Maximum Likelihood Drift Estimation
for Gaussian Process with Stationary Increments

Yuliya Mishura
Taras Shevchenko
National University of Kyiv

Kostiantyn Ralchenko
Taras Shevchenko
National University of Kyiv

Sergiy Shklyar
Taras Shevchenko
National University of Kyiv

Abstract

The paper deals with the regression model $X_t = \theta t + B_t$, $t \in [0, T]$, where $B = \{B_t, t \geq 0\}$ is a centered Gaussian process with stationary increments. We study the estimation of the unknown parameter θ and establish the formula for the likelihood function in terms of a solution to an integral equation. Then we find the maximum likelihood estimator and prove its strong consistency. The results obtained generalize the known results for fractional and mixed fractional Brownian motion.

Keywords: Gaussian process, stationary increments, discrete observations, continuous observations; maximum likelihood estimator, strong consistency, fractional Brownian motion, mixed fractional Brownian motion.

1. Introduction

We study the problem of the drift parameter estimation for the stochastic process

$$X_t = \theta t + B_t, \tag{1}$$

where $\theta \in \mathbb{R}$ is an unknown parameter, and $B = \{B_t, t \geq 0\}$ is a centered Gaussian process with stationary increments, $B_0 = 0$. In the particular case when $B = B^H$ is a fractional Brownian motion, this model has been studied by many authors. Mention the paper by Norros, Valkeila, and Virtamo (1999) that treats the maximum likelihood estimation by continuous observations of the trajectory of X on the interval $[0, T]$ (see also Le Breton (1998)). Further, Hu, Nualart, Xiao, and Zhang (2011) investigate the exact maximum likelihood estimator by discrete observations at the points $tk = kh$, $k = 1, 2, \ldots, N$; Bertin, Torres, and Tudor (2011) consider the maximum likelihood estimation in the discrete scheme of observations, where the trajectory of X is observed at the points $t_k = \frac{k}{N}$, $k = 1, 2, \ldots, N^\alpha$, $\alpha > 1$. For hypothesis testing of the drift parameter sign in the model (1) driven by a fractional Brownian motion, see Stiburek (2017). The paper Cai, Chigansky, and Kleptsyna (2016) treats the likelihood function for Gaussian processes not necessarily having stationary increments. However, on the one hand, our approach is different from their one, it cannot be deduced from their general formulas and on the other hand, gives rather elegant representations. The construction of the maximum likelihood estimator in the case where B is the sum of two fractional Brow-

nian motions was studied in Mishura (2016) and Mishura and Voronov (2015). A similar non-Gaussian model driven by the Rosenblatt process was considered in Bertin *et al.* (2011).

As already mentioned, we consider the case when B is a centered Gaussian processes with stationary increments. We construct the maximum likelihood estimators for both discrete and continuous schemes of observations. The assumptions on the process in the continuous case are formulated in terms of the second derivative of its covariance function, see Assumptions 1 and 2. The exact formula for the maximum likelihood estimator contains a solution of an integral equation with the kernel obtained after the differentiation. We give the sufficient conditions for the strong consistency of the estimators. Several examples of the process B are considered.

The paper is organized as follows. Section 2 is devoted to the case of the discrete observations. The maximum likelihood estimation for continuous time is studied in Section 3.

2. Maximum likelihood estimation by discrete observations

We start with the construction of the likelihood function and the maximum likelihood estimator in the case of discrete observations. In the next section these results will be used for the derivation of the likelihood function in the continuous-time case, see the proof of Theorem 3.3.

Let the process X defined by formula (1) be observed at the points $t_k, k = 0, 1, \ldots, N$,

$$0 = t_0 < t_1 < \ldots < t_N \leq T. \tag{2}$$

The problem is to estimate the parameter θ by the observations X_{t_k}, $k = 0, 1, \ldots, N$ of the process X_t.

2.1. Likelihood function and construction of the estimator

Denote
$$\Delta X^{(N)} = \left(X_{t_k} - X_{t_{k-1}} \right)_{k=1}^N, \quad \Delta B^{(N)} = \left(B_{t_k} - B_{t_{k-1}} \right)_{k=1}^N.$$

Note that in our model $X_{t_0} = X_0 = 0$, and the N-dimensional vector $\Delta X^{(N)}$ is a one-to-one function of the observations. The vectors $\Delta B^{(N)}$ and $\Delta X^{(N)}$ are Gaussian with different means (except the case $\theta = 0$) and the same covariance matrix. We denote this covariance matrix by $\Gamma^{(N)}$. The next maximum likelihood estimator coincides with the least square estimator considered in Rao (2002, eq. (4a.1.5)).

Lemma 2.1. *Assume that the Gaussian distribution of the vector $(B_{t_k})_{k=1}^N$ is nonsingular. Then one can take the function*

$$L^{(N)}_{\Delta X^{(N)}=x}(\theta) = \frac{f_\theta(x)}{f_0(x)} = \exp\left\{ \theta z^\top \left(\Gamma^{(N)} \right)^{-1} x - \frac{\theta^2}{2} z^\top \left(\Gamma^{(N)} \right)^{-1} z \right\}, \tag{3}$$

where $z = (t_k - t_{k-1})_{k=1}^N$, as a likelihood function in the discrete-time model. MLE is linear with respect to the observations and equals

$$\hat{\theta}^{(N)} = \frac{z^\top \left(\Gamma^{(N)} \right)^{-1} \Delta X^{(N)}}{z^\top \left(\Gamma^{(N)} \right)^{-1} z}. \tag{4}$$

Proof. The pdf of $\Delta B^{(N)}$ with respect to the Lebesgue measure equals

$$f_\theta(x) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \Gamma^{(N)}}} \exp\left\{ -\frac{1}{2}(x - \theta z)^\top \left(\Gamma^{(N)} \right)^{-1} (x - \theta z) \right\}.$$

The density of the observations for given θ with respect to the distribution of the observations for $\theta = 0$ is taken as a likelihood function. □

Remark 2.2. Let the process X be observed on a regular grid, i.e., at the points $t_k = kh$, $k = 1, \ldots, N$, where $h > 0$. Then $\Gamma^{(N)}$ is a Toeplitz matrix, that is

$$\Gamma^{(N)}_{k+l,l} = \Gamma^{(N)}_{l,k+l} = \mathsf{E}\left(B_{(k+l)h} - B_{(k+l-1)h}\right)\left(B_{lh} - B_{(l-1)h}\right) = \mathsf{E}\,B_{(k+1)h}B_h - \mathsf{E}\,B_{kh}B_h$$

does not depend on l due to the stationarity of increments. This simplifies the numerical computation of MLE.

