Large deviations of regression parameter estimator in continuous-time models with sub-Gaussian noise

Alexander V. Ivanov, Igor V. Orlovskyi[*]

National Technical University of Ukraine, «Igor Sikorsky Kyiv Polytechnic Institute», Department of Mathematical Analysis and Probability Theory, Peremogi avenue 37, 03056 Kiev, Ukraine

alexntuu@gmail.com (A.V. Ivanov), orlovskyi@matan.kpi.ua (I.V. Orlovskyi)

Abstract A continuous-time regression model with a jointly strictly sub-Gaussian random noise is considered in the paper. Upper exponential bounds for probabilities of large deviations of the least squares estimator for the regression parameter are obtained.

1 Introduction

Theory of large deviations in mathematical statistics and statistics of stochastic processes deals with the asymptotic behaviour of tails of distribution functions of parametric and nonparametric statistical estimators. Concerning parametric estimators it is necessary to refer to the monograph of Ibragimov and Has'minskii [6] where the exponential convergence rate of probabilities of large deviations for maximum likelihood estimator was obtained. This result led to the appearance of a large number of publications on the subject of large deviations of statistical estimators.

Further we will speak about least squares estimators (l.s.e.'s) for parameters of a nonlinear regression model. In the paper of Ivanov [8] a statement was proved on

---

[*]Corresponding author.

VTeX

the power decreasing rate for probabilities of large deviations of l.s.e. for a scalar parameter in the nonlinear regression model with i.i.d. observation errors having moments of finite order. Prakasa Rao [13] obtained a similar result with the exponential decreasing rate in the Gaussian nonlinear regression.

In the paper of Sieders and Dzhaparidze [15] a general Theorem 2.1 on probabilities of large deviations for M-estimators based on a data set of any structure was proved that generalizes the mentioned result in [6] with an application to l.s.e. for parameters of the nonlinear regression with pre-Gaussian and sub-Gaussian i.i.d. observation errors (Theorems 3.1 and 3.2 in [15]). Some results in this direction are obtained by Ivanov [9].

The results on probabilities of large deviations of an l.s.e. in a nonlinear regression model with correlated observations can be found in the works of Ivanov and Leonenko [11], Prakasa Rao [14], Hu [4], Yang and Hu [16], Huang et al. [5].

Upper exponential bounds for probabilities of large deviations of an l.s.e. for a parameter of the nonlinear regression in discrete-time models with a jointly strictly sub-Gaussian (j.s.s.-G.) random noise were obtained in Ivanov [10]. In the present paper we extend some results of [10] to continuous-time observation models.

Consider a regression model

$$X(t) = a(t, \theta) + \varepsilon(t), \quad t \geq 0, \tag{1}$$

where $a(t, \tau)$, $(t, \tau) \in \mathbb{R}_+ \times \Theta^c$, is a continuous function, a true parameter value $\theta = (\theta_1, \ldots, \theta_q)'$ belongs to an open bounded convex set $\Theta \subset \mathbb{R}^q$ and a random noise $\varepsilon = \{\varepsilon(t), t \in \mathbb{R}\}$ satisfies the following condition.

N1. ε is a mean-square and almost sure (a.s.) continuous stochastic process defined on the probability space $(\Omega, \mathfrak{F}, P)$, $E\varepsilon(t) = 0, t \in \mathbb{R}$.

We shall write $\int = \int_0^T$.

Definition 1. Any random vector $\theta_T = (\theta_{1T}, \ldots, \theta_{qT})' \in \Theta^c$ having the property

$$Q_T(\theta_T) = \inf_{\tau \in \Theta^c} Q_T(\tau), \ Q_T(\tau) = \int \big[X(t) - a(t, \tau)\big]^2 dt.$$

is said to be the l.s.e. for an unknown parameter θ, obtained by the observations $\{X(t), t \in [0, T]\}$.

Under assumptions introduced above there exists at least one such random vector θ_T [12].

In the asymptotic theory of nonlinear regression in the problem of normal approximation of the distribution of an l.s.e., the difference $\theta_T - \theta$ is normed by diagonal matrix [11]

$$d_T(\theta) = \text{diag}\big(d_{iT}(\theta), \ i = \overline{1, q}\big), \qquad d_{iT}^2(\theta) = \int \left(\frac{\partial}{\partial \theta_i} a(t, \theta)\right)^2 dt.$$

Further it is supposed that the function $a(t, \cdot) \in C^1(\Theta)$ for any $t \geq 0$.

The paper is organized in the following way. In Section 2 an upper exponential bound is obtained for large deviations of $d_T(\theta)(\theta_T - \theta)$ in the regression model (1) with a j.s.s.-G. random noise ε. In Section 3 the results of Section 2 are applied to a stationary j.s.s.-G. noise ε. Section 4 contains examples of regression functions a and noise ε satisfying the conditions of our theorems.

2   Large deviations in models with a jointly strictly sub-Gaussian noise

Definition 2. A random vector $\xi = (\xi_1, \ldots, \xi_n)' \in \mathbb{R}^n$ is called strictly sub-Gaussian (s.s.-G.) if for any $\Delta = (\Delta_1, \ldots, \Delta_n)' \in \mathbb{R}^n$

$$E \exp\{\langle \xi, \Delta \rangle\} \leq \exp\left\{\frac{1}{2}\langle B\Delta, \Delta \rangle\right\},$$

where $\langle \xi, \Delta \rangle = \sum_{i=1}^n \xi_i \Delta_i$, $B = (B(i, j))_{i,j=1}^n$ is the covariance matrix of ξ, that is $B(i, j) = E\xi_i\xi_j, i, j = \overline{1, n}$, $\langle B\Delta, \Delta \rangle = \sum_{i,j=1}^n B(i, j)\Delta_i \Delta_j$.

Note that we obtain from Definition 2 the definition of an s.s.-G. random variable (r.v.) ξ taking $n = 1$.

Definition 3. $\{\xi(t), \ t \in \mathbb{R}\}$ is said to be jointly strictly sub-Gaussian (j.s.s.-G.) stochastic process, if for any $n \geq 1$, and any $t_1, \ldots, t_n \in \mathbb{R}$ the random vector $\xi_n = (\xi(t_1), \ldots, \xi(t_n))'$ is s.s.-G.

