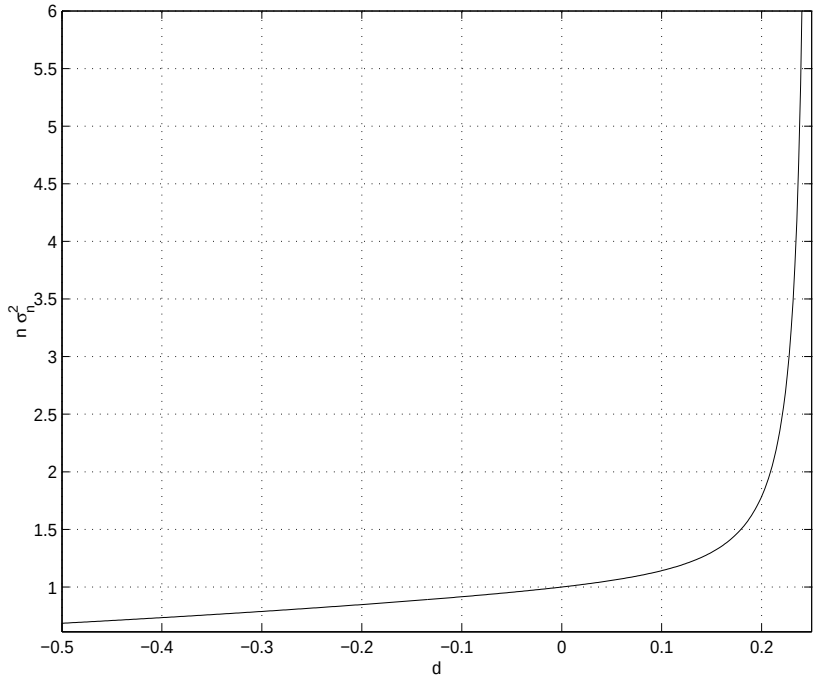

Figure 4.3. A plot of $n\sigma_n^2$ as a function of the difference parameter d for the fractional ARIMA$(0, d, 0)$ case.

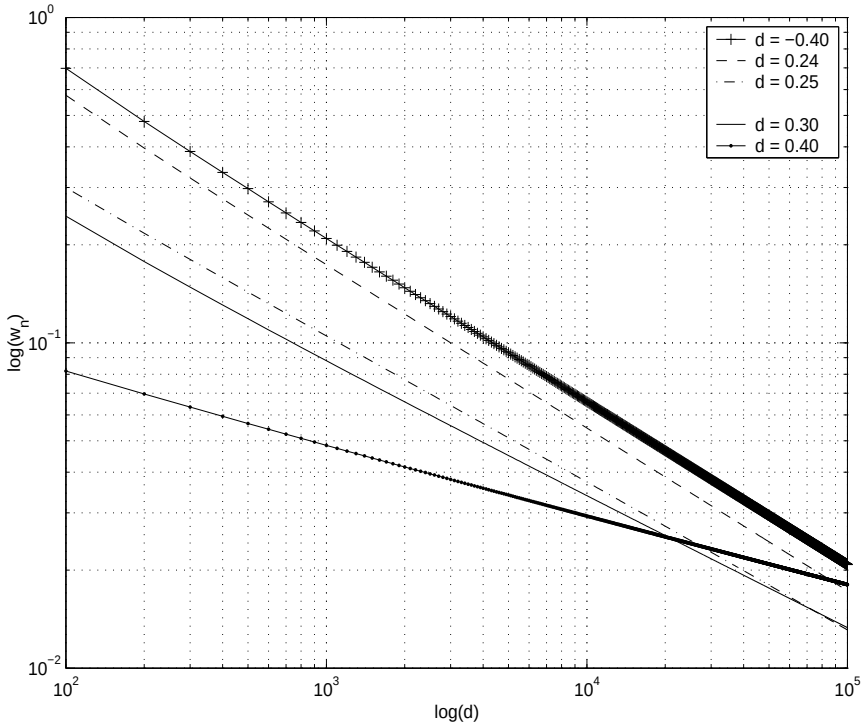

Figure 4.4. A plot of the width of the 95% confidence intervals for different d values

4.3.1. Comments on the Confidence Intervals

For known d, the 95% confidence interval of the estimate $\hat{d}_n$ is $[d_-, d_+]$, with d_- and d_+ as in (4.17), with σ_n as in Figure 4.3 if $H \in (\frac{-1}{2}, \frac{1}{4})$, as in (4.16) if $H = \frac{1}{4}$ and as in (4.18) if $H \in (\frac{3}{4}, 1)$. Let w_n denote the width of such intervals, i.e.,

$$\begin{aligned} w_n &= d_+ - d_- \\ &= \frac{3.92\sigma_n}{(1+\rho(1)-(1-\rho(1))n^{2d-1})^2 - 3.84\sigma_n^2} \end{aligned}$$