2.2. Properties of the estimator

Since $\Delta X^{(N)} = \Delta B^{(N)} + \theta z$, the maximum likelihood estimator (4) equals

$$\hat{\theta}^{(N)} = \theta + \frac{z^\top \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)}}{z^\top \left(\Gamma^{(N)}\right)^{-1} z}.$$

Lemma 2.3. *Under the assumptions of Lemma 2.1, the estimator $\hat{\theta}^{(N)}$ is unbiased and normally distributed. Its variance equals*

$$\operatorname{var}\hat{\theta}^{(N)} = \frac{1}{z^\top \left(\Gamma^{(N)}\right)^{-1} z}.$$

Proof. The estimator $\hat{\theta}^{(N)}$ is unbiased and normally distributed because $\hat{\theta}^{(N)} - \theta$ is linear and centered Gaussian vector $\Delta B^{(N)}$. The variance of the estimator is equal to

$$\operatorname{var}\hat{\theta}^{(N)} = \frac{\operatorname{var}\left(z^\top \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)}\right)}{\left(z^\top \left(\Gamma^{(N)}\right)^{-1} z\right)^2} = \frac{z^\top \left(\Gamma^{(N)}\right)^{-1} \operatorname{var}\left(\Delta B^{(N)}\right)\left(\Gamma^{(N)}\right)^{-1} z}{\left(z^\top \left(\Gamma^{(N)}\right)^{-1} z\right)^2}$$

$$= \frac{z^\top \left(\Gamma^{(N)}\right)^{-1} \Gamma^{(N)} \left(\Gamma^{(N)}\right)^{-1} z}{\left(z^\top \left(\Gamma^{(N)}\right)^{-1} z\right)^2} = \frac{1}{z^\top \left(\Gamma^{(N)}\right)^{-1} z}. \qquad \square$$

To prove the consistency of the estimator, we need the following technical result.

Lemma 2.4. *If $A \in \mathbb{R}^{N \times N}$ is a positive definite matrix, $x \in \mathbb{R}^N$, $x \neq 0$ is a non-zero vector, then*

$$x^\top A^{-1} x \geq \frac{\|x\|^4}{x^\top A x}.$$

Proof. As the matrix A is positive definite, $x^\top A x > 0$ and there exists a positive definite matrix $A^{1/2}$ (and so the matrix $A^{1/2}$ is symmetric and nonsingular) such that $(A^{1/2})^2 = A$. By the Cauchy-Schwarz inequality,

$$\|x\|^4 = \left(x^\top A^{1/2} A^{-1/2} x\right)^2 \leq \left\|A^{1/2}x\right\|^2 \left\|A^{-1/2}x\right\|^2 = \left(x^\top A x\right)\left(x^\top A^{-1} x\right),$$

whence the desired inequality follows. $\qquad \square$

In the rest of this section we assume that the process X is observed on a regular grid, at the points $t_k = kh$, $k = 1, \ldots, N$, for some $h > 0$. We also assume that for any N the Gaussian distribution of the vector $(B_{kh})_{k=1}^N$ is nonsingular.

Theorem 2.5. *Let $h > 0$, and*

$$\mathsf{E}\left(B_{(k+1)h} - B_{kh}\right) B_h \to 0 \quad as \quad k \to \infty.$$

Let $\hat{\theta}^{(N)}$ be the ML estimator of parameter θ of the model (1) by the observations X_{kh}, $k = 1, \ldots, N$. Then the estimator $\hat{\theta}^{(N)}$ is mean-square consistent, i.e.,

$$\mathsf{E}\left(\hat{\theta}^{(N)} - \theta\right)^2 \to 0 \quad as \quad N \to \infty.$$

Proof. By Remark 2.2, the matrix $\Gamma^{(N)}$ has Toeplitz structure,

$$\Gamma^{(N)}_{l,m} = \Gamma^{(N)}_{l,m} = \mathsf{E}\left(B_{(|l-m|+1)h} - B_{|l-m|h}\right)B_h.$$

Moreover, $\Gamma^{(N)}_{k,l}$ does not depend on N as soon as $N \geq \max(k,l)$. By Toeplitz theorem,

$$\frac{1}{N^2}\sum_{l=1}^{N}\sum_{m=1}^{N}\Gamma^{(N)}_{l,m} = \frac{1}{N}\mathsf{E}(B_h)^2 - \sum_{k=2}^{N}\frac{2(N+1-k)}{N^2}\mathsf{E}\left(B_{kh} - B_{(k-1)h}\right)B_h \rightarrow$$

$$\rightarrow \lim_{k\to\infty}\mathsf{E}\left(B_{kh} - B_{(k-1)h}\right)B_h = 0 \quad \text{as} \quad N \to \infty.$$

For a regular grid we have that $z = (h,\ldots,h)^\top$. Hence, in this case,

$$z^\top \Gamma^{(N)} z = h^2 \sum_{l=1}^{N}\sum_{m=1}^{N}\Gamma^{(N)}_{l,m}, \qquad \|z\| = h\sqrt{N}.$$

Finally, with the use of Lemma 2.3,

$$\mathsf{E}\left(\hat{\theta}^{(N)} - \theta\right)^2 = \frac{1}{z^\top\left(\Gamma^{(N)}\right)^{-1}z} \leq \frac{z^\top\Gamma^{(N)}z}{\|z\|^4} = \frac{1}{h^2 N^2}\sum_{l=1}^{N}\sum_{m=1}^{N}\Gamma^{(N)}_{l,m} \rightarrow 0 \quad \text{as} \quad N \to \infty. \quad \square$$

To prove the strong consistency, we need the following auxiliary statement.