These definitions and a more detailed information on sub-Gaussian r.v.'s, vectors and stochastic processes can be found in the monograph [1] by Buldygin and Kozachenko.

Concerning the random noise ε in the model (1) we introduce the following assumption.

N2(i) ε is a j.s.s.-G. stochastic process with the covariance function $B(t, s) = E\varepsilon(t)\varepsilon(s), t, s \in \mathbb{R}$.

    (ii) For any $T > 0$, $\Delta(\cdot) \in L_2([0, T])$

$$\langle B\Delta, \Delta \rangle_T = \int \int B(t, s)\Delta(t)\Delta(s)dtds \leq d_0\|\Delta\|_T^2 \tag{2}$$

for some constant $d_0 > 0$, $\|\Delta\|_T = (\int \Delta^2(t)dt)^{\frac{1}{2}}$.

For a fixed T the exact bound in (2) is

$$\langle B\Delta, \Delta \rangle_T \leq \|B\|_T \|\Delta\|_T^2,$$

where $\|B\|_T$ is the norm of a self-adjoint positive semidefinite operator B in $L_2([0, T])$. Note that $\|B\|_T$ is a nondecreasing function of $T > 0$, so there exists

$$\lim_{T \to \infty} \|B\|_T \leq d_0 < \infty,$$

if (2) is fulfilled.

Example 2.1. Assume the covariance function $B(t, s)$ is such that

$$1) \ \ b_1^2 = \int_0^\infty \int_0^\infty B^2(t, s)dtds < \infty \qquad \text{or} \quad 2) \ \ b_2 = \sup_{t \in \mathbb{R}_+} \int_0^\infty |B(t, s)|ds < \infty.$$

Then using condition 1) and the Fubini theorem we get

$$\langle B\Delta, \Delta \rangle_T = \int \left(\int B(t, s)\Delta(s)ds \right) \Delta(t)dt$$

$$\leq \left(\int \left(\int B(t, s)\Delta(s)ds \right)^2 dt \right)^{\frac{1}{2}} \|\Delta\|_T \leq b_1\|\Delta\|_T^2,$$

and we can take $d_0 = b_1$. On the other hand,

$$\langle B\Delta, \Delta \rangle_T \leq \int \int |B(t, s)| \Delta^2(t) dt ds \leq b_2 \|\Delta\|_T^2,$$

and we can take $d_0 = b_2$.

Let $\Delta(t)$, $t \in \mathbb{R}_+$, be a continuous function. Then condition **N1** implies the existence of the integral

$$I(T) = \int \Delta(t)\varepsilon(t) dt, \tag{3}$$

determined for almost all paths of the process $\varepsilon(t)$, $t \in [0, T]$, as the Riemann integral. Consider partitions $r(n)$

$$0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = T$$

of the interval $[0, T]$ such that $\max_{1 \leq k \leq n}(t_k^{(n)} - t_{k-1}^{(n)}) \to 0$, as $n \to \infty$, and the corresponding integral sums

$$I_n(T) = \sum_{k=1}^{n} u_k^{(n)} \varepsilon(t_k^{(n)}), \qquad u_k^{(n)} = \Delta(t_k^{(n)})(t_k^{(n)} - t_{k-1}^{(n)}), \qquad k = \overline{1, n}.$$

Then

$$I_n(T) \to I(T) \text{ a.s.,} \quad \text{as } n \to \infty. \tag{4}$$

It is obvious also that

$$E I_n^2(T) \to E I^2(T) = \langle B\Delta, \Delta \rangle_T, \quad \text{as } n \to \infty. \tag{5}$$

Lemma 1. *Under conditions* **N1**, **N2** *integral* (3) *is an s.s.-G. r.v., for any* $T > 0$.

Proof. From Definition 3 it follows that the process $\varepsilon(t), t \in [0, T]$, is j.s.s.-G., if for any $n \geq 1$, and $t_1, \ldots, t_n \in [0, T], u_1, \ldots, u_n \in \mathbb{R}, \lambda \in \mathbb{R}$,

$$E \exp\left\{\lambda \sum_{k=1}^{n} u_k \varepsilon(t_k)\right\} \leq \exp\left\{\frac{1}{2}\lambda^2 \sum_{i,j=1}^{n} B(t_i, t_j) u_i u_j\right\}.$$

Taking $t_k = t_k^{(n)}, u_k = u_k^{(n)}, k = \overline{1, n}$, we obtain

$$E \exp\{\lambda I_n(T)\} \leq \exp\left\{\frac{1}{2}\lambda^2 E I_n^2(T)\right\}.$$

Due to (4) and (5) by the Fatou lemma (see, for example, [3])

$$E \exp\{\lambda I(T)\} = E \lim_{n\to\infty} \exp\{\lambda I_n(T)\} \leq \liminf_{n\to\infty} E \exp\{\lambda I_n(T)\}$$

$$\leq \lim_{n\to\infty} \exp\left\{\frac{1}{2}\lambda^2 E I_n^2(T)\right\} = \exp\left\{\frac{1}{2}\lambda^2 E I^2(T)\right\}. \qquad \square$$

The following statement on the exponential bound for distribution tails of integrals (3) plays an important role in subsequent proofs.

