4th WSEAS International Conference on ELECTRONICS, CONTROL and SIGNAL PROCESSING, Miami, Florida, USA, 17-19 November, 2005 (pp.193-198)

Hurst Parameter Estimation from Noisy Observations of Data Traffic Traces

G. R. DATTATREYA
Department of Computer Science
University of Texas at Dallas
Richardson, Texas 75083-0688

Abstract: Data traffic traces are known to be bursty with long range dependence. The exact self-similarity model of long range dependence can pose analytical and practical problems at very small and very large time lags. In our model, the time series of the traffic trace (referred to as the signal) is assumed to possess an autocovariance profile corresponding to exact self-similarity over a range of lags, $\{k\}$, satisfying $M < k < L$. At lower lags, exact self-similarity may breakdown, or additive moving average type noise (inaccuracies) may corrupt the autocovariances. At very high lags, far beyond the number of observed samples, the autocovariance structure is irrelevant and may be assumed to be infinite summable. Therefore, L can be as large as desired. Applications of such a model are discussed. The mean, variance, and the Hurst parameter of the signal, as well as the autocovariances of any independent zero mean moving average type additive noise are assumed to be unknown. A class of linear combinations of sample average second order statistics of noisy observations is constructed. They are unbiased estimates of their corresponding expectations. These expectations are shown to be devoid of the noise statistics. The ratio of two such expectations eliminates the signal variance. The ratio is a well behaved monotonic function of the only remaining unknown, the Hurst parameter. Equating the ratio of these expectations to the ratio of the corresponding sample averages from the noisy observations leads to a very easily solvable nonlinear equation with a unique root. The result and related issues are discussed.

Key-Words: Parameter estimation, Hurst parameter, Fractal dimension, Traffic engineering, Traffic modeling and analysis, Self-similarity.

1 Introduction

Data traffic in telecommunication networks depicts vastly different characteristics from those of voice traffic. Several studies have traced the origin of this distinction to the widely fluctuating heavy tailed probability density function of data file sizes, as opposed to the memoryless property of voice call holding time. Many different approaches have been used to model such "bursty" data traffic. Some of these are (a) cumulative amount of traffic received at a point as a Self-Similar process (Park and Willinger, Chapter 1 in [1]), (b) aggregation (merging at a point) of numerous ON-OFF data sources with at least one of the ON or OFF time period exhibiting a heavy-tailed probability density function (Brichet, Simonian, Massoulie, and Veitch, Chapter 5 in [1]), and (c) a Poisson sequence of arrivals of data units, each data unit exhibiting a heavy-tailed behavior (Makowski and Parulekar, Chapter 9 in [1]). Inter-relationships between these models and some of their variations have also been studied. For example, the Fractional Brownian Motion (Mandelbrot and Van Ness [2]) is an example of the family of self-similar processes. Relaxation of some strict properties of the exact self-similar process leads to long-memory processes (Beran [3]) and $\frac{1}{f}$ processes (Wornell and Oppenheim [4]). The above cited monograph [1] edited by Park and Willinger is a collection of articles on the topics of measurement, modeling, performance analysis, and control of self-similar data traffic. These articles also bring out the above mentioned inter-relationships. Data traffic is one (and a more recent) application of such a class of bursty model processes. Earlier studies of applications of such processes include long range Nile river data [3], error clusters in communication systems [5], and natural scenery [6]. All the known representations of such bursty processes have one common parameter value. This parameter has different but equivalent representations and names. That is, the Hurst parameter H, the fractal dimension D, and the the exponent γ of the $\frac{1}{f}$ process' power spectrum are related as $\gamma = 2H - 1 = 5 - 2D$. This paper develops an estimator for the Hurst parameter starting from a time series trace of such a bursty signal, possibly corrupted by independent and short-term (moving average type) correlated additive noise. The mean of the signal is assumed to be unknown. The estimator is shown to be robust to such noise. The estimator is based on sums and differences of unbiased estimates of second order statistics obtained from the noisy-bursty signal. The term signal (rather than data) is used for the time series to avoid confusion with the term data in "data traffic." The signal can be from any bursty process. For example, in data traffic, measurements may be a sequence of the amounts of data bits received in successive equal time intervals.