Lemma 2.6. *Let $h > 0$, and $\hat{\theta}^{(N)}$ be the ML estimator of parameter θ of the model* (1) *by the observations X_{kh}, $k = 1,\ldots,N$. Then the random process $\hat{\theta}^{(N)}$ has independent increments.*

Proof. In the next paragraph, $N_2 \leq N_3$ are positive integers, $(I,0) = (I_{N_2}, 0_{N_2\times(N_3-N_2)})$ is $N_2 \times N_3$ diagonal matrix with ones on the diagonal, and its transpose is denoted by $\binom{I}{0}$. The vector $\Delta B^{(N_2)} = (B_h,\ldots,B_{(N_2-1)h} - B_{N_2 h})^\top$ is the beginning of vector $\Delta B^{(N_3)} = (B_h,\ldots,B_{(N_3-1)h} - B_{N_3 h})^\top$; the vector $z_{N_2} = (h,\ldots,h)^\top \in \mathbb{R}^{N_2}$ is the beginning of vector $z_{N_3} = (h,\ldots,h)^\top \in \mathbb{R}^{N_3}$, so

$$\Delta B^{(N_2)} = (I,0)\,\Delta B^{(N_3)}, \qquad z_{N_2} = (I,0)\,z_{N_3}.$$

Then

$$\mathsf{E}\,\Delta B^{(N_3)}\left(\Delta B^{(N_2)}\right)^\top = \mathsf{E}\,\Delta N^{(N_3)}\left(\Delta B^{(N_3)}\right)^\top \binom{I}{0} = \Gamma^{(N_3)}\binom{I}{0},$$

$$\mathsf{E}\,\hat{\theta}^{(N_3)}\hat{\theta}^{(N_2)} = \frac{\mathsf{E}\,z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}\Delta B^{(N_3)}\left(\Delta B^{(N_2)}\right)^\top\left(\Gamma^{(N_2)}\right)^{-1}z_{N_2}}{z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}z_{N_3}\,z_{N_2}^\top\left(\Gamma^{(N_2)}\right)^{-1}z_{N_2}} =$$

$$= \frac{z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}\Gamma^{(N_3)}\binom{I}{0}\left(\Gamma^{(N_2)}\right)^{-1}z_{N_2}}{z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}z_{N_3}\,z_{N_2}^\top\left(\Gamma^{(N_2)}\right)^{-1}z_{N_2}} =$$

$$= \frac{z_{N_3}^\top\binom{I}{0}\left(\Gamma^{(N_2)}\right)^{-1}z_{N_2}}{z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}z_{N_3}\,z_{N_2}^\top\left(\Gamma^{(N_2)}\right)^{-1}z_{N_2}} =$$

$$= \frac{z_{N_2}^\top\left(\Gamma^{(N_2)}\right)^{-1}z_{N_2}}{z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}z_{N_3}\,z_{N_2}^\top\left(\Gamma^{(N_2)}\right)^{-1}z_{N_2}} = \frac{1}{z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}z_{N_3}}.$$

For $N_1 \leq N_2 \leq N_3$

$$\mathsf{E}\,\hat{\theta}^{(N_3)}\left(\hat{\theta}^{(N_2)} - \hat{\theta}^{(N_1)}\right) = \mathsf{E}\,\hat{\theta}^{(N_3)}\hat{\theta}^{(N_2)} - \mathsf{E}\,\hat{\theta}^{(N_3)}\hat{\theta}^{(N_1)}$$

$$= \frac{1}{z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}z_{N_3}} - \frac{1}{z_{N_3}^\top\left(\Gamma^{(N_3)}\right)^{-1}z_{N_3}} = 0;$$

therefore for $N_1 \leq N_2 \leq N_3 \leq N_4$

$$\mathsf{E} \left(\hat{\theta}^{(N_4)} - \hat{\theta}^{(N_3)} \right) \left(\hat{\theta}^{(N_2)} - \hat{\theta}^{(N_1)} \right) = 0.$$

Thus, the Gaussian process $\{\hat{\theta}^{(N)}, \ N = 1, 2, \ldots\}$ is proved to have uncorrelated increments. Hence its increments are independent. □

Theorem 2.7. *Under the assumptions of Theorem 2.5, the estimator $\theta^{(N)}$ is strongly consistent, i.e. $\hat{\theta}^{(N)} \to \theta$ as $N \to \infty$ almost surely.*

Proof. By Theorem 2.5 $\mathrm{var}\, \hat{\theta}^{(N)} \to 0$ as $N \to \infty$, so

$$\mathrm{var} \left(\hat{\theta}^{(N)} - \hat{\theta}^{(N_0)} \right) = \mathrm{var}\, \hat{\theta}^{(N)} + \mathrm{var}\, \hat{\theta}^{(N_0)} - 2\sqrt{\mathrm{var}\, \hat{\theta}^{(N)}\, \mathrm{var}\, \hat{\theta}^{(N_0)}}\, \mathrm{corr} \left(\hat{\theta}^{(N)}, \hat{\theta}^{(N_0)} \right) \to \mathrm{var}\, \hat{\theta}^{(N_0)}$$

as $N \to \infty$. The process $\hat{\theta}^{(N)}$ has independent increments. Therefore by Kolmogorov's inequality, for $\epsilon > 0$ and $N \in \mathbb{N}$

$$\mathsf{P} \left(\sup_{N \geq N_0} \left| \hat{\theta}^{(N)} - \hat{\theta}^{(N_0)} \right| > \frac{\epsilon}{2} \right) \leq \frac{4}{\epsilon^2} \lim_{N \to \infty} \mathrm{var} \left(\hat{\theta}^{(N)} - \hat{\theta}^{(N_0)} \right) = \frac{4}{\epsilon^2} \mathrm{var}\, \hat{\theta}^{(N_0)}.$$

