Statistical Methodology 6 (2009) 513–526

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

# Non-asymptotic sequential confidence regions with fixed sizes for the multivariate nonlinear parameters of regression

Andrey V. Timofeev *

*Department of Statistics, Speech Technology Center, 4 Krasutskogo Street, St.-Petersburg, 196084, Russia*

ARTICLE INFO

*Article history:*
Received 25 October 2008
Received in revised form
10 May 2009
Accepted 12 May 2009

*Keywords:*
Non-asymptotic nonlinear estimation
Sequential analysis
Confidence regions
Guaranteed accuracy

ABSTRACT

In this paper we consider a sequential design for the estimation of nonlinear parameters of regression with guaranteed accuracy. Non-asymptotic confidence regions with fixed sizes for the least squares estimates are used. The obtained confidence region is valid for finite numbers of data points when the distributions of the observations are unknown.

© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

In this paper we investigate some non-asymptotic properties of the least squares estimates for nonlinear parameters of regression. Lately there has been considerable interest in nonlinear models in different practical fields, especially in economics. The primary reason behind this interest is the fact that linear models, while having a wide variety of practical applications, did not quite meet the expectations. The asymptotic properties of nonlinear least squares estimates are well investigated and discussed [1,5–8,12,14]. At the same time, only few results addressing the finite sample properties exist, whereas the non-asymptotic solution for the problem of the parameter estimation for regression is practically important because the sample volume is always limited from above. Non-asymptotic estimation of the scalar parameter of nonlinear regression by means of confidence regions was examined by Timofeev [12]. A similar estimation of the multivariate parameter was researched by Timofeev [13]. In this paper a sequential design is suggested that will make it possible to solve the problem of nonlinear estimation of multivariate parameter of regression by means of confidence

* Tel.: +7 812 550 67 19; fax: +7 812 555 89 82.
*E-mail address:* timofeev.andrey@gmail.com.

1572-3127/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2009.05.002

regions in the non-asymptotic setting. As opposed to the method suggested by Timofeev [13], the solution presented in this paper does not use explicit expressions for the gradients of the loss function. The solution was obtained under the condition that distributions of observations are unknown. The mean observation time was estimated in the suggested sequential design.

## 2. Statement of the problem

A stochastic process $\mathbf{X} = \{X(k)|k \geq 0\}$ satisfies the following equation

$$X(k) = A(k, \theta^*) + \xi(k), \quad k \geq 1, \tag{1}$$

where $X(k), A(k, \theta^*), \xi(k) \in R^m$. The nonlinear functions $\{A(k, \theta^*)\,|k \geq 0\}$ are defined. A vector sequence $\{\xi(k)\,|k \geq 1\}$ with unknown distribution is such that

$$\forall k \geq 0 : \left(\mathbf{E}\,(\xi(k)) = 0, \mathbf{E}\left(\xi(k)\xi(k)^{\mathrm{T}}\right) = L(k)\right),$$

where $\{L(k)\,|k \geq 0\}$ is a known sequence of non-random diagonal matrixes $m \times m$ that can be described as follows:

$$\forall k \geq 0 : L(k) = \operatorname{diag}\left(l_1^2(k), \ldots, l_m^2(k)\right), \quad \exists K : \forall_{j=1}^{m} l_j^2(k) < K < \infty.$$

An unknown parameter $\theta^*$ belongs to a closed ball $\Theta$ embedded in an $m$-dimensional Euclidean space $R^m$.

Let $\Delta = \{\delta_1, \delta_2, \ldots, \delta_m\} \subseteq R^1, \forall \delta_j \in \Delta : \delta_j \in ]0, \infty[$ be a set of required dimensions of the confidence parallelepiped. $P_c \in ]0, 1[$ is the required value of the confidence coefficient.

We need to develop a sequential design for confidence estimation of the parameter $\theta^*$ that would determine:

- the stochastic stopping time $\tau \geq 1$,
- a rule for building a confidence rectangular parallelepiped $\Xi(\tau)$ in the compact $\Theta$,

that meet the following conditions:

$$\mathbf{P}_{\theta^*}\left(\theta^* \in \Xi(\tau)\right) \geq P_c, \qquad \mathbf{P}_{\theta^*}(\tau < \infty) = 1,$$
$$\forall \theta_1, \theta_2 \in \Xi(\tau) : \forall_{j=1}^{m} \left|\langle\theta_1\rangle_j - \langle\theta_2\rangle_j\right| \leq \delta_j, \quad \delta_j \in \Delta.$$

From now on, $\|\cdot\|$ stands for a norm of the space in which the compact $\Theta$ is embedded. $\langle a\rangle_i$ from now on stands for the $i$th component of the vector $a$. The brackets will be omitted when no ambiguity arises. For the sake of clarity, the following shorthand notation will be used throughout the rest of the paper: $\mathbf{P}$ (instead of $\mathbf{P}_\theta$) and $\mathbf{E}$ (instead of $\mathbf{E}_\theta$).