2 Development of Noisy-Bursty Signal Model

2.1 Background and motivation for the model

All the models of bursty data traffic and similar bursty signals conform, at least to some extent, to the following property of self-similarity, seminally explored in Mandelbrot and Van Ness [2]. Let $Z(t)$ be a real valued stochastic process with a continuous parameter t and δ be a positive number. If the processes $Z(t_0 + \tau) - Z(t_0)$ and $\delta^{-H}[Z(t_0 + \delta\tau) - Z(t_0)],\ H > 0$ have the same finite joint distributions, $Z(t)$ is said to be (exactly) self-similar with the (Hurst) parameter H. An example of such an exact self-similar process is the Fractional Brownian motion, $Y(t)$, defined with the help of its increment process, as follows. If the increment $Y(t + t_0) - Y(t_0)$ is a zero mean, stationary Gaussian process with variance t^{2H} and $0.5 \leq H < 1$, then $Y(t)$ is called Fractional Brownian Motion (FBM). The derivative of FBM is a zero mean Gaussian random variable with infinite variance. In addition to this, Mandelbrot and Van Ness [2] also show the following. (a) FBM has continuous sample paths with probability 1, but is almost surely not differentiable. (b) Any non-constant mean-square continuous exactly self-similar process must be FBM. (c) Non-Gaussian exactly self-similar processes must necessarily have increments with infinite variance. These properties imply that the most useful exactly self-similar process is the FBM. Even FBM presents mathematical difficulties. It is not mean square differentiable. However, its increments over any nonzero time intervals have finite variances. The time series of FBM increments over successive (contiguous) non-overlapping intervals is called the discrete Fractional Gaussian noise (FGN). It is a stationary sequence. Let $V(i)$ be an FGN sequence with mean μ and variance of σ^2. The autocovariance sequence of $V(i)$ is known to be (Beran [3])

$$C_j = \frac{\sigma^2}{2}\left[|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\right] \qquad (1)$$

and the normalized (with respect to the variance or the zeroth autocovariance σ^2) autocovariance coefficients are

$$c_j = \frac{1}{2}\left[|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\right]. \qquad (2)$$

While the discrete FGN overcomes having to deal with the problems of observations over time periods tending to zero, it still poses both mathematical and practical problems. The sum of autocovariances is not finite. That is, $\sum_{j=0}^{\infty} c_j = \infty$. Therefore, the Fourier transform of c_j is infinity at the zero frequency point. It is also easy to show that (see Kulkarni [7]) such an FGN sequence cannot be realized as the output of a linear, shift invariant, causal, and stable system [8]. The above are some simple examples to demonstrate the problems associated with exact self-similarity.

2.2 The model

Mandelbrot [5] emphasizes that the properties of self-similarity were never meant to hold rigidly when the time lag is extremely small or extremely large. That is, practical discrete time bursty signals exhibiting approximate self-similarity will possess autocovariances that will deviate from the exact profile. for very large j. The discretization can be such that the autocovariances of the bursty signal follow (1) for low (positive) values of j. Any deviation from the profile (1) can be equivalently assumed to be due to the addition of independent zero mean noise sequence whose autocovariances are zero for a lag beyond M. These arguments lead us to consider practical bursty signals to exhibit "limited scale self-similarity" only, and by this, we assume that the profile of autocovariances of the noisy signal follow the profile (1) over a finite range $M < j \leq L$, but L can be as large as desired. Unless otherwise mentioned, we assume that the infinite sequence of autocovariances of the practical signals sum to a finite quantity. In the above, practical signals are stationary stochastic processes and autocovariances are expectations. Sample (observed) signals and sample autocovariances are defined later. Since exact self-similarity is not assumed, the bursty signal (which is an increment process) can be non-Gaussian and still possess finite variance. Therefore, unless otherwise mentioned, the bursty signal is not restricted to be Gaussian.

2.3 Applications of the model

The relaxation of the exact self-similarity profile of autocovariances at smaller lags, $k \leq M$ has particular applications in data traffic engineering, in addition to the possible breakdown of exact self-similarity at low lags. The process of measurement of the data traffic trace may cause short-term correlated perturbations, as follows. The number of bits over successive equal time intervals in the traffic trace may be self-similar but measurements may be made in terms of number of packet fragments over time intervals. There may be a jitter in the time intervals themselves necessitated by having to wait for a packet fragment to end. Transformations in models to generate self-similar traffic may introduce discrepancies in autocovariances at low lags, for example, as follows. Dattatreya and Kulkarni [9] use a high order autoregressive model to synthetically generate a time series representing a bursty traffic trace. The time series (with all positive values) can be used as the number of bytes of bursty traffic over successive time intervals.

