

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

July 13, 2005

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, $\sqrt{n}/\ln n$. 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

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

where $0 < H < 1$ and $C = \text{Var}B_H(1)$. 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

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

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

[†]Keywords and phrases: fractional Gaussian noise, self-similarity, 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 likelihood 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 mean squared error (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^Δ 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

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

\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

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

where

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

and

$$\sigma_{HH} = 8\pi^2 \left\{ 2\pi \int_{-\pi}^{\pi} F^2(x, H) dx - \left(\int_{-\pi}^{\pi} F(x, H) dx \right)^2 \right\}^{-1}, \quad (2.4)$$

$$\sigma_{HC} = -4\pi C \int_{-\pi}^{\pi} F(x, H) dx \left\{ 2\pi \int_{-\pi}^{\pi} F^2(x, H) dx - \left(\int_{-\pi}^{\pi} F(x, H) dx \right)^2 \right\}^{-1}, \quad (2.5)$$

$$\sigma_{CC} = 4\pi C^2 \int_{-\pi}^{\pi} F^2(x, H) dx \left\{ 2\pi \int_{-\pi}^{\pi} F^2(x, H) dx - \left(\int_{-\pi}^{\pi} F(x, H) dx \right)^2 \right\}^{-1}, \quad (2.6)$$

where

$$F(x, H) = \frac{2 \int_{\mathbb{R}} (1 - \cos s) |s|^{-(1+2H)} \ln |s| ds}{\int_{\mathbb{R}} (1 - \cos s) |s|^{-(1+2H)} ds} - \frac{2 \sum_{k \in \mathbb{Z}} |x + 2k\pi|^{-(1+2H)} \ln |x + 2k\pi|}{\sum_{k \in \mathbb{Z}} |x + 2k\pi|^{-(1+2H)}}. \quad (2.7)$$

If H is known, then

$$\sqrt{n}(\hat{C}^{(1)} - C) \xrightarrow{\mathcal{D}} X_C \sim N(0, 2C^2). \quad (2.8)$$

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

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

The spectral density of fractional Gaussian noise is given by

$$f(x; H, C) = C \left(\int_{-\infty}^{\infty} (1 - \cos s) |s|^{-(1+2H)} ds \right)^{-1} (1 - \cos x) \sum_{k=-\infty}^{\infty} |x + 2k\pi|^{-(1+2H)}, \quad (2.9)$$

which satisfies the conditions (A0)-(A9) (Fox and Taqqu (1986), Dahlhaus (1989)). By taking the derivative of the log of (2.9), one gets

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

Let

$$F(x, H) = \partial \ln f(x; H, C) / \partial H,$$

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 (2.8) can be obtained in the same way by letting $\theta = C$. \square

Remarks:

- The function F in (2.7) 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 \ln in the numerator ensures that

$$F(x, 1) = -\infty. \quad (2.10)$$

This fact will be used in the sequel.

- The two ratios in Relation (2.7) defining $F(x, H)$ are monotone decreasing w.r.t H . Indeed, the first ratio can be expressed as

$$\frac{2 \int_{\mathbb{R}} (1 - \cos s) |s|^{-(1+2H)} \ln |s| ds}{\int_{\mathbb{R}} (1 - \cos s) |s|^{-(1+2H)} ds} = -\frac{\partial}{\partial H} \ln \left(\int_{-\infty}^{\infty} (1 - \cos s) |s|^{-(1+2H)} ds \right).$$

Set $f_1(H) = \ln \left(\int_{-\infty}^{\infty} (1 - \cos s) |s|^{-(1+2H)} ds \right)$. Then by the Cauchy-Schwartz inequality,

$$f_1(H_1) + f_1(H_2) \geq 2f_1((H_1 + H_2)/2),$$

for all $H_1, H_2 \in (0, 1)$, i.e, $f_1(H)$ is convex w.r.t. H . Thus the first ratio is monotone decreasing. Similar arguments can be used to show the monotonicity of the second ratio.

