Теорія ймовір. та матем. статист.
Вип. 95, 2016, стор. 92–100

Teor. Ĭmovir. ta Matem. Statyst.
No. 95, 2016, pp. 92–100

LARGE DEVIATIONS OF REGRESSION PARAMETER ESTIMATE IN THE MODELS WITH STATIONARY SUB-GAUSSIAN NOISE

UDC 519.21

A.V. IVANOV

Abstract. Exponential bounds for probabilities of large deviations of nonlinear regression parameter least squares estimate in the models with jointly strictly sub-Gaussian random noise are obtained.

1. Introduction

Probabilities of large deviations of the normed least squares estimate (l.s.e.) of unknown nonlinear regression parameter have been discussed earlier in statistical literature. So, in [1] it was proved a statement on the l.s.e. probabilities of large deviations with power decreasing rate of scalar parameter in nonlinear regression model with i.i.d. observation errors having moments of finite order, and in [2] a similar result was obtained with exponential decreasing rate in Gaussian nonlinear regression.

In the paper [3] a general Theorem 2.1 on probabilities of large deviations for M-estimates based on a data set of any structure was proved that generalizes the result of monograph [4] with application to l.s.e. of nonlinear regression parameters with pre-Gaussian and sub-Gaussian i.i.d. observation errors (Theorems 3.1 and 3.2 in [3]). Some theorems in this direction are proved in [5] also. The results on l.s.e. probabilities of large deviations in nonlinear regression models with correlated observations one can find in [6]–[10].

Suppose a random sequence

$$X_t = a_t(\theta) + \varepsilon_t, \quad t \geq 1, \tag{1}$$

is observed, where $a_t(\theta)$, $\theta = (\theta_1, \ldots, \theta_q) \in \Theta^c \subset \mathbb{R}^q$, $t \geq 1$ are continuous functions, true parameter value θ belongs to bounded open convex set Θ, $\varepsilon = \{\varepsilon_t, t \in \mathbb{Z}\}$ is a time series defined on probability space $(\Omega, \mathfrak{F}, P)$, $\mathsf{E}\,\varepsilon_t = 0$. We will write $\sum = \sum_{t=1}^{T}$.

Definition 1. Any random vector $\widehat{\theta}_T = (\widehat{\theta}_{1T}, \ldots, \widehat{\theta}_{qT}) \in \Theta^c$ having the property

$$Q_T(\widehat{\theta}_T) = \min_{\tau \in \Theta^c} Q_T(\tau), \quad Q_T(\tau) = \sum (X_t - a_t(\tau))^2$$

is called the l.s.e. of unknown parameter θ obtained by observations X_t, $t = 1, \ldots, T$.

Under suppositions introduced above there exists at least one such a random vector $\widehat{\theta}_T$ [11].

In asymptotic theory of nonlinear regression in the problem of l.s.e. distribution normal approximation we norm the difference $\widehat{\theta_T} - \theta$ by diagonal matrix [5, 6]

$$d_T(\theta) = diag(d_{iT}(\theta),\ i = \overline{1,q}), \quad d_{iT}^2(\theta) = \sum \left(\frac{\partial}{\partial \theta_i} a_t(\theta) \right)^2 .$$

2000 *Mathematics Subject Classification.* Primary 60G50, 65B10, 60G15; Secondary 40A05.

Key words and phrases. Large deviations, least squares estimate, nonlinear regression, discrete white sub-Gaussian noise, spectral density.

Further it is assumed that functions $a_t(\tau)$, $t \geq 1$, are continuously differentiable in $\tau \in \Theta$.

The paper is organized in the following way. In the Section 2 a bound is obtained for large deviations of $d_T(\theta)(\widehat{\theta}_T - \theta)$ in the regression model (1) with dependent, strictly sub-Gaussian errors ε_t. In the Section 3 the results of Section 2 are applied to ARMA processes with strictly sub-Gaussian innovations. The Section 4 contains an example of regression function satisfying the conditions of our theorems.

2. Large deviations in the presence of jointly strictly sub-Gaussian noise

The next concepts one can find in [12].

Definition 2. Random vector $\eta = (\eta_1, \dots, \eta_n)' \in \mathbb{R}^n$ is said to be strictly sub-Gaussian, if for any $\Delta = (\Delta_1, \dots, \Delta_n)' \in \mathbb{R}^n$

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

where $\langle \eta, \Delta\rangle = \sum_{i=1}^n \eta_i \Delta_i$, $\mathsf{B} = (\mathsf{B}(i,j))_{i,j=1}^n$ is the covariance matrix of η, that is $\mathsf{B}(i,j) = \mathsf{E}\,\eta_i\eta_j$, $i,j = 1\dots,n$, $\langle \mathsf{B}\Delta, \Delta\rangle = \sum_{i,j=1}^n \mathsf{B}(i,j)\Delta_i\Delta_j$.

