MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS OF FRACTIONAL BROWNIAN TRAFFIC WITH GEOMETRICAL SAMPLING

Attila Vidács
High Speed Networks Laboratories
Dept. of Telecommunications and Telematics
Technical University of Budapest
Pázmány Péter sétány 1/D,
H-1117 Budapest, Hungary
vidacs@ttt-atm.ttt.bme.hu

Jorma T. Virtamo
Lab. of Telecommunications Technology
Helsinki University of Technology
P.O.B. 3000, FIN-02015 HUT, Finland
jorma.virtamo@hut.fi

Abstract Traffic model based on the fractional Brownian motion (fBm) contains three parameters: the mean rate m, variance parameter a and the Hurst parameter H. The estimation of these parameters by the maximum likelihood (ML) method is studied. Explicit expressions for the ML estimates $\hat{m}$ and $\hat{a}$ in terms of H are given, as well as the expression for the log-likelihood function from which the estimate $\hat{H}$ is obtained as the maximizing argument. A geometric sequence of sampling points, $t_i = \alpha^i$, is introduced in order to see the scaling behaviour of the traffic with fewer samples. It is shown that by a proper 'descaling' the traffic process is stationary on this grid leading to a Toeplitz-type covariance matrix. Approximations for the inverted covariance matrix and its determinant are introduced. The accuracy of the estimation algorithm is studied by simulations. Comparisons with corresponding estimates obtained with linear grid show that the geometrical sampling indeed improves the accuracy of the estimate $\hat{H}$ with a given number of samples.

Keywords: Traffic modeling, fractional Brownian motion, Maximum Likelihood Estimation.

1. INTRODUCTION

One of the simplest and most studied models for aggregated data traffic is the fractional Brownian motion (fBm) model [7], which is a model for truly self-similar Gaussian traffic. An important feature of the fBm model is its parsimonity [3]: in its basic form the model contains only three parameters, the mean rate m, the variance parameter a and the Hurst parameter H describing the scaling behaviour of the traffic. The estimation of even a small number

D. H. K. Tsang et al. (eds.), *Broadband Communications*

of parameters poses a problem for long range dependent traffic. Some early work [7] suggested that to obtain a reasonable accuracy a very large number of sample points may be required. As H describes the scaling behaviour of the traffic variability, the sample points have to cover several time scales, i.e., the total time range must be several orders of magnitude greater than the finest time resolution in the measurement.

In this paper we show that by an appropriate choice of the sampling instants, the number of sampling points can be considerably reduced. In particular, we will introduce a grid of geometrically distributed sampling points $t_i = \alpha^{i-1}$, $i = 1, \ldots, n$ where α is some constant less than one. The geometrical grid, being 'self-similar' fits well with the traffic process and gives rise to a simple structure in the covariance matrix.

Throughout this work we apply the maximum likelihood estimation (MLE) method [1]. MLE method has previously been applied to this problem by Deriche and Tewfik [2] and Ninness [5] using ordinary linear sampling. Explicit formulas for the estimators of m and a are given along with the log-likelihood function for determining the estimator for H. A major difficulty in this method is the calculation of the inverse and determinant of the covariance matrix. For the original fBm process the increment process is stationary. We show that another stationary process is obtained from the fBm process by 'descaling' and changing the process index to logarithmic time, i.e., on the geometrical sampling grid the descaled process is stationary. It turns out that the elements of the inverse covariance matrix far from the diagonal are small, enabling us to derive a simple approximation for the inverse matrix directly without using e.g. Whittle's method [1] based on the spectral analysis.

We compare the effectiveness of the MLE estimator based on ordinary evenly spaced sampling grid with that obtained with a geometrical grid by simulations.

The rest of this paper is organized as follows. In Section 2 we review the fractional Brownian motion traffic model with its three parameters. The general problem of the estimation of these parameters by the maximum likelihood method is considered in Section 3. The idea of geometrical sampling and the descaled process, along with an approximate form of the MLE, are introduced in Section 4. For comparison, in Section 5 we present the MLE method for the case of ordinary linear sampling. In Section 6, we present results for estimating the fBm parameters with the described methods from simulated realizations of the process. Section 7 concludes the paper.

