Optimal estimation of a signal perturbed by a fractional Brownian noise

Artemov A. V.  
Burnaev E. V.

12 February 2015

Abstract

We consider the problem of optimal estimation of the value of a vector parameter \( \theta = (\theta_0, \ldots, \theta_n)^\top \) of a drift term in a fractional Brownian motion represented by a finite sum \( \sum_{i=0}^{n} \theta_i \varphi_i(t) \) over known functions \( \varphi_i(t) \), \( i = 0, \ldots, n \). For the value of the parameter \( \theta \), we obtain a maximum likelihood estimate as well as Bayesian estimates for normal and uniform prior distributions.

Keywords: fractal Brownian motion, maximum likelihood estimate, Bayesian estimate, sequential estimation, optimal stopping

1 Problem definition

Let \( \xi = (\xi_t)_{0 \leq t \leq T} \) be a stochastic process defined on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) and represented by

\[ \xi_t = a(t) + \sigma(t)B_t^H, \]

where \( B^H = (B_t^H)_{0 \leq t \leq T} \) is a fractional Brownian motion with Hurst index \( H \in (0, 1) \), and let drift \( a(t) \) and diffusion \( \sigma(t) \) coefficients satisfy the conditions \( \int_0^T |a(t)| \, dt < \infty \) and \( \int_0^T |\sigma(t)|^2 \, dt < \infty \), respectively. The function \( \sigma(t) \) is assumed to be known. Let the drift term \( a(t) \) be represented by a sum

\[ a(t) = \sum_{i=0}^{n} \theta_i \varphi_i(t) \]

over the known functions \( \varphi_i(t) \), satisfying \( \int_0^T |\varphi_i(t)| \, dt < \infty \), \( i = 0, \ldots, n \), with unknown parameters \( \theta_i \), \( i = 0, \ldots, n \). For brevity we consider vector-valued variables \( \theta = (\theta_0, \ldots, \theta_n)^\top \) and \( \varphi(t) = (\varphi_0(t), \ldots, \varphi_n(t))^\top \), such that

\[ a(t) = \theta^\top \varphi(t). \]

We consider the problem of finding a sequential estimate of \( \theta \) given observations \( \{\xi_s, 0 \leq s \leq t\} \) available up to time \( t \) using the maximum likelihood and the Bayesian approaches. Within the maximum likelihood approach, \( \theta \) is considered as an unknown nonrandom vector-valued parameter, and we seek to find an estimate \( \hat{\theta}_{\text{ML}} = \hat{\theta}_{\text{ML}}(t) \) maximizing the likelihood of the observed process.

In the Bayesian case, we assume \( \theta \) to be a random vector taking values in \( \mathbb{R}^{n+1} \) according to some known prior distribution \( p(\theta|x), x \in \mathbb{R}^{n+1} \). We then consider the problem of finding a sequential estimation rule \( \hat{\theta}_{\text{BAYES}} = (\tau_{\text{BAYES}}, \theta_{\text{BAYES}}) \) such that

\[ \inf_{\delta \in \mathbb{D}} \mathbb{E} \left[ c\tau + \|\theta - \hat{\theta}\|^2 \right] = \mathbb{E} \left[ c\tau_{\text{BAYES}} + \|\theta - \hat{\theta}_{\text{BAYES}}\|^2 \right], \]

where \( \mathbb{D} = \{\delta : \delta = (\tau, \hat{\theta})\} \) is a class of stopping rules with finite stopping times \( \tau \leq T < \infty \) w.r.t. filtration \( \mathcal{F}_T = \sigma(\{\xi_s, 0 \leq s \leq t\}) \). The constant \( c > 0 \) is interpreted as a cost of the observations. The Bayesian estimation strategy consists in stopping the observations at a time \( \tau_{\text{BAYES}} \) and declaring \( \hat{\theta}_{\text{BAYES}} \) to be the optimal estimate of \( \theta \).

The problem of extracting a deterministic signal from observations perturbed by a fractional Gaussian noise has attracted little attention in literature devoted to optimal estimation. The only work in this direction known to us is [1] where an optimal Bayesian estimate for the parameter \( \mu \) of the fractional Bayesian motion with a linear drift \( a(t) = \mu t \) is derived assuming that \( \mu \) is a normally distributed random variable with known mean and variance.
2 Fractional Brownian motion

The process of fractional Brownian motion (FBM) was introduced by Kolmogorov [7] and later constructively defined by Mandelbrot [4]. We use notations from [3].