Definition 3. Time series $\varepsilon = \{\varepsilon_t, t \in \mathbb{Z}\}$ is said to be jointly strictly sub-Gaussian, if for any $n \geq 1$, and any $t_1, \dots, t_n \in \mathbb{Z}$ random vector $\varepsilon(n) = (\varepsilon_{t_1}, \dots, \varepsilon_{t_n})'$ is strictly sub-Gaussian.

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

N.1.(i) Time series ε is jointly strictly sub-Gaussian, $\mathsf{B}(t,s) = \mathsf{E}\,\varepsilon_t\varepsilon_s$, $t,s \in \mathbb{Z}$.

(ii) For any $T \geq 1$, $\Delta = (\Delta_1, \dots, \Delta_T) \in \mathbb{R}^T$,

$$\langle \mathsf{B}\Delta, \Delta\rangle \leq d_0\|\Delta\|^2 \tag{2}$$

for some constant $d_0 > 0$, $\|\Delta\| = \left(\sum \Delta_t^2\right)^{1/2}$, and $\mathsf{B} = (B(t,s))_{t,s=1}^T$.

For fixed T the exact bound in (2) is

$$\langle \mathsf{B}\Delta, \Delta\rangle \leq \lambda_{max}(T)\|\Delta\|^2,$$

where $\lambda_{max}(T)$ is the maximal eigenvalue of symmetric positive semi-definite matrix B (the norm of selfajoint positive semi-definite operator B in $\mathbb{R}^T$). Note that $\lambda_{max}(T)$ is monotonically nondecreasing number sequence, so there exists

$$\lim_{T\to\infty} \lambda_{max}(T) \leq d_0 < \infty.$$

Below we give some examples of constant d_0.

Further in this paper the following statement on exponential bound of the weighted sums of jointly strictly sub-Gaussian random variables (r.v.-s) distributions tails plays an important role. Write $S_T = \sum \Delta_t\varepsilon_t$.

Lemma 1. *Under condition* **N.1** *for any* $x > 0$

$$P\{S_T \geq x\} \leq G_T(x),\ P\{S_T \leq -x\} \leq G_T(x), \tag{3}$$

$$P\{|S_T| \geq x\} \leq 2G_T(x), \tag{4}$$

where

$$G_T(x) = \exp\left\{-\frac{x^2}{2d_0\|\Delta\|^2}\right\}. \tag{5}$$

Proof. Proof is obvious (see, for example, [12]). For any $x > 0$, $\lambda > 0$ by Chebyshev-Markov inequality and (2)

$$P\{S_T \geq x\} \leq \exp\{-\lambda x\}\exp\left\{\frac{\lambda^2}{2}\langle \mathsf{B}\Delta, \Delta\rangle\right\} \leq \exp\left\{\frac{1}{2}\lambda^2 d_0\|\Delta\|^2 - \lambda x\right\}. \tag{6}$$

Minimization of the right hand side of (6) in λ proves the 1st inequality in (3). The proof of the 2-nd inequality in (3) is the same. The inequality (4) follow from (3). □

To formulate conditions on regression function $a_t(\tau)$ in the spirit of [3] (see also [5] and [6]) we need in some notation. Write $U_T(\theta) = d_T(\theta)(\Theta^c - \theta)$, $\Gamma_{T,\theta,R} = U_T(\theta) \cap \{u\colon R \leq \|u\| \leq R+1\}$, $u = (u_1, \dots, u_q) \in \mathbb{R}^q$. Denote by **G** the family of all functions $g = g_T(R)$, $T \geq 1$, $R > 0$, having the following properties:

1) for fixed T $g_T(R) \uparrow \infty$, as $R \to \infty$;

2) for any $r > 0$

$$\lim_{R\to\infty, T\to\infty} R^r \exp\{-g_T(R)\} = 0.$$

Let $\gamma(R)$ be, generally speaking, different polynomials of R with coefficients that do not depend on values T, R, θ, u, v, where $\gamma(R)$ appear. Set also $\Delta_t(u) = a_t(\theta + d_T^{-1}(\theta)u) - a_t(\theta),\ t = \overline{1,T}$,

$$\Phi_T(u,v) = \sum \left(\Delta_t(u) - \Delta_t(v)\right)^2, \ u, v \in U_T(\theta).$$

Assume the existence of function $g \in \mathbf{G}$, constants $\delta \in \left(0, \frac{1}{2}\right)$, $\varkappa > 0$, $\rho \in (0,1]$, polynomials $\gamma(R)$ such that for sufficiently large T, R (we will write $T > T_0$, $R > R_0$) the next conditions are fulfilled.