A log-log plot of w_n versus the number of samples n is given in Figure 4.4 for

different values of d. As in the second-order self-similar case, it is remarkable to see the plot resembling a straight line for each value of d. Thus, the width w_n can be written as

$$w_n \approx an^{-b}, \tag{4.19}$$

where a and b are constants for fixed d value. The values of these constants are given in Table 4.2.

It is interesting to note that the width w_n is upper bounded by the w_n at the value $d = -0.4$. Hence in the case when d is not known (which is the typical case with real data), we choose the confidence interval centered around $\hat{d}_n$ with width

$$w_n = \frac{6.67}{\sqrt{n}}. \tag{4.20}$$

4.3.2. Summary of the Algorithm

In what follows, we present a summary of the new method:

- Let $X_1, X_2, \ldots, X_n$ be a realization of a fractional ARIMA$(0, d, 0)$ process,
- Compute $\hat{\rho}_n(1)$ as in (2.15),
- Compute $\hat{d}_n$ as in (4.12), which is the estimated difference parameter,
- The 95% confidence interval of d is centered around the estimate $\hat{d}_n$ with width as in (4.20).

d	a	b
-0.40	6.67	0.50
-0.30	5.96	0.50
-0.20	5.26	0.50
-0.10	4.59	0.50
0.00	3.98	0.50
0.10	3.48	0.50
0.20	3.55	0.50
0.24	6.04	0.50
0.25	2.33	0.45
0.30	1.50	0.41
0.40	0.21	0.21

Table 4.2. The values of the constants a and b in (4.19) for different d values.

4.4. Illustrative Examples

In this section, we apply the optimization method and the wavelet method to second-order self-similar artificial data (Section 4.4.1), real data (Section 4.4.2), and compare the results. In all analyses to follow, we use the wavelet method with three vanishing moments.

4.4.1. Second-Order Self-Similar Artificial Data

For each H $= 0.10, 0.20, \ldots, 0.90$, we generate 100 realizations of a fractional Gaussian noise with the corresponding nominal values H $= 0.10, 0.20, \ldots, 0.90$. The length of each realization is $n = 4000$ points. For a given estimation method, we obtain $L = 100$ estimated values of H. Call these estimates $\hat{H}_n^{(k)}$, $k = 1, 2, \ldots, 100$. We compute their sample mean and compare it to the corresponding nominal H value.

We then compute empirical 95% confidence intervals of the estimates so that 95 of the estimates $\hat{H}_n^{(k)}$ fall in this interval. We also provide the theoretical 95% confidence intervals of the estimates resulted through the optimization method as described in Sections 4.2.2 and 4.2.2 for the sake of comparison with the empirical ones. The result of the application of the optimization and the wavelet methods to these data sets is given in Tables 4.3 and 4.4, respectively.

From both Tables 4.3 and 4.4, it is observed that the confidence intervals obtained through the optimization method CI_o are narrower than those obtained through the wavelet method. The width of the empirical confidence intervals for the optimization method is about 0.05 versus 0.11 to 0.13 for those obtained through the wavelet method.

The theoretical and actual confidence intervals are almost the same for the optimization method. This similarity holds even for $H \geq 0.80$, where we assumed normality although we knew that the distribution of $\hat{\rho}_n$ is not normal.

For $H < 0.80$, we see that the mean of the estimated Hurst parameter obtained by the optimization method $\hat{H}_n$ is the same as the true value H. For $H < 0.40$, the mean of the estimated Hurst parameter obtained by the wavelet method $\hat{H}_w$ is far from the true H and the latter does not fall in the 95% confidence interval. For $0.40 \leq H \leq 0.70$, the mean of $\hat{H}_n^w$ is closer to H and the theoretical confidence intervals contain the true value.

For $H = 0.80$, the optimization method under estimates the true value, with the mean of the estimates $\hat{H}_n$ being 0.79. The wavelet method, on the other hand, over estimates the true Hurst parameter value by the same quantity, namely the mean of the estimates $\hat{H}_n^w$ is 0.81. For $H = 0.90$, the estimates produced by the wavelet method over estimate the true value by the same quantity, namely the mean of the estimates $\hat{H}_n^w$ is 0.91. The optimization method, on the other hand,

under estimates H, with the mean of the estimates $\hat{H}_n$ being 0.87. In this case, on average, $\hat{H}_n^w$ are closer to the true value than $\hat{H}_n$. However, the empirical confidence intervals of the wavelets are much larger than those of the optimization method, with the width of the former is almost double the latter. It is also worth noting that for $H > 0.20$, the confidence intervals of the estimates from the optimization method are contained in those resulted through the wavelet method.