The standard fractional Brownian motion \( B^H_t \) on \([0, T]\) with Hurst index \( H \in (0, 1) \) is a Gaussian process with continuous sample path such that

\[
B^H_0 = 0, \quad \mathbb{E}_B^H = 0, \quad \mathbb{E}_{B^H} = \frac{1}{2} (s^{2H} - t^{2H} + |t - s|^{2H}).
\]

When \( H = 1/2 \), FBM reduces to an ordinary Brownian motion, however, when \( H \neq 1/2 \), FBM is not a martingale. Let us denote for \( 0 \leq s < t \leq T \)

\[
\kappa_H = 2H \left( \frac{3}{2} - H \right) \Gamma \left( \frac{1}{2} + H \right), \quad \lambda_H = \frac{2H(3 - 2H)\Gamma(1/2 + H)}{(3/2 - H)}, \quad w_H(t) = \lambda_H^{-1}t^{2-2H},
\]

and define the process \( M^H_t = (M^H_t)_{0 \leq t \leq T} \) according to the relation

\[
M^H_t = \int_0^t \kappa_H(t, s) dB^H_s.
\]

The process \( M^H_t \) defined in this way is a Gaussian martingale and has the quadratic variation \( \langle M^H_t \rangle \) equal to \( w_H(t) \) (see [3], [6]). For convenience we also define the process \( m^H_t = (m^H_t)_{0 \leq t \leq T} \) by the relation \( m^H_t = M^H_t / w_H(t) \).

3 The Girsanov theorem for the FBM

In this section, we cite a result from [3] regarding the likelihood process for the fractional Brownian motion with a drift. Let \( Y = (Y_t)_{0 \leq t \leq T} \) be a process defined on the filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) and let its stochastic differential satisfy the relation

\[
dY_t = C(t) dt + D(t) dB_t^H,
\]

where \( B^H = (B^H_t)_{0 \leq t \leq T} \) is a FBM with Hurst index \( H \in (0, 1) \), and functions \( C(t) \) and \( D(t) \) are such that the function \( Q_H(t) \) is properly defined by the relation

\[
Q_H(t) = \frac{d}{dw^H_t} \int_0^t k_H(t, s)D^{-1}(s)C(s) ds.
\]

In this formula, differentiation w.r.t. \( dw^H_t \) is understood in the following way:

\[
\frac{df(t)}{dw^H_t} = \lambda_H \frac{\kappa_H}{2 - 2H} t^{2H-1} \frac{df(t)}{dt}.
\]

Defining the function \( Q_H(t) \) allows one to formulate an analogue of the Girsanov theorem for the process \( Y \).

**Theorem 1** (see [3]) Let \( Q_H(t) \) belong to the space \( L^2([0, T], dw^H_t) \), where the quantity \( dw^H_t \) is defined by (5). Let us define a random process \( \Lambda^H_t = (\Lambda^H_t)_{0 \leq t \leq T} \) by the relation

\[
\Lambda^H_t = \exp \left\{ \int_0^t Q_H(s) dM^H_s - \frac{1}{2} \int_0^t (Q_H(s))^2 dw^H_s \right\}.
\]

The \( \mathbb{E}_H = 1 \) and the distribution of \( Y \) w.r.t. the measure \( \mathbb{P}^Y = \Lambda^H \mathbb{P} \) coincides with the distribution of \( \int_0^T D(s) dB^H_s \) w.r.t. \( \mathbb{P} \).

The random process \( \Lambda^H_t \) is called the likelihood process or the Radon-Nikodym derivative \( d\mathbb{P}^Y / d\mathbb{P} \) of the measure \( \mathbb{P}^Y \) w.r.t. the measure \( \mathbb{P} \).
4 The maximum likelihood estimate of the drift parameter

Let us consider the problem of finding the maximum likelihood estimate for the drift parameter $\theta$ defined in (1). According to (1)-(2), the process $\xi_t$ satisfies the equation

$$\xi_t = \sum_{i=0}^{n} \theta_i \varphi_i(t) + \sigma(t)B_t^H,$$

while its stochastic differential satisfies the relation

$$d\xi_t = \sum_{i=0}^{n} \theta_i \varphi'_i(t) dt + \sigma(t) dB_t^H.$$

The structure of the likelihood process and the corresponding estimate is described by the following theorem.