Lemma 2. *Under conditions* **N1** *and* **N2** *for any* $T > 0$, $x > 0$,

$$P\{I(T) \geq x\} \leq G_T(x), \qquad P\{I(T) \leq -x\} \leq G_T(x),$$
$$P\{|I(T)| \geq x\} \leq 2G_T(x), \tag{6}$$

where

$$G_T(x) = \exp\left\{-\frac{x^2}{2d_0\|\Delta\|_T^2}\right\}.$$

Proof. The proof is obvious (see, for example, [1]). For any $x > 0$, $\lambda > 0$ by the Chebyshev–Markov inequality, (2), and Lemma 1

$$P\{I(T) \geq x\} \leq \exp\{-\lambda x\} \exp\left\{\frac{1}{2}\lambda^2 \langle B\Delta, \Delta\rangle_T\right\} \leq \exp\left\{\frac{1}{2}\lambda^2 d_0\|\Delta\|_T^2 - \lambda x\right\}. \tag{7}$$

Minimization of the right-hand side of (7) in λ gives the first inequality in (6). The proof of the second inequality is similar. The third inequality follows from the previous ones. □

We need some notation to formulate conditions on the regression function $a(t, \theta)$ using the approach of the paper [15] (see also [9, 11]). Write

$$U_T(\theta) = d_T(\theta)(\Theta^c - \theta), \qquad \Gamma_{T,\theta,R} = U_T(\theta) \cap \{u : R \leq \|u\| \leq R + 1\},$$

$u = (u_1, \ldots, u_q)' \in \mathbb{R}^q$. Denote by G the family of all functions $g = g_T(R)$, $T > 0$, $R > 0$, having the following properties:
1) for fixed T, $g_T(R) \uparrow \infty$, as $R \to \infty$;
2) for any $r > 0$,

$$R^r \exp\{-g_T(R)\} \to 0, \quad \text{as } R, T \to \infty.$$

Let $\gamma(R)$ be polynomials of R (possibly different) with coefficients that do not depend on values T, θ, u, v. Set also

$$\Delta(t, u) = a(t, \theta + d_T^{-1}(\theta)u) - a(t, \theta), \quad t \in [0, T],$$
$$\Phi_T(u, v) = \int (\Delta(t, u) - \Delta(t, v))^2 dt, \quad u, v \in U_T(\theta).$$

Assume the existence of a function $g \in G$, constants $\delta \in (0, \frac{1}{2})$, $\varkappa > 0$, $\rho \in (0, 1]$ and polynomials $\gamma(R)$ such that for sufficiently large T, R (we will write $T > T_0$, $R > R_0$) the following conditions are fulfilled.
R1(i) For any $u, v \in \Gamma_{T,\theta,R}$ such that $\|u - v\| \leq \varkappa$

$$\Phi_T(u, v) \leq \gamma(R)\|u - v\|^{2\rho}; \tag{8}$$

(ii) for any $u \in \Gamma_{T,\theta,R}$ $\Phi_T(u, 0) \leq \gamma(R)$.
R2. For any $u \in \Gamma_{T,\theta,R}$

$$\Phi_T(u, 0) \geq 2d_0\delta^{-2}g_T(R). \tag{9}$$

Theorem 1. *If conditions* **N1**, **N2**, **R1** *and* **R2** *are fulfilled, then there exist constants* B_0, $b_0 > 0$ *such that for* $T > T_0$, $R > R_0$

$$P\{\|d_T(\theta)(\theta_T - \theta)\| \geq R\} \leq B_0 \exp\{-b_0 g_T(R)\}, \tag{10}$$

where for any $\beta > 0$ *the constant* B_0 *can be chosen so that*

$$b_0 \geq \frac{\rho}{\rho + q} - \beta. \tag{11}$$

Proof. Set

$$I(T, u) = \int \Delta(t, u)\varepsilon(t)dt, \qquad \zeta_T(u) = I(T, u) - \frac{1}{2}\Phi_T(u, 0).$$

To prove the theorem it is sufficient to check the fulfilment of assumptions (M1) and (M2) of the Theorem 2.1 in [15], reformulated in the manner similar to that used in the proof of Theorem 3.1, ibid.:

(M1) for any $m > 0$ and $u, v \in \Gamma_{T,\theta,R}$ such that $\|u - v\| \leq \varkappa$,

$$E|\zeta_T(u) - \zeta_T(v)|^m \leq \gamma(R)\|u - v\|^{\rho m}; \tag{12}$$

(M2)

$$P\left\{\zeta_T(u) - \zeta_T(0) \geq -\left(\frac{1}{2} - \delta\right)\Phi_T(u, 0)\right\} \leq \exp\{-g_T(R)\}. \tag{13}$$

From the first inequality in (6) of Lemma 2 for $\Delta(t) = \Delta(t, u)$, $x = \delta\Phi_T(u, 0)$, condition **R2**, taking into account that $\zeta_T(0) = 0$ in our particular case, we obtain

$$P\left\{\zeta_T(u) - \zeta_T(0) \geq -\left(\frac{1}{2} - \delta\right)\Phi_T(u, 0)\right\} = P\{I(T, u) \geq \delta\Phi_T(u, 0)\}$$
$$\leq \exp\{-\delta^2(2d_0)^{-1}\Phi_T(u, 0)\}$$
$$\leq \exp\{-g_T(R)\},$$

i.e. (13) is true.

On the other hand,

$$E|\zeta_T(u) - \zeta_T(v)|^m$$
$$\leq \max(1, 2^{m-1}) \cdot (E|I(T, u) - I(T, v)|^m + 2^{-m}|\Phi_T(u, 0) - \Phi_T(v, 0)|^m), \tag{14}$$

$$|\Phi_T(u, 0) - \Phi_T(v, 0)|$$
$$\leq \int |\Delta(t, u) - \Delta(t, v)| \cdot |\Delta(t, u) + \Delta(t, v)|dt$$
$$\leq \Phi_T^{\frac{1}{2}}(u, v) \cdot (\Phi_T^{\frac{1}{2}}(u, 0) + \Phi_T^{\frac{1}{2}}(v, 0))$$
$$\leq 2(\gamma(R))^{\frac{1}{2}}\|u - v\|^{\rho}(\gamma(R))^{\frac{1}{2}}$$
$$\leq (\gamma(R) + \gamma(R))\|u - v\|^{\rho} = \gamma(R)\|u - v\|^{\rho}$$

according to **R1** (polynomials $\gamma(R)$ are different in the last two lines). Thus we obtain the bound

$$\left|\Phi_T(u,0) - \Phi_T(v,0)\right|^m \le \gamma(R)\|u - v\|^{\rho m}. \tag{15}$$