The number of flops is 2.5×10^4 for the optimization method and 1.3×10^7 for the wavelet method. Thus, the former is 5200 times faster than the latter. In general, it is apparent that the optimization method gives more accurate and reliable results and is much faster than the wavelet method.

4.4.2. Real Data

In this section, we apply both methods to the standard measurements discussed in Section 3.3.3, the Nile river data (Figure 3.3), and Bellcore data (Figure 3.4). A summary of the application of both methods is provided in Table 4.5.

An application of the wavelet method to the Nile rever data results in the estimated Hurst parameter $\hat{H} = 0.81$ with 95% confidence interval $[0.63, 0.98]$. The optimization method, on the other hand, results in $\hat{H} = 0.83$ with 95% confidence interval $[0.73, 0.93]$. We note that the variance-time, R/S and periodogram methods resulted in $\hat{H} = 0.87, 0.94, 0.84$, respectively [5].

H	Mean of $\hat{H}_n$	Theoretical CI_o	Empirical CI_o
0.10	0.10	[0.07,0.13]	[0.07,0.13]
0.20	0.20	[0.17,0.23]	[0.17,0.23]
0.30	0.30	[0.27,0.33]	[0.27,0.32]
0.40	0.40	[0.38,0.42]	[0.37,0.43]
0.50	0.50	[0.48,0.52]	[0.47,0.52]
0.60	0.60	[0.58,0.62]	[0.58,0.63]
0.70	0.70	[0.68,0.72]	[0.68,0.72]
0.80	0.79	[0.77,0.81]	[0.77,0.82]
0.90	0.87	[0.86,0.90]	[0.85,0.90]

Table 4.3. Results of empirical and theoretical study of the optimization method using 100 independent realizations.

H	Mean of $\hat{H}_n^w$	Empirical CI_w
0.10	0.00	[-0.05,0.06]
0.20	0.16	[0.10,0.22]
0.30	0.28	[0.21,0.34]
0.40	0.39	[0.32,0.44]
0.50	0.50	[0.44,0.56]
0.60	0.60	[0.55,0.67]
0.70	0.70	[0.65,0.75]
0.80	0.81	[0.75,0.87]
0.90	0.91	[0.84,0.95]

Table 4.4. Results of empirical study of the wavelet method using 100 independent realizations.

Data	$\hat{H}_n$	CI_o	$\hat{H}_n^w$	CI_w
Nile	0.83	[0.73,0.93]	0.81	[0.63,0.98]
BC	0.81	[0.78,0.84]	0.79	[0.75,0.82]

Table 4.5. A summary of the results of the application of the optimization and wavelet methods to different sets of real data.

Passing the Bellcore data through the wavelet method gives $\hat{H} = 0.79$ with a relatively small 95% confidence interval of $[0.75, 0.82]$. The optimization method, on the other hand, results in $\hat{H} = 0.81$ with 95% confidence interval $[0.78, 0.84]$. The variance-time, R/S, and periodogram methods resulted in $\hat{H} = 0.80, 0.79, 0.82$, respectively [24].

In short, it is clear that the optimization method and the wavelet method give close estimates in the first two cases. The estimates in the third case are not close, but that of the optimization and variance-time methods are close enough. It is also noted that in all cases considered in this section, the optimization method's estimates fall in the 95% confidence intervals of the wavelet method. Moreover, the confidence intervals of the optimization method are contained in these of the wavelet method.

4.5. Summary

In this chapter, we have presented a new tool to estimate the long-range dependence parameter in local area network traffic. We considered the two models of long-range dependence: Second-order self-similar and fractional ARIMA processes. A summary of the algorithm for both models is presented in Sections 4.2.2 and 4.3.2, respectively.

The confidence intervals and biasedness of the estimates from this new method are obtained. This new method is then applied to pseudo-random data and to real LAN traffic data. We compared the performance of the new method to that of the widely-used wavelet method. We demonstrated that the former is much faster and produces smaller confidence intervals of the Hurst parameter estimates. Furthermore, the confidence intervals of the estimates from the new method were shown to be contained in the confidence intervals of the wavelet method. The width of these confidence intervals was shown to decay hyperbolically with respect to the number of sample points.