2. FRACTIONAL BROWNIAN TRAFFIC

A normalized *fractional Brownian motion* with Hurst-parameter $H \in [0.5, 1)$, denoted by $Z(t)$, $(t \in \mathbb{R})$, is characterized by the following properties [6]:

1. $Z(t)$ has stationary increments;

2. $Z(0) = 0$, and $\mathrm{E}[Z(t)] = 0$ for all t;
3. $\mathrm{Var}[Z(t)] = \mathrm{E}\left[Z(t)^2\right] = |t|^{2H}$ for all t;
4. $Z(t)$ has continuous paths;
5. $Z(t)$ is a Gaussian process, i.e., all its finite-dimensional marginal distributions are Gaussian.

In the special case $H = 0.5$, $Z(t)$ is the standard Brownian motion. It follows from the above properties that $Z(t)$ is a self-similar process whose scaling behaviour is defined by the Hurst-parameter H as follows:

$$Z(\alpha t) \sim \alpha^H Z(t). \tag{1}$$

The covariance structure of the process is given by

$$\mathrm{Cov}\left[Z(t_1), Z(t_2)\right] = \frac{1}{2}\left\{t_1^{2H} + t_2^{2H} - |t_2 - t_1|^{2H}\right\}. \tag{2}$$

Fractional Brownian motion is a popular model for long-range dependent traffic. Norros [6] has suggested the following model

$$X(t) = mt + \sqrt{a}Z(t), \tag{3}$$

where $X(t)$ represents the amount of traffic arrived in $(0, t)$. The model has three parameters, m, a and H with the following interpretations and intervals for allowed values: $m > 0$ is the mean input rate, $a > 0$ is a variance parameter, and $H \in [0.5, 1)$ is the self-similarity parameter of $Z(t)$.

3. EXACT GAUSSIAN MLE

We use the notation of Beran [1]. Assume the traffic has been observed at n time instants forming the vector $\mathbf{t} = (t_1, \ldots, t_n)^{\mathrm{t}}$ where $(\cdot)^{\mathrm{t}}$ denotes the transpose. And let $\mathbf{X} = (X(t_1), \ldots, X(t_n))^{\mathrm{t}}$ be the vector of observed traffic values at these instants. Since $X(t)$ is Gaussian, the joint probability density function of $\mathbf{X}$ is

$$h(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\mathbf{\Gamma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{t}}\mathbf{\Gamma}^{-1}(\mathbf{x}-\mathbf{m})}, \tag{4}$$

where $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{t}} \in \mathrm{I\!R}^n$, $\mathbf{m} = m\mathbf{t}$, and $|\mathbf{\Gamma}|$ is the determinant of the covariance matrix

$$\mathbf{\Gamma} = \mathrm{Cov}\left[\mathbf{X}, \mathbf{X}^{\mathrm{t}}\right] = \mathrm{E}\left[\mathbf{X}\mathbf{X}^{\mathrm{t}}\right] - \mathrm{E}[\mathbf{X}]\,\mathrm{E}\left[\mathbf{X}^{\mathrm{t}}\right]. \tag{5}$$

The MLE for m is obtained by maximizing $\log h(\mathbf{X}; m)$ with respect to m, resulting in the estimator

$$\hat{m} = \hat{m}(H) = \frac{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}^{-1}\mathbf{X}}{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}^{-1}\,\mathbf{t}}. \tag{6}$$

Note, that the estimate is unbiased, irrespective whether our estimate for H is correct or not. The variance of $\hat{m}$ can also be calculated with the assumption that H is known exactly, $\hat{H} = H$. With straightforward calculations we get

$$\mathrm{Var}\,[\hat{m}] = \frac{a}{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}^{-1}\,\mathbf{t}}. \tag{7}$$

The variance of our estimator is smaller than for the estimator based on the sample mean, by the factor in the denominator (which is close to one).