By the formula for the moments of nonnegative r.v. (see, for example, [2] and compare with [4]) and the third inequality of Lemma 2 being applied to $\Delta(t) = \Delta(t, u) - \Delta(t, v), t \in [0, T]$, it holds

$$
\begin{aligned}
E\left|I(T; u, v)\right|^m &= m \int_0^\infty x^{m-1} P\{|I(T; u, v)| \ge x\}dx \\
&\le 2m \int_0^\infty x^{m-1} \exp\left\{-\frac{x^2}{2d_0\Phi_T(u, v)}\right\}dx \\
&= \sqrt{2\pi}md_0^{\frac{m}{2}}\Phi_T^{\frac{m}{2}}(u, v)E|Z|^{m-1},
\end{aligned} \tag{16}
$$

where $I(T; u, v) = I(T, u) - I(T, v)$, Z is a standard Gaussian r.v.,

$$E|Z|^{m-1} = \pi^{-\frac{1}{2}}2^{\frac{m-1}{2}}\Gamma\left(\frac{m}{2}\right), \quad m > 0. \tag{17}$$

Relations (16), (17), and (8) lead to the bound

$$E\left|I(T; u, v)\right|^m \le 2^{\frac{m}{2}}m\Gamma\left(\frac{m}{2}\right)d_0^{\frac{m}{2}}\Phi_T^{\frac{m}{2}}(u, v) \le \gamma(R)\|u - v\|^{\rho m}. \tag{18}$$

From (14), (15), and (18) it follows (12). $\qquad\square$

Suppose there exist a diagonal matrix $s_T = \text{diag}(s_{iT}, \ i = \overline{1, q})$ with elements that do not depend on $\tau \in \Theta$, and constants $0 < \underline{c}_i < \overline{c}_i < \infty, i = \overline{1, q}$, such that uniformly in $\tau \in \Theta$ for $T > T_0$

$$\underline{c}_i < s_{iT}^{-1}d_{iT}(\tau) < \overline{c}_i, \quad i = \overline{1, q}. \tag{19}$$

Then instead of the matrix $d_T(\theta)$ (at least in the framework of the topic of this paper) it is possible to consider, without loss of generality, the normalizing matrix s_T.

The next condition is more restrictive than **R1** and **R2**, however it is simpler due to requirement (19).

R3. There exist numbers $0 < c_0 < c_1 < \infty$ such that for any $u, v \in U_T(\theta) = s_T(\Theta^c - \theta)$ and $T > T_0$,

$$c_0\|u - v\|^2 \le \Phi_T(u, v) \le c_1\|u - v\|^2. \tag{20}$$

It goes without saying that in the expression for $\Phi_T(u, v)$ in (20) we use the matrix s_T^{-1} instead of $d_T^{-1}(\theta)$.

A condition of the type (20) has been introduced in [8] and used in [13, 15, 4] and other works. The next theorem generalizes Theorem 3.2 from [15].

Theorem 2. *Under conditions* **N1***,* **N2** *and* **R3** *there exist constants* $B, b > 0$ *such that for* $T > T_0, R > R_0$

$$P\{\|s_T(\theta_T - \theta)\| \geq R\} \leq B \exp\{-bR^2\}, \tag{21}$$

moreover for any $\beta > 0$ *the constant* B *can be chosen so that*

$$b \geq \frac{c_0}{8d_0(1+q)} - \beta. \tag{22}$$

Proof. We will show that **R3** implies conditions **R1** and **R2**. Inequality (8) of the condition **R1(i)** follows from the right-hand side of inequality (20), if we take $\rho = 1$, $\gamma(R) = c_1$. Inequality of the condition **R1(ii)** follows as well from the right-hand side of (20), if we take $\upsilon = 0$, $\gamma(R) = c_1(R+1)^2$.

To check the fulfilment of condition **R2** we should rewrite the left-hand side of (20) for $\upsilon = 0$:

$$\Phi_T(u, 0) \geq c_0\|u\|^2 \geq 2d_0\delta^{-2}\left(\frac{1}{2}\delta^2 d_0^{-1} c_0 R^2\right),$$

i.e., in the inequality (9) one can take $g_T(R) = \frac{1}{2}\delta^2 d_0^{-1} c_0 R^2$. In this case for the exponent in (10) we have

$$-b_0 g_T(R) = -\left(\frac{1}{2}\delta^2 b_0 d_0^{-1} c_0\right)R^2.$$

Now, since $\rho = 1$ in (11), for any $\beta > 0$ in (21) we can take

$$b = b_\delta = \frac{1}{2}\delta^2 b_0 d_0^{-1} c_0 \geq \frac{\delta^2 c_0}{2d_0(1+q)} - \beta.$$

By **R3** and **N1**, **N2**, inequality (9) with $g_T(R) = \frac{1}{2}\delta^2 d_0^{-1} c_0 R^2$ holds for any $\delta \in (0, \frac{1}{2})$. We get inequality (22) as $\delta \to \frac{1}{2}$. $\qquad\square$

3   Large deviations in the case of a stationary jointly strictly sub-Gaussian noise

We impose an additional restriction on the noise process ε.

N3. The stochastic process ε is stationary with the covariance function $B(t) = E\varepsilon(0)\varepsilon(t)$, $t \in \mathbb{R}$, and the bounded spectral density $f(\lambda)$, $\lambda \in \mathbb{R}$:

$$f_0 = \sup_{\lambda \in \mathbb{R}} f(\lambda) < \infty.$$

Under assumption **N3** the following corollaries of the theorems proved in Section 2 are true.

Corollary 1. *If conditions* **N1***,* **N2(i)***,* **N3***,* **R1** *and* **R2** *are fulfilled, then the statement of Theorem 1 is true with* $d_0 = 2\pi f_0$.

Proof. We just need to show that condition **N2(ii)** is fulfilled. Indeed, by the Plancherel identity,

$$\langle B\Delta, \Delta \rangle_T = \int_{-\infty}^{\infty} f(\lambda) \left| \int e^{i\lambda t} \Delta(t)dt \right|^2 d\lambda \leq 2\pi f_0 \|\Delta\|_T^2. \qquad \square$$

Corollary 2. *Under conditions* **N1**, **N2(i)**, **N3** *and* **R3** *the statement of Theorem 2 is true with inequality (22) rewritten in the form*

$$b \geq \frac{c_0}{16\pi f_0(1+q)} - \beta. \qquad (23)$$

Our next assumption is a particularization of the requirements **N2** and **N3**.