A.1. (i) For any $u, v \in \Gamma_{T,\theta,R}$ such that $\|u - v\| \leq \varkappa$

$$\Phi_T(u,v) \leq \|u-v\|^{2\rho}\gamma(R). \tag{7}$$

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

A.2. For any $u \in \Gamma_{T,\theta,R}$

$$\Phi_T(u,0) \geq 2d_0\delta^{-2}g_T(R). \tag{8}$$

Theorem 1. *Under conditions* **N.1**, **A.1** *and* **A.2** *there exist constants* $B_0, b_0 > 0$ *such that for* $T > T_0$, $R > R_0$

$$P\left\{\|d_T(\theta)(\widehat{\theta}_T - \theta)\| \geq R\right\} \leq B_0 \exp\left\{-b_0 g_T(R)\right\}, \tag{9}$$

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

$$b_0 \geq \frac{\rho}{\rho + q} - \beta. \tag{10}$$

Proof. It is sufficient to check the fulfilment of assumptions (M1) and (M2) of the mentioned Theorem 2.1 from [3]. Inequalities (11) and (17) below are just these assumptions reformulated in the manner similar to the used in the proof of Theorem 3.1 in [3]. Set $S_T(u) = \sum \Delta_t(u)\varepsilon_t$, $\zeta_T(u) = S_T(u) - \frac{1}{2}\Phi_T(u,0)$. Following the line of the Theorem 3.1 [3] proof we will derive for any $m > 0$ and $u, v \in \Gamma_{T,\theta,R}$ the inequality

$$\mathsf{E}\left|\zeta_T(u) - \zeta_T(v)\right|^m \leq \|u-v\|^{\rho m}\gamma(R). \tag{11}$$

We have

$$\begin{gathered}\mathsf{E}\left|\zeta_T(u) - \zeta_T(v)\right|^m \leq \\ \max(1, 2^{m-1})\left(\mathsf{E}\left|S_T(u) - S_T(v)\right|^m + 2^{-m}\left|\Phi_T(u,0) - \Phi_T(v,0)\right|^m\right),\end{gathered} \tag{12}$$

$$\begin{gathered}\left|\Phi_T(u,0) - \Phi_T(v,0)\right| \leq \sum\left|\Delta_t(u) - \Delta_t(v)\right|\left|\Delta_t(u) + \Delta_t(v)\right| \leq \\ 2^{1/2}\Phi_T^{1/2}(u,v)\left(\Phi_T^{1/2}(u,0) + \Phi_T^{1/2}(v,0)\right) \leq 2^{3/2}\|u-v\|^{\rho}\left(\gamma(R)\right)^{1/2}\left(\gamma(R)\right)^{1/2} \leq \\ 2^{1/2}\|u-v\|^{\rho}(\gamma(R) + \gamma(R)) = \|u-v\|^{\rho}\gamma(R)\end{gathered}$$

according to **A.1** (polynomials $\gamma(R)$ are different!). Thus we obtained the bound

$$\left|\Phi_T(u,0) - \Phi_T(v,0)\right|^m \leq \|u-v\|^{\rho m}\gamma(R). \tag{13}$$

On the other hand by the formula for the moments of nonnegative r.v.-s (see, for example, [13], and compare with [8]) and Lemma 1 being applied to $\Delta_t = \Delta_t(u) - \Delta_t(v)$, $t = 1, \dots, T$, $S_T = S_T(u,v) = S_T(u) - S_T(v)$,

$$\mathsf{E}\,|S_T(u,v)|^m = m \int_0^\infty x^{m-1} P\left\{|S_T(u,v)| \geq x\right\} dx \leq$$

$$2m \int_0^\infty x^{m-1} \exp\left\{-\frac{x^2}{2d_0 \Phi_T(u,v)}\right\} dx = \sqrt{2\pi} m d_0^{\frac{m}{2}} \Phi_T^{\frac{m}{2}}(u,v)\, \mathsf{E}\,|z|^{m-1}, \tag{14}$$

where z is standard Gaussian r.v. and

$$\mathsf{E}\,|z|^{m-1} = \pi^{-1/2} 2^{\frac{m-1}{2}} \Gamma(\frac{m}{2}), \quad m > 0. \tag{15}$$

Relations (14) and (15) give together with (7) the bound

$$\mathsf{E}\,|S_T(u,v)|^m \leq 2^{\frac{m}{2}} m \Gamma(\frac{m}{2}) d_0^{\frac{m}{2}} \Phi_T^{\frac{m}{2}}(u,v) \leq \|u - v\|^{\rho m} \gamma(R). \tag{16}$$

From (12), (13) and (16) it follows (11).