3 Hurst Parameter Estimation Literature

Traditional techniques to estimate the Hurst parameter have been largely graphical, except the Whittle's estimator, which is analytically developed. Taqqu, Teverovsky, and Willinger [10] is an empirical study of them. The autocovariance sequence of a bursty signal decays approximately as a negative power of the lag of the autocovariance. Therefore, the correlogram of observed signals plotted on a log-log scale would follow a straight line (approximately), whose slope determines the Hurst parameter. Similarly, the power spectrum of bursty signals varies as a power of the frequency. A log-log plot of the periodogram of the observed bursty signal would be approximately linear and again, the slope determines the Hurst parameter. The Whittle's estimator fits the observed power spectrum to a model spectrum and determines the Hurst parameter by minimizing an error criterion. In practice, the spectrum of the observed signal is computed at discrete points through the use of Discrete Fourier Transform.

Wornell and Oppenheim [4] study estimation of fractal signals from noisy measurements using wavelets. They use the following alternative way to avoid the complications due to the problems of exact self similarity at extremely small and extremely large time lags. The $\frac{1}{f}$ processes are generally defined as processes whose empirical power spectra are of the form $S(\omega) \sim \frac{\sigma^2}{|\omega|^\gamma}$ over several decades of frequency ω, where $0 < \gamma < 2$ and typically γ is around 1. Nearly $\frac{1}{f}$ processes are defined as those whose spectra deviate from that of the exact $\frac{1}{f}$ process by no more than a constant upper and lower bound factors. They use an earlier result that one can construct a class of nearly $\frac{1}{f}$ processes using wavelet expansions in terms of uncorrelated transform coefficients with a particular variance progression profile. They note that a corresponding analysis result appears to exist, at least empirically. That is, for a reasonably arbitrary choice of wavelet, there is strong empirical evidence that the wavelet coefficients from these nearly $\frac{1}{f}$ processes obey a particular variance progression and also turn out to be only weakly correlated. They conclude that the wavelet transform is effective in removing strong long-range dependence from the process. Assuming that the wavelet coefficients have been extracted, they proceed with an EM algorithm [11] to estimate the variance of the signal, the variance of the additive white noise, and the parameter γ (which is affine-related to the Hurst parameter H) from the functional form of the wavelet coefficients. They point out that in this case, the likelihood function has multiple minima, all of which, except the desired solution, are avoidable pathological saddle points.

Abry and Veitch [12] conduct wavelet-based analysis of long-range dependence and develop an explicit closed form estimator for the Hurst parameter. The quantities appearing in the estimator function include inner products of the signal (a function of time) with several shifted and dilated templates of the dual mother wavelet. They mention that these inner products can be efficiently computed by a fast recursive filter-bank-based pyramidal algorithm whose computational cost is extremely low. They point out the following. (*a*) If multiple parameters are unknown, it is important to estimate the Hurst parameter first. (*b*) Under certain conditions, their estimator is asymptotically unbiased, and in practice has very low bias. (*c*) The correlation structure of the data represented by the wavelet coefficients is not long-range-dependent, in contrast to the original data. (*d*) Under Gaussian and quasidecorrelation of the wavelet coefficient hypothesis and in the asymptotic limit, it can be shown that the variance of the estimator is the smallest possible, that is, equal to the Cramer Rao bound, for a given parameter value occurring in their analysis. (*e*) Finite length data pollutes the wavelet coefficients, if better estimation is attempted through a higher value of a particular parameter (in their analysis), and a practical compromise is in order.

Kettani and Gubner [13] propose a simple estimator for the Hurst parameter based on the known relation between the normalized first autocovariance coefficient of exactly self-similar signal and the Hurst parameter. The autocovariance coefficients in equation (2) are already normalized, since $c_0 = 1$. Using $c_1 = 2^{2H-1}$ in equation (2), they propose $\hat{H} = 0.5\left[1 + \log_2(1 + \hat{c}_1)\right]$ as an estimate for H based on the normalized sample average autocovariance coefficient $\hat{c}_1$. They point out a result due to Hosking [14] that for $0 < H < \frac{3}{4}$, the estimate $\hat{c_1}$ is asymptotically normal and then obtain confidence intervals from the same.