Theorem 2 Let the drift coefficient $a(t)$ of the fractional Brownian motion have the form (2)-(3). Then the maximum likelihood estimate $\hat{\theta}_{ML}$ for the drift parameter $\theta$ is defined by

$$\hat{\theta}_{ML} = R_H^{-1}(t)\psi_H^t,$$

where $R_H(t)$ is a nonrandom matrix with elements defined by

$$(R_H(t))_{ij} = \int_0^t \psi_i(s)\psi_j(s) d\psi_H^s, \quad i,j = 0,\ldots,n,$$

and $\psi_H^t = (\psi_H^t)_{0\leq t \leq T}$ is a stochastic process taking values in $\mathbb{R}^{n+1}$ with coordinates defined by

$$(\psi_H^t)_i = \int_0^t \psi_i(s) dM^H_s, \quad i = 0,\ldots,n,$$

where the functions $\psi_i(t), i = 0,\ldots,n$, are given by

$$\psi_i(t) = \frac{d}{ds} \int_0^t k_H(t,s)\sigma^{-1}(s)\varphi'_i(s) ds, \quad i = 0,\ldots,n,$$

and $M^H_t$ is defined by (6) with $\xi = (\xi_t)_{0\leq t \leq T}$ instead of $B^H = (B^H_t)_{0\leq t \leq T}$.

Proof. The general form of $Q^\theta_H(t)$ function is defined by (7). Using the notation from (13) for the functions $\psi_i(t), i = 0,\ldots,n$, we obtain

$$Q^\theta_H(t) = \sum_{i=0}^{n} \theta_i \frac{d}{ds} \int_0^t k_H(t,s)\sigma^{-1}(s)\varphi'_i(s) ds = \sum_{i=0}^{n} \theta_i \psi_i(t).$$

The likelihood process $\Lambda^H$ is then defined as (see (8)):

$$\Lambda^H(\theta) = \exp \left\{ \sum_{i=0}^{n} \theta_i \int_0^t \psi_i(s) dM^H_s - \frac{1}{2} \int_0^t \left( \sum_{i=0}^{n} \theta_i \psi_i(s) \right)^2 d\psi_H^s \right\}.$$

The process $\Lambda^H$ defines the Radon-Nikodym derivative of the measure generated by the observations $\xi$ from (9) w.r.t. the measure of the process $\xi_t = B^H_s, s \leq t$. Using the vector notation from (3), one can write the formula in (14) more compactly:

$$\Lambda^H(\theta) = \exp \left\{ \theta^\top \psi_H^t - \frac{1}{2} \theta^\top R_H(t) \theta \right\},$$

where the elements of the $n \times n$ matrix $R_H(t)$ and the components of the $(n+1)$-dimensional process $\psi_H^t$ are defined by (11) and (12), respectively. The maximum likelihood estimate $\hat{\theta}_{ML} = \arg \max_{\theta} \Lambda^H(\theta)$ is obtained as a solution of the system of linear equations

$$\int_0^t \psi_i(s) dM^H_s - \sum_{j=0}^{n} \theta_j \int_0^t \psi_i(s)\psi_j(s) d\psi_H^s = 0, \quad i = 0,\ldots,n,$$

which could be written in a vector form

$$\psi_H^t - R_H(t)\theta = 0.$$

If the matrix $R_H(t)$ is invertible for every $t \geq 0$, then the solution of the system is $\hat{\theta}_{ML}$ from (10).
Corollary 1 (the case of a polynomial drift) Let $\varphi_i(t) = t^i, i = 0, \ldots, n$, and assume the diffusion coefficient to be constant $\sigma(t) = \sigma$. Then the observable process has the structure $\xi_t = \sum_{i=0}^n \theta_i t^i + \sigma B^H_t$, functions $\psi_i(t) = \beta_H(i)/\sigma^{i+1}, i = 0, \ldots, n$, whereas the components of the vector-valued stochastic process $\psi^H$ from (12) and the elements of the matrix $R_H(t)$ from (11) are defined by

