Impact of the sampling rate on the estimation of the parameters of fractional Brownian motion

Zhengyuan Zhu and Murad S. Taqqu

Technical Report #2004-14
January 28, 2004

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

Impact of the sampling rate on the estimation of the parameters of fractional Brownian motion *†

Zhengyuan Zhu
University of North Carolina
Chapel Hill

Murad S. Taqqu
Boston University

January 28, 2004

Abstract

Fractional Brownian motion is a mean zero self-similar Gaussian process with stationary increments. Its covariance depends on two parameters, the self-similar parameter H and the variance C. Suppose that one wants to estimate optimally these parameters by using n equally spaced observations. How should these observations be distributed? We show that the spacing of the observations does not affect the estimation of H (this is due to the self-similarity of the process), but the spacing does affect the estimation of the variance C. For example, if the observations are equally spaced on $[0, n]$ (unit-spacing), the rate of convergence of the maximum likelihood estimator (MLE) of the variance C is $\sqrt{n}$. However, if the observations are equally spaced on $[0, 1]$ ($1/n$-spacing), or on $[0, n^2]$ (n-spacing), the rate is slower, $\sqrt{n}/\log n$. We also determine the optimal choice of the spacing Δ when it is constant, independent of the sample size n. While the rate of convergence of the MLE of C is $\sqrt{n}$ in this case, irrespective of the value of Δ, the value of the optimal spacing depends on H. It is 1 (unit-spacing) if $H = 1/2$ but is very large if H is close to 1.

1 Introduction

Fractional Brownian motion $B_H(t), t \in \mathbb{R}$, is the only Gaussian self-similar process with stationary increments. It satisfies $B_H(0) = 0$, has mean 0 and covariance

$$EB_H(t)B_H(s) = \frac{C}{2}\Big(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\Big),$$

where $0 < H < 1$ and $C = \mathrm{Var}B_H(1)$. Self-similarity means that for any constant $a > 0$, $\{B_H(at)\}_{t\in\mathbb{R}}$ and $\{a^H B_H(t)\}t \in \mathbb{R}$ have identical finite-dimensional distributions (see Samorodnitsky and Taqqu (1994) or Embrechts and Maejima (2003) for additional information).

Suppose that we want to estimate the unknown parameters H and C on the basis of n equally spaced observations, that is by using the vector

$$\mathbf{Z}_n^{\Delta} = \Big(B_H(\Delta_n) - B_H(0), B_H(2\Delta_n) - B_H(\Delta_n), \ldots, B_H(n\Delta_n) - B_H((n-1)\Delta_n)\Big),$$

*AMS Subject classification. Primary 60G18; secondary 62D05, 62F12.

†Keywords and phrases: fractional Gaussian noise, self-simularity, long-range dependence, maximum likelihood estimation

where Δ_n is the spacing between observations. For example,

$$\Delta_n = 1,\ \Delta_n = a \neq 1,\ \Delta_n = 1/n,\ \Delta_n = n,\ \Delta_n = e^{-n},\ \Delta_n = e^n.$$

Since we are interested in asymptotic results, we also allow Δ_n to depend on n. We want to find out how the choice of Δ_n affects the quality of the maximum likelihhod estimators $\hat{H}_n$ and $\hat{C}_n$ of H and C respectively. For ease of notation we will suppress the index n and write simply $\hat{H}$ and $\hat{C}$ instead of $\hat{H}_n$ and $\hat{C}_n$.

The case $\Delta_n = 1/n$ corresponds to asymptotics based on increasingly dense observations in a fixed and bounded region, which is known in spatial statistics as "infill asymptotics" (Cressie (1993)) or "fixed-domain asymptotics" (Stein (1999)). Zhu and Stein (2002) derived results for a moment estimator of H and C of two-dimensional fractional Brownian surfaces for $\Delta_n = 1/n$, which is consistent with the results we obtained here.

The problem we are interested in is in the same spirit as that of understanding the effect of aggregation on long-range dependent data considered by Hannig, Marron and Riedi (2001) and Smith and Trovero (2002). Hannig, Marron and Riedi (2001) show that at low frequencies, the aggregated spectral density behaves like a scaled version of the original one. Smith and Trovero (2002) focus on the effect of aggregation on the local Whittle estimator and show that for a finite aggregation level, neither the MSE nor the value of the upper frequency that is used for estimation are affected by the fixed level of aggregation. These issues concern the estimation of H. Since we consider here fractional Brownian motion, which is self-similar, the MLE of H is not affected by the sampling rate. Our goal is to understand how the choice of sampling rate affects the estimation of the variance C.