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^Δ is related to $\mathbf{Z}_n^{(1)}$ by the equality

$$\mathbf{Z}_n^\Delta \stackrel{\mathcal{D}}{=} (\Delta_n)^H \mathbf{Z}_n^{(1)}, \quad (2.11)$$

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

$$\mathbf{R}_n^\Delta = (\Delta_n)^{2H} \mathbf{R}_n^{(1)} = (\Delta_n)^{2H} C \mathbf{R}_n^H, \quad (2.12)$$

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 (2.11) and (2.12). 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 (2.8). 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^Δ 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 (2.12), \mathbf{Z}_n^Δ has the covariance matrix $\mathbf{R}_n^\Delta = \Delta_n^{2H} C \mathbf{R}_n^H$, the log-likelihood of \mathbf{Z}_n^Δ is

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

For a given H , the corresponding MLE for C is

$$\hat{C}(H, \mathbf{Z}_n^\Delta) = \frac{1}{n\Delta_n^{2H}} (\mathbf{Z}_n^\Delta)^T (\mathbf{R}_n^H)^{-1} (\mathbf{Z}_n^\Delta). \quad (2.13)$$

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

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

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

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

where the prime denotes a derivative with respect to H . Denote the root 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 $k = \Delta_n^{-H}$,

$$\hat{H} = f(\mathbf{Z}_n^\Delta) = f(\Delta_n^{-H}\mathbf{Z}_n^\Delta) \stackrel{\mathcal{D}}{=} f(\mathbf{Z}_n^{(1)}) = \hat{H}^{(1)}, \quad (2.14)$$

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

By replacing H by \hat{H} , the MLE \hat{C} of C in (2.13) 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 (2.11) and (2.14). \square

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 (2.8) that

$$\sqrt{n}(\hat{C} - C) \xrightarrow{\mathcal{D}} X_C \sim N(0, 2C^2) \quad (2.15)$$

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

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

Theorem 3.1

- (a) If $|\ln \Delta_n| \gg \sqrt{n}$, then the estimator \hat{C} is not consistent.
- (b) If $\ln \Delta_n / \sqrt{n} \rightarrow A$, then

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

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

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

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

$$\frac{\ln \Delta_n}{\sqrt{n}} \rightarrow \pm\infty,$$

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

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

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

If $\ln \Delta_n / \sqrt{n} \rightarrow A$, then

$$\hat{C} = \hat{C}^{(1)} \Delta_n^{-2(\hat{H}^{(1)}-H)} \rightarrow C e^{AN(0, 4\sigma_{HH})} \quad \square$$

Observe that $|\ln \Delta_n| \gg \sqrt{n}$ 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

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

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

$$\left(\begin{array}{c} \sqrt{n}(\hat{H} - H) \\ \frac{\sqrt{n}}{|\ln \Delta_n|}(\hat{C} - C) \end{array} \right) \xrightarrow{\mathcal{D}} N(0, \Sigma), \quad (3.1)$$

with

$$\Sigma = \left(\begin{array}{cc} \sigma_{HH} & \pm 2C\sigma_{HH} \\ \pm 2C\sigma_{HH} & 4C^2\sigma_{HH} \end{array} \right), \quad (3.2)$$

where the sign of the covariance in (3.2) is positive when $\ln \Delta_n < 0$, and negative when $\ln \Delta_n > 0$.

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 $\frac{\ln \Delta_n}{\sqrt{n}} \rightarrow 0$,

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

so that

$$\hat{C} - C = -2\hat{C}^{(1)}\frac{\ln \Delta_n}{\sqrt{n}}\sqrt{n}(\hat{H}^{(1)} - H) + \frac{1}{\sqrt{n}}\sqrt{n}(\hat{C}^{(1)} - C) + o\left(\frac{\ln \Delta_n}{\sqrt{n}}\right). \quad (3.3)$$

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

$$|\ln \Delta_n| \rightarrow \infty.$$

Therefore

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

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 from (2.2). \square

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 the above normalization are asymptotically perfectly correlated. The correlation is +1 if $\ln \Delta_n < 0$ (example: $\Delta_n = 1/n$) and it is -1 if $\ln \Delta_n > 0$ (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 $\sqrt{n}/|\ln n|$, 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 $\sqrt{n}/\ln n$.