Then, using the unbiasedness of the estimator, we get

$$\mathsf{P} \left(\sup_{N \geq N_0} \left| \hat{\theta}^{(N)} - \theta \right| \geq \epsilon \right) \leq \mathsf{P} \left(\left| \hat{\theta}^{(N_0)} - \theta \right| \geq \frac{\epsilon}{2} \right) + \mathsf{P} \left(\sup_{N \geq N_0} \left| \hat{\theta}^{(N)} - \hat{\theta}^{(N_0)} \right| \geq \frac{\epsilon}{2} \right) \leq$$

$$\leq \frac{4}{\epsilon^2} \mathrm{var}\, \hat{\theta}^{(N_0)} + \frac{4}{\epsilon^2} \mathrm{var}\, \hat{\theta}^{(N_0)} = \frac{8}{\epsilon^2} \mathrm{var}\, \hat{\theta}^{(N_0)} \to 0 \quad \text{as} \quad N_0 \to \infty,$$

whence $|\hat{\theta}^{(N)} - \theta| \to 0$ as $N \to \infty$ almost surely. □

Example 2.8. Let us consider the model (1) with $B_t = B_t^{H_1} + B_t^{H_2}$, where $B_t^{H_1}$ and $B_t^{H_2}$ are two independent fractional Brownian motions with Hurst indices $H_1, H_2 \in (0, 1)$, i.e. centered Gaussian processes with covariance functions

$$\mathsf{E}\, B_t^{H_i} B_s^{H_i} = \frac{1}{2} \left(t^{2H_i} + s^{2H_i} - |t - s|^{2H_i} \right), \quad t \geq 0, \ s \geq 0, \ i = 1, 2.$$

These processes have stationary increments, and

$$\mathsf{E} \left(B_{(k+1)h}^{H_i} - B_{kh}^{H_i} \right) B_h^{H_i} \sim h^{2H_i} H_i \left(2H_i - 1 \right) k^{2H_i - 2} \to 0, \quad \text{as } k \to \infty,$$

see e. g. (Mishura 2008, Sec. 1.2). Taking into account the independence of centered processes $B_t^{H_1}$ and $B_t^{H_2}$, we obtain that

$$\mathsf{E}(B_{(k+1)h} - B_{kh})B_h = \mathsf{E} \left(B_{(k+1)h}^{H_1} - B_{kh}^{H_1} \right) B_h^{H_1} + \mathsf{E} \left(B_{(k+1)h}^{H_2} - B_{kh}^{H_2} \right) B_h^{H_2} \to 0 \quad \text{as } k \to \infty.$$

Thus, the assumptions of Theorem 2.5 are satisfied.

3. Maximum likelihood estimation by continuous observations

Let the process X be observed on the whole interval $[0, T]$. It is required to estimate the unknown parameter θ by these observations.

3.1. Likelihood function and construction of the estimator

In this section we construct a formula for continuous-time MLE, similar to the formula (4) for the discrete case.

Assumption 1. The covariance function of B_t has a mixed derivative

$$\frac{\partial^2}{\partial s\, \partial t} \left(\mathsf{E}\, B_t B_s \right) = K(t - s),$$

where $K(t)$ is an even function, $K \in L_1[-T, T]$.

Lemma 3.1. *Under Assumption 1, the integral $\int_0^T f(t)\, dB_t$ exists as the mean square limit of the corresponding Riemann sums for any $f \in L_2[0, T]$. Moreover,*

$$\mathsf{E}\left[\int_0^T f(t)\, dB_t \int_0^T g(s)\, dB_s \right] = \int_0^T f(t) \int_0^T K(t - s) g(s)\, ds\, dt \tag{5}$$

for any $f, g \in L_2[0, T]$.

Proof. According to Huang and Cambanis (1978) (see also Cramér and Leadbetter 2004, Sec. 5.3), the integral $\int_0^T f(t)\, dB_t$ exists if and only if the double Riemann integral $\int_0^T \int_0^T f(t) f(s) K(t - s)\, ds\, dt$ exists. Moreover, if the both integrals $\int_0^T f(t)\, dB_t$ and $\int_0^T g(s)\, dB_s$ exist, then formula (5) holds. However, using the properties of a convolution, one can prove that

$$\int_0^T \int_0^T f(t) f(s) K(t - s)\, ds\, dt \leq \|K\|_{L_1[-T,T]} \, \|f\|_{L_2[0,T]}^2 < \infty. \qquad \square$$

Define a linear operator $\Gamma_T \colon L_2[0, T] \to L_2[0, T]$ by

$$\Gamma_T f(t) = \int_0^T K(t - s) f(s)\, ds. \tag{6}$$

It follows from (6) that

$$\mathsf{E}\left[\int_0^T f(t)\, dB_t \int_0^T g(s)\, dB_s \right] = \int_0^T \Gamma_T f(t) g(t)\, dt. \tag{7}$$

For a fixed set of points $t_1, \ldots, t_N$ which satisfy (2) define mutually adjoint linear operators $M \colon L_2[0, T] \to \mathbb{R}^N$ and $M^* \colon \mathbb{R}^N \to L_2[0, T]$. If $f \in \mathbb{R}^N$, then let $Mf \in \mathbb{R}^N$ be a vector whose k-th element is equal to $\int_{t_{k-1}}^{t_k} f(s)\, ds$. The adjoint operator M^* is

$$M^* x = \sum_{k=1}^N x_k \, \mathbf{1}_{[t_{k-1}, t_k]}, \qquad x \in \mathbb{R}^N.$$

The basic properties of the operator Γ_T are collected in the following evident lemma.