In Section 2, we consider the case $\Delta_n = 1$ and in Section 3 we consider values of Δ_n that either decrease to 0 or increase to ∞ with n . We also consider the case $\Delta_n = a$, that is, a spacing independent of n and determine its optimal value. This optimal value is not $\Delta_n = 1$ if $H \neq 1/2$.

2 Preliminary results

Consider first the case where $\Delta_n = 1$, that is when the n observations are equally spaced on the interval $[0, n]$. The corresponding vector $\mathbf{Z}_n^\Delta$ is denoted in this case $\mathbf{Z}_n^{(1)}$, and is commonly called *fractional Gaussian noise.* It is a stationary Gaussian vector with covariance matrix $\mathbf{R}_n^{(1)} = C\mathbf{R}_n^H$ and

$$(\mathbf{R}_n^H)_{i,j} = \frac{1}{2}\{|i-j+1|^{2H} - 2|i-j|^{2H} + |i-j-1|^{2H}\}. \tag{2.1}$$

$\mathbf{R}_n^H$ is the correlation matrix.

It follows from a result of Dahlhaus (1989) that the MLE $\hat{H}^{(1)}$ and $\hat{C}^{(1)}$ of H and C are asymptotically Gaussian. More precisely,

Lemma 2.1

$$\sqrt{n}\begin{pmatrix} \hat{H}^{(1)} - H \\ \hat{C}^{(1)} - C \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} X_H \\ X_C \end{pmatrix} \sim N(0, \Sigma^{(1)}), \tag{2.2}$$

where

$$\Sigma^{(1)} = \begin{pmatrix} \sigma_{HH} & \sigma_{HC} \\ \sigma_{HC} & \sigma_{CC} \end{pmatrix}, \tag{2.3}$$

and

$$\begin{aligned}
\sigma_{HH} &= 8\pi^2 \left\{ 2\pi \int_{-\pi}^{\pi} F^2(x,H)dx - \left(\int_{-\pi}^{\pi} F(x,H)dx \right)^2 \right\}^{-1}, \\
\sigma_{HC} &= -4\pi C \int_{-\pi}^{\pi} F(x,H)dx \left\{ 2\pi \int_{-\pi}^{\pi} F^2(x,H)dx - \left(\int_{-\pi}^{\pi} F(x,H)dx \right)^2 \right\}^{-1}, \\
\sigma_{CC} &= 4\pi C^2 \int_{-\pi}^{\pi} F^2(x,H)dx \left\{ 2\pi \int_{-\pi}^{\pi} F^2(x,H)dx - \left(\int_{-\pi}^{\pi} F(x,H)dx \right)^2 \right\}^{-1},
\end{aligned}$$

where

$$F(x,H) = \frac{2\int_{\mathbb{R}} (1-\cos s)|s|^{-(1+2H)} \log|s| ds}{\int_{\mathbb{R}} (1-\cos s)|s|^{-(1+2H)} ds} - \frac{2\sum_{k\in\mathbb{Z}} |x+2k\pi|^{-(1+2H)} \log|x+2k\pi|}{\sum_{k\in\mathbb{Z}} |x+2k\pi|^{-(1+2H)}}. \tag{2.4}$$

If H is known, then

$$\sqrt{n}(\hat{C}^{(1)} - C) \xrightarrow{\mathcal{D}} X_C \sim N(0, 2C^2). \tag{2.5}$$

Proof: Let $f(x;\theta)$ be the spectral density of a random process that satisfies conditions (A0)-(A9) in Dahlhaus (1989) and let $\hat{\theta}$ be the exact MLE of θ. Then by Theorem 3.2 of Dahlhaus (1989),

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, \Gamma(\theta_0)^{-1}),$$

where

$$\Gamma(\theta_0) = \frac{1}{4\pi} \int_{-\pi}^{\pi} (\nabla \log f(x;H,C))(\nabla \log f(x;H,C))^T dx.$$