To accomplish the proof we have to apply the 1st inequality in (3) of Lemma 1 for $\Delta_t = \Delta_t(u)$ and $x = \delta \Phi_T(u,0)$. Then from **A.2** we obtain

$$P\left\{S_T(u) \geq \delta \Phi_T(u,0)\right\} \leq \exp\left\{-\frac{\delta^2}{2d_0} \Phi_T(u,0)\right\} \leq \exp\{-g_T(R)\}. \tag{17}$$

As it follows from (11) and (17) the theorem is proved. □

The next condition and Theorem 2 one can consider as a simplification of the conditions **A.1**, **A.2** of Theorem 1. Theorem 2 generalizes Theorem 3.2 from [3].

A.3. There exist numbers $0 < c_0(\theta) < c_1(\theta) < \infty$ such that for any $u, v \in U_T(\theta)$ and $T > T_0$

$$c_0(\theta) \|u - v\|^2 \leq \Phi_T(u,v) \leq c_1(\theta) \|u - v\|^2. \tag{18}$$

The condition of the type (18) has been introduced in [1] and used in [2, 3, 8] and other works.

Theorem 2. *Under conditions* **N.1** *and* **A.3** *there exist constants B_0 and b such that for $T > T_0$, $R > R_0$*

$$P\left\{\left\|d_T(\theta)\left(\widehat{\theta}_T - \theta\right)\right\| \geq R\right\} \leq B_0 \exp\{-bR^2\}, \tag{19}$$

and what's more for any $\beta > 0$ constant B_0 can be chosen so that

$$b \geq \frac{c_0(\theta)}{8d_0(1+q)} - \beta. \tag{20}$$

Proof. We shall verify the fulfilment of conditions **A.1** and **A.2**. Then the conclusion of the theorem will follow from Theorem 1.

Inequality (7) of the condition **A.1**(i) follows from the right hand side of inequality (18), if we will take in (7) $\rho = 1$, $\gamma(R) = c_1(\theta)$. Inequality of the condition **A.1**(ii) follows as well from the right hand side of (18), if we will take $v = 0$, $\gamma(R) = c_1(\theta)(R+1)^2$.

To check the fulfilment of condition **A.2** we will rewrite the left hand side of (18) for $v = 0$:

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

that is in the inequality (8) on can take

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

In this case in (9) the exponent $-b_0 g_T(R) = -\left(\frac{1}{2}\delta^2 b_0 d_0^{-1} c_0(\theta)\right) R^2$. Since now in (10) $\rho = 1$, then for any $\beta > 0$ in (19) one can choose

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

We get inequality (20) when $\delta \to \frac{1}{2}$. □

3. Partial cases of jointly strictly sub-Gaussian noise

We use below the partial case of definition 2 for n=1.

Definition 4. A r.v. η is said to be strictly sub-Gaussian if for any $\Delta \in \mathbb{R}$

$$\mathsf{E}\exp\{\Delta\eta\} \leq \exp\left\{\frac{1}{2}\sigma_\eta^2\Delta^2\right\}$$

with $\sigma_\eta^2 = \mathsf{E}\,\eta^2$.

Let $\{\xi_j, j \in \mathbb{Z}\}$ be a sequence of i.i.d. strictly sub-Gaussian r.v.-s. It is naturally to call such a sequence (discrete) white sub-Gaussian noise.

N.2. Random noise ε in the model (1) is of the form

$$\varepsilon_t = \sum_{j=-\infty}^{\infty} \Psi_{tj}\xi_j, \ t \in \mathbb{Z}, \tag{21}$$

where $\{\xi_j, j \in \mathbb{Z}\}$ is a white sub-Gaussian noise ($\mathsf{E}\,\xi_j = 0$, $\mathsf{E}\,\xi_j^2 = \sigma_\xi^2$), and

$$\|\Psi\|_2^2 = \sum_{t=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} \Psi_{tj}^2 < \infty.$$