Next, consider the estimator for a. $\mathbf{\Gamma}$ is a simple linear function of a, $\mathbf{\Gamma} = a\,\mathbf{\Gamma}_H$, where $\mathbf{\Gamma}_H$ is independent of a and is given by

$$\mathbf{\Gamma}_H = \mathrm{E}\left[\mathbf{Z}\mathbf{Z}^{\mathrm{t}}\right] = \left[\mathrm{Cov}\,[Z(t_i), Z(t_j)]\right]_{i,j=1,\ldots,n}. \tag{8}$$

The MLE of a is obtained by maximizing the log-likelihood function $\log h(\mathbf{X}; a)$ with respect to a. If we do not know the mean input rate m in advance, $\mathbf{m}$ should be replaced by $\hat{m}\mathbf{t}$, and we get:

$$\hat{a}(H) = \frac{1}{n}\frac{(\mathbf{X}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{t}}. \tag{9}$$

Again, assuming for the time being that H is known correctly the expectation and variance of $\hat{a}$ can be calculated and finally we have

$$\mathrm{E}\,[\hat{a}] = \frac{n-1}{n}a, \qquad \mathrm{Var}\left[\frac{n}{n-1}\hat{a}\right] = \frac{2a^2(n-1)}{n^2}. \tag{10}$$

Thus $\hat{a}$ has the "normal" $(n-1)/n$ bias.

Finally, we are left with the maximization of the H-dependent part of the log-likelihood function, i.e., essentially we have to minimize

$$\tilde{L}(\mathbf{X}; H) = \log|\mathbf{\Gamma}_H| + n\log\frac{(\mathbf{X}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X})(\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{t}) - (\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{X})^2}{\mathbf{t}^{\mathrm{t}}\,\mathbf{\Gamma}_H^{-1}\,\mathbf{t}}. \tag{11}$$

The minimum is obtained for some value $\hat{H}$ which is the MLE estimate; the corresponding MLE estimates for m and a are $\hat{m} = m(\hat{H})$ and $\hat{a} = a(\hat{H})$.

4. GEOMETRICAL SAMPLING

The Hurst parameter H describes the scaling behaviour of the traffic. Therefore, in order to determine its value from measured traffic, the sample points have to cover several time scales, i.e., the total time range of the measurements has to be many orders of magnitude greater than the finest resolution (smallest interval between the sampling points). With the ordinary linear sampling, i.e., sampling points at constant intervals, this leads to the requirement of very large number of sampling points. In order to use the measurements more efficiently we introduce a geometric sequence of sampling points, $t_i = \alpha^i$, $i = 1, \dots, n$, with some $0 < \alpha < 1$.

In addition to distributing the sampling points in a better way on different time scales, geometric sampling fits neatly with the self-similar behaviour of the fBm traffic. We show first that by a simple transformation we can obtain from the fBm process another process which is a stationary process of logarithmic time. As a geometric sequence corresponds to equidistant points in logarithmic time, it follows that the samples of the modified process constitute a stationary sequence. This leads to a simple Toeplitz-type structure of the covariance matrix and allows us to develop approximations to the inverse and determinant of the covariance matrix.

4.1 DESCALED PROCESS

$Z(t)$ has the self-similar property $Z(\alpha t) \sim \alpha^H Z(t)$. Now consider the 'descaled' process $\breve{Z}(t) \overset{d}{=} t^{-H} Z(t)$ which has the scaling property

$$\breve{Z}(\alpha t) \sim (\alpha t)^{-H} Z(\alpha t) = t^{-H} Z(t) = \breve{Z}(t). \tag{12}$$

Further let us take a new time variable $u = -\log t$ and denote $\tilde{Z}(u) \overset{d}{=} \breve{Z}(e^{-u}) = \breve{Z}(t)$. Now we have

$$\tilde{Z}(u - \log\alpha) = \breve{Z}(e^{-u+\log\alpha}) = \breve{Z}(\alpha e^{-u}) = \breve{Z}(\alpha t) \sim \breve{Z}(t) = \tilde{Z}(u). \tag{13}$$

Thus the process $\tilde{Z}(u)$ is stationary and has the following covariance structure:

$$\mathrm{Cov}\left[\tilde{Z}(u_1), \tilde{Z}(u_2)\right] = \frac{1}{2} e^{H(u_2-u_1)} \left\{1 + e^{-2H(u_2-u_1)} - \left(1 - e^{-(u_2-u_1)}\right)^{2H}\right\}, \tag{14}$$

so the descaled process $\tilde{Z}(u)$ is short range dependent.