The spectral density of fractional Gaussian noise is given by

$$f(x;H,C) = C \left(\int_{-\infty}^{\infty} (1-\cos s)|s|^{-(1+2H)} ds \right)^{-1} (1-\cos x) \sum_{k=-\infty}^{\infty} |x+2k\pi|^{-(1+2H)}, \tag{2.6}$$

which satisfies the conditions (A0)-(A9) (Fox and Taqqu (1986), Dahlhaus (1989)). By taking the derivative of the log of (2.6), one gets

$$\begin{aligned}
\frac{\partial}{\partial C} \log f(x;H,C) &= \frac{1}{C}, \\
\frac{\partial}{\partial H} \log f(x;H,C) &= \frac{2\int_{\mathbb{R}} (1-\cos s)|s|^{-(1+2H)} \log|s| ds}{\int_{\mathbb{R}} (1-\cos s)|s|^{-(1+2H)} ds} - \frac{2\sum_{k\in\mathbb{Z}} |x+2k\pi|^{-(1+2H)} \log|x+2k\pi|}{\sum_{k\in\mathbb{Z}} |x+2k\pi|^{-(1+2H)}}.
\end{aligned}$$

Let $F(x,H) = \partial \log f(x;H,C)/\partial H$, and $\theta = (H,C)$. Then

$$\begin{aligned}
\Gamma(\theta_0) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(F(x,H),\frac{1}{C}\right)^T\left(F(x,H),\frac{1}{C}\right)dx \\
&= \begin{pmatrix} \frac{1}{4\pi}\int_{-\pi}^{\pi}F^2(x,H)dx & \frac{1}{4\pi C}\int_{-\pi}^{\pi}F(x,H)dx \\ \frac{1}{4\pi C}\int_{-\pi}^{\pi}F(x,H)dx & \frac{1}{2C^2} \end{pmatrix}.
\end{aligned}$$

Relation (2.3) follows by taking inverse of $\Gamma(\theta_0)$. Relation (2.5) can be obtained in the same way by letting $\theta = C$. □

Remark. The function F in (2.4) is well-defined for all $H \in (0,1)$. Indeed, the integrals and the sums defining F converge as their argument tends to infinity because $-(1+2H)+1 = -2H < 0$. Furthermore, the integrals defining F converge as their argument tends to zero because $1-\cos s \sim s^2/2$ and $2-(1+2H)+1 = 2-2H > 0$. Since the integrals diverge if $H = 1$, the presence of the log in the numerator ensures that $F(x,1) = -\infty$. This fact will be used in the sequel.

Lemma 2.1 states what happens when the spacing between observations is $\Delta_n = 1$. We shall now consider what happens for a general Δ_n.

Because of the self-similarity of fractional Brownian motion, the corresponding vector $\mathbf{Z}_n^{\Delta}$ is related to $\mathbf{Z}_n^{(1)}$ by the equality

$$\mathbf{Z}_n^{\Delta} \stackrel{\mathcal{D}}{=} (\Delta_n)^H \mathbf{Z}_n^{(1)}, \tag{2.7}$$

which holds in the sense of finite-dimensional distributions. This relation will be used extensively. If $\mathbf{R}_n^{\Delta}$ denotes the covariance matrix of $\mathbf{Z}_n^{\Delta}$, then

$$\mathbf{R}_n^{\Delta} = (\Delta_n)^{2H}\mathbf{R}_n^{(1)} = (\Delta_n)^{2H} C \mathbf{R}_n^{H}, \tag{2.8}$$

where $\mathbf{R}_n^H$ is given in (2.1).

If H is known and one has only to estimate C, then the MLE $\hat{C}(H,\mathbf{Z}_n^{\Delta})$ of C is

$$\begin{aligned}
\hat{C}(H,\mathbf{Z}_n^{\Delta}) &= \frac{1}{n\Delta_n^{2H}}(\mathbf{Z}_n^{\Delta})^T(\mathbf{R}_n^H)^{-1}(\mathbf{Z}_n^{\Delta}) \\
&\stackrel{\mathcal{D}}{=} \frac{1}{n}(\mathbf{Z}_n^{(1)})^T(\mathbf{R}_n^H)^{-1}(\mathbf{Z}_n^{(1)}) \\
&= \hat{C}^{(1)},
\end{aligned}$$

by using (2.7) and (2.8). Hence when H is known the estimator $\hat{C}(H,\mathbf{Z}_n^{\Delta})$ is identical in distribution to $\hat{C}^{(1)}$, and one gets

$$\sqrt{n}(\hat{C}(H,\mathbf{Z}_n^{\Delta}) - C) \xrightarrow{\mathcal{D}} N(0, 2C^2),$$

by (2.5). The rate $\sqrt{n}$ and the limiting variance do not depend on the spacing Δ_n.