4 Development of the Estimator

4.1 The approach and its justification

The first and second order statistics of long-range dependent data such as from a bursty signal model introduced here have been widely cited as possessing poor statistical properties because time average is performed over strongly correlated data. These second order statistics are also biased and this is often cited as another reason not to use such sample statistics to estimate the Hurst parameter. As noted in the above section, sophisticated mathematical transformations and analyses have been developed by researchers to attempt to decorrelate the data. The analyses do show that in their full capabilities (often in the limit, in multiple senses) the proposed estimators attain the best possible behaviors. They also demonstrate by a variety of means that even with practical restrictions, their estimators perform close to their theoretically attainable. It is not clear, in general, that even well developed estimators based on the second order statistics of the observations must necessarily perform much worse than the theoretically attainable. This hypothesis does not contradict the known problem of poor second order statistics. It only suggests that due to the strong and long range persisting correlations in the original signal, the theoretically attainable performance criteria may not be much better than those attainable by good estimators based on second order statistics. In the following, such an estimation equation is developed through an exact analysis of the statistical properties of sample average autocovariances of the noisy-bursty signal.

4.2 Second order statistics of the signal

The following background results are used. Starting from (2) it is easy to show that

$$s_n = \sum_{j=0}^{n} c_j = \frac{1 + (n+1)^{2H} - n^{2H}}{2}, \tag{3}$$

by induction. Similarly, by induction, it is easy to show that

$$t_m = \sum_{n=0}^{m} s_n = \frac{(m+1)^{2H} + m + 1}{2}. \tag{4}$$

Note that $c_0 = s_0 = t_0 = 1$. Let $x_1, ..., x_n, ...$ be the outcomes of the noise-free bursty signal. Let $X_1, ..., X_n, ...$ be the sequence of corresponding random variables with mean μ, variance σ^2, and autocovariances satisfying (1). Let

$u_1, ..., u_n, ...$ be the actual noise values added to the noise-free bursty signal. Let $U_1, .., U_n, ...$ be the corresponding stationary, zero mean additive noise random variables with autocovariances $\rho_0, ..., \rho_M, 0, ..., 0, ...$ That is, the autocovariances of the additive noise are zero beyond a lag of M samples. The bursty signal sequence X_i and the additive noise sequence U_i are independent. $Y_i = X_i + U_i$, $i = 1, ..., n, ...$ are the random variables corresponding to the observed sequence y_i from which H is required to be a estimated. The noise and the bursty signal are independent. Therefore, the autocovariances of the noisy bursty signal sequence work out to be

$$R_k = \sigma^2 c_k + \rho_k. \quad (5)$$

The autocovariance sequence is an even function; that is, $R_k = R_{-k}$. The sample mean of the observed noisy-bursty signal is

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i. \quad (6)$$

The sample average autocovariances are

$$\hat{R}_k = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})(y_{i+k} - \hat{y}). \quad (7)$$

Substituting (6) in (7), simplifying, and taking expectation, we obtain

$$E[\hat{R}_k] = R_k - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(R_{i+k-j}) \quad (8)$$

where R_k is the autocovariance of the noisy-bursty signal. We are interested in $k > M$, above the lag M, so that the contribution of autocovariances of the noise to R_k in (5) is zero. To simplify the double summation in (8), the square grid $j = 1, ..., n;\ i = 1, ..., n$ is split into four regions as in Figure 1. Each region includes the boundary; in each region, the sign of $(i + k - j)$ is unchanging. Using this approach and after cumbersome algebraic manipulation, we obtain

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} R_{|i+k-j|} &= \sum_{l=k}^{n+k-1}(R_k + ... + R_l) \\ &+ \sum_{l=1}^{k-1}(R_l + ... + R_{k-1}) \\ &+ (n-k)(R_0 + ... + R_{k-1}) \\ &+ \sum_{l=1}^{n-k-1}(R_1 + ... + R_l). \quad (9)\end{aligned}$$

As in (5), for all l, $R_l = \sigma^2 c_l + \rho_l$. Separate each summation above into two parts, one containing only c_l and the other containing ρ_l only. After algebraic manipulation,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{|i+k-j|} &= t_{n+k-1} + t_{n-k-1} \\ &- 2t_{k-1} - (n-k). \quad (10)\end{aligned}$$