N4(i). The random noise ε is of the form

$$\varepsilon(t) = \int_{-\infty}^{t} \psi(t-s)d\xi(s) = \int_{0}^{\infty} \psi(s)d\xi(t-s), \qquad (24)$$

where $\xi = \{\xi(t), \ t \in \mathbb{R}\}$ is a mean-square continuous j.s.s.-G. stochastic process with orthogonal increments, $E\xi(t) = 0$,

$$E\left|\xi(t+s) - \xi(t)\right|^2 = s, \quad t \in \mathbb{R}, \ s > 0;$$

$\psi(t) = 0$ as $t < 0$ and

$$\int_{0}^{\infty} \psi^2(t)dt < \infty.$$

The stochastic integral in (24) is understood as a mean-square Stieltjes integral [3]. The process ξ is an integrated white noise, ε can be considered as a stationary process at the output of a physically realizable filter with the covariance function (see ibid.)

$$B(t) = \int_{0}^{\infty} \psi(t+u)\psi(u)du$$

and the spectral density

$$f(\lambda) = \left|h(i\lambda)\right|^2, \qquad h(i\lambda) = (2\pi)^{-\frac{1}{2}} \int_{0}^{\infty} \psi(t)e^{-i\lambda t}dt.$$

N4(ii). $f_0 = \sup_{\lambda \in \mathbb{R}} |h(i\lambda)|^2 < \infty.$
Obviously **N4(ii)** holds, if $\int_{0}^{\infty} |\psi(t)|dt < \infty.$

Lemma 3. *If condition* **N4(i)** *holds, then* ε *in (24) is a j.s.s.-G. process.*

Proof. Let $n \geq 1$ be a fixed number and $t_1, \ldots, t_n, \Delta_1, \ldots, \Delta_n$ be arbitrary real numbers. It is necessary to prove that

$$E \exp\left\{\sum_{k=1}^{n} \Delta_k \varepsilon(t_k)\right\} \leq \exp\left\{\frac{1}{2}\sum_{i,j=1}^{n} B(t_i - t_j)\Delta_i \Delta_j\right\}. \tag{25}$$

Formula (24) can be rewritten in the form

$$\varepsilon(t) = \int_{-\infty}^{\infty} \psi(t - s)d\xi(s), \quad \psi(t) = 0, \text{ as } t < 0.$$

Denote $\psi_k(s) = \psi(t_k - s), k = \overline{1, n}$. Then

$$\varepsilon(t_k) = \int_{-\infty}^{\infty} \psi_k(s)d\xi(s), \quad k = \overline{1, n}.$$

Let a sequence of simple functions

$$\psi_k^{(m)}(s) = \sum_{l=1}^{r(k,m)} c_{kl}^{(m)} \chi_{\pi_{kl}^{(m)}}(s), \quad m \geq 1,$$

where $\pi_{kl}^{(m)} = [\alpha_{kl}^{(m)}, \beta_{kl}^{(m)}), k = \overline{1, n}, l = \overline{1, r(k, m)}, \chi_A(s)$ is indicator of a set A, approximate the function $\psi_k(s)$ in $L_2(\mathbb{R})$:

$$\int_{-\infty}^{\infty} |\psi_k(s) - \psi_k^{(m)}(s)|^2 ds \to 0, \quad \text{as } m \to \infty.$$

Then the sequences of random variables

$$\varepsilon_k^{(m)} = \sum_{l=1}^{r(k,m)} c_{kl}^{(m)}\left(\xi(\beta_{kl}^{(m)}) - \xi(\alpha_{kl}^{(m)})\right) = \int_{-\infty}^{\infty} \psi_k^{(m)}(s)d\xi(s)$$

mean-square converge to $\varepsilon(t_k)$ in $L_2(\Omega)$:

$$E\left|\varepsilon(t_k) - \varepsilon_k^{(m)}\right|^2 \to 0, \quad k = \overline{1, n}, \quad \text{as } m \to \infty. \tag{26}$$

For any fixed m, the random vector with coordinates $\varepsilon_k^{(m)}, k = \overline{1, n}$ is s.s.-G. Indeed,

$$\sum_{k=1}^{m} \Delta_k \varepsilon_k^{(m)} = \sum_{k=1}^{m} \Delta_k \sum_{l=1}^{r(k,m)} c_{kl}^{(m)}\left(\xi(\beta_{kl}^{(m)}) - \xi(\alpha_{kl}^{(m)})\right)$$

$$= \sum_{k'=1}^{n'(m)} u_{k'}^{(m)} \xi(\eta_{k'}^{(m)}),$$

where $u_{k'}^{(m)}$ are real numbers and $\eta_{k'}^{(m)}$ are different real numbers. By condition **N4(i)** the random vector with coordinates $\xi(\eta_{k'}^{(m)})$, $k' = \overline{1, n'(m)}$, is s.s.-G., and therefore,

$$
\begin{aligned}
E \exp\left\{ \sum_{k=1}^{m} \Delta_k \varepsilon_k^{(m)} \right\} &= E \exp\left\{ \sum_{k'=1}^{n'(m)} u_{k'}^{(m)} \xi\left(\eta_{k'}^{(m)}\right) \right\} \\
&\leq \exp\left\{ \frac{1}{2} E \left(\sum_{k'=1}^{n'(m)} u_{k'}^{(m)} \xi\left(\eta_{k'}^{(m)}\right) \right)^2 \right\} \\
&= \exp\left\{ \frac{1}{2} \left(\sum_{k=1}^{m} \Delta_k \varepsilon_k^{(m)} \right)^2 \right\}.
\end{aligned}
\tag{27}
$$

From (26) it follows that $\varepsilon_k^{(m)} \xrightarrow{P} \varepsilon_{t_k}$, $k = \overline{1, n}$, as $m \to \infty$, and thus there exists some subsequence of indexes $m' \to \infty$, independent of k, such that $\varepsilon_k^{(m')} \to \varepsilon_{t_k}$ a.s., $k = \overline{1, n}$, as $m' \to \infty$.