Lemma 3.2. *Let Assumption 1 hold. Then*

(i) *The operator Γ_T is bounded ($\|\Gamma_T\| \leq \|K\|_{L_1[-T,T]}$) and self-adjoint;*

(ii) *The following relation between the operator Γ_T and the covariance matrix $\Gamma^{(N)}$ from Proposition 2.1 holds:*
$$M \Gamma_T M^* = \Gamma^{(N)},$$
i. e., the matrix of the operator $M \Gamma_T M^* \colon \mathbb{R}^N \to \mathbb{R}^N$ equals $\Gamma^{(N)}$.

Now we are ready to formulate our key assumption on the kernel K (in terms of the operator Γ_T).

Assumption 2. For all $T > 0$, the constant function $\mathbf{1}_{[0,T]}(t) = 1$, $t \in [0, T]$, belongs to the range of the operator Γ_T, i. e. there exists a function $h_T \in L_2[0, T]$ such that

$$\Gamma_T h_T = \mathbf{1}_{[0,T]}.$$

Theorem 3.3. *If all finite-dimensional distributions of the process* $\{B_t, \ t \in (0,T]\}$, *are nonsingular and Assumptions 1 and 2 hold, then*

$$L(\theta) = \exp\left\{\theta \int_0^T h_T(s)\, dB_s - \frac{\theta^2}{2} \int_0^T h_T(s)\, ds\right\} \tag{8}$$

is a likelihood function.

Proof. Let us show that the function $L(\theta)$ defined in (8) is a density function for the distribution of the process X_t for a given θ with respect to the density function of a distribution of the process B_t (it coincides with X_t when $\theta = 0$). In other words, we need to prove that

$$dP_\theta = L(\theta)\, dP_0,$$

where P_θ is the probability measure that corresponds to the value of the parameter θ. It suffices to show that for all partitions $0 = t_0 < t_1 < \ldots < t_N \le T$ of the interval $[0,T]$ and for all cylinder sets $A \in \mathcal{F}_N$ the following equality holds:

$$\int_A dP_\theta = \int_A L(\theta)\, dP_0, \tag{9}$$

where $\mathcal{F}_N$ is the σ-algebra, generated by the values B_{t_k} of the process B_t at the points t_k, $k = 1, \ldots, N$. We have

$$\int_A dP_\theta = \int_A L^{(N)}(\theta)\, P_0,$$

$$\int_A L(\theta)\, dP_0 = \int_A \mathsf{E}_0[L(\theta) \mid \mathcal{F}_N]\, dP_0,$$

where $L^{(N)}$ is the likelihood function (3) for the discrete-time model, and $\mathsf{E}_0[\cdot \mid \mathcal{F}_N]$ is the conditional expectation corresponding to the probability measure P_0. To prove (9), it suffices to show that

$$L^{(N)}(\theta) = \mathsf{E}_0[L(\theta) \mid \mathcal{F}_N].$$

If $\theta = 0$, then $X_t = B_t$,

$$\mathsf{E}_0[L(\theta) \mid \mathcal{F}_N] = \mathsf{E}\exp\left\{\theta \int_0^T h_T(s)\, dB_s - \frac{\theta^2}{2} \int_0^T h_T(s) ds\right\}.$$

Due to joint normality of $\int_0^T h_T(s)\, dB_s$ and $\Delta B^{(N)}$, the conditional distribution of $\int_0^T h_T(s)\, dB_s$ with respect to $\mathcal{F}_N$ is Gaussian (Anderson 2003, Theorem 2.5.1); its conditional variance is nonrandom. Let us find its parameters. By the least squares method,

$$\mathsf{E}[B_t \mid \mathcal{F}_N] = \mathsf{E}\left[B_t \mid \Delta B^{(N)}\right] = \mathrm{cov}\left(B_t, \Delta B^{(N)}\right)\left(\mathrm{cov}\left(\Delta B^{(N)}, \Delta B^{(N)}\right)\right)^{-1}\Delta B^{(N)}.$$

We have $\mathrm{cov}\left(\Delta B^{(N)}, \Delta B^{(N)}\right) = \Gamma^{(N)}$. Calculate $\mathrm{cov}\left(B_t, \Delta B^{(N)}\right)$:

$$\mathrm{cov}\left(B_t, \ B_{t_k} - B_{t_{k-1}}\right) = \int_{s=0}^t \int_{u=t_{k-1}}^{t_k} K(s-u)\, du\, ds,$$

and for any vector $x = (x_k)_{k=1}^N \in \mathbb{R}^N$

$$\mathrm{cov}\left(B_t, \Delta B^{(N)}\right) x = \sum_{k=1}^N \mathrm{cov}\left(B_t, \ B_{t_k} - B_{t_{k-1}}\right) x_k = \int_{s=0}^t \sum_{k=1}^N \int_{u=t_{k-1}}^{t_k} K(s-u)x_k\, du\, ds =$$

$$= \int_0^t \int_0^T K(s-u)M^*x(u)\, du\, ds = \int_0^T \mathbf{1}_{[0,t]}(s)\, \Gamma_T M^*x(s)\, ds =$$

$$= \int_0^T \Gamma_T \mathbf{1}_{[0,t]}(s)\, M^*x(s)\, ds = \left(M\Gamma_T \mathbf{1}_{[0,t]}\right)^\top x,$$

whence we get

$$\operatorname{cov}\left(B_t, \Delta B^{(N)}\right) = \left(M \Gamma_T \, \mathbf{1}_{[0,t]}\right)^\top.$$

Therefore

$$\mathsf{E}[B_t \mid \mathcal{F}_N] = \left(M \Gamma_T \, \mathbf{1}_{[0,t]}\right)^\top \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)} = \int_0^t \left(\Gamma_T M^* \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)}\right)(s) \, ds.$$

Then

$$\begin{aligned}
\mathsf{E}\left[\int_0^T h_T(t) \, dB_t \Big| \mathcal{F}_N\right] &= \int_0^T h_T(s) \left(\Gamma_T M^* \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)}\right)(s) \, ds \\
&= \int_0^T (\Gamma_T h_T)(s) \left(M^* \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)}\right)(s) \, ds \\
&= (M \Gamma_T h_T)^\top \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)},
\end{aligned}$$

where we have used that the operator Γ_T is self-adjoint.