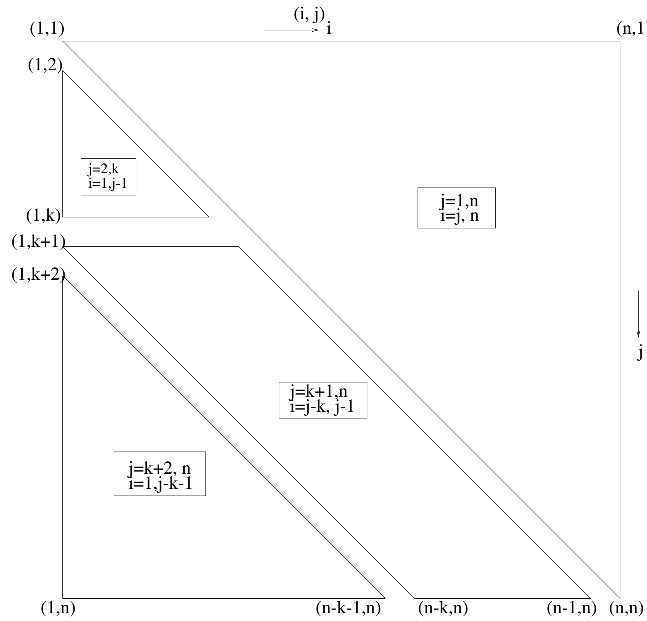

Figure 1. Regions with unchanging signs of $(i + k - j)$

The contributions of the autocovariances of the additive noise in (9) get simplified differently (due to $\rho_k = 0$, $k > M$). They work out to be

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{|i+k-j|} &= \sum_{l=1}^{M}\sum_{j=l}^{M}\rho_j \\ &+ n\sum_{j=0}^{M}\rho_j - k\sum_{j=0}^{M}\rho_j \\ &+ \sum_{l=1}^{M}\sum_{j=1}^{l}\rho_j \\ &+ (n-1-M)\sum_{j=1}^{M}\rho_j \\ &- k\sum_{j=1}^{M}\rho_j \quad (11) \\ &= \eta - k\zeta \quad (12)\end{aligned}$$

where η is that part of RHS of (11) which is invariant to k and

$$\zeta = \sum_{j=1}^{M}\rho_j + \sum_{j=0}^{M}\rho_j \quad (13)$$

is the sum of all coefficients of $-k$ in (11). Combining all the simplifications, and noting that $k > M$ so that $\rho_k = 0$, we have

$$\begin{aligned}E[\hat{R}_k] &= -\frac{t_{n+k-1} + t_{n-k-1} - 2t_{k-1} - n + k}{n^2}\sigma^2 \\ &+ \sigma^2 c_k + \frac{\eta - k\zeta}{n^2} \quad (14)\end{aligned}$$

4.3 Elimination of noise statistics and signal variance

The idea is to add and subtract several sample average autocovariance estimates and eliminate the contributions of the noise statistics in the combined expectation. Thereafter, the proportionality constant σ^2 remains. This can be eliminated by using ratios of two such sums and differences. For any sequence

$$g_k = a + bk, \tag{15}$$

it is easy to verify that

$$\sum_{k=N+1}^{2N} g_k - \sum_{k=3N+1}^{4N} g_k - \sum_{k=5N+1}^{6N} g_k + \sum_{k=7N+1}^{8N} g_k = 0. \tag{16}$$

Similarly,

$$\sum_{k=2N+1}^{3N} g_k - \sum_{k=4N+1}^{5N} g_k - \sum_{k=6N+1}^{7N} g_k + \sum_{k=8N+1}^{9N} g_k = 0. \tag{17}$$

Therefore, define the following linear combinations of sample average autocovariance estimates of the noisy-bursty signal

$$\hat{R}_T = \sum_{k=N+1}^{2N} \hat{R}_k - \sum_{k=3N+1}^{4N} \hat{R}_k - \sum_{k=5N+1}^{6N} \hat{R}_k + \sum_{k=7N+1}^{8N} \hat{R}_k \tag{18}$$

and

$$\hat{R}_B = \sum_{k=2N+1}^{3N} \hat{R}_k - \sum_{k=4N+1}^{5N} \hat{R}_k - \sum_{k=6N+1}^{7N} \hat{R}_k + \sum_{k=8N+1}^{9N} \hat{R}_k. \tag{19}$$

Their expectations are as follows.