We consider now the situation where both H and C are not known and need to be estimated. The following lemma relates the estimators $\hat{H}$ and $\hat{C}$ based on $\mathbf{Z}_n^{\Delta}$ to the estimators $\hat{H}^{(1)}$ and $\hat{C}^{(1)}$ based on $\mathbf{Z}_n^{(1)}$.

Lemma 2.2

$$\begin{pmatrix} \hat{H} \\ \hat{C} \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} \hat{H}^{(1)} \\ \hat{C}^{(1)} \Delta_n^{-2(\hat{H}^{(1)}-H)} \end{pmatrix}$$

PROOF: Since by (2.8), $\mathbf{Z}_n^{\Delta}$ has the covariance matrix $\mathbf{R}_n^{\Delta} = \Delta_n^{2H} C \mathbf{R}_n^H$, the log-likelihood of $\mathbf{Z}_n^{\Delta}$ is

$$\begin{aligned} l(H, C; \mathbf{Z}_n^{\Delta}) &= -\frac{1}{2} \log |\mathbf{R}_n^{\Delta}| - \frac{1}{2} (\mathbf{Z}_n^{\Delta})^T (\mathbf{R}_n^{\Delta})^{-1} (\mathbf{Z}_n^{\Delta}) - \frac{n}{2} \log 2\pi \\ &= -\frac{1}{2} \log |\mathbf{R}_n^H| - Hn \log \Delta_n - \frac{n}{2} \log C - \frac{1}{2C\Delta_n^{2H}} (\mathbf{Z}_n^{\Delta})^T (\mathbf{R}_n^H)^{-1} (\mathbf{Z}_n^{\Delta}) - \frac{n}{2} \log 2\pi. \end{aligned}$$

For a given H, the corresponding MLE for C is

$$\hat{C}(H, \mathbf{Z}_n^{\Delta}) = \frac{1}{n\Delta_n^{2H}} (\mathbf{Z}_n^{\Delta})^T (\mathbf{R}_n^H)^{-1} (\mathbf{Z}_n^{\Delta}). \tag{2.9}$$

Replacing C by $\hat{C}(H, \mathbf{Z}_n^{\Delta})$ in the log-likelihood of H and C, one gets the profile likelihood of H as

$$\begin{aligned} l^*(H; \mathbf{Z}_n^{\Delta}) &= l(H, \hat{C}(H, \mathbf{Z}_n^{\Delta}); \mathbf{Z}_n^{\Delta}) \\ &= -\frac{1}{2} \log |\mathbf{R}_n^H| - Hn \log \Delta_n - \frac{n}{2} \log (\mathbf{Z}_n^{\Delta})^T (\mathbf{R}_n^H)^{-1} (\mathbf{Z}_n^{\Delta}) + \frac{n}{2} \log n\Delta_n^{2H} - \frac{n}{2} - \frac{n}{2} \log 2\pi \\ &= -\frac{1}{2} \log |\mathbf{R}_n^H| - \frac{n}{2} \log (\mathbf{Z}_n^{\Delta})^T (\mathbf{R}_n^H)^{-1} (\mathbf{Z}_n^{\Delta}) + \frac{n}{2} (\log(n/2\pi) - 1). \end{aligned}$$

The estimator $\hat{H}$ is defined as the root of

$$(l^*)'(H, \mathbf{Z}_n^{\Delta}) = 0.$$

Denote it as $\hat{H} = f(\mathbf{Z}_n^{\Delta})$. For any constant $k \in \mathbb{R}$, the root of $(l^*)'(H, k\mathbf{Z}_n^{\Delta}) = 0$ is $f(k\mathbf{Z}_n^{\Delta})$. Since

$$(l^*)'(H, \mathbf{Z}_n^{\Delta}) = (l^*)'(H, k\mathbf{Z}_n^{\Delta}),$$

one gets

$$f(\mathbf{Z}_n^{\Delta}) = f(k\mathbf{Z}_n^{\Delta}), \text{ for any } k \in \mathbb{R},$$

and thus, setting $k = \Delta_n^{-H}$,

$$\hat{H} = f(\mathbf{Z}_n^{\Delta}) = f(\Delta_n^{-H} \mathbf{Z}_n^{\Delta}) \stackrel{\mathcal{D}}{=} f(\mathbf{Z}_n^{(1)}) = \hat{H}^{(1)}, \tag{2.10}$$

which shows that the distribution of estimator of H does not depend on the choice of Δ_n.