In view of the above, we believe that this method can be used as an on-line estimation tool for H and thus be exploited in the new TCP algorithms that exploit the known self-similar (and therefore long-range dependent) nature of network traffic. We mention for example, TCP – Traffic Prediction proposed in [13].

CHAPTER 5

SUMMARY AND CONCLUDING REMARKS

In this thesis, we have presented a new tool to decide whether or not a process has a given parametric correlation structure. The cutoff value of the error function to decide whether the given process follows the prescribed model or not was found empirically so that the probability of false alarm is less than 0.05. The new method is tested on pseudo random data over various ranges of the long-range dependence parameter and on real data. The new method was shown to successfully answer the question of whether the studied process follows the prescribed model or not.

In the case when the data is known to have a given parametric correlation structure, we have developed a new parameter estimation method. This method allows us to obtain the distribution and confidence intervals of the estimate. The width of these confidence intervals was shown to decay hyperbolically with respect to the number of sample points. We applied the optimization method to pseudo random data to estimate the Hurst parameter of the process. We then compared the performance of the optimization method with that of the wavelet method. This comparison demonstrated that the optimization method produced more accurate and more reliable results than the wavelet method. Moreover, the optimization method was also shown to be much faster than the wavelet method.

In view of the above, we believe that the optimization method can be used as an on-line estimation tool for H and thus be exploited in the new TCP algorithms that exploit the known self-similar (and therefore long-range dependent) nature of network traffic. We mention, for example, TCP – Traffic Prediction proposed in [13]. On the other hand, it would be of interest to use the new method to find the variables affecting the long-range dependence parameter β (e.g., number of users, amount of data, relation between current TCP algorithms and β, topology of the network, etc.)

If the traffic is believed to exhibit both short and long-range dependence, then this traffic can be thought of as a sum of two processes: A short-range dependent process and a long-range dependent process. In this case, an extension of the new method can be considered to test the validity of the model and estimate the corresponding parameters.

More recent results suggest that wide area network traffic is multifractal [7]. In short, a multifractal process is a long-range dependent process with parameter $\beta(t)$ that is a function of time. Thus, such process is not stationary. The case when $\beta(t) = \beta$ is a constant is referred to as "monofractal," which we have been dealing with throughout this thesis. Hence, for a multifractal process, the auto-correlation function is a function of time too. The function $\beta(t)$ is commonly taken to be piecewise constant. In such case, the optimization method is still applicable on the regions where $\beta(t)$ is constant.

APPENDIX A
OTHER METHODS IN THE LITERATURE PROPOSED TO ESTIMATE THE LONG-RANGE DEPENDENCE PARAMETER

A.1. Introduction

In Section 2.7 we discussed in detail the wavelet method, which is a widely used method in the Networking community to test for long-range dependence and estimate the corresponding parameter. In this appendix, we summarize the other methods proposed so far by different researchers to test the long-range dependence phenomena and estimate its parameter β. We mainly discuss in Sections A.2 to A.8 the following methods: Rescaled Adjusted Range, Variance-Time, Residuals of Regression, Higuchi's, Correlogram, periodogram, and Whittle Estimator, respectively.

All these methods (except the variance-time method) are asymptotic in the sense of (2.10) and (2.11). All the methods (except Whittle estimator) are graphical. An empirical study of the estimates obtained using methods in Section A.2 to A.8 are available in [38].

A.2. Rescaled Adjusted Range Method

This method is known more simply as the R/S method. It was introduced by Hurst [19], and it is one of the best known methods. Let X_i, Y_n and $\hat{\sigma}_n^2$ be the increment process, the cumulative process, and the sample mean of X_i as defined in Section 2.6. The R/S statistic is then given by

$$R/S(n) = \frac{1}{\hat{\sigma}_n}\Big[\max_{0\le j\le n}\{Y(j) - \frac{j}{n}Y(n)\} - \min_{0\le j\le n}\{Y(j) - \frac{j}{n}Y(n)\}\Big]. \tag{A.1}$$

If X_i is long-range dependent, then

$$E[R/S(n)] \sim cn^H \quad \text{as} \quad n \to \infty, \tag{A.2}$$

where c is a positive constant, and H is the Hurst parameter.

Thus, if the process X_i is long-range dependent, a plot of $\log(E[R/S(n)])$ versus $\log n$ results in a straight line for large values of n, the sample size. The slope of this line is the value of H. See [28, 5] for details.