## 3. Solution method

Let us assume that stochastic vector functions $\{A(k, \theta)\,|k \geq 0\}$ on $\Theta$ meet the following condition:

$$\forall\,(\theta_1, \theta_2 \in \Theta, n \geq 1) : \sum_{k=0}^{n-1} R_k\,(\theta_1, \theta_2)\,(nm)^{-1} > 0. \tag{2}$$

Here $R_k\,(\theta_1, \theta_2) = [A(k, \theta_1) - A(k, \theta_2)]\, L^{-1}(k)\, [A(k, \theta_1) - A(k, \theta_2)]^{\mathrm{T}}, R_k\,(\theta_1, \theta_2) \in R^1$. Let $\Theta^* = \{\theta_n|n \geq 1\}$ be a sequence of estimators of the parameter $\theta^*$ which is defined as follows:

$$\forall n \geq 1 : \theta_n = \arg \inf_{\theta \in \Theta} I(n, \theta) \tag{3}$$

where $I(n, \theta) = \sum_{k=1}^{t(n)} [X(k) - A(k, \theta)]\, L^{-1}(k)\, [X(k) - A(k, \theta)]^{\mathrm{T}}\,(t(n)m)^{-1}$.

Here $\{t(n)\} \subseteq N, N = \{0, 1, 2, \ldots\}, \lim_{n\to\infty} t(n) \to \infty, I(n, \theta) \in R^1$. For each $n \geq 1$ define a functional $\Phi\,(n, \theta, \theta_n)$ and a closed set sequence $\{\Xi(n)\,|n \geq 1\} \subseteq \Theta$ so that

$$\Phi\,(n, \theta, \theta_n) = I(n, \theta) - I(n, \theta_n), \quad \forall\,(n \geq 1, \theta \in \Xi(n)) : \Phi\,(n, \theta, \theta_n) \leq c(n).$$

Here $\{c(n) \,|n \geq 1\}$ is a known sequence of non-stochastic functions. For each $n \geq 1$ define such elements $\theta_L(n), \theta_U(n) \in \Theta$ that

$$\forall\, (n \geq 1, \theta \in \Xi(n)) : \forall_{j=1}^{m} \left[\langle \theta_L(n) \rangle_j \leq \langle \theta \rangle_j \leq \langle \theta_U(n) \rangle_j\right].$$

The sequence of sets $\{\Xi^*(n) \,|n \geq 1\} \subseteq \Theta$ is such that

$$\forall n \geq 1 : \Xi^*(n) = \left\{ \theta \,\middle|\, \theta \in \Theta, \underset{j=1}{\overset{m}{\forall}} \left[\langle \theta_L(n) \rangle_j \leq \langle \theta \rangle_j \leq \langle \theta_U(n) \rangle_j\right] \right\}.$$

The sequential design for confidence estimation of the parameter $\theta^*$ will be regarded as a pair $(\gamma(n), \tau)$ where

$$\forall n \geq 1 : \gamma(n) = \theta_U(n) - \theta_L(n), \quad \tau = \inf\left\{ n \geq 1 \,\middle|\, \forall_{j=1}^{m} \left(\langle \gamma(n) \rangle_j \leq \delta_j\right) \right\}.$$

The properties of the sequential design are described by the following theorem:

**Theorem 1.** *Assume that the following statements are true:*

(1) $\forall_{j=1}^{m} \delta_j > 0, \delta_j \in \Delta$;
(2) *the sequence* $\{A(k, \theta)|k \geq 1\}$ *meets condition* (2);
(3) *with probability* $1 \forall k \geq 0 \exists L_0 \in ]0, \infty[ : \forall \theta_1, \theta_2 \in \Theta$:

$$\underset{j=1}{\overset{m}{\forall}} \left|\langle A(k, \theta_1) \rangle_j - \langle A(k, \theta_2) \rangle_j\right| l_j^{-2}(k) \leq L_0 \left\|\theta_1 - \theta_2\right\|, \quad \|\cdot\| \textit{ is a norm in } R^m;$$