By using $\hat{H} = f(\mathbf{Z}_n^{\Delta})$, the MLE $\hat{C}$ of C in (2.9) is

$$\hat{C} = C(\hat{H}, \mathbf{Z}_n^{\Delta}) = \frac{1}{n\Delta_n^{2\hat{H}}} (\mathbf{Z}_n^{\Delta})^T (\mathbf{R}_n^{\hat{H}})^{-1} (\mathbf{Z}_n^{\Delta}).$$

By replacing each $\hat{H}$ by $f(\Delta_n^{-H}\mathbf{Z}_n^\Delta)$ one gets

$$\begin{aligned}\hat{C} &= \frac{1}{n\Delta_n^{2f(\Delta_n^{-H}\mathbf{Z}_n^\Delta)}}(\Delta_n^{-H}\mathbf{Z}_n^\Delta)^T(\mathbf{R}_n^{f(\Delta_n^{-H}\mathbf{Z}_n^\Delta)})^{-1}(\Delta_n^{-H}\mathbf{Z}_n^\Delta)\\ &\stackrel{\mathcal{D}}{=} \frac{1}{n}(\mathbf{Z}_n^{(1)})^T(\mathbf{R}_n^{\hat{H}^{(1)}})^{-1}(\mathbf{Z}_n^{(1)})\Delta_n^{-2(\hat{H}^{(1)}-H)}\\ &= \hat{C}^{(1)}\Delta_n^{-2(\hat{H}^{(1)}-H)}\end{aligned}$$

by using (2.7) and (2.10). □

Remark. When H is known, one takes $\hat{H}^{(1)} = H$, and Lemma 2.2 implies in this case that $\hat{C} \stackrel{\mathcal{D}}{=} \hat{C}^{(1)}$, that is, $\hat{C}$ does not depend on the sampling rate Δ_n. It then follows from (2.5) that

$$\sqrt{n}(\hat{C} - C) \xrightarrow{\mathcal{D}} X_C \sim N(0, 2C^2) \qquad (2.11)$$

when H is known. The next section describes what happens when H is not known.

3 Impact of the sampling rate

The first theorem deals with situations where Δ_n is either too large or too small. We write $a_n \ll b_n$ or $b_n \gg a_n$ if $a_n = o(b_n)$, that is if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

Theorem 3.1
(a) If $|\log \Delta_n| \gg \sqrt{n}$, *then the estimator* $\hat{C}$ *is not consistent.*
(b) If $\log \Delta_n/\sqrt{n} \to A$, *then*

$$\hat{C} \xrightarrow{\mathcal{D}} Ce^{AN(0,4\sigma_{HH})},$$

that is, $\hat{C}$ *converges to a random variable with a lognormal distribution.* $\hat{C}$ *is also not consistent in this case.*

Proof: One has $\hat{C}^{(1)} \xrightarrow{\text{a.s.}} C$ by the law of large numbers, and

$$\Delta_n^{-2(\hat{H}^{(1)}-H)} = \exp\left\{\frac{\log \Delta_n}{\sqrt{n}} 2\sqrt{n}(\hat{H}^{(1)} - H)\right\},$$

where $\sqrt{n}(\hat{H}^{(1)} - H) \xrightarrow{\mathcal{D}} N(0, \sigma_{HH})$, and

$$\frac{\log \Delta_n}{\sqrt{n}} \to \pm\infty,$$

according to whether $\Delta_n \ll e^{-\sqrt{n}}$ or $\Delta_n \gg e^{-\sqrt{n}}$. Therefore,

$$\hat{C} = \hat{C}^{(1)}\Delta_n^{-2(\hat{H}^{(1)}-H)} \to \infty \text{ or } 0,$$

showing that $\hat{C}$ is not a consistent estimator of C.

If $\log \Delta_n/\sqrt{n} \to A$, then

$$\hat{C} = \hat{C}^{(1)} \Delta_n^{-2(\hat{H}^{(1)}-H)} \to Ce^{AN(0,4\sigma_{HH})} \qquad \square$$

Observe that $|\log \Delta_n| \gg \sqrt{n}$ implies that either

$$\Delta_n \ll e^{-\sqrt{n}} \quad \text{or} \quad \Delta_n \gg e^{\sqrt{n}}.$$

Theorem 3.1 shows that C cannot be consistently estimated if Δ_n either converges to 0 too fast or diverges to ∞ too fast. The next theorem considers the case where Δ_n tends either to 0 or ∞ but does so at a slower rate, namely

$$e^{-\sqrt{n}} \ll \Delta_n \ll e^{\sqrt{n}} \quad \text{with} \quad \Delta_n \to 0 \text{ or } \infty.$$