A.3. Variance-Time Analysis

Also known as the aggregated variance method. This method is based on (2.9). Let $X^{(m)}$ be the aggregated process of the increment process X as defined by (2.6). Let $(\sigma^{(m)})^2$ be the variance of this aggregated process and $(\hat{\sigma}^{(m)})^2$ be its sample

variance. Since $(\hat{\sigma}^{(m)})^2$ is an estimate of $(\sigma^{(m)})^2$, based on (2.9), a log-log plot of $(\hat{\sigma}^{(m)})^2$ versus m will result in a straight line with slope β. See [5] for details.

A modification of this method is known as *the Absolute Values of the Aggregated Process or Series*. In this case, instead of computing the sample variance $(\hat{\sigma}^{(m)})^2$, the average of the sum of the absolute values of the aggregated process is computed, namely

$$\frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k)|. \tag{A.3}$$

It turns out that if the process X_i is long-range dependent, the plot of the logarithm of the preceding average versus the logarithm of the aggregation level m results in a line with slope $H - 1$. See [5] for details.

A.4. Residuals of Regression

This method was suggested by Peng *et al.* [35]. We first break up the process X_i into blocks of size m. Then, the partial sums of the series are calculated within each block. Let $Z(j)$, $j = 1, 2, \ldots, m$, denote these partial sums. Then, fit a least-squares line to the $Z(i)$ and compute the sample variance of the residuals.

This procedure is repeated for each of the blocks, and an average of the resulting variances is obtained. Since the blocks have the same size m, this is equivalent to calculating the sample variance of the entire series X_i. If the process X_i is long-range dependent, then a log-log plot of the result versus m results in a straight line with slope $2H$ as proven by Taqqu *et al.* [38].

A.5. Higuchi's Method

This method was presented by Higuchi [16]. This method involves calculating the length of a path or its fractal dimension D (see (2.4) for the definition of D). We first calculate the partial sums $Y(j) = \sum_{i=1}^{j} X_i$ (i.e., constructing the cumulative process from the increment process X_i). Then, we find the normalized length of the curve, namely

$$L(m) = \frac{n-1}{m^3} \sum_{i=1}^{m} \lfloor \frac{n-i}{m} \rfloor^{-1} \sum k = 1 \lfloor (n-i)/m \rfloor |Y(i+km) - Y(i+(k-1)m)|, \tag{A.4}$$

Where n is the number of samples, m is the aggregation level or the block size and $\lfloor\rfloor$ denotes the greatest integer (floor) function. Then $EL(m) \sim cm^{-D}$, where c is a positive constant and $D = 2 - H$. Hence, as before, a log-log plot $EL(m)$ versus m produces a straight line of slope $D = 2 - H$.

A.6. Correlogram

This method is based on (2.10), from which if X_i is long-range dependent, a log-log plot of the sample correlations of X_i versus the lag k results in a straight line for large lag values. The slope of this line is $-\beta$. See [5] for details.

A.7. Periodogram Method

Let $\hat{f}(\lambda)$ be the sample spectral density of X_i, namely

$$\hat{f}(\lambda) = \frac{1}{2\pi n} |\sum_{j=1}^{n} X_j e^{ij\lambda}|^2, \tag{A.5}$$

where λ is the frequency and n is the number of samples. Then, if X_i is long-range dependent, based on (2.11), a log-log plot of the $\hat{f}(\lambda)$ against $|\lambda|$ gives a straight line as $|\lambda|$ approaches the origin. The slope of this line is $-\alpha$. See [5] for details.

A.8. Whittle Estimator

This method was first proposed by Whittle [43] in the context of short-range dependence, see [5] for the application of this method to long-range dependence

processes. As the periodogram method, this method is also based on (2.11). Define

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{\hat{f}(\lambda)}{f(\lambda;\eta)} d\lambda, \tag{A.6}$$

where $f(\lambda;\eta)$ is the spectral density at frequency λ, and the vector of unknown parameters η (in our case, this is a scaler equals H for second-order self-similar case and d for fractional ARIMA$(0,d,0)$ case,) and $\hat{f}(\lambda)$ is the sample spectral density. The integral in (A.6) can be replaced by the sum over the frequencies $\frac{2\pi}{n}, \frac{4\pi}{n}, \ldots, \frac{2(n-1)\pi}{n}$. The estimator of η denoted by $\hat{\eta}$ is chosen to minimize the function $Q(\eta)$. The variance of the estimator is then $\frac{4\pi}{n} Q(\hat{\eta})$.

APPENDIX B

PROGRAMS

In this Appendix we provide MATLAB program files that can be used to incorporate the ideas illustrated in this thesis.