$$
(\psi^H_i)_i = \frac{\beta_H(i)}{\sigma} \int_0^t s^{-1} dM^H_s, \quad (R_H(t))_{ij} = \frac{\alpha_H(i,j)}{\sigma^2} t^{i+j-2H}
$$

respectively, where

$$
\alpha_H(i,j) = \lambda_{H}^{-1} \beta_H(i) \beta_H(j) \frac{2 - 2H}{i+j-2H}, \\
\beta_H(i) = i \frac{2 - 2H + i - 1}{2 - 2H} \frac{\Gamma(3 - 2H)}{\Gamma(3/2 - H + i - 1)} \frac{\Gamma(3/2 - H)}{\Gamma(3/2 - H)}, \quad i, j = 0, \ldots, n.
$$

The maximum likelihood estimate $\hat{\theta}_{ML}$ is obtained as a solution to the equation $\psi^H_t - R_H(t)\theta = 0$.

We note that for $n = 1$ the observable process satisfies the stochastic differential equation $d\xi_t = \theta_0 dt + \sigma dB^H_t$ and the likelihood process has the form $\Lambda^H_1(\theta) = \exp\{\theta_1^2 \sigma^2 M^H_1 - \theta_1^2 \sigma^2 \lambda_{H}^{-1} t^{2-2H}/2\}$. Therefore the maximum likelihood estimate $(\hat{\theta}_1)_{ML}$ of $\theta_1$ has the form

$$
(\hat{\theta}_1)_{ML} = \frac{\sigma M^H_1}{w_H(t)}.
$$

This particular result (for $\sigma = 1$) has been obtained in [6].

5 The Bayesian estimate of the drift parameter

Consider the problem of finding the Bayesian estimate of the parameter $\theta \in \mathbb{R}^{n+1}$ assuming that $\theta$ has a prior distribution $P^\theta$ with density $p^\theta(x), x = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1}$.

According to the generalized Bayes rule (see [1], [8]), the conditional distribution density of $\theta$ given observations $F^\xi_t = \sigma(\{\xi_s, 0 \leq s \leq t\})$ is represented by

$$
p^\theta(x, F^\xi_t) = \frac{dP(\theta_0 \leq x_0, \ldots, \theta_n \leq x_n|F^\xi_t)}{dx_0 \cdots dx_n} = \frac{p^\theta(x) \Lambda^H_1(x)}{\int_{\mathbb{R}^{n+1}} p^\theta(z) \Lambda^H_1(z) dz}, \quad x \in \mathbb{R}^{n+1},
$$

where $\Lambda^H_1(x)$ is the likelihood process previously described in section 4. We further consider two special cases where the prior distribution of $\theta$ is either normal or uniform.

5.1 The case of a normal prior distribution

The main result of this section is presented in the following theorem.

Theorem 3 Let $\theta$ be a multivariate normal random variable with mean $m$ and covariance matrix $\Sigma$. Then the optimal Bayesian estimate $\hat{\theta}_{BAYES}$ for the value of $\theta$ is the posterior mean

$$
\hat{\theta}_{BAYES} = E[\theta | F^\xi_t] = (R_H(t) + \Sigma^{-1})^{-1}(\psi^H_t + \Sigma^{-1}m).
$$

The estimation error $E(\|\theta - \hat{\theta}_{BAYES}\|^2 | F^\xi_t)$ is defined by the trace of the posterior covariance matrix

$$
cov[\theta | F^\xi_t] = (R_H(t) + \Sigma^{-1})^{-1}.
$$

Proof. It is well known that the optimal least squares estimate for the value of the vector $\theta$ conditioned upon the observations history $\{\xi_s, 0 \leq s \leq t\}$ up to the moment $t$ is defined by the conditional expectation $E[\theta | F^\xi_t]$. The estimation error is defined by the trace of the conditional covariance matrix $\text{cov}[\theta | F^\xi_t]$. In what follows we show that these quantities are easy to compute in the normal case.
Using the formulae (15) and (17) and writing the multivariate normal density as $p^\theta(x) = (2\pi)^{-(n+1)/2}(\det\Sigma)^{-1/2}\exp\{-\frac{1}{2}(x - m)^T\Sigma^{-1}(x - m)/2\}$, we obtain the following formula for the conditional distribution of $\theta$ given $\mathcal{F}_t^\xi$ is a constant, the conditional covariance matrix has the form