Further, $M \Gamma_T h_T = M \, \mathbf{1}_{[0,T]} = z$, where the vector z is defined after (3). Hence,

$$\mathsf{E}\left[\int_0^T h_T(t) \, dB_t \Big| \mathcal{F}_N\right] = z^\top \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)}.$$

In order to calculate the variance we apply the partition-of-variance equality

$$\operatorname{var}\left[\int_0^T h_T(t) \, dB_t\right] = \operatorname{var}\left(\mathsf{E}\left[\int_0^T h_T(t) \, dB_t \Big| \mathcal{F}_N\right]\right) + \operatorname{var}\left[\int_0^T h_T(t) \, dB_t \Big| \mathcal{F}_N\right].$$

We have

$$\operatorname{var}\left[\int_0^T h_T(t) \, dB_t\right] = \int_0^T (\Gamma_T h_T)(t) \, h_T(t) \, dt = \int_0^T \mathbf{1}_{[0,T]}(t) h_T(t) \, dt = \int_0^T h_T(t) \, dt,$$

and

$$\operatorname{var}\left(\mathsf{E}\left[\int_0^T h_T(t) \, dB_t \Big| \mathcal{F}_N\right]\right) = \operatorname{var}\left(z^\top \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)}\right) = z^\top \left(\Gamma^{(N)}\right)^{-1} z.$$

Hence,

$$\operatorname{var}\left[\int_0^T h_T(t) \, dB_t \Big| \mathcal{F}_N\right] = \int_0^T h_T(t) \, dt - z^\top \left(\Gamma^{(N)}\right)^{-1} z. \tag{10}$$

Applying the formula for the mean of the log-normal distribution, we obtain

$$\begin{aligned}
\mathsf{E}_0[L(\theta) \mid \mathcal{F}_N] = \mathsf{E} \exp\Big\{ \theta z^\top \left(\Gamma^{(N)}\right)^{-1} \Delta B^{(N)} \\
+ \frac{\theta^2}{2}\left(\int_0^T h_T(t) \, dt - z^\top \left(\Gamma^{(N)}\right)^{-1} z\right) - \frac{\theta^2}{2} \int_0^T h_T(s) ds \Big\} = L^{(N)}(\theta).
\end{aligned}$$

Thus, (9) is proved. $\qquad\square$

Corollary 3.4. *The maximum likelihood estimator of θ by continuous observations is given by*

$$\hat{\theta}_T = \frac{\int_0^T h_T(t) \, dX_t}{\int_0^T h_T(t) \, dt}. \tag{11}$$

3.2. Properties of the estimator

It follows immediately from (11) that the maximum likelihood estimator $\hat{\theta}_T$ is equal to

$$\hat{\theta}_T = \theta + \frac{\int_0^T h_T(t) \, dB_t}{\int_0^T h_T(t) \, dt}. \tag{12}$$

Proposition 3.5. *The estimator $\hat\theta_T$ is unbiased and normally distributed. Its variance is equal to*

$$\operatorname{var}\hat\theta_T = \mathsf{E}\left(\hat\theta_T - \theta\right)^2 = \frac{1}{\int_0^T h_T(t)\,dt}. \tag{13}$$

Proof. Unbiasedness and normality follows from the fact that $\hat\theta - \theta$ is a linear functional of centered Gaussian process B. By (7),

$$\operatorname{var}\left(\int_0^T h_T(t)\,dB_t\right) = \int_0^T \Gamma_T h(t)\,h_T(t)\,dt = \int_0^T \mathbf{1}_{[0,\mathrm{T}]}(t)\,h_T(t)\,dt = \int_0^T h_T(t)\,dt.$$

Thus, equation (13) immediately follows from (12). $\qquad\square$

Corollary 3.6. *Let the process $B = \{B_t,\ t \ge 0\}$ satisfy Assumptions 1 and 2. If*

$$\int_0^T h_T(t)\,dt \to \infty, \quad as \quad T \to +\infty, \tag{14}$$

then the maximum likelihood estimator $\hat\theta_T$ is mean-square consistent, i.e., $\mathsf{E}(\hat\theta_T - \theta)^2 \to 0$, as $T \to +\infty$.

It can be hard to verify the condition (14). The following result gives sufficient conditions for the consistency in terms of the autocovariance function of B.

Theorem 3.7. *Let the process $B = \{B_t,\ t \ge 0\}$ satisfy Assumptions 1 and 2. If the covariance function of the increment process $B_N - B_{N-1}$ tends to 0:*

$$\mathsf{E}\left(B_{N+1} - B_N\right)B_1 \to 0 \quad as \quad N \to \infty,$$

then the maximum likelihood estimator $\hat\theta_T$ is mean-square consistent.

Proof. The estimator $\hat\theta^{(N)}$ from the discrete sample $\{X_1, \ldots, X_N\}$ is mean-square consistent by Theorem 2.5. The estimator from the continuous-time sample $\{X_t,\ t \in [0, T]\}$ is unbiased. Now compare the variances of the discrete and continuous-time estimators.

The desired inequalities are got from the proof of Theorem 2.5. Suppose that $T \ge 1$, N is an integer such that $N \le T < N + 1$. By equation (10) we have

$$\int_0^T h_T(t)\,dt \ge z^\top \left(\Gamma^{(N)}\right)^{-1} z. \tag{15}$$

As $\operatorname{var}\hat\theta_T = \frac{1}{\int_0^T h_T(t)\,dt}$, $\operatorname{var}\hat\theta^{(N)} = \left(z^\top\left(\Gamma^{(N)}\right)^{-1}z\right)^{-1}$, we have $\operatorname{var}\hat\theta_T \le \operatorname{var}\hat\theta^{(N)}$, and

$$\lim_{T\to+\infty} \mathsf{E}\left(\hat\theta_T - \theta\right)^2 = \lim_{N\to\infty} \mathsf{E}\left(\hat\theta^{(N)} - \theta\right)^2 = 0. \qquad\square$$

To prove the strong consistency of $\hat\theta_T$, we need the following auxiliary result.