If we 'descale' the process $X(t)$ by the factor t^{-H} and use u as the process index, the covariance matrix $\tilde{\mathbf{\Gamma}}$ of the descaled samples $\tilde{\mathbf{X}} = (\tilde{X}(u_1), \tilde{X}(u_2), \dots, \tilde{X}(u_n))^{\mathrm{t}}$ with $u_i = -\log t_i = (1 - i)\log\alpha$ can be written as

$$\tilde{\mathbf{\Gamma}} = \mathrm{E}\left[\tilde{\mathbf{X}}\tilde{\mathbf{X}}^{\mathrm{t}}\right] = a \cdot \mathrm{E}\left[\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^{\mathrm{t}}\right]. \tag{15}$$

Note, that our geometrical grid is now equally spaced with regard to u. Thus, if we use the notation $\tilde{Z}_i = \tilde{Z}(u_i)$ the process $\tilde{\mathbf{Z}} = (\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_n)$ is a stationary process in discrete time with zero mean and unit variance and its auto-correlation function $\rho(k)$ can be defined as

$$\rho(i - j) = \frac{1}{2}\alpha^{-H|i-j|} \left\{1 + \alpha^{2H|i-j|} - \left(1 - \alpha^{|i-j|}\right)^{2H}\right\}, \tag{16}$$

and thus

$$\tilde{\mathbf{\Gamma}}_{ij} = a\rho(i - j), \quad i, j = 1, 2, \dots, n. \tag{17}$$

4.2 APPROXIMATE MLE

In practice, the exact MLE poses computational problems. And this is not just because of the computation time needed in case of large data sets, but because of the evaluation of the inverse and the determinant of the covariance matrix may be numerically unstable. To avoid these problems, one can use approximate methods for the calculations. In [1], several possible approaches are discussed, among them the well known Whittle's approximate MLE.

In our case we focus on the properties of the covariance matrix $\mathbf{\Gamma}_H$, trying to take advantage of the stationarity and short range dependent properties of the descaled process. Using the 'descaling matrix' $\mathbf{D} = \mathrm{diag}(t_1^{-H}, \dots, t_n^{-H})$ we can easily derive $\tilde{\mathbf{\Gamma}} = \mathbf{D}\,\mathbf{\Gamma}\,\mathbf{D}$, and from this we get

$$\mathbf{\Gamma}_H^{-1} = \mathbf{D}\tilde{\mathbf{\Gamma}}_H^{-1}\mathbf{D}; \qquad |\mathbf{\Gamma}_H| = \alpha^{Hn(n-1)}|\tilde{\mathbf{\Gamma}}_H|. \tag{18}$$

The elements of the autocorrelation matrix $\tilde{\mathbf{\Gamma}}_H$ can be written as

$$(\tilde{\mathbf{\Gamma}}_H)_{i,j} = \alpha^{-H|i-j|}g(\alpha^{|i-j|}), \quad i,j = 1,2,\dots,n \tag{19}$$

with

$$g(x) = \frac{1}{2}\left(1 + x^{2H} - (1-x)^{2H}\right). \tag{20}$$

It is interesting to note, that $g(x)$ is nearly completely linear for $x \in (0,1)$. Figure 1 shows the difference of $g(x) - x$ for different values of H. It can be seen from the plot that the largest absolute difference is less than 0.02 for each value of H. This observation gives us the idea to use the approximation $g(x) \approx x$. So $\tilde{\mathbf{\Gamma}}_H$ can be approximated as $\tilde{\mathbf{\Gamma}}_H \approx \mathbf{R}$, where $\mathbf{R}$ is a Toeplitz-type matrix of the form $[\mathbf{R}]_{ij} = \gamma^{|i-j|}$, $i,j = 1,2,\dots n$, with $\gamma = \alpha^{1-H}$.

The inverse of $\mathbf{R}$ can be easily calculated as [8]

$$\mathbf{R}^{-1} = \frac{1}{\frac{1}{\gamma} - \gamma}\begin{pmatrix} \gamma^{-1} & -1 & 0 & \cdots & 0 \\ -1 & \gamma+\gamma^{-1} & -1 & \ddots & \vdots \\ 0 & -1 & \gamma+\gamma^{-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & \gamma^{-1} \end{pmatrix}, \tag{21}$$

and the determinant of $\mathbf{R}$ is given as $|\mathbf{R}| = (1-\gamma^2)^{n-1}$ [8].