$$
\mathbb{E}[\theta | \mathcal{F}_t^\xi] = \mathbf{A}^{-1}b = (\mathbf{R}_H(t) + \Sigma^{-1})^{-1}(\psi_t^H + \Sigma^{-1}m),
$$

$$
\text{cov}[\theta | \mathcal{F}_t^\xi] = \mathbf{A}^{-1} = (\mathbf{R}_H(t) + \Sigma^{-1})^{-1},
$$

respectively. The quantity $\mathbb{E}[\|\theta - \hat{\theta}_{BAYES}\|^2 | \mathcal{F}_t^\xi]$ representing the conditional mean squared estimation error in the normal case has the form

$$
\mathbb{E}[\|\theta - \hat{\theta}_{BAYES}\|^2 | \mathcal{F}_t^\xi] = \mathbb{E}[\text{tr}((\mathbf{R}_H(t) + \Sigma^{-1})^{-1})] = \text{tr}((\mathbf{R}_H(t) + \Sigma^{-1})^{-1}).
$$

**Corollary 2** Let the conditions of theorem 3 be satisfied. Then the optimal stopping time for (4) is nonrandom.

**Proof.** In order to determine the optimal stopping time $\tau_{BAYES}$ in (4) we have to solve the following optimal stopping problem:

$$
\tau_{BAYES} = \arg\inf_{\tau \in \mathbb{D}} \mathbb{E}\left[ct + \mathbb{E}(\|\theta - \hat{\theta}_{BAYES}\|^2 | \mathcal{F}_t^\xi)\right] = \arg\inf_{t \in [0,T]} F_H(t),
$$

where the function

$$
F_H(t) = ct + \mathbb{E}(\|\theta - \hat{\theta}_{BAYES}\|^2 | \mathcal{F}_t^\xi) = ct + \text{tr}((\mathbf{R}_H(t) + \Sigma^{-1})^{-1}), \quad t \in [0, T],
$$

is nonrandom.

**Corollary 3** (the case of a polynomial drift) Let $\varphi_i(t) = t^i$, $i = 0, \ldots, n$, assume the diffusion coefficient to be constant, $\sigma(t) = \sigma$, and let the covariance matrix $\Sigma$ be diagonal meaning that all $\theta_0, \ldots, \theta_n$ are independent of each other. Then the function $F_H(t)$ from (20) has a single minimum for $t \in [0, T]$.

**Proof.** Let $\Sigma = \text{diag}(\gamma_0^2, \ldots, \gamma_n^2)$, where $\gamma_i^2 = \text{var}(\theta_i), i = 0, \ldots, n$. Then the trace of the conditional covariance matrix has the form

$$
\text{tr}((\mathbf{R}_H(t) + \Sigma^{-1})^{-1}) = \prod_{i=0}^n \left(\frac{\alpha_H(i, i)^{2i-2H}}{\sigma^4} + \gamma^{-2}\right)^{-1}
$$
and is a strictly increasing function for \( t > 0 \). Thus \( F_H(t) \) in (20) is a sum of a strictly increasing and a strictly decreasing function and has a unique minimum at some \( t \in [0, T] \).

A similar result can be obtained for the case of a linear trend \( a(t) = \mu t \) [1]. Fig. 1 presents the graph of \( F_H(t) \) for the case of a quadratic trend and values \( H = 0.2, c = 0.02 \).

Note that if the observable process \( \xi \) satisfies a linear stochastic differential equation \( d\xi = \theta_1 \, dt + \sigma \, dB^H \), where \( \theta_1 \) is a normally distributed random variable with expectation \( m \) and variance \( \gamma^2 \), then its conditional distribution is normal with density

\[
p_{\theta_1}(x \mid \mathcal{F}_t^c) = \sqrt{\frac{w_H(t)/\sigma^2 + 1/\gamma^2}{2\pi}} \exp\left\{ -\left( x - \frac{M_H^t/\sigma + m/\gamma^2}{w_H(t)/\sigma^2 + 1/\gamma^2} \right)^2 \frac{w_H(t)/\sigma^2 + 1/\gamma^2}{2} \right\}.
\]