It follows from the condition **N.2** that series (21) converges almost sure (see, for example, [14]), and time series $\varepsilon = \{\varepsilon_t, t \in \mathbb{Z}\}$ is determined almost sure. Covariance function of ε is

$$B(t,s) = \sigma_\xi^2 \sum_{j=-\infty}^{\infty} \Psi_{tj}\Psi_{sj}, \quad t,s \in \mathbb{Z}.$$

Lemma 2. *Under condition* **N.2** *ε is jointly strictly sub-Gaussian time series.*

Proof. For vector $\varepsilon(n) = (\varepsilon_{t1}, \dots, \varepsilon_{tn})'$, $\Delta = (\Delta_1, \dots, \Delta_n)' \in \mathbb{R}^n$ from definitions 2 and 3

$$\langle \varepsilon(n), \Delta\rangle = \sum_{k=1}^{n} \varepsilon_{t_k}\Delta_k = \sum_{j=-\infty}^{\infty}\left(\sum_{k=1}^{n}\Delta_k\Psi_{t_k,j}\right)\xi_j.$$

Using independence of r.v.-s ξ_j we obtain then

$$\mathsf{E}\exp\{\langle\varepsilon(n),\Delta\rangle\} = \mathsf{E}\exp\left\{\textstyle\sum_{j=-\infty}^{\infty}\left(\sum_{k=1}^{n}\Delta_k\Psi_{t_k,j}\right)\xi_j\right\}$$
$$\textstyle\prod_{j=-\infty}^{\infty}\mathsf{E}\exp\left\{\left(\sum_{k=1}^{n}\Delta_k\Psi_{t_k,j}\right)\xi_j\right\} \leq \prod_{j=-\infty}^{\infty}\exp\left\{\frac{1}{2}\left(\sum_{k=1}^{n}\Delta_k\Psi_{t_k,j}\right)^2\sigma_\xi^2\right\}$$
$$\exp\left\{\tfrac{1}{2}\sigma_\xi^2\textstyle\sum_{j=-\infty}^{\infty}\left(\sum_{k=1}^{n}\Delta_k\Psi_{t_k,j}\right)^2\right\} = \exp\left\{\tfrac{1}{2}\textstyle\sum_{k,l=1}^{n}\left(\sigma_\xi^2\sum_{j=-\infty}^{\infty}\Psi_{t_k,j}\Psi_{t_l,j}\right)\Delta_k\Delta_l\right\}$$
$$\exp\left\{\tfrac{1}{2}\textstyle\sum_{k,l=1}^{n}B(t_k,t_l)\Delta_k\Delta_l\right\} = \exp\left\{\tfrac{1}{2}\langle B\Delta,\Delta\rangle\right\}$$

with $B = (B(t_k,t_l))_{k,l=1}^n$. □

Lemma 3. *If condition* **N.2** *is fulfilled, then*

$$\langle B\Delta, \Delta\rangle \leq \sigma_\xi^2\|\Psi\|_2^2\|\Delta\|^2.$$

Proof. Really, from the proof of previous lemma it follows

$$\langle B\Delta,\Delta\rangle = \sigma_\xi^2 \sum_{j=-\infty}^{\infty} \left(\sum_{k=1}^{n} \Psi_{t_k,j}\Delta_k\right)^2 \le \sigma_\xi^2 \sum_{j=-\infty}^{\infty} \left(\sum_{k=1}^{n} \Psi_{t_k,j}^2 \sum_{k=1}^{n} \Delta_k^2\right) = \sigma_\xi^2 \sum_{k=1}^{n} \sum_{j=-\infty}^{\infty} \Psi_{t_k,j}^2 \|\Delta\|^2 \le \sigma_\xi^2 \|\Psi\|_2^2 \|\Delta\|^2.$$

□

We will note that when $B = (B(t,s))_{t,s=1}^{T}$, $\Delta = (\Delta_1,\dots,\Delta_T)'$, then the bound of Lemma 3 takes the form

$$\langle B\Delta,\Delta\rangle \le \sigma_\xi^2 \|\widetilde{\Psi}\|_2^2 \|\Delta\|^2,$$

with

$$\|\widetilde{\Psi}\|_2^2 = \sum_{t=1}^{\infty} \sum_{j=-\infty}^{\infty} \Psi_{tj}^2.$$

Thus from condition **N.2** it follows **N.1** with $d_0 = \sigma_\xi^2 \|\widetilde{\Psi}\|_2^2$ in inequality (2).

Corollary 1. *Under conditions* **N.2** *and* **A.2** *with* $d_0 = \sigma_\xi^2 \|\widetilde{\Psi}\|_2^2$ *in* (8) *the conclusion of Theorem 1 is true.*

Corollary 2. *Under conditions* **N.2** *and* **A.3** *the conclusion of Theorem 2 is true with* $d_0 = \sigma_\xi^2 \|\widetilde{\Psi}\|_2^2$ *in the inequality* (20).