Using the fact $\mathbf{t}^t\,\mathbf{D}\mathbf{R}^{-1}\mathbf{D}\,\mathbf{t} = 1$ and $\mathbf{t}^t\,\mathbf{D}\mathbf{R}^{-1}\mathbf{D} = (1,0,\dots,0)$, we get $\hat{m} = X(1)$ so using the above approximation the MLE estimate for m reduces simply to the sample mean. As for the estimate for a we get

$$\hat{a}(H) = \frac{1}{n}\left(\mathbf{X}^t\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} - X_1^2\right). \tag{22}$$

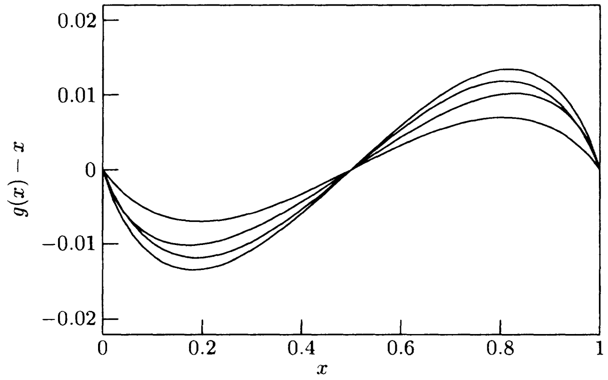

Figure 1 Error of approximation $g(x) \approx x$ for $H = 0.6, 0.7, 0.8$ and 0.9.

Finally, to get an estimate for H we have to minimize the function

$$L(\mathbf{X}; H) = \frac{n-1}{n} \log\left(\alpha^{nH}(1-\alpha^{2-2H})\right) + \log\left(\mathbf{X}^{t}\,\mathbf{D}\,\mathbf{R}^{-1}\mathbf{D}\,\mathbf{X} - X_1^2\right). \tag{23}$$

It should be noted that though the linear approximation to $g(x)$ is rather accurate, the resulting inverse matrix $\mathbf{R}^{-1}$ of Eq.(21) is rather poor an approximation to $\tilde{\mathbf{\Gamma}}^{-1}$ for large n. Nevertheless, the use of $\mathbf{R}^{-1}$ in the log-likelihood function (23), as we will see, yields a good estimate for H, while the accuracy of the estimate $\hat{a}$ suffers more from this approximation.

4.3 IMPROVED APPROXIMATION

Since the matrix $\tilde{\mathbf{\Gamma}}$ is a Toeplitz-type matrix with decreasing elements as we go farther from the diagonal, we expect that its inverse can be well approximated with a band matrix of the form:

$$\mathbf{C} = \begin{pmatrix} c_1 & \cdots & c_p & 0 & \cdots & 0 \\ \vdots & c_1 & \vdots & c_p & \ddots & \vdots \\ c_p & \cdots & c_1 & \vdots & \ddots & 0 \\ 0 & c_p & \cdots & c_1 & \cdots & c_p \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & c_p & \cdots & c_1 \end{pmatrix} \tag{24}$$

so that $\tilde{\mathbf{\Gamma}}_H^{-1} \approx \mathbf{C}$. Our aim is to set the p parameters $c_1, \ldots, c_p$ to get $\mathbf{C}\tilde{\mathbf{\Gamma}}_H \approx \mathbf{E}$. For example, this can be achieved by solving the equation

$$(c_p, \ldots, c_2, c_1, c_2, \ldots, c_p) \cdot \mathbf{G} = (0, \ldots, 0, 1, 0, \ldots, 0), \tag{25}$$

where $\mathbf{G} = (\tilde{\mathbf{\Gamma}}_H)_{(2p-1)\times(2p-1)}$ and from this we have

$$c_i = \mathbf{G}^{-1}_{p(p+i-1)}, \quad i = 1, 2, \ldots, p. \tag{26}$$

With this approximation we only need to calculate the inverse of a $(2p-1)$-by-$(2p-1)$ matrix.[1] (To improve the approximate inverse, its elements in the upper-left and lower-right corners can be further corrected.)