Therefore the Bayesian estimate \( (\theta_1)_{\text{BAYES}} = \mathbb{E} [\theta_1 \mid \mathcal{F}_t^c] \) of \( \theta_1 \) and its corresponding estimation error \( \mathbb{E} [(\theta_1 - (\theta_1)_{\text{BAYES}})^2 \mid \mathcal{F}_t^c] \) are given by the relations

\[
(\theta_1)_{\text{BAYES}} = \frac{M_H^t/\sigma + m/\gamma^2}{w_H(t)/\sigma^2 + 1/\gamma^2} \quad \text{and} \quad \mathbb{E} [(\theta_1 - (\theta_1)_{\text{BAYES}})^2 \mid \mathcal{F}_t^c] = \frac{1}{w_H(t)/\sigma^2 + 1/\gamma^2}
\]

respectively, a result previously obtained by Norros [6] (for \( \sigma = 1 \)).

### 5.2 The case of a uniform prior distribution

Consider the problem of finding a Bayesian estimate for the value of the parameter \( \theta \in \mathbb{R}^{n+1} \), given that it is uniformly distributed on the \((n+1)\)-dimensional cube \( r = \prod_{i=0}^{n}[a_i, b_i] \).

The density of the prior distribution \( \theta \) is specified by

\[
p^\theta(x) = \prod_{i=0}^{n} \frac{1}{b_i - a_i} 1_{[a_i, b_i]}(x_i) = \frac{1}{|r|} 1_r(x),
\]

where \( 1_r(x) = \prod_{i=0}^{n} 1_{[a_i, b_i]}(x_i), |r| = \prod_{i=0}^{n}(b_i - a_i)^{-1} \). The corresponding posterior density is given by the expression

\[
p^\theta(x \mid \mathcal{F}_t^c) = \frac{1}{Z_H^t} 1_r(x) \exp \left( x^\top \psi_H^t - \frac{1}{2} x^\top R_H(t)x \right) = \frac{1}{Z_H^t} 1_r(x) \Lambda_H^t(x),
\]

where the process \( Z_H^t = (Z_H^t)_{0 \leq t \leq T} \), specified by the equality \( Z_H^t = \int_0^t \Lambda_H^s(x) \, d^n x \), plays the role of the normalization factor, and \( \Lambda_H^t(x) \) is the likelihood process defined according to (14).

An analytic derivation of the normalization factor \( Z_H^t \), the conditional mean \( \mathbb{E} [\theta \mid \mathcal{F}_t^c] \) and the covariance matrix \( \text{cov} [\theta \mid \mathcal{F}_t^c] \) is a difficult problem for an arbitrary value of \( n \) (these quantities can be numerically computed using, e.g., the algorithm from [2]). We shall dwell upon the derivation

![Figure 1: Graph of the cost function $F_H(t)$ for $n = 2$, $H = 0.2$, $c = 0.02$.](image)
of the estimate for an important particular case of a linear trend, where the observable process $\xi$ satisfies the following stochastic differential equation

$$d\xi_t = \theta_1 \, dt + \sigma \, dB_t^H,$$

where $\theta_1 \sim U(a, b)$. We present the result of our derivation in the following theorem.

**Theorem 4** Let $\theta_1$ from (21) be a random variable uniformly distributed on $[a, b]$, and independent of $B_t^H$. Then the optimal Bayesian estimate of the value of the parameter $\theta_1$ has the form

$$\hat{\theta}_1^{\text{BAYES}} = m_t^H + [\Lambda_H(\theta_1)]^{-1}[\Lambda_H(a) - \Lambda_H(b)],$$

the conditional mean square estimation error is given by

$$\gamma_t^H = \mathbb{E} \left( (\theta_1 - \hat{\theta}_1^{\text{BAYES}})^2 \mid F_t^\xi \right)$$

where

$$Z_t^H = \sqrt{\frac{2\pi}{w_H(t)}} \exp \left( \frac{1}{2} \left( (m_t^H)^2 w_H(t) \right) \right),$$

$$C_t^H = \Phi \left( (b - m_t^H) \sqrt{w_H(t)} \right) - \Phi \left( (a - m_t^H) \sqrt{w_H(t)} \right).$$

**Proof.** The conditional distribution $p^\theta_1(x \mid F_t^\xi)$ is easy to obtain using a direct computation, it is given by a formula

$$p^\theta_1(x \mid F_t^\xi) = \frac{1}{Z_t^H} \mathbf{1}_{[a, b]}(x) \exp \left( w_H(t) \left( xm_t^H - \frac{x^2}{2} \right) \right),$$

where the process $Z_t^H$ is defined according to (24). The conditional mean and variance are similarly obtained by computing the corresponding integrals.