$$E[\hat{R}_T] = \frac{\sigma^2}{n^2} \sum_{k\in T} \alpha_k \left(n^2 c_k - t_{n+k-1} - t_{n-k-1} + 2t_{k-1}\right) \tag{20}$$

where the summation in (20) is taken over a set T given by

$$\{N{+}1, .., 2N, 3N{+}1, .., 4N, 5N{+}1, .., 6N, 7N{+}1, .., 8N\} \tag{21}$$

and α_k is a sign variable as follows.

$$\begin{aligned} \alpha_k &= +1, \ k \in \{N+1, ..., 3N, \ 7N+1, ..., 9N\} \\ &= -1, \ k \in \{3N+1, ..., 7N\}. \end{aligned} \tag{22}$$

Similarly,

$$E[\hat{R}_B] = \frac{\sigma^2}{n^2} \sum_{k\in B} \alpha_k \left(n^2 c_k - t_{n+k-1} - t_{n-k-1} + 2t_{k-1}\right) \tag{23}$$

where the summation in (23) is taken over the set B given by

$$\{2N{+}1, ., 3N, \ 4N{+}1, ., 5N, \ 6N{+}1, ., 7N, \ 8N{+}1, ., 9N\} \tag{24}$$

and the sign function α_k is as defined in (22) above. The parts of the sums and differences

$$\sum_{k\in T} \alpha_k c_k \quad \text{and} \quad \sum_{k\in B} \alpha_k c_k \tag{25}$$

can be simplified using (3). For example,

$$\sum_{k=mN+1}^{(m+1)N} c_k = \frac{[(m+1)N+1]^{2H} - [(m+1)N]^{2H}}{2} - \frac{(mN+1)^{2H} - (mN)^{2H}}{2}. \tag{26}$$

Obviously, $\hat{R}_T$ and $\hat{R}_B$ are unbiased estimates for $E[\hat{R}_T]$ and $E[\hat{R}_B]$, respectively. The ratio of these expectations eliminates σ^2 and is a function of H, n, and N only. Of these, n and N are chosen for data analysis depending on the number of samples available, M, the lag beyond which the noise autocovariance is known to vanish, and N is limited by the number of samples available and the largest lag for which the autocovariance profile is known to follow (1). To estimate H, the ratio of $\hat{R}_T$ to $\hat{R}_B$ is equated to the ratio of their corresponding expectations. This results in a nonlinear function of H being equated to a statistic obtained from the noisy-bursty signal observations. This nonlinear equation in H is given by

$$\frac{\sum_{k\in T} \alpha_k \left(n^2 c_k - t_{n+k-1} - t_{n-k-1} + 2t_{k-1}\right)}{\sum_{k\in B} \alpha_k \left(n^2 c_k - t_{n+k-1} - t_{n-k-1} + 2t_{k-1}\right)} = \frac{\hat{R}_T}{\hat{R}_B}. \tag{27}$$

In an estimation experiment, the RHS of (27) is a number and H is the argument of the LHS of (27). The LHS is well behaved. Figure 2 shows a plot of the LHS of (27) as a function of $H \in (0.5, \ 1.0)$ for three different combinations of $(n, \ N)$. Both $E[\hat{R}_T]$ and $E[\hat{R}_B]$ are zero at the extreme values of $H = 0.5$ and 1.0. Therefore, the function is undefined at these extreme values. The determination of the estimate for H reduces to selecting the value of H for which the observed statistic $\frac{\hat{R}_T}{\hat{R}_B}$ corresponds to the ordinate of the LHS of (27) drawn as a function of H. This is easily accomplished either as a numerical solution to the nonlinear equation (27) or through a table look up.

5 Conclusion

Exact self-similarity of signals is very restrictive and presents mathematical and practical difficulties at extremely small and extremely large time lags. Our model allows the profile of autocovariances to deviate from that of exact self similarity for small and very large time lags. This overcomes four difficulties as follows. (*a*) The breakdown of exact self-similarity at very small lags due to its requirement of unbounded instantaneous variance is overcome. (*b*) Additive short-term correlated (moving average type) perturbations in the measurement are dealt with. (*c*) The autocovariances of the bursty signal for lags beyond any practical significance can be assumed to fall off in a way that their