5. LINEAR SAMPLING

Let $\mathbf{X} = (X(t_1), X(t_2), \ldots, X(t_n))^{\mathrm{t}}$ be the vector of observed traffic values at instances

$$t_i = \frac{i}{n}, \quad i = 1, 2, \ldots, n. \tag{27}$$

The increment sequence $(Y_1, Y_2, \ldots)$ with $Y_i = X(t_i) - X(t_{i-1})$ (substituting $X(t_0) \equiv X(0) = 0$) is a strongly correlated stationary sequence with

$$\mathrm{Cov}\,[Y_i, Y_j] = \frac{1}{2} a n^{-2H} \left(|i-j+1|^{2H} + |i-j-1|^{2H} - 2|i-j|^{2H} \right) \tag{28}$$

for $i, j = 1, 2, \ldots, n$. The formulas for the exact Gaussian MLE for this increment process are nearly the same as in Section 3, we only need to replace the covariance matrix $\mathbf{\Gamma}$ with $\mathbf{\Sigma} = [\mathrm{Cov}\,[Y_i, Y_j]]_{i,j=1,2,\ldots,n}$, and the vector $\mathbf{t}$ with the vector $(1/n, 1/n, \ldots, 1/n)^{\mathrm{t}}$. After some minor simplifications we get an estimate for m

$$\hat{m} = \hat{m}(H) = \frac{\mathbf{1}^{\mathrm{t}}\, \mathbf{\Sigma}^{-1}\, \mathbf{Y}}{\mathbf{1}^{\mathrm{t}}\, \mathbf{\Sigma}^{-1}\, \mathbf{1}} \cdot n \tag{29}$$

where $\mathbf{1}$ is a vector of ones, and $\mathbf{\Sigma} = a \mathbf{\Sigma}_H$. For a we have the estimator

$$\hat{a}(H) = \frac{1}{n} \left(\mathbf{Y}^{\mathrm{t}}\, \mathbf{\Sigma}_H^{-1}\, \mathbf{Y} - \frac{(\mathbf{1}^{\mathrm{t}}\, \mathbf{\Sigma}^{-1}\, \mathbf{Y})^2}{\mathbf{1}^{\mathrm{t}}\, \mathbf{\Sigma}^{-1}\, \mathbf{1}} \right). \tag{30}$$

Again, finally we have to minimize

$$\tilde{L}(\mathbf{Y}; H) = \log |\mathbf{\Sigma}_H| + n \log \left(\mathbf{Y}^{\mathrm{t}}\, \mathbf{\Sigma}_H^{-1}\, \mathbf{Y} - \frac{(\mathbf{1}^{\mathrm{t}}\, \mathbf{\Sigma}^{-1}\, \mathbf{Y})^2}{\mathbf{1}^{\mathrm{t}}\, \mathbf{\Sigma}^{-1}\, \mathbf{1}} \right). \tag{31}$$

The minimum is obtained for some value $\hat{H}$ which is the MLE estimate.

However, to calculate the inverse and the determinant of $\mathbf{\Sigma}_H$ the same problems arise as in the case of geometrical sampling with the covariant matrix $\tilde{\mathbf{\Gamma}}_H$. Since $\mathbf{\Sigma}_H$ is also a Toeplitz type matrix, the same method as described in Section 4.3 can be used to approximate $\mathbf{\Sigma}_H^{-1}$ with $\mathbf{C}$ of Eq.(24).

[1]To be more exact, because of the symmetric structure we only need to calculate the inverse of a p-by-p matrix using slightly more complicated formulas.

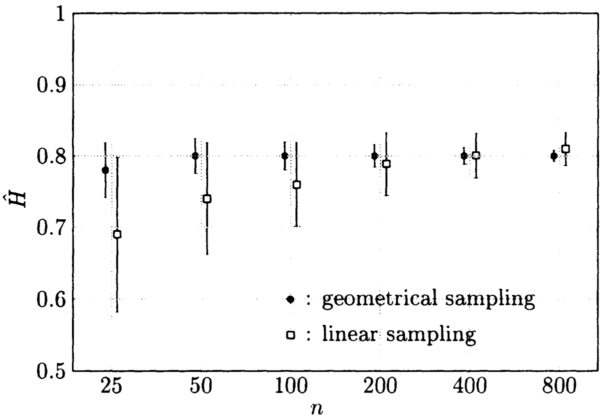

Figure 2 Estimates of H using geometrical and linear sampling.