Assume that in (21) $\Psi_{tj} = \Psi_{t-j}$, and $\Psi_{t-j} = 0$ when $j > t$. Thus instead of **N.2** we introduce the following condition.

N.3. Errors of observations in regression model (1) have the form

$$\varepsilon_t = \sum_{j=-\infty}^{t} \Psi_{t-j}\xi_j = \sum_{j=0}^{\infty} \Psi_j \xi_{t-j},\ t \in \mathbb{Z}, \tag{22}$$

where $\{\xi_j, j \in \mathbb{Z}\}$ is a white sub-Gaussian noise ($\mathsf{E}\,\xi_j = 0$, $\mathsf{E}\,\xi_j^2 = \sigma_\xi^2$), with

$$\|\Psi\|_{l_2}^2 = \sum_{j=0}^{\infty} \Psi_j^2 < \infty. \tag{23}$$

Condition **N.3** means that ε_t is a reaction of linear homogeneous and physically realizable system on the random impulse sequence $\{\xi_j,\ j \in \mathbb{Z}\}$ (physically realizable filter) [14].

As far as time series ε in condition **N.3** is a partial case of time series ε in condition **N.2**, then the sequence (22) is a jointly strictly sub-Gaussian time-series along Lemma 2. Additionally, as it is well known, (22) is a stationary time series with covariance function

$$B(t) = \sigma_\xi^2 \sum_{j=0}^{\infty} \Psi_j \Psi_{j+|t|},\ t \in \mathbb{Z},$$

and spectral density

$$f(\lambda) = \frac{\sigma_\xi^2}{2\pi} \left| \sum_{j=0}^{\infty} \Psi_j e^{ij\lambda} \right|^2,\ \lambda \in [-\pi,\pi]. \tag{24}$$

N.4. Time series (22) has a bounded spectral density:

$$f_0 = \sup_{\lambda \in [-\pi,\pi]} f(\lambda) < \infty.$$

Under conditions **N.3**, **N.4** by Gerglotz' theorem

$$\langle B\Delta,\Delta\rangle = \sum_{t,s=1}^{T} B(t-s)\Delta_t\Delta_s = \int_{-\pi}^{\pi} f(\lambda) \left|\sum e^{i\lambda t}\Delta_t\right|^2 d\lambda \le f_0 \int_{-\pi}^{\pi} |\sum e^{i\lambda t}\Delta_t|^2 dt = 2\pi f_0 \|\Delta\|^2,$$

and therefore on can take in (2) and (5) $d_0 = 2\pi f_0$.

Corollary 3. *Under conditions* **N.3**, **N.4**, *and* **A.2** *with* $d_0 = 2\pi f_0$ *the statement of Theorem 1 is true.*

Corollary 4. *Under conditions* **N.3**, **N.4**, *and* **A.3** *the statement of Theorem 2 is true with inequality*

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

in the capacity of inequality (20).

ARMA(p,k) processes are important examples of time series (22). These processes are defined by system of recurrence relations (see, for example, [15])

$$\varepsilon_t - a_1\varepsilon_{t-1} - \cdots - a_p\varepsilon_{t-p} = \xi_t + b_1\xi_{t-1} + \cdots + b_k\xi_{t-k},\ t \in \mathbb{Z}, \tag{25}$$

where $\{\xi_t,\ t \in \mathbb{Z}\}$ is white sub-Gaussian noise. If S is backward shift operator, the relations (25) can be rewritten in the form

$$a(S)\varepsilon_t = b(S)\xi_t,$$

where

$$a(z) = 1 - a_1 z - \cdots - a_p z^p,\ b(z) = 1 + b_1 z + \cdots + b_k z^k.$$

If polynomials $a(z)$, $b(z)$ have no joint roots and $a(z) \neq 0$, $b(z) \neq 0$ for $|z| \leq 1$, then $\{\xi_t, t \in \mathbb{Z}\}$ is a stationary ARMA(p,k)-process.