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*

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

with

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

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

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

and the corresponding asymptotic variance is in this case,

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

PROOF: By (3.3) and (2.2),

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

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

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

and the asymptotic variance of \hat{C} is

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

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. \square

Remarks:

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

	H	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
a^*	0.1	0.36	0.38	0.40	0.42	0.43	0.45	0.46	0.48	0.49	0.50
RE		1.26	1.26	1.26	1.25	1.24	1.24	1.23	1.22	1.22	1.21
a^*	0.2	0.52	0.53	0.54	0.56	0.57	0.58	0.59	0.61	0.62	0.63
RE		1.20	1.19	1.18	1.18	1.17	1.16	1.15	1.14	1.13	1.12
a^*	0.3	0.65	0.66	0.67	0.69	0.70	0.71	0.73	0.74	0.76	0.78
RE		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
RE		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
RE		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	2.05
RE		1.08	1.10	1.12	1.15	1.18	1.22	1.26	1.31	1.37	1.44
a^*	0.7	2.17	2.32	2.48	2.66	2.87	3.12	3.40	3.74	4.15	4.64
RE		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	32.82
RE		3.42	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	151.10	243.20	406.45	703.53	1255.97	2298.64	4280.19
RE		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 $RE = \sigma_{CC}/(\sigma_C^*)^2$ for different values of H

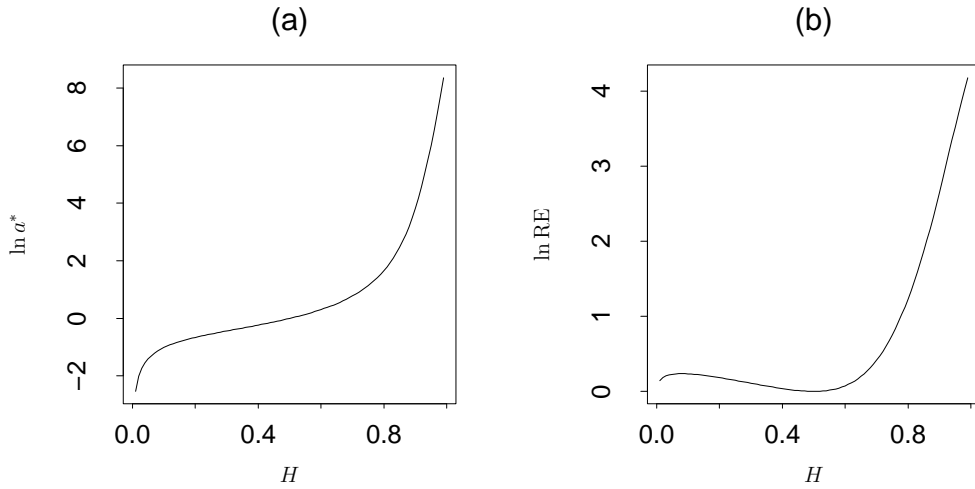


Figure 1: Plots of $\ln a^*$ and $\ln RE$ against H .

2. 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 $H \neq 1/2$.

3. When $H = 1/2$, $a^* = 1$ as

$$\int_{-\pi}^{\pi} F(x, 1/2) dx = 2 \int_{\mathbb{R}} (1 - \cos s) \frac{\ln |s|}{s^2} ds - 2 \int_{-\pi}^{\pi} (1 - \cos x) \sum_{k=-\infty}^{\infty} \frac{\ln |x + 2k\pi|}{|x + 2k\pi|^{1+2H}} dx = 0.$$

The first equation holds because $\int_{\mathbb{R}} (1 - \cos s) |s|^{-2} ds = 1/2\pi$ and $\sum_{k=-\infty}^{\infty} |x + 2k\pi|^{-2} = (1 - \cos x)^{-1}$, which follows from $f(x, 1/2, C) = C/2\pi$ and (2.9).