Lemma 3.8. *Let the process B satisfy the conditions of Theorem 3.3. Then the estimator process $\hat\theta = \{\hat\theta_T, T \ge 0\}$ has independent increments.*

Proof. Let $T_2 \le T_3$. Then

$$\mathsf{E}\left[\int_0^{T_3} h_{T_3}(t)\,dB_t \int_0^{T_2} h_{T_2}(s)\,dB_s\right] = \int_0^{T_2} \Gamma_{T_3} h_{T_3}(t)\,h_{T_2}(t)\,dt$$

$$= \int_0^{T_2} \mathbf{1}_{[0,T_3]}(t)\,h_{T_2}(t)\,dt = \int_0^{T_2} h_{T_2}(t)\,dt.$$

Thus, if $0 < T_1 \le T_2 \le T_3 \le T_4$, then

$$
\begin{aligned}
&\mathsf{E}(\hat{\theta}_{T_4} - \hat{\theta}_{T_3})(\hat{\theta}_{T_2} - \hat{\theta}_{T_1}) \\
&= \mathsf{E}\left(\frac{\int_0^{T_4} h_{T_4}(t)\, dB_t}{\int_0^{T_4} h_{T_4}(t)\, dt} - \frac{\int_0^{T_3} h_{T_3}(t)\, dB_t}{\int_0^{T_3} h_{T_3}(t)\, dt} \right)\left(\frac{\int_0^{T_2} h_{T_2}(t)\, dB_t}{\int_0^{T_2} h_{T_2}(t)\, dt} - \frac{\int_0^{T_1} h_{T_1}(t)\, dB_t}{\int_0^{T_1} h_{T_1}(t)\, dt} \right) = \\
&= \frac{\mathsf{E}\left[\int_0^{T_4} h_{T_4}(t)\, dB_t \int_0^{T_2} h_{T_2}(t)\, dB_t \right]}{\int_0^{T_4} h_{T_4}(t)\, dt \int_0^{T_2} h_{T_2}(t)\, dt} - \frac{\mathsf{E}\left[\int_0^{T_3} h_{T_3}(t)\, dB_t \int_0^{T_2} h_{T_2}(t)\, dB_t \right]}{\int_0^{T_3} h_{T_3}(t)\, dt \int_0^{T_2} h_{T_2}(t)\, dt} - \\
&\quad - \frac{\mathsf{E}\left[\int_0^{T_4} h_{T_4}(t)\, dB_t \int_0^{T_1} h_{T_1}(t)\, dB_t \right]}{\int_0^{T_4} h_{T_4}(t)\, dt \int_0^{T_1} h_{T_1}(t)\, dt} + \frac{\mathsf{E}\left[\int_0^{T_3} h_{T_3}(t)\, dB_t \int_0^{T_1} h_{T_1}(t)\, dB_t \right]}{\int_0^{T_3} h_{T_3}(t)\, dt \int_0^{T_1} h_{T_1}(t)\, dt} = \\
&= \frac{1}{\int_0^{T_4} h_{T_4}(t)\, dt} - \frac{1}{\int_0^{T_3} h_{T_3}(t)\, dt} - \frac{1}{\int_0^{T_4} h_{T_4}(t)\, dt} + \frac{1}{\int_0^{T_3} h_{T_3}(t)\, dt} = 0.
\end{aligned}
$$

Similarly to the proof of Lemma 2.6, the random process $\hat{\theta}_T$ is Gaussian and its increments are proved to be uncorrelated so they are independent. $\qquad\square$

Theorem 3.9. *Under conditions of Theorem 3.7 the estimator $\hat{\theta}_T$ is strongly consistent.*

Proof. By Kolmogorov's inequality, for any $\epsilon > 0$ and $t_0 > 0$

$$
\mathsf{P}\left(\sup_{T > t_0} |\hat{\theta}_T - \theta| > \epsilon \right) \le \mathsf{P}\left(|\hat{\theta}_{t_0} - \theta| > \frac{\epsilon}{2} \right) + \mathsf{P}\left(\sup_{T > t_0} |\hat{\theta}_T - \theta_{t_0}| > \frac{\epsilon}{2} \right) \le
$$

$$
\le \frac{4}{\epsilon^2}\, \mathrm{var}\, \hat{\theta}_{t_0} + \frac{4}{\epsilon^2} \lim_{T \to +\infty} \mathrm{var}\left(\hat{\theta}_T - \hat{\theta}_{t_0} \right) = \frac{8}{\epsilon^2}\, \mathrm{var}\, \hat{\theta}_{t_0}.
$$

By Theorem 3.7,

$$
\lim_{t_0 \to +\infty} \mathsf{P}\left(\sup_{T > t_0} |\hat{\theta}_T - \theta| > \epsilon \right) = 0 \quad \text{for all} \quad \epsilon > 0,
$$

whence the strong consistency follows. $\qquad\square$

Remark 3.10. The Brownian motion does not satisfy Assumption 1 (as for covariance function $\max(s,t)$ of Wiener process, $\frac{\partial \max(s,t)}{\partial t}$ is not continuous in s). So we extend our model such that it can handle Wiener process. Let the process B be a sum of two independent random processes,

$$
B_t = B_t^{\mathrm{C}} + W_t, \tag{16}
$$

where B^{C} satisfies Assumption 1, and W is a standard Wiener process. Let us look at the changes of the statements if the process B admits representation (16) (Assumption 1 for B is dropped). Lemma 3.1 changes as follows:

$$
\mathsf{E}\left[\int_0^T f(t)\, dB_t \int_0^T g(s)\, dB_s \right] = \int_0^T f(t) \int_0^T K(t - s) g(s)\, ds\, dt + \int_0^T f(t) g(t)\, dt.
$$

Equation (7) will stand true, if we set

$$
\Gamma_T f(t) = f(t) + \Gamma_T^{\mathrm{C}} f(t) = f(t) + \int_0^T K(t - s) f(s)\, ds.
$$

Lemma 3.2 stands true (with $\|\Gamma_T\| \le \|K\|_{L_1[-T,T]} + 1$). Theorem 3.3 holds true; minor changes in the proof are required.