Finally, by the Fatou lemma and (27)

$$
\begin{aligned}
E \exp\left\{ \sum_{k=1}^{n} \Delta_k \varepsilon(t_k) \right\} &= E \lim_{m' \to \infty} \exp\left\{ \sum_{k=1}^{n} \Delta_k \varepsilon_k^{(m')} \right\} \\
&\leq \liminf_{m' \to \infty} E \exp\left\{ \sum_{k=1}^{n} \Delta_k \varepsilon_k^{(m')} \right\} \\
&\leq \lim_{m' \to \infty} \exp\left\{ \frac{1}{2} E \left(\sum_{k=1}^{n} \Delta_k \varepsilon_k^{(m')} \right)^2 \right\} \\
&= \exp\left\{ \frac{1}{2} E \left(\sum_{k=1}^{n} \Delta_k \varepsilon(t_k) \right)^2 \right\} \\
&= \exp\left\{ \frac{1}{2} \sum_{i,j=1}^{n} B(t_i - t_j) \Delta_i \Delta_j \right\}.
\end{aligned}
$$

So, we have obtained (25). □

Corollary 3. *If conditions* **N1**, **N4(i)**, **N4(ii)**, **R1** *and* **R2** *are fulfilled, then the conclusion of Theorem 1 is true with $d_0 = 2\pi f_0$.*

Corollary 4. *If conditions* **N1**, **N4(i)**, **N4(ii)** *and* **R3** *are fulfilled, then the conclusion of Theorem 2 is true with a constant b satisfying (23).*

Assume

$$
\liminf_{T \to \infty} T^{-\frac{1}{2}} s_{iT} > 0, \quad i = \overline{1, q}.
\tag{28}
$$

Corollary 5. *Under conditions of Theorem 2 or Corollaries 2, 4, and (28) for any $\rho > 0$, $\nu \in [0, \frac{1}{2})$, and $T > T_0$*

$$
P\left\{ \left\| T^{-\frac{1}{2}} s_T (\theta_T - \theta) \right\| \geq \rho T^{-\nu} \right\} \leq B \exp\left\{ -b\rho T^{1-2\nu} \right\}.
\tag{29}
$$

Proof. To show (29) it is sufficient to take $R = \rho T^{\frac{1}{2}-\nu}$ in (21). $\qquad\square$

For $\nu = 0$ we arrive at a quite strong result on the weak consistency of l.s.e. Similarly, in the conditions of Corollary 5 the following result on probabilities of moderate deviations for l.s.e. holds: for any $h > 0, T > T_0$

$$P\left\{\left\|s_T(\theta_T - \theta)\right\| \geq h \ln^{\frac{1}{2}} T\right\} \leq BT^{-bh^2}.$$

Obviously, Gaussian stochastic processes ε are j.s.s.-G. ones, and all the previous results are valid for them.

4   Two examples

In this section, we consider an example of a regression function satisfying the condition **R3** and an example of the j.s.s.-G. process ξ from expression (24) in condition **N4(i)**.

Example 4.1. Suppose

$$a(t, \tau) = \exp\{\langle \tau, y(t)\rangle\},$$

where $\langle \tau, y(t)\rangle = \sum_{i=1}^{q} \tau_i y_i(t)$, regressors $y(t) = (y_1(t), \ldots, y_q(t))'$, $t \geq 0$, take values in a compact set $Y \subset \mathbb{R}^q$.

Let us assume

$$J_T = \left(T^{-1}\int\limits_0^T y_i(t)y_j(t)dt\right)_{i,j=1}^{q} \quad \rightarrow \quad J = (J_{ij})_{i,j=1}^{q}, \quad \text{as } T \to \infty, \qquad (30)$$

J is a positive definite matrix. In this case the regression function $a(t, \tau)$ satisfies condition **R3**. Indeed, let

$$H = \max_{y\in Y,\ \tau\in\Theta^c} \exp\{\langle y, \tau\rangle\}, \qquad L = \min_{y\in Y,\ \tau\in\Theta^c} \exp\{\langle y, \tau\rangle\}$$

Then for any $\delta > 0$ and $T > T_0$

$$L^2(J_{ii} - \delta) < T^{-1}d_{iT}^2(\theta) < H^2(J_{ii} + \delta), \quad i = \overline{1, q},$$

and according to (19) we can take $s_T = T^{\frac{1}{2}}\mathbb{I}_q$ with the identity matrix $\mathbb{I}_q$ of order q.

For a fixed t

$$\exp\{\langle y(t), \theta + T^{-\frac{1}{2}}u\rangle\} - \exp\{\langle y(t), \theta + T^{-\frac{1}{2}}v\rangle\}$$

$$= T^{-\frac{1}{2}}\sum_{i=1}^{q} y_i(t)\exp\{\langle y(t), \theta + T^{-\frac{1}{2}}(u + \eta(v - u))\rangle\}(u_i - v_i), \quad \eta \in (0, 1),$$

and therefore for any $\delta > 0$ and $T > T_0$

$$\Phi_T(u, v) \leq H^2\left(T^{-1}\int \|y(t)\|^2 dt\right)\|u - v\|^2 < H^2(\operatorname{Tr} J + \delta)\|u - v\|^2,$$

So we obtain the right-hand side of (20) with the constant $c_1 > H^2 \operatorname{Tr} J$.