Theorem 3.2 *If $|\log \Delta_n| \ll \sqrt{n}$, and $\Delta_n \ll 1$ or $\Delta_n \gg 1$, then*

$$\begin{pmatrix} \sqrt{n}(\hat{H} - H) \\ \frac{\sqrt{n}}{|\log \Delta_n|}(\hat{C} - C) \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, \Sigma), \tag{3.1}$$

with

$$\Sigma = \begin{pmatrix} \sigma_{HH} & \pm 2C\sigma_{HH} \\ \pm 2C\sigma_{HH} & 4C^2\sigma_{HH} \end{pmatrix}, \tag{3.2}$$

where the sign of the covariance is positive when $\log \Delta_n < 0$, and negative when $\log \Delta_n > 0$.

PROOF: In view of Lemma 2.2,

$$\begin{aligned} \hat{C} - C &= \hat{C}^{(1)} \Delta_n^{-2(\hat{H}^{(1)}-H)} - C \\ &= \hat{C}^{(1)}(\Delta_n^{-2(\hat{H}^{(1)}-H)} - 1) + \hat{C}^{(1)} - C. \end{aligned}$$

Since $\frac{\log \Delta_n}{\sqrt{n}} \to 0$,

$$\begin{aligned} \Delta_n^{-2(\hat{H}^{(1)}-H)} - 1 &= \exp\{-2(\log \Delta_n)(\hat{H}^{(1)} - H)\} - 1 \\ &= -2\frac{\log \Delta_n}{\sqrt{n}}\sqrt{n}(\hat{H}^{(1)} - H) + o(\frac{\log \Delta_n}{\sqrt{n}}), \end{aligned}$$

so that

$$\hat{C} - C = -2\hat{C}^{(1)}\frac{\log \Delta_n}{\sqrt{n}}\sqrt{n}(\hat{H}^{(1)} - H) + \frac{1}{\sqrt{n}}\sqrt{n}(\hat{C}^{(1)} - C) + o(\frac{\log \Delta_n}{\sqrt{n}}). \tag{3.3}$$

The assumptions $\Delta_n \ll 1$ or $\Delta_n \gg 1$ imply

$$|\log \Delta_n| \to \infty.$$

Therefore

$$\begin{aligned}\frac{\sqrt{n}}{|\log\Delta_n|}(\hat{C}-C) &= -2\frac{\log\Delta_n}{|\log\Delta_n|}\hat{C}^{(1)}\sqrt{n}(\hat{H}^{(1)}-H)+\frac{1}{|\log\Delta_n|}\sqrt{n}(\hat{C}^{(1)}-C)+o(1)\\ &\xrightarrow{\mathcal{D}} -2\frac{\log\Delta_n}{|\log\Delta_n|}CX_H \sim \pm 2CN(0,\sigma_{HH}),\end{aligned}$$

where we used (2.2) and the fact that $\hat{C}^{(1)} \xrightarrow{\text{a.s.}} C$ by the law of large numbers. The result follows. □

Remarks:

- Σ is a singular matrix. The correlation function corresponding to Σ is
$$\begin{pmatrix} 1 & \pm 1 \\ \pm 1 & 1 \end{pmatrix},$$
i.e., $\hat{H}$ and $\hat{C}$ under above normalization are asymptotically perfectly correlated. The correlation is $+1$ if $\log\Delta_n < 0$ (example: $\Delta_n = 1/n$) and it is -1 if $\log\Delta_n > 0$ (example: $\Delta_n = n$).
- The normalization factors for $\hat{H}$ and $\hat{C}$ in (3.1) are not the same. If one chooses a common normalization for $\hat{H}$ and $\hat{C}$, this normalization has to be $\sqrt{n}/|\log n|$, and Σ in this case is the matrix
$$\begin{pmatrix} 0 & 0 \\ 0 & 4C^2\sigma_{HH} \end{pmatrix}.$$

Example. If the observations are equally spaced on $[0,n]$, that is, $\Delta_n = 1$, then the rate of convergence of $\hat{C}-C$ is $\sqrt{n}$. However, if the observations are equally spaced on $[0,1]$, that is $\Delta_n = 1/n$, or if the observations are equally spaced on $[0,n^2]$, that is, $\Delta_n = n$, then the rate of convergence of $\hat{C}-C$ is slower, namely $\sqrt{n}/\log n$.