Table 1: The means and variances of $\hat{\theta}_T$

H	T	Sample Mean	Sample Variance	Theoretical Variance
0.6	1	2.0344	1.9829	1.8292
	10	2.0047	0.2584	0.2356
0.7	1	1.9995	1.9956	1.9692
	10	2.0296	0.3484	0.3270
0.8	1	2.0165	2.0435	1.9930
	10	2.0071	0.5155	0.4392
0.9	1	2.0117	2.0376	1.9984
	10	2.0099	0.7710	0.5867

3.3. Examples

Example 3.11. Let B be a fractional Brownian motion with the Hurst index $H \in (1/2, 1)$. Then $K(t) = \frac{H(2H-1)}{|t|^{2-2H}}$. We denote by Γ_T^H the corresponding operator Γ_T. Then for the function

$$h_T(s) = C_H s^{1/2-H}(T-s)^{1/2-H},$$

$C_H = \left(H(2H-1)\mathrm{B}\left(H - \frac{1}{2}, \frac{3}{2} - H\right)\right)^{-1}$, we have that $\Gamma_T^H h_T = \mathbf{1}_{[0,T]}$, see Norros *et al.* (1999). The maximum likelihood estimator is given by

$$\hat{\theta}_T = \frac{T^{2H-2}}{\mathrm{B}(3/2 - H,\ 3/2 - H)} \int_0^T s^{1/2-H}(T-s)^{1/2-H}\,dX_s.$$

Example 3.12. Consider the following model:

$$X_t = \theta t + W_t + B_t^H, \tag{17}$$

where W is a standard Wiener process, B^H is a fractional Brownian motion with Hurst index H, and random processes W_t and B_t^H are independent. The process $W_t + B_t^H$ admits representation (16) with $B^C = B^H$. Corresponding operator Γ_T is $\Gamma_T = I + \Gamma_T^H$ (see Example 3.11 for the definition of Γ_T^H). The operator Γ_T^H is self-adjoint and positive semi-definite. Hence, the operator Γ_T is invertible. Thus Assumption 2 holds true.

The function $h_T = \Gamma_T^{-1}\,\mathbf{1}_{[0,T]}$ can be evaluated iteratively

$$h_T = \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\left\|\Gamma_T^H\right\| I - \Gamma_T^H\right)^k \mathbf{1}_{[0,T]}}{\left(1 + \frac{1}{2}\left\|\Gamma_T^H\right\|\right)^{k+1}}. \tag{18}$$

3.4. Simulations

We illustrate the behavior of the maximum likelihood estimator for Example 3.12 by means of simulation experiments. For $T = 1$ and $T = 10$ and various values of H we find h_T iteratively by (18). Then for $\theta = 2$ we simulate 1000 realizations of the process (17) for each H and compute the estimates by (11). The means and variances of these estimates are reported in Table 1. The theoretical variances calculated by (13) are also presented. We see that these simulation studies confirm the theoretical properties of $\hat{\theta}_T$, especially unbiasedness and consistency.

References

Anderson TV (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley, Hoboken NJ.

Bertin K, Torres S, Tudor CA (2011). "Maximum-likelihood Estimators and Random Walks in Long Memory Models." *Statistics*, **45**(4), 361–374.

Cai C, Chigansky P, Kleptsyna M (2016). "Mixed Gaussian Processes: A Filtering Approach." *Ann. Probab.*, **44**(4), 3032–3075.

Cramér H, Leadbetter MR (2004). *Stationary and Related Stochastic Processes*. Dover Publications, Inc., Mineola, NY. Sample function properties and their applications, Reprint of the 1967 original.

Hu Y, Nualart D, Xiao W, Zhang W (2011). "Exact Maximum Likelihood Estimator for Drift Fractional Brownian Motion at Discrete Observation." *Acta Math. Sci. Ser. B Engl. Ed.*, **31**(5), 1851–1859.

Huang ST, Cambanis S (1978). "Stochastic and Multiple Wiener Integrals for Gaussian Processes." *Ann. Probab.*, **6**(4), 585–614.

Le Breton A (1998). "Filtering and Parameter Estimation in a Simple Linear System Driven by a Fractional Brownian Motion." *Statist. Probab. Lett.*, **38**(3), 263–274.

Mishura Y (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes*, volume 1929. Springer Science & Business Media.

Mishura Y (2016). "Maximum Likelihood Drift Estimation for the Mixing of Two Fractional Brownian Motions." In *Stochastic and Infinite Dimensional Analysis*, pp. 263–280. Springer.

Mishura Y, Voronov I (2015). "Construction of Maximum Likelihood Estimator in the Mixed Fractional–Fractional Brownian Motion Model with Double Long-range Dependence." *Mod. Stoch. Theory Appl.*, **2**(2), 147–164.

Norros I, Valkeila E, Virtamo J (1999). "An Elementary Approach to a Girsanov Formula and Other Analytical Results on Fractional Brownian motions." *Bernoulli*, **5(4)**, 571–587.

Rao CR (2002). *Linear Statistical Inference and its Applications*. Second edition. Wiley, New York. Wiley Series in Probability and Statistics.

Stiburek D (2017). "Statistical Inference on the Drift Parameter in Fractional Brownian Motion with a Deterministic Drift." *Comm. Statist. Theory Methods*, **46**(2), 892–905.

Affiliation:

Yuliya Mishura, Kostiantyn Ralchenko, Sergiy Shklyar
Department of Probability Theory, Statistics and Actuarial Mathematics
Taras Shevchenko National University of Kyiv
64 Volodymyrska
01601 Kyiv, Ukraine
E-mail: myus@univ.kiev.ua, k.ralchenko@gmail.com, shklyar@univ.kiev.ua