6. SIMULATION RESULTS

The fBm samples were generated using the fact $\mathbf{Z} \sim \mathbf{\Gamma}_H^{1/2}\mathbf{N}$ where $\mathbf{N}$ is a vector of independent standard Gaussian variables. The model parameters were set as $m = 1$, $a = 1$ and $H = 0.8$ as an example, but similar results were obtained using different values of the parameters. The parameter α for the geometrical grid was chosen so that the difference between the nearest two measurement time instants (the 'resolution' of the measurement) was 10^{-6}. Figure 2 shows the results of H estimates as a function of the number of sample points using both geometrical and linear sampling. In the geometrical case Eq.(23) was minimized while for the linear sampling we used the formula Eq.(31) where the inverse of Σ_H was approximated with a band matrix of Eq.(24) with $p = 2$. The 95% confidence interval was obtained by repeating the simulations 100 times and calculating the sample variance of the estimates. The results show that the estimates using geometrical sampling have much smaller variance and are unbiased for sample sizes larger than 25. However, the variance of the estimates is always higher than in the geometrical case. For example, the variance for 800 samples using linear sampling is nearly the same as for only 50 geometrically sampled points.

The next question was how the two different sampling methods affect the estimates for the variance parameter a. Figure 3 displays the results, assuming that H is known. These simulations were useful to test whether our approximations in calculating the inverse and determinant of the covariance matrices are adequate or not. Figure 3 presents two different approximations for the geometrical sampling. First, we used the simple approximate inverse covariance matrix of Eq.(21) in Eq.(9) using Eq.(18) (denoted by light gray dots and

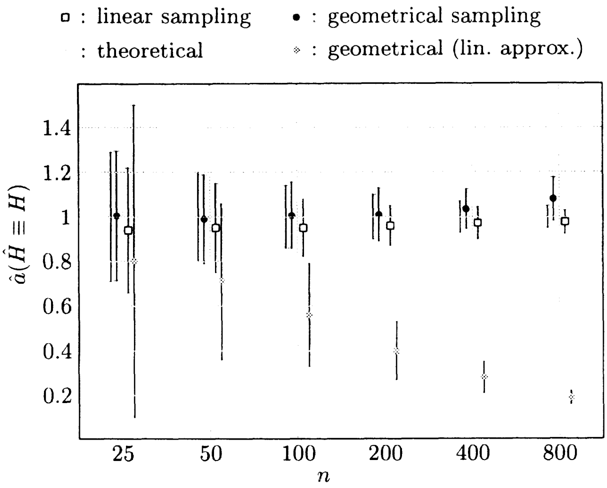

Figure 3 Estimates of a using geometrical and linear sampling and different approximations, assuming H is known.

labeled 'linear approximation' in the figure). As can be seen, the estimates of a are strongly biased and the bias is getting larger as the number of samples increases. So this estimate is clearly inadequate, the approximation of Eq.(21) had to be refined. Next, we used the approximation of Eq.(24) for $\tilde{\Gamma}_H^{-1}$ with five parameters ($p = 5$). As we see from Figure 3, the strong bias from the $\hat{a}$ estimates disappeared and the variance of the estimates is only slightly higher than the theoretical value that can be calculated using Eq.(10). (Note, however, that the bias for sample sizes of 400 and 800 seems to be slightly increased.) Finally, the linear sampling method was used. Its estimates are asymptotically unbiased and have approximately the same variance as expected. The approximate inverse matrix used was as in Eq.(24) with only two parameters ($p = 2$). Figure 4 shows the MLE $\hat{a}$ estimates without any *a priori* knowledge about the model parameters. All the approximations used here were the same as in the previous cases. Since H is not known and can only be estimated with a given variance, the estimates of a have larger variances than in the previous simulations. The question is how robust those estimates are when $\hat{H}$ can have a slight bias (see Figure 2). As for the geometrical sampling, the bias of $\hat{a}$ gets smaller and its variance is also decreasing rapidly as the sample size increases. On the other hand, for the linear sampling case the estimates seem to be biased for larger sample sizes and their variance does not seem to decrease. The reason for this behaviour lies in the fact that the linear sampling for estimating H is less accurate than the geometrical sampling. The bias in $\hat{H}$ together with its higher variance is responsible for the bias and variance of $\hat{a}$, even if the linear