4. Table 1 displays the optimal value a^* and the relative efficiency

$$RE = \sigma_{CC}/(\sigma_C^*)^2 = \sigma_{CC}/2C^2 \geq 1$$

for different values of H . The relative efficiency compares the asymptotic variance σ_{CC} of \hat{C} at $a = 1$ with that at $a = a^*$ (see (2.6) and (3.7) respectively). As noted in (2.10), $F(x, 1) = -\infty$. Consequently, 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.

5. Figure 1.a displays the plot of $\ln a^*$ versus H , which suggests that a^* is monotone increasing with H , and $\ln a^* \rightarrow -\infty$ as $H \rightarrow 0$. One can prove analytically that $\ln a^* = -\int_{-\pi}^{\pi} F(x, H) dx / 4\pi \sim \ln H$ as $H \rightarrow 0$, while the problem of whether a^* is a monotone function of H is open.
6. Figure 1.b displays the plot of $\ln RE$ versus H . It suggests that the relationship between RE and H is not monotone: RE equals 1 (its minimum value) when $H = 1/2$ (this is so, because $a^* = 1$ when $H = 1/2$), and RE tends to 1 when H tends to zero. This second fact can be proved analytically by showing that $\int_{-\pi}^{\pi} F(x, H) dx \sim -4\pi \ln H$ and $\int_{-\pi}^{\pi} F^2(x, H) dx \sim 4\pi/H$ as $H \rightarrow 0$, which together with (2.6) implies that σ_{CC} tends to $2C^2$ as H tends to 0.
7. The optimal choice $\Delta_n = a^*$ has the same asymptotic variance $2C^2$ as in the case (2.8), where H was assumed known.
8. 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.
9. The function $F(x, H)$ given in (2.7) enters in the expression (3.6) and (3.7) of a^* and $\sigma_{CC}/(\sigma_C^*)^2$. It involves integrals of

$$f_1(s; H) = (1 - \cos s) |s|^{-(1+2H)}$$

and

$$f_2(s; H) = (1 - \cos s)|s|^{-(1+2H)} \ln |s|$$

over \mathbb{R} , which need to be evaluated numerically. For H close to 0, both f_1 and f_2 decrease slowly as $s \rightarrow \infty$, and the numerical routines found in MATLAB and SPLUS are not adequate and give erroneous results. For example, when the SPLUS numerical integration routine `integrate()` is used for $H = 0.1$, the highest subdivision limit allowed by the machine is reached before a desirable precision can be obtained, and the absolute error bound is as high as 0.72. The error bound for a^* is even higher, as further operations on the integral, including division, integration, and exponentiation, are necessary to compute a^* .

To evaluate $\int_0^\infty f_1(s; H)ds$, we first transform it into

$$\int_0^{2\pi} (1 - \cos s) \sum_0^\infty |s + 2k\pi|^{-(1+2H)} ds$$

by change of variable. The infinite sum in the integrand is then approximated by the sum of the first $k + 1$ terms and $(s + (2k + 1)\pi)^{-2H}/(4\pi H)$, which is an approximation of the remaining terms using integration. This approximation is highly accurate for moderately large k even when H is small. For example, with $H = 0.1$, the maximum absolute difference between the approximation of $k = 100$ and $k = 1000$ is less than 4.3×10^{-7} for $s \in (0, 2\pi)$. Because the integration now is over a finite interval $[0, 2\pi]$ with no singular point, it is much easier to do the numerical integration, and the absolute error bound of the integration for $H = 0.1$ is 1.8×10^{-8} using SPLUS routine. For $f_2(s; H)$, a similar transformation is done, and the infinite sum in the integrand is approximated by the sum of the first $k + 1$ terms and $(\ln(s + (2k + 1)\pi) + 1/(2H))(s + (2k + 1)\pi)^{-2H}/(4\pi H)$. All the entries in Table 1 are produced using the aforementioned transformation method to evaluate the integrals, with $k = 1000$ for approximating the infinite sums.

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