The next result deals with the situation where the spacing Δ_n does not depend on n.

Theorem 3.3 *If $\Delta_n = a$, a constant independent of n, then*

$$\sqrt{n}\begin{pmatrix} \hat{H}-H \\ \hat{C}-C \end{pmatrix} \xrightarrow{\mathcal{D}} N(0,\Sigma), \tag{3.4}$$

with

$$\Sigma = \begin{pmatrix} \sigma_{HH} & -2(\log a)C\sigma_{HH}+\sigma_{CC} \\ -2(\log a)C\sigma_{HH}+\sigma_{CC} & 4(\log a)^2C^2\sigma_{HH}+\sigma_{CC}-4(\log a)C\sigma_{HC} \end{pmatrix}. \tag{3.5}$$

The value a^* of a which minimizes the asymptotic variance of $\hat{C}$ is given by

$$a^* = \exp\{\sigma_{HC}/2C\sigma_{HH}\} = \exp\left\{-\frac{1}{4\pi}\int_{-\pi}^{\pi} F(x,H)dx\right\}, \tag{3.6}$$

and the corresponding asymptotic variance is in this case,

$$(\sigma_C^*)^2 = \sigma_{CC} - \frac{\sigma_{HC}^2}{\sigma_{HH}} = 2C^2. \tag{3.7}$$

	H	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
a^*	0.1	0.68	0.59	0.54	0.51	0.50	0.50	0.50	0.50	0.51	0.52
RF		1.04	1.08	1.11	1.14	1.17	1.18	1.19	1.19	1.19	1.19
a^*	0.2	0.53	0.54	0.55	0.56	0.57	0.58	0.60	0.61	0.62	0.63
RF		1.19	1.18	1.18	1.17	1.16	1.16	1.15	1.14	1.13	1.12
a^*	0.3	0.65	0.66	0.67	0.69	0.70	0.72	0.73	0.74	0.76	0.78
RF		1.11	1.11	1.10	1.09	1.08	1.07	1.07	1.06	1.05	1.04
a^*	0.4	0.79	0.81	0.83	0.85	0.87	0.89	0.91	0.93	0.95	0.97
RF		1.04	1.03	1.03	1.02	1.02	1.01	1.01	1.00	1.00	1.00
a^*	0.5	1.00	1.03	1.05	1.08	1.12	1.15	1.19	1.22	1.27	1.31
RF		1.00	1.00	1.00	1.01	1.01	1.02	1.02	1.03	1.05	1.06
a^*	0.6	1.36	1.41	1.47	1.53	1.60	1.67	1.75	1.84	1.94	2.04
RF		1.08	1.10	1.12	1.15	1.18	1.22	1.26	1.31	1.37	1.43
a^*	0.7	2.17	2.32	2.48	2.66	2.87	3.12	3.40	3.74	4.15	4.64
RF		1.52	1.61	1.71	1.83	1.96	2.12	2.31	2.53	2.78	3.08
a^*	0.8	5.24	5.99	6.94	8.15	9.73	11.83	14.70	18.69	24.41	32.81
RF		3.43	3.84	4.33	4.91	5.61	6.46	7.47	8.70	10.18	11.98
a^*	0.9	45.55	65.46	97.57	151.06	243.10	406.31	703.21	1255.60	2297.60	4279.79
RF		14.17	16.82	20.02	23.87	28.46	33.89	40.24	47.56	55.87	65.13

Table 1: The optimal a^* and the relative efficiency RF$=\sigma_{CC}/(\sigma_C^*)^2$ for different values of H

PROOF: By (3.3) and (2.2),

$$\sqrt{n}(\hat{C}-C) \xrightarrow{\mathcal{D}} -2(\log a)CX_H + X_C.$$

In view of Lemmas 2.1 and 2.2, Relation (3.4) holds with asymptotic covariance

$$\mathrm{Cov}(X_H, -2(\log a)CX_H + X_C) = -2(\log a)C\sigma_{HH} + \sigma_{CC},$$

and the asymptotic variance of $\hat{C}$ is

$$\begin{aligned}
\tilde{\sigma}_C^2 &= \mathrm{Var}\left(-2(\log a)CX_H + X_C\right) \\
&= 4(\log a)^2C^2\sigma_{HH} + \sigma_{CC} - 4(\log a)C\sigma_{HC} \\
&= 4C^2\sigma_{HH}\left(\log a - \frac{\sigma_{HC}}{2C\sigma_{HH}}\right)^2 + \sigma_{CC} - \frac{\sigma_{HC}^2}{\sigma_{HH}}.
\end{aligned}$$

So the optimal choice of a is (3.6) and the corresponding asymptotic variance is given by (3.7). The second equalities in (3.6) and in (3.7) follow from Lemma 2.1. □

Remarks:

- The optimal choice $\Delta_n = a^*$ depends on H but not on C.