B.1. Pseudo Random Number Generator

B.1.1. Second-Order Self-Similar Case

```
function R = Rho(h,N)
% function R = Rho(h,N)
% This code constructs an NxN covariance matrix R
% which is second-order self-similar.
% h is the Hurst parameter,
% N is the number of data points
% The resultant R is to be used at "mvnrnd" command

H = 2*h;
k = 1:N;
c = 0.5*((k+1).^H-2*k.^H+(k-1).^H);
c = [1 c];
R = toeplitz(c);
```

B.1.2. Fractional ARIMA$(0, d, 0)$ Case

```
function R = Rho2(d,N)
% function R = Rho2(d,N)
% This code constructs an NxN covariance matrix R
% which is fractional ARIMA(0,d,0).
% d is the difference parameter,
% N is the number of data points
% The resultant R is to be used at "mvnrnd" command


k=1:N;
c = gammaln(k+d)-gammaln(1+k-d);
c = (gamma(1-d)/gamma(d))*exp(c);
c = [1 c];
R = toeplitz(c);
```

B.2. Testing Method

B.2.1. Second-Order Self-Similar Case

```
function [H,E] = hurst2(X)
% Given Nx1 observation vector X vector of observations
% in bytes or packets, this code checks whether the given
% process is soss and finds its Hurst parameter H
% by picking H minimizing the sum of the squares of the
```

```
% difference of the sample covariances of observations and the model.
mu = mean(X);
N = length(X);
Xmu =  X-mu; %centering X around the mean mu
%we could go all the way to n,
%but we are trying to reduce the "edge effect"
K = 50; % This may be varied
for i=0:K
c(i+1) = sum(Xmu(1:N-i).*Xmu(i+1:N))/(length(Xmu(1:N-i))-1);
end  %c(N) = cov(N-1), where cov is the covariance of X
h = .01:.01:1; %the stepsize .01 is arbitrary
nc = length(c)-1;
for k =1:nc
rho = c(k+1)/c(1);
f(k,:) = (((1/2).*((k+1).^(2.*h)-2.*k.^(2.*h)+(k-1).^(2.*h))-rho).^2);
end
f = sum(f);
[minf,lo] = min(f);
H = lo/100;
E = minf/(4*K);
% The condition is that E < 0.001, otherwise X is not soss.
```

The following code is used to determine the empirical cutoff e in Figure 3.1.

```
% This code provides the cutoff e for the specified
% probability of false alarm in soss case.

N = 4000;  % Number of data points
L = 10000; % Number of realizations
ii=0;
for h=0.12:0.01:0.99
clear H E R X
h
R = Rho(h,N-1); X = mvnrnd(zeros(1,N), R,L);

for i=1:L
[H(i),E(i)] = hurst2(X(i,:));
end

E = sort(E);
ii=ii+1;
e(ii) = E(.95*L) % e is chosen so that P_FA < 0.05
end
```

B.2.2. Fractional ARIMA$(0, d, 0)$ Case

```
function [D,E] = farima2(X)
```

```
% Given Nx1 observation vector X vector of observations
% in bytes or packets, this code checks whether the given
% process is fractional ARIMA(0,d,0) and finds its difference parameter d
% by picking d minimizing the sum of the squares of the
% difference of the sample covariances of observations and the model.
mu = mean(X);
N = length(X);
Xmu =  X-mu; %centering X around the mean mu
%we could go all the way to n,
%but we are trying to reduce the "edge effect"
K = 50; %.01*N;
for i=0:K
c(i+1) = sum(Xmu(1:N-i).*Xmu(i+1:N))/(length(Xmu(1:N-i))-1);
end   %c(N) = gamma(N-1), where gamma is the covariance of X
d = -.49:.01:0.50; %the stepsize .01 is arbitrary
nc = length(c)-1;
for k =1:nc
rho = c(k+1)/c(1);
f(k,:) = ((gamma(1-d).*gamma(k+d))./(gamma(d).*gamma(1+k-d))-rho).^2;
end
f = sum(f);
[minf,lo] = min(f);
D = d(lo);
```

```
E = minf/(4*K);
% The condition is that E < 0.001, otherwise X is not farima(0,d,0).
```

The following code is used to determine the empirical cutoff e in Figure 3.2.

```
% This code provides the cutoff e for the specified
% probability of false alarm in farima(0,d,0) case.

N = 4000;  % Number of data points
L = 10000; % Number of realizations
ii=0;
for d=-.38:0.01:0.49
d
clear R X
R = Rho2(d,N-1); X = mvnrnd(zeros(1,N), R,L);

for i=1:L
[D(i),E(i)] = farima2(X(i,:));
end

E = sort(E);
ii=ii+1;
e1(ii) = E(0.95*L)
end
```