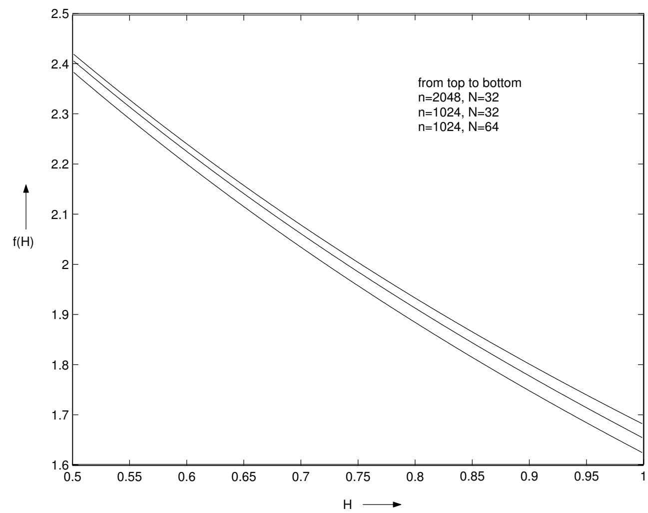

Figure 2. $\frac{E[\hat{R}_T]}{E[\hat{R}_B]}$ as a function of H; $n = 1024$ and $N = 64$

infinite sum is finite, to overcome analytical difficulties. (*d*) The variance of stationary bursty signal can be finite without requiring the signal to be Gaussian.

Applications of the model are discussed. Simple linear combinations (sums and differences) of sample average second order statistics are used to eliminate the noise statistics in their expectations. The ratio of expectations of two such sample averages eliminates the signal variance. This results in a smooth monotonically decreasing function of the single unknown, H, being equated to the single computed statistic from the noisy-bursty signal. Solving for H from the resulting equation is very simple. The number of required observation samples is $n + 9N$. The largest lag for which the sample average autocovariance is sought is $9N$ and hence the model autocovariance profile in (1) must hold for upto $9N$.

The development here illustrates a particular choice of sums and differences of sample average autocovariances. Other choices can be explored to increase the dynamic range of the ratio of expectations. The variance of the estimate of H is not studied here and is definitely an important one. Derivation of the exact variance of the final H estimate appears to be very difficult. Other approaches to assess the quality of the final estimate and comparisons to other estimators in the literature are also important studies. With the estimator developed for H here, approaches to estimate the autocovariances of the additive noise do not appear to be very difficult. The smoothing of the effect of noise with the help of these estimates is another important problem. Finally, data analysis experiments with simulated data and Ethernet data traffic trace are also important.

References

[1] K. Park and W. Willinger (eds.), *Self-Similar Network Traffic and Performance Evaluation*. NY: John Wiley and Sons, 2000.

[2] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motion, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422-437, Oct. 1968.

[3] J. Beran, *Statistics for Long-Memory Processes*. NY: Chapman and Hall, 1994.

[4] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Signal Processing*, vol. 40, pp. 611 - 623, Mar. 1992.

[5] B. B. Mandelbrot, "Self-similar error clusters in communication systems and the concept of conditional stationarity," *IEEE Transactions on Communication Technology*, vol. 13, pp. 71-90, Mar. 1965.

[6] B. B. Mandelbrot, *The Fractal Geometry of Nature*. NY: W. H. Freeman, 1983.

[7] S. S. Kulkarni, "Adaptive load-balancing over multiple routes in mobile ad hoc networks," Ph. D. dissertation, Department of Computer Science, University of Texas at Dallas, Richardson, TX, Dec. 2002.

[8] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[9] G. R. Dattatreya and S. S. Kulkarni, "Performance of communication networks fielding bursty data traffic," Annual Review of Communications, vol. 57, pp. 1259 - 1273, Nov. 2004.

[10] M. S. Taqqu, V. Teverovsky, W. Willinger, "Estimators for long-range dependence: an empirical study," *Fractals*, vol. 3, pp. 785 - 798, 1995.

[11] N. M. Laird, A. P. Dempster, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Ann. Roy. Stat. Soc.*, pp. 1 - 38, Dec. 1977.

[12] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Information Theory*, vol. 44, pp. 1 - 15, Jan. 1998.

[13] H. Kettani and J. A. Gubner, "A novel approach to the estimation of the Hurst parameter in self-similar traffic," Proceedings of the 27th Annual IEEE Conference on Local Computer Networks (LCN 2002), Tampa, Florida, November 2002, pp. 160 - 165.

[14] J. R. M. Hosking, "Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series," *J. Econometrics*, vol. 73, pp. 261 - 284, 1996.