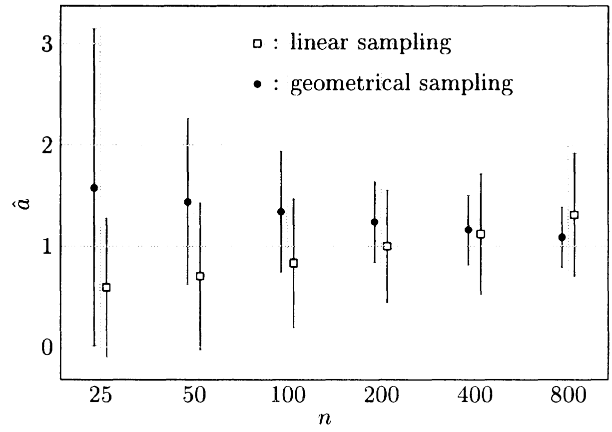

Figure 4 Estimates of a (when H is also estimated) using geometrical and linear sampling and different approximations.

sampling seems to be a better choice to estimate a than the geometrical one for known H (see Figure 3).

As for the MLE estimates for m the geometrical sampling does not give any extra advantage or disadvantage compared to the linear sampling. In fact, the MLE estimate gives almost negligible reduction in the variance of $\hat{m}$ when compared to the sample mean as an estimate for m.

7. CONCLUSION

In this paper we have introduced the idea of using geometrical sampling for the ML estimation of the parameters of fractional Brownian traffic. The intention with this sampling is to reduce the number of sampling points required for a given predefined confidence level. Intuitively, the geometrical sampling distributes the sampling points advantageously at different time scales, whereas linear sampling stresses the finest time scale.

We have derived expressions for the estimators of m and a and the log-likelihood function from which the estimator of H can be derived. Approximations were developed for the inverse and the determinant of the covariance matrix, needed for the calculation of the estimates. With these approximations the evaluation of the log-likelihood function is fast and the maximization with respect to H can easily be made.

The experiments with simulated traffic showed that the geometrical sampling does indeed give a better estimate for H leading to a reduction of sample points. In one example the number of required points was reduced from 800 to 50. For the estimation of a the geometrical sampling does not give any direct advantage, but as the estimator $\hat{a}$ actually depends on the estimator $\hat{H}$, the overall accuracy obtained is better. For the estimation of m, different sampling schemes give essentially the same result, the estimate is basically the observed average rate.

Acknowledgments

This work was done within the Com^2 project funded by the Academy of Finland. Useful discussions with Prof. V. Sharma are gratefully acknowledged.

References

[1] Beran, J. *Statistics for Long-Memory Processes.* Chapman & Hall, One Penn Plaza, New York, NY 10119, 1994.

[2] Deriche, M. and Tewfik, A.H. Maximum Likelihood Estimation of the Parameters of Discrete Fractionally Differenced Gaussian Noise Process. *IEEE Transactions on Signal Processing*, 41(10):2977–2989, October, 1993.

[3] Erramilli, A., Pruthi, P. and Willinger, W. Self-Similarity in High-Speed Network Traffic Measurements: Fact or Artifact? *Proc. in 12th Nordic Teletraffic Seminar—NTS12*, (Eds, Norros, I. and Virtamo, J.), pages 299–310, Espoo, Finland, 22-24 August, 1995.

[4] Krishnan, K.R., Neidhardt, A.L. and Erramilli, A. Scaling Analysis in Traffic Management of Self-Similar Processes. *Proc. of 15th International Teletraffic Congress—ITC15*, (Eds, Ramaswami, V. and Wirth, P.E.), pages 1087–1096, Washington, DC, USA, 22-27 June, 1997.

[5] Ninness, B. Maximum likelihood estimation of the parameters of fractional Brownian motions. *Proc. of the 34th IEEE Conference on Decision and Control*, 4:4018–4023, 1995.

[6] Norros, I. A storage model with self-similar input. *Queueing Systems*, 16:387–396, 1994.

[7] Roberts, J., Mocci, U. and Virtamo, J. eds. Broadband Network Teletraffic—Performance Evaluation and Design of Broadband Multiservice Networks. *Final Report of Action COST 242*, pages 354–369, Springer, 1996.

[8] Rózsa, P. *Linear Algebra and its Applications.* University Press, Budapest, 1974. (in Hungarian)