B.3. Estimation Method

B.3.1. Second-Order Self-Similar Case

```
% function [H,CI] = hurst(X)
% Given Nx1 observation vector X vector of observations
% in bytes or packets, this code  finds the Hurst parameter H
% of the process which is assumed to be soss.
% This code is a modified version of hurst2.m code,
% It checks cov(1) only to deduce H.
function [H,rho] = hurst(X)
mu = mean(X);
n = length(X);
Xmu =  X-mu; %centering X around the mean mu
c0 = sum(Xmu.*Xmu)/(n); % Variance
c1 = sum(Xmu(1:n-1).*Xmu(2:n))/(n); % cov(1)
rho = c1/c0; % autocorrelation of the process
H = .5*(1+log(1+c1/c0)/log(2));
CI = 5/sqrt(n);
```

B.3.2. Fractional ARIMA$(0, d, 0)$ Case

```
% function [d,CI] = farima(X)
% Given Nx1 observation vector X vector of observations
% in bytes or packets, this code  finds the difference parameter d
```

```
% of the process which is assumed to be fractional ARIMA(0,d,0).
% This code is a modified version of hurst2.m code,
% It checks cov(1) only to deduce d.
function [d,rho] = farima(X)
mu = mean(X);
n = length(X);
Xmu =  X-mu; %centering X around the mean mu
c0 = sum(Xmu.*Xmu)/(n); % Variance
c1 = sum(Xmu(1:n-1).*Xmu(2:n))/(n); % cov(1)
rho = c1/c0; % autocorrelation of the process
d = rho/(1+rho);
CI = 6.67/sqrt(n);
```

BIBLIOGRAPHY

[1] P. Abry and D. Veitch (1998). Wavelet Analysis of Long-Range Dependent Traffic. *IEEE Transactions on Information Theory*, 44(1):2–15.

[2] P. Abry, P. Flandrin, M. S. Taqqu and D. Veitch (2000). Wavelets for the Analysis, Estimation, and Synthesis of Scaling Data *Self-Similar Network Traffic and Performance Evaluation.* K. Park and W. Willinger (editors), John Wiley & Sons, New York, New York.

[3] T. W. Anderson (1958). *The Statistical Analysis of Time Series.* John Wiley & Sons, New York, New York.

[4] Baccelli, F. and Hong, D. (2002) "AIMD, Fairness and Fractal Scaling of TCP Traffic" Infocom, June 2002,

[5] J. Beran (1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York, New York.

[6] M. Corvella and A. Bestavros (1997). Self-similarity in World Wide Web traffic: evidenceand possible causes. *IEEE/ACM Transactions on Networking,* 5:835–846.

[7] A. Feldmann, A. C. Gilbert, and W. Willinger (1998). Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic, *ACM Computer Communication Review,* 28:42–55.

[8] I. S. Gradshteyn and I. M. Ryshik (1965) *Tables of Integrals, Series and Products.* Academic Press, London.

[9] C. W. J. Granger and R. Joyeux, (1980). An Introduction to Long-Memory Models and Fractional Differencing *Journal of Time Series Analysis*, 1:15–29.

[10] G. R. Gimmett and D. R. Stirzaker (1994). *Probability and Random Processes.* Clarendon Press, Oxford, United Kingdom.

[11] M. Grossglauser and J. Bolot (1996). On the Relevance of Long-Range Dependence in Network Traffic. *Computer Communication Review*, 26(4):15–24.

[12] J. A. Gubner (2002) Fallacies in the Theory of Long-Range-Dependent Processes. *submitted to IEEE Transactions on Information Theory.*

[13] G. He, Y. Gao, J. C. Hou and K. Park (2002). A Case for Exploiting Self-Similarity of Internet Traffic in TCP Congestion Control. *Submitted to IEEE/ACM Transactions on Networking.*

[14] H. Heeke (1991). Statistical Multiplexing Gain for Variable Bit Rate Codecs in ATM Networks. *International Journal of Digital and Analog Communication Systems,* 3:261–268.

[15] D. Heyman, A. Tabatabai and T. V. Lakshman (1991). Statistical Analysis and Simulation of Video Teleconferencing in ATM Networks. *IEEE Transactions on Circuits and Systems Video Technology,* 2:49–59.