We further present several asymptotic properties of the obtained Bayesian filter (22).

For $a \to -\infty, b \to +\infty$ (i.e. when $\theta_1$ is arbitrary) the Bayesian estimate in (22) coincides with the maximum likelihood estimate. Indeed, as $x \to \pm \infty$ we obtain $\Lambda_H(x) \to 0$ so the second term in (22) vanishes as $x \to \pm \infty$, meaning that $\hat{\theta}_1^{\text{BAYES}} \to m_t^H$.

As $t \to \infty$, the Bayesian estimate in (22) also coincides with the maximum likelihood estimate. Indeed, for $t \to \infty$ we have $w_H(t) \to \infty$, meaning that the second term in (22) vanishes as $t \to \infty$, and $\hat{\theta}_1^{\text{BAYES}} \to m_t^H$.

Consider the problem of finding the optimal stopping time in (4). The cost function in this problem is given by

$$\mathbb{E} \left[ c_T + \mathbb{E} \left( (\theta_1 - \hat{\theta}_1^{\text{BAYES}})^2 \mid F_T^\xi \right) \right] = \mathbb{E} \left[ c_T + \gamma_T^H \right],$$

where the random process $\gamma_T^H = (\gamma_t^H)_{0 \leq t \leq T}$ is given by the relation (23). Note that as $t \to \infty$, the following relation holds: $\gamma_t^H \to 0$. To determine the optimal stopping time, it is necessary to solve

$$\tau_{\text{BAYES}} = \arg \inf_T \mathbb{E} \left[ c_T + \gamma_T^H \right].$$

Since formulas (22), (23) and (24) are very complicated, an analytic solution for $\tau_{\text{BAYES}}$ from (25) is infeasible, meaning that only a numerical estimation of the stopping time is possible (see approaches in [5]).

6 Acknowledgements

The research was carried out at IITP RAS and supported by RSF grant No. 14-50-00150 only.

Artemov A.V. is with Lomonosov Moscow State University, Complex Systems Modelling Laboratory, Moscow, Russia; Yandex Data Factory, Moscow, Russia (artemov@physics.msu.ru).

Burnaev E.V. is with A.A. Kharkkevich Institute of Information Transmission Problems, RAS, Moscow, Russia (burnaev@iitp.ru).
References

[1] Çetin U., Novikov A., Shiryaev A. N. Bayesian sequential estimation of a drift of fractional Brownian motion. — Sequential Analysis, 2013, v. 32, No 3, p. 288–296.

[2] Genz A. Numerical computation of multivariate normal probabilities. — J. Comput. Graph. Statist., 1992, v. 1, No 2, p. 141–149.

[3] Kleptsyna M. L., Le Breton A., Roubaud M.-C. Parameter estimation and optimal filtering for fractional type stochastic systems. — Statist. Inference Stochastic Processes, 2000, v. 3, No 1–2, p. 173–182.

[4] Mandelbrot B. B., Van Ness J. W. Fractional Brownian motions, fractional noises and applications. — SIAM Rev., 1968, v. 10, No 4, p. 422–437.

[5] Peskir G., Shiryaev A. Optimal Stopping and Free-Boundary Problems. ETH Zürich: Springer, 2006, 502 p.

[6] Norros I., Valkeila E., Virtamo J. An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. — Bernoulli, 1999, v. 5, No 4, p. 571–587.

[7] Kolmogorov A. N. The Wiener spiral and some other interesting curves in Hilbert space. — Dokl. Akad. Nauk SSSR, 1940, v. 26, No 2, p. 115–118.

[8] Liptser R. S., Shiryaev A. N. Statistics of stochastic processes. Moscow: Nauka, 1974, 696 p.