(4)

$$\underset{j=1}{\overset{m}{\forall}} P\left( \sum_{k \geq 0} k^{-2} \mathbf{E}\left(\xi_{\mathbf{j}}^{\mathbf{4}}(k) l_j^{-4}(k)\right) < \infty \right) = 1;$$

(5) *the sequence of stochastic functions*

$$\left\{ \sum_{k=1}^{n} R_k(\theta_1, \theta)(nm)^{-1} \right\}_{n \geq 1}$$

*converges to a limiting function* $R(\theta_1 \theta)$ *on* $\Theta$ *when* $\theta_1$ *is fixed; in addition* $R(\theta_1, \theta)$ *is a continuous function on* $\Theta$ *and* $R(\theta_1, \theta) > 0$ *when* $\theta_1 \neq \theta$;
(6) $t(n) = \left[n^{2+r}\right] + 1, n \geq 1$, $[a]$- *integral part of the* $a$, $r > 0$.

*Further, if the sequence* $\{c(n)|n \geq 1\}$ *is such that*

$$\forall n \geq 1 : c(n) = \frac{4L_0\, q}{m(1 - P_c)^{0.5}} \cdot n^{-r/2} \left(\frac{\pi^2}{6}\right)^{0.5},$$

*the following assertions hold true:*

(1) $\forall \theta^* \in \Theta : \mathbf{P}_{\theta^*}(\theta^* \in \Xi^*(\tau)) \geq P_c$;
(2) $\forall \theta^* \in \Theta : \mathbf{P}_{\theta^*}(\tau < \infty) = 1$.

The proof of Theorem 1 appears in the Appendix. It is necessary to note that when $t(n) = \tau$, the obtained confidence region will have an a priori fixed size which is determined by set $\Delta$. Conditions (3), (4) and (5) are used in the proof of the second statement of Theorem 1 only.

## 4. Evaluation of the average observation time in the considered sequential design

For practical purposes it is important to evaluate $\mathbf{E}_{\theta^*}\tau$—the average observation time for the sequential design $(\gamma(n), \tau)$. Let us estimate the value $\mathbf{E}_{\theta^*}\tau$ from above.

Let us set $\delta^* = \inf\{\delta_i \,|\delta_i \in \Delta\}$, where $\Delta = \{\delta_1, \delta_2, \dots \delta_m\} \subseteq R^1$. For some $z \in ]0, \infty[$ define the following value:

$$\rho = \inf\left\{n \geq s \left| g(\delta^*/2)^2 - 2c(n) \geq z \right.\right\},$$

where constants $s, g \in ]0, \infty[$ are such that

$$\forall \left(\theta, \theta' \in \Theta, n \geq s\right) : \mathbf{G}\left(n, \theta, \theta'\right) \geq g \left\|\theta - \theta'\right\|^2 \quad a.s. \tag{4}$$

where $\mathbf{G}\left(n, \theta, \theta'\right) = \sum_{k=1}^{t(n)} R_k\left(\theta, \theta'\right)(t(n)m)^{-1}, n \geq 1$.

**Theorem 2.** *Assume that conditions* (1)–(4) *of* Theorem 1 *are met; besides there are known values* $\pi > 0, p > 0, W > 0$ *and*

$$\mathbf{E}_{\theta*}\left(\sup_{\theta_1, \theta_2 \in \Theta}(\eta^4(n, \theta_1, \theta_2))\right) \leq \Pi n^{-p}, \quad \sum_{n\geq 1} n^{-p} \leq W, \text{ where}$$

$$\eta(n, \theta_1, \theta_2) = \sum_{k=1}^{t(n)} 2\xi(k)L^{-1}(k)\left(A(k, \theta_2) - A(k, \theta_1)\right)^{\mathrm{T}} / (t(n)m), \quad \eta(n, \theta_1, \theta_2) \in R^1.$$

*In this case, if condition* (4) *holds true, it is safe to assume that:*

(1) $\mathbf{P}_{\theta^*}(\tau < \infty) = 1$.
(2) $\mathbf{E}_{\theta^*}\tau \leq \rho + 2W\Pi z^{-4}$.

The proof of Theorem 2 appears in the Appendix.