On the other hand, for a fixed t

$$
\begin{aligned}
\big(\Delta(t, u) &- \Delta(t, v)\big)^2 \\
&= \big(\exp\{\langle y(t), \theta + T^{-\frac{1}{2}}u\rangle\} - \exp\{\langle y(t), \theta + T^{-\frac{1}{2}}v\rangle\}\big)^2 \\
&= \exp\{2\langle y(t), \theta + T^{-\frac{1}{2}}v\rangle\}\big(\exp\{\langle y(t), T^{-\frac{1}{2}}(u - v)\rangle\} - 1\big)^2.
\end{aligned}
$$

Since $(e^x - 1)^2 \geq x^2$, $x \geq 0$, and $(e^x - 1)^2 \geq e^{2x}x^2$, $x < 0$, it holds

$$
\big(\exp\{\langle y(t), T^{-\frac{1}{2}}(u - v)\rangle\} - 1\big)^2 \geq L_t T^{-1}\langle y(t), u - v\rangle^2,
$$

with $L_t = \min\{1, \exp\{2\langle y(t), T^{-\frac{1}{2}}(u - v)\rangle\}\}$, and

$$
\begin{aligned}
\big(\Delta(t, u) &- \Delta(t, v)\big)^2 \\
&\geq \min\big\{\exp\{2\langle y(t), \theta + T^{-\frac{1}{2}}v\rangle\}, \ \exp\{2\langle y(t), \theta + T^{-\frac{1}{2}}u\rangle\}\big\}T^{-1}\langle y(t), u - v\rangle^2 \\
&> L^2 T^{-1}\langle y(t), u - v\rangle^2.
\end{aligned}
$$

Thus for any $\delta > 0$ and $T > T_0$

$$
\Phi_T(u, v) \geq L^2\langle J_T(u - v), u - v\rangle > L^2\big(\lambda_{\min}(J) - \delta\big)\|u - v\|^2,
$$

and we have obtained the left-hand side of (20) with the constant $c_0 < L^2\lambda_{\min}(J)$, where $\lambda_{\min}(J)$ is the least eigenvalue of the positive definite matrix J.

The next fact is a reformulation of Corollary 4 for the regression function $a(t, \tau)$ of our example.

Corollary 3.4′. *Under conditions* **N1**, **N4(i)**, **N4(ii)** *and* (30) *there exist constants* B, $b > 0$ *such that for* $T > T_0$, $R > R_0$

$$
P\big\{\big\|T^{\frac{1}{2}}(\theta_T - \theta)\big\| \geq R\big\} \leq B \exp\{-bR^2\}.
$$

Moreover for any $\beta > 0$ *the constant* B *can be chosen so that*

$$
b \geq \frac{L^2\lambda_{\min}(J)}{16\pi f_0(1 + q)} - \beta.
$$

Example 4.2. Here we offer an example of the j.s.s.-G. stochastic process ξ with orthogonal increments in the formula (24) using the Ito–Nicio series (see [7] and references therein).

Consider any orthonormal basis $\{\varphi_k, \ k \geq 1\}$ in $L_2(\mathbb{R}_+)$ and a sequence $\{Z_k, \ k \geq 1\}$ of independent $N(0, 1)$ r.v.'s. Then

$$
w_0(t) = \sum_{k=1}^{\infty} Z_k \int_0^t \varphi_k(u)du, \quad t \geq 0,
$$

is a standard Wiener process with covariances $E w_0(t) w_0(s) = \min\{t, s\}$. We need some kind of the Wiener process on the real line $\mathbb{R}$. Let $\{w_1(t),\ t \geq 0\}$, $\{w_2(t),\ t \geq 0\}$ be two independent Wiener processes of the following form:

$$w_i(t) = \sum_{k=1}^{\infty} Z_{ik} \int_0^t \varphi_k(u) du, \quad t \geq 0,\ i = 1, 2,$$

and $\{Z_{ik},\ k \geq 1,\ i = 1, 2\}$ be independent $N(0, 1)$ r.v.'s. Then the required Wiener process on $\mathbb{R}$ can be defined as $w(t) = w_1(t),\ t \geq 0$, and $w(t) = w_2(|t|),\ t < 0$. For any real $t_1 < t_2 \leq t_3 < t_4$

$$E\big(w(t_2) - w(t_1)\big)\big(w(t_4) - w(t_3)\big) = 0, \tag{31}$$

i.e. increments are orthogonal. On the other hand, for any $t > s$

$$E\big(w(t) - w(s)\big)^2 = t - s.$$

Let $\{\xi_{ik},\ k \geq 1,\ i = 1, 2\}$ be i.i.d. s.s.-G. r.v.'s (and non-Gaussian!) with unit variance. Some examples of s.s.-G. r.v.'s can be found in [1]. Thus the Bernoulli r.v. and the r.v. uniformly distributed in $[-\sqrt{3},\ \sqrt{3}]$ are s.s.-G. and have unit variances.

Let us introduce the stochastic processes

$$\xi_i(t) = \sum_{k=1}^{\infty} \xi_{ik} \int_0^t \varphi_k(u) du, \quad t \geq 0,\ i = 1, 2,$$

$\xi(t) = \xi_1(t),\ t \geq 0$, and $\xi(t) = \xi_2(|t|),\ t < 0$. Then $\xi = \{\xi(t),\ t \in \mathbb{R}\}$ is a process with orthogonal increments and is not a Gaussian one.

However, it is a j.s.s.-G. process. To prove this statement consider arbitrary numbers $t_1 < \cdots < t_n$, where the first m numbers, $0 \leq m \leq n$, are negative and the rest $n - m$ numbers are positive. Let $\Delta = (\Delta_1, \ldots, \Delta_n)' \in \mathbb{R}^n$ be any vector. Then

$$\Sigma_2 = \sum_{k=1}^{N} \xi_{2k} \left(\sum_{i=1}^{m} \Delta_i \int_0^{|t_i|} \varphi_k(u) du \right) \to \sum_{i=1}^{m} \Delta_i \xi_2\big(|t_i|\big) \text{ a.s.}, \quad \text{as } N \to \infty,$$

$$\Sigma_1 = \sum_{k=1}^{N} \xi_{1k} \left(\sum_{i=m+1}^{n} \Delta_i \int_0^{t_i} \varphi_k(u) du \right) \to \sum_{i=m+1}^{n} \Delta_i \xi_1(t_i) \text{ a.s.}, \quad \text{as } N \to \infty.$$