- The choice $\Delta_n = a = 1$ yields

$$\sqrt{n}(\hat{C} - C) \xrightarrow{\mathcal{D}} X_C \sim N(0, \sigma_{CC}),$$

which is not optimal when $H \neq 1/2$.

- Table 1 gives the a^* and the relative efficiency $\sigma_{CC}/(\sigma_C^*)^2 = \sigma_{CC}/2C^2 \geq 1$ for different values of H. The relative efficiency compares the asymptotic variance of $\hat{C}$ at $a = 1$ with that at $a = a^*$. Recall that $F(x, 1) = -\infty$. As H tends toward 1, the optimal spacing a^*, which is given by (3.6), becomes much larger than 1 and, if one uses instead of a^* the value $a = 1$, the asymptotic variance of $\hat{C}$ becomes very large.

- The optimal choice $\Delta_n = a^*$ has the same asymptotic variance $2C^2$ as in the case (2.5), where H was assumed known.

- Since a^* depends on H, one needs to estimate a^* by using an estimator $\hat{H}$ of H. One can use for example the MLE $\hat{H}$ of Theorem 3.3 or, more generally, any estimator $\hat{H}$ satisfying $\hat{H} \xrightarrow{P} H$. Let $\hat{\sigma}_C^2$ be the asymptotic variance of $\hat{C}$ for $a = a^*(\hat{H})$. As $a^*(H)$ is a continuous function of H, one has $\hat{\sigma}_C^2 \xrightarrow{P} (\sigma_C^*)^2$, i.e., the estimator of C based on $a^*(\hat{H})$is asymptotically efficient.

- The function $F(x, H)$ given in (2.4) enters in the expression (3.6) and (3.7) of a^* and $\sigma_{CC}/(\sigma_C^*)^2$. It involves integrals that need to be evaluated numerically. As $s \downarrow 0$, the integrands $(1 - \cos s)|s|^{-(1+2H)} \log |s|$ and $(1 - \cos s)|s|^{-(1+2H)}$ in (2.4) tend to $-\infty$ and ∞ respectively, if $H > 1/2$. This present difficulties for H close to 1. For H close to 0, the difficulties stem from the slow decrease of these integrands as $s \to \infty$. The numerical routines found in MATLAB and SPLUS are not adequate and give erroneous results. We wrote our own routines to generate the entries of Table 1.

Acknowledgements

Zhengyuan Zhu would like to thank Professor Michael Stein for valuable conversations on this subject. Murad S. Taqqu thanks the University of North Carolina at Chapel Hill and the Statistical and Applied Mathematical Sciences Institute (SAMSI) for providing a great working environment. This research was partially supported by the NSF Grant DMS-0102410 at Boston University.

References

Cressie, N. (1993), *Statistics for Spatial Data: revised ed*, John Wiley, New York.

Dahlhaus, R. (1989), 'Efficient parameter estimation for self similar processes', *The Annals of Statistics* **17**(4), 1749–1766.

Embrechts, P. & Maejima, M. (2003), *Selfsimilar Processes*, Academic Press.

Fox, R. & Taqqu, M. S. (1986), 'Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series', *The Annals of Statistics* **14**, 517–532.

Hannig, J., Marron, J. S. & Riedi, R. H. (2001), 'Zooming statistics: inference across scales', *Journal of the Korean Statistical Society* **30**, 327–345.

Samorodnitsky, G. & Taqqu, M. S. (1994), *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York, London.

Smith, R. L. & Trovero, M. (2002), The effect of aggregation on estimates of long-range dependence, Preprint.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer-Verlag, New York.

Zhu, Z. & Stein, M. L. (2002), 'Parameter estimation for fractional Brownian surfaces', *Statistica Sinica* **12**(3), 863–883.

Zhengyuan Zhu
Department of Statistics and Operations Research
CB#3260, New West
The University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3260
zhuz@email.unc.edu

Murad S. Taqqu
Department of Mathematics and Statistics
111 Cummington Street
Boston University
Boston, MA 02215, USA
murad@math.bu.edu