[16] T. Higuchi (1988). Approach to an Irregular Time Series on the Basis of the Fractal Theory. *Physica D,* 31:277–283.

[17] J. R. M. Hosking (1981). Fractional Differencing. *Biometrika,* 68:165–176.

[18] J. R. M. Hosking (1981). Asymptotic Distribution of the Sample Mean, Autocovariances, and Autocorrelations of Long-memory Time Series. *Journal of Econometrics,* 73:261–284.

[19] H. E. Hurst (1951). Long-term Storage Capacity of Reservoirs *Transactions of the American Society of Civil Engineers,* 116:770–779.

[20] S. Karlin, H. M. Taylor (1975). *A First Course in Stochastic Processes.* Second Edition. Academic Press.

[21] H. Kettani and W. Feng (2001). *A New Method for Estimating the Hurst Parameter in Self-Similar Traffic,* Los Alamos National Laboratory Unclassified Report 01-0905, Los Alamos, New Mexico.

[22] H. Kettani and W. Feng (2001). *On the Stationarity and Variation of Self-Similarity in Network Traffic,* Los Alamos National Laboratory Unclassified Report 01-0906, Los Alamos, New Mexico.

[23] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. (1993) On the Self-Similar Nature of Ethernet Traffic. In *Proceedings of ACM SIGCOMM'93.*

[24] W. Leland, M. Taqqu, W. Willinger, and D. Wilson (1994). On the Self-

Similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transactions on Networking*, 2(1):1–15.

[25] N. Likhanov (2000). Bounds on the Buffer Occupancy Probability with Self-Similar Input Traffic. in K. Park and W. Willinger, Self-Similar Network Traffic and Performance Evaluation pp. 193-213, Wiley-Interscience.

[26] M. Loève (1977) *Probability Theory,* Vol. 1 (4th edition). Springer, Berlin.

[27] D. G. Luenberger (1973). *Introduction to Linear and Nonlinear Programming.* Addison-Wesley Publishing Co., Reading, MA.

[28] B. B. Mandelbrot (1983). *The Fractal Geometry of Nature.* W. H. Freeman and Co., New York, New York.

[29] R. Morris and D. Lin (2000). Variance of Aggregated Web Traffic. In *Proceedings of IEEE INFOCOM 2000.*

[30] K. Park, G. Kim, and M. Crovella (1996). On the Relationship Between File Sizes, Transport Protocols, and Self-Similar Network Traffic. In *Proceedings of the 4th International Conference on Network Protocols.*

[31] K. Park, G. Kim, and M. Crovella (1997). On the Effect of Traffic Self-Similarity on Network Performance. In *Proceedings of the SPIE International Conference on Performance and Control of Network Systems.*

[32] K. Park and W. Willinger (2000). Self-Similar Network Traffic: An Overview

Self-Similar Network Traffic and Performance Evaluation. K. Park and W. Willinger (editors), John Wiley & Sons, New York, New York.

[33] V. Paxson and S. Floyd (1995). Wide-Area Traffic: The Failure of Poisson Modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244.

[34] V. Paxson (1995). Fast Approximation of Self-Similar Network Traffic. *technical report LBL-36750/UC-405.*

[35] C. K. Peng, S. V. Buldyrev, M. Simons, H. E. Stanley and A. L. Goldberger (1994). Mosaic Organization of DNA Nucleotides. *Physical Review E,* 49:1685–1689

[36] E. W. Stacy (1962) A Generalization of the Gamma Distribution. *Annals of Mathematical Statistics,* 33(3): 1187–1192.

[37] L. L. Scharf (1991) *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis* Addison-Wesley Publishing Company.

[38] M. S. Taqqu, V. Teverovsky and W. Willinger (1995). Estimators for Long-range Dependence: An Empirical Study *Fractals*, 3(4):785–788.

[39] O. Tousson (1925). Mémoire Sur L'Histoire Du Nil. *Mémoires de l'institut d'Egypte.*

[40] W. Vervaat (1985). Sample Path Properties of Self-similar Processes With Stationary Increments. *Annals of Probability,* 13:1–27.

[41] W. Vervaat (1987). Properties of General Self-similar Processes. *Bulletin of the International Statistical Institute,* 52(4):199–216.

[42] C. Walck (1998) *Hand-Book on Statistical Distributions for Experimentalists.* Internal Report SUF–PFY/96-01, Particle Physics Group Fysikum, University of Stockholm, Stockholm, Sweden.

[43] P. Whittle (1953). Estimation and Information in Stationary Time Series *Arkiv for Matematik,* 2:423–434.