Sometimes it is convenient to rewrite ARMA(p,k)-process as pure moving average process MA(∞) in the form

$$\varepsilon_t = \Psi(S)\xi_t,\ \Psi(z) = \frac{b(z)}{a(z)} = \sum_{j=0}^{\infty} \Psi_j z^j, \tag{26}$$

similarly to (22). If the series (23) converges, then spectral density (24) can be written due (26) as

$$f(\lambda) = \frac{\sigma_\xi^2}{2\pi} \frac{\left|b(e^{i\lambda})\right|^2}{\left|a(e^{i\lambda})\right|^2},\ \lambda \in [-\pi, \pi]. \tag{27}$$

Since polynomial $a(z)$ has no roots on unit circle, then $f(\lambda),\ \lambda \in [-\pi, \pi]$, is continuous. We will write

$$f_{max} = \max_{-\pi \leq \lambda \leq \pi} \frac{\left|b(e^{i\lambda})\right|^2}{\left|a(e^{i\lambda})\right|^2}. \tag{28}$$

N.5. Errors of observations in regression model (1) form ARMA(p,k)-process (25) with white sub-Gaussian noise $\{\xi_t, t \in \mathbb{Z}\}$.

In this case, Corollary 4, for example, can be restated as

Corollary 5. *Under conditions* **N.5** *and* **A.3** *the conclusion of Theorem 2 is true with*

$$b \geq \frac{c_0(\theta)}{8\sigma_\xi^2 f_{max}(1+q)} - \beta.$$

Really, under **N.5** due to (27) and (28) $d_0 = 2\pi f_0 = \sigma_\xi^2 f_{max}$.

Assume

$$\liminf_{T\to\infty} T^{-1/2} d_{iT}(\theta) > 0,\ i = \overline{1,q}. \tag{29}$$

Corollary 6. *Under conditions of Theorem 2 or Corollaries 2, 4, 5, and* (29) *for any* $\rho > 0$, $\nu \in [0, 1/2)$, *and* $T > T_0$

$$P\{\|T^{-1/2} d_T(\theta)(\widehat{\theta}_T - \theta)\| \geq \rho T^{-\nu}\} \leq B_0 \exp\{-b\rho T^{1-2\nu}\}. \tag{30}$$

To proof (30) it is sufficient to take $R = T^{1/2-\nu}$ in (20). For $\nu = 0$ we arrive at quite strong result on l.s.e. weak consistency.

Similarly in conditions of Corollary 6 the following result on probabilities of moderate deviations for l.s.e. holds: for any $c > 0$, and $T > T_0$

$$P\{\|d_T(\theta)(\widehat{\theta}_T - \theta)\| \geq c \ln^{1/2} T\} \leq B_0 T^{-bc}.$$

Gaussian time series, obviously, are jointly strictly sub-Gaussian, and all the paper results are valid for them.

4. **Example**

Consider an example of regression model (1)

$$X_T = \exp\left\{\sum_{i=1}^{q} \theta_i y_i(t)\right\} + \varepsilon_t, \quad t \geq 1, \tag{31}$$

were regression $y(t) = (y_1(t), \dots, y_q(t))$, $t \geq 1$, take values in a compact domain $Y \subset \mathbb{R}^q$. So, in (31)

$$a_t(\theta) = \exp\{\langle \theta, y(t)\rangle\}, \quad \langle \theta, y(t)\rangle = \textstyle\sum_{i=1}^{q} \theta_i y_i(t). \tag{32}$$

Suppose

$$J_T = \left(T^{-1} \textstyle\sum y_i(t) y_j(t)\right)_{i,j=1}^{q} \longrightarrow J = (J_{ij})_{i,j=1}^{q}, \text{ as } T \to \infty. \tag{33}$$

In this case the regression function (32) satisfies condition **A.3**. Really, let

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

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

$$L^2(J_{ii} - \delta) \leq T^{-1} d_{iT}^2(\theta) \leq H^2(J_{ii} + \delta), \quad i = \overline{1,q}. \tag{34}$$

Relations (34) mean that without loss of generality one can take the normalizing matrix $T^{1/2}\mathbb{I}_q$ instead of $d_T(\theta)$, in all the formulations of the paper, $\mathbb{I}_q$ is identity matrix of order q.

For fixed t

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

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

$$\Phi_T(u,v) = \textstyle\sum(\exp\{\langle y(t), \theta + T^{-1/2}u\rangle\} - \exp\{\langle y(t), \theta + T^{-1/2}v\rangle\})^2 \leq$$
$$H^2 T^{-1} \textstyle\sum \|y(t)\|^2 \|u - v\|^2 \leq H^2(TrJ + \delta)\|u - v\|^2.$$

So we have obtained the right hand side of (18) with not depending on θ constant $c_1 > H^2 TrJ$.