By the Fatou lemma

$$E \exp\left\{ \sum_{i=1}^{n} \Delta_i \xi(t_i) \right\} = E \lim_{N \to \infty} \exp\{\Sigma_2 + \Sigma_1\}$$

$$\leq \liminf_{N \to \infty} \prod_{k=1}^{N} E \exp\left\{ \xi_{2k} \left(\sum_{i=1}^{m} \Delta_i \int_0^{|t_i|} \varphi_k(u) du \right) \right\}.$$

$$\cdot \prod_{k=1}^{N} E \exp\left\{\xi_{1k}\left(\sum_{i=m+1}^{n} \Delta_i \int_0^{t_i} \varphi_k(u)du\right)\right\} \leq$$

$$\leq \lim_{N\to\infty} \exp\left\{\frac{1}{2}\left(\sum_{k=1}^{N}\left(\sum_{i=1}^{m} \Delta_i \int_0^{|t_i|} \varphi_k(u)du\right)^2\right.\right.$$

$$\left.\left.+ \sum_{k=1}^{N}\left(\sum_{i=m+1}^{n} \Delta_i \int_0^{t_i} \varphi_k(u)du\right)^2\right)\right\}$$

$$= \lim_{N\to\infty} \exp\left\{\frac{1}{2}\left(\sum_{i,j=1}^{m}\left(\sum_{k=1}^{N} \int_0^{|t_i|} \varphi_k(u)du \int_0^{|t_j|} \varphi_k(u)du\right)\Delta_i \Delta_j\right.\right.$$

$$\left.\left.+ \sum_{i,j=m+1}^{n}\left(\sum_{k=1}^{N} \int_0^{t_i} \varphi_k(u)du \int_0^{t_j} \varphi_k(u)du\right)\Delta_i \Delta_j\right)\right\}.$$

By Parseval's identity

$$\lim_{N\to\infty} \sum_{k=1}^{N} \int_0^{|t_i|} \varphi_k(u)du \int_0^{|t_j|} \varphi_k(u)du$$

$$= \sum_{k=1}^{\infty} \int_0^{\infty} \chi_{[0,\,|t_i|]}(u)\varphi_k(u)du \int_0^{\infty} \chi_{[0,\,|t_j|]}(u)\varphi_k(u)du$$

$$= \int_0^{\infty} \chi_{[0,\,|t_i|]}(u)\chi_{[0,\,|t_j|]}(u)du = \min\{|t_i|,\,|t_j|\}.$$

Similarly

$$\lim_{N\to\infty} \sum_{k=1}^{N} \int_0^{t_i} \varphi_k(u)du \int_0^{t_j} \varphi_k(u)du = \min\{t_i,\,t_j\}.$$

It means that

$$E \exp\left\{\sum_{i=1}^{n} \Delta_i \xi(t_i)\right\} \leq \exp\left\{\frac{1}{2}\langle B\Delta,\,\Delta\rangle\right\}$$

with

$$B = \begin{pmatrix} B_2 & 0 \\ 0 & B_1 \end{pmatrix},$$

where $B_2 = (\min\{|t_i|,\,|t_j|\})_{i,j=1}^{m}$, $B_1 = (\min\{t_i,\,t_j\})_{i,j=m+1}^{n}$, and the process ξ is j.s.s.-G.

References

[1] Buldygin, V.V., Kozachenko, Y.V.: Metric Characterization of Random Variables and Random Processes. AMS, Providence (2000) MR1743716

[2] Feller, W.: An Introduction to Probability Theory and Its Applications vol. 2, 2nd edn. Willey, New York (1957) MR0088081

[3] Gikhman, I.I., Skorokhod, A.V.: Introduction to the Theory of Random Processes, 1st English edn. Sounders (W.B.) Co Ltd (1969) MR0247660

[4] Hu, S.H.: A large deviation result for the least squares estimators in nonlinear regression. Stochastic Process and their Applications **47**, 345–352 (1993). MR1239845. https://doi.org/10.1016/0304-4149(93)90022-V

[5] Huang, X., Tang, X., Deng, X., Wang, X.: The large deviation for the least squares estimator of nonlinear regression model based on WOD errors. J. Inequal. Appl. **125** (2016). MR3489878. https://doi.org/10.1186/s13660-016-1064-6

[6] Ibragimov, I.A., Has'minskii, R.Z.: Statistical Estmation: Asymptptic Theory. Springer, New York (1981) MR0620321

[7] Ito, K., Nisio, M.: On the convergence of sums of independent Banach space valued random variables. Osaka J. of Math **5**(1), 35–48 (1968) MR0235593

[8] Ivanov, A.V.: An asymtotic expansion for the distribution of the least squares estimator of the non-linear regression parameter. Theory. Probab. Appl. **21**(3), 557–570 (1977)

[9] Ivanov, A.V.: Asymptotic Theory of Nonlinear Regression. Kluwer Academic Publishers, Dordecht/Boston/London (1997). MR1472234. https://doi.org/10.1007/978-94-015-8877-5

[10] Ivanov, A.V.: Large deviations of regression parameter estimate in the models with stationary sub-Gaussian noise. Theor. Probability and Math. Statist. **95**, 99–108 (2017). MR3631646. https://doi.org/10.1090/tpms/1024

[11] Ivanov, A.V., Leonenko, N.N.: Statistical Analysis of Random Fields. Kluwer Academic Publishers, Dordecht/Boston/London (1989). MR1009786. https://doi.org/10.1007/978-94-009-1183-3

[12] Pfanzagl, J.: On the measurability and consistency of minimum contrast estimates. Metrika **14**, 249–272 (1969)

[13] Rao, B.L.S.P.: On the exponential rate convergence of the least squares estimator in the nonlinear regression model with Gaussian errors. Statist. Probab. Lett. **2**, 139–142 (1984). MR0747613. https://doi.org/10.1016/0167-7152(84)90004-X

[14] Rao, B.L.S.P.: The rate of convergence for the least squares estimator in a non-linear regression model with dependent errors. J. Multivariate Analysis **14**(3), 315–322 (1984). MR0747260. https://doi.org/10.1016/0047-259X(84)90036-8

[15] Sieders, A., Dzhaparidze, K.O.: A large deviation result for parameter estimators and its application to nonlinear regression analysis. Ann. Statist. **15**(3), 1031–1049 (1987). MR0902244. https://doi.org/10.1214/aos/1176350491

[16] Yang, W.Z., Hu, S.H.: Large deviation for a least squares estimator in a nonlinear regression model. Stat. Probab. Lett. **91**, 135–144 (2014). MR3208127. https://doi.org/10.1016/j.spl.2014.04.022