On the other hand for fixed t

$$\Delta_t^2(u) = \left(\exp\{\langle y(t), \theta + T^{-1/2}u\rangle\} - \exp\{\langle y(t), \theta\rangle\}\right)^2 =$$
$$\exp\{2\langle y(t), \theta\rangle\} \left(\exp\{\langle y(t), T^{-1/2}u\rangle\} - 1\right)^2.$$

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

$$\left(\exp\{\langle y(t), T^{-1/2}u\rangle\} - 1\right)^2 \geq L_t T^{-1} \langle y(t), u\rangle^2,$$

with

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

and

$$\Delta_t^2(u) \geq \min(\exp\{2\langle y(t), \theta\rangle\}, \exp\{2\langle y(t), \theta + T^{-1/2}u\rangle\}) \cdot T^{-1}\langle y(t), u\rangle^2 \geq$$
$$L^2 T^{-1} \langle y(t), u\rangle^2, \quad t = \overline{1,T}.$$

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

$$\Phi_T(u, 0) \geq L^2 \langle J_T u, u \rangle \geq L^2(\lambda_{min}(J) - \delta)\|u\|^2,$$

and we have obtained the left hand side of inequality (18) for $v = 0$, that was used in fact in the proof of Theorem 2, with constant $c_0 < L^2\lambda_{min}(J)$ not depending on θ. We denoted by $\lambda_{min}(J)$ the least eigenvalue of positive definite matrix J.

The next fact is a reformulation of the Corollary 4 for regression model (31).

Statement 1. Under conditions **N.3**, **N.4**, and (33) there exist constants B_0 and b such that for $T > T_0$, $R > R_0$

$$P\{\|T^{1/2}(\widehat{\theta}_T - \theta)\| \geq R\} \leq B_0 \exp\{-bR^2\}.$$

Moreover for any $\beta > 0$ constant B_0 can be chosen so that

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

In subsequent publication we are going to consider time continuous regression model with jointly strictly sub-Gaussian random process in the capacity of noise.

References

1. A.V. Ivanov, *An asymtotic expansion for the distribution of the least squares estimator of the non-linear regression parameter*, Theory. Probab. Appl., **21** (1977), No3, 557–570.
2. Prakasa Rao, B.L.S., *On the exponential rate convergence of the least squares estimator in the nonlinear regression model with Gaussian errors*, Statist. Probab. Lett., **2** (1984), 139–142.
3. A. Sieders, K. O. Dzhaparidze, *A large deviation result for parameter estimators and its application to nonlinear regression analysis* , Ann. Statist., **15** (1987), No 3, 1031–1049.
4. I.A. Ibragimov, R.Z. Has'minskii, *Statistical estmation: Asymptptic Theory*, Springer, New York, 1981.
5. A.V. Ivanov, *Asymptotic Theory of Nonlinear Regression*, Kluwer Academic Publishers, Dordecht/Boston/London, 1997.
6. A.V. Ivanov, N.N. Leonenko, *Statistical Analysis of Random Fields*, Kluwer Academic Publishers, Dordecht/Boston/London, 1989.
7. Prakasa Rao, B.L.S., *The rate of convergence for the least squares estimator in a non-linear regression model with dependent errors*, J. Multivariate Analysis, **14** (1984), No3, 315–322.
8. S.H. Hu, *A large deviation result for the least squares estimators in nonlinear regression*, Stochastic Process and their Applications, **47** (1993), 345–352.
9. W.Z. Yang, S.H. Hu, *Large deviation for a least squares estimator in a nonlinear regression model*, Stat. Probab. Lett., **91** (2014), 135–144.
10. X. Huang et al., *The large deviation for the least squares estimator of nonlinear regression model based on WOD errors*, J. Inequal. Appl. **125** (2016), (DOI: 10.1186/s13660-016-1064-6).
11. J. Pfanzagl, *On the measurability and consistency of minimum contrast estimates*, J. Metrika, **14** (1969), 249–272.
12. V.V. Buldygin, Yu.V. Kozachenko, *Metric characterization of random variables and random processes*, AMS, Providence, 2000.
13. W. Feller, *An introduction to probability theory and its applications*, vol. **2**, 2-nd Edition, Wiley, New York, 1957.
14. I.I. Gikhman, A.V. Skorokhod, *Introduction to the Theory of Random Processes*, Dover Publications, Inc., 1996.
15. P.J. Brockwell, R.A. Davis, *Introduction to Time Series and Forecasting*, 2-nd Edition, Springer, New York, 2002.

Department of Mathematical Analysis and Probability Theory, Faculty of Physics and Mathematics, NTUU"KPI", Kyiv, Ukraine

E-mail address: alexntuu@gmail.com

Received 29/08/2016