## 5. Discussion

The main idea of the method is based on the research of the behavior of a loss function $I(n, \theta)$ in the neighbourhood of the minimum. For each $t(n)$ the size of the confidence set for the nonlinear parameter $\theta$ is determined by the value $c(n)$ and by the sensitivity of the loss function $I(n, \theta)$ to the variations of this parameter. The assumption of existence on the compact $\Theta$ of the Lipschitz constant for functions $\{A(k, \theta)\}_{k>0}$ is not a strong condition, and is met for most nonlinear functions that are used in regression analysis. Similar assumptions were used by Wu [14] and Skouras [11] in a proof of the strong consistency of least squares estimates in nonlinear regression models. The bigger the following values are ($q$—size of the compact $\Theta$, $K$—the upper bound of the noise dispersion values, $P_c$—confidence coefficient and $L_0$—the Lipschitz constant), the bigger the value $c(n)$ is, and therefore, the bigger the size of the confidence set $\Xi^*(n)$ found for a particular $t(n)$. Condition (4) is a requirement of a strongly convex function $\left(\mathbf{G}\left(n, \theta, \theta'\right)\right)^{0.5}, n \geq 1$ on the compact $\Theta$ with a convexity parameter $2(gq)^{0.5}$, which follows from the strong convex variant of the Weierstrass Theorem. The greater the absolute value of the parameter $g$ is, the greater the sensitivity of the function $\mathbf{G}\left(n, \theta, \theta'\right)$ to the variations of the parameter $\theta$ is, and the smaller sample volume $t(n)$ can be used in order to find a confidence set of given dimensions.

Let us consider a very simple example. Skouras [11] considers the following model $y(t) = \exp(-\theta x(t)) + \xi(t)$, where $\theta \in [\alpha, \beta] = \Theta, \alpha > 0$, $\{x(t)\}$ is a sequence of bounded positive regressors. In addition, we assume that $P(x(t) \in [A, B]) = 1, \forall t \geq 0 : (\mathbf{E}\xi(t) = 0, \mathbf{E}\xi^2(t) = 1, \mathbf{E}(\xi_{\mathbf{j}}^{\mathbf{4}}(t)) = \mu_4)$, $\{x(t)\}$ are independent random variables and there exists a constant $Z$ such that

$$\forall \theta_1, \theta_2 \in [\alpha, \beta] : \mathbf{E}\left(\exp(-\theta_2 x(k)) - \exp(-\theta_1 x(k))\right)^2 \leq Z < \infty.$$

In this case,

$$R_k(\theta_1, \theta_2) = (\exp(-\theta_2 x(k)) - \exp(-\theta_1 x(k)))^2,$$
$$\eta(n, \theta_1, \theta_2) = \sum_{k=1}^{t(n)} 2\xi(k)\,(\exp(-\theta_2 x(t)) - \exp(-\theta_1 x(t)))\,t^{-1}(n).$$

According to Skouras [11], for every $\theta_2, \theta_1 \in [\alpha, \beta]$ there exist constants $c_2, c_1 > 0$ such that

$$c_1 |\theta_2 - \theta_1| x(t) \le |\exp(-\theta_2 x(t)) - \exp(-\theta_1 x(t))| \le c_2 |\theta_2 - \theta_1| x(t) \le c_2 B |\theta_2 - \theta_1|.$$

It is easy to see that

$$\sum_{k=1}^{\infty} \frac{\mathbf{E}\,(\exp(-\theta_2 x(k)) - \exp(-\theta_1 x(k)))^2}{k^2} \le \sum_{k=1}^{\infty} \frac{Z}{k^2} = \frac{Z\pi^2}{6}.$$

Thus using Kolmogorov's strong law of large numbers [4], we have

$$\lim_{n\to\infty} \sum_{k=1}^{n} (\exp(-\theta_2 x(k)) - \exp(-\theta_1 x(k)))^2\, n^{-1}$$
$$= \lim_{n\to\infty} \sum_{k=1}^{n} \mathbf{E}\,(\exp(-\theta_2 x(k)) - \exp(-\theta_1 x(k)))^2\, n^{-1} = R(\theta_2, \theta_1).$$

Here function $R(\theta, \theta_1)$ has on $[\alpha, \beta]$ the Lipschitz constant $L = c_2 B$, and therefore is a continuous function on $\Theta$. Besides, $R(\theta_1, \theta) > 0$ when $\theta_1 \neq \theta$. Further,

$$P\left(\sum_{k\ge 0} k^{-2}\mathbf{E}\left(\xi_{\mathbf{j}}^{\mathbf{4}}(k)\right) = \mu_4\pi^2/6 < \infty\right) = 1;$$

and the assumption (4) of Theorem 1 is met. Further, the next inequality is obvious

$$\mathbf{G}\left(n, \theta, \theta'\right) = \sum_{k=1}^{t(n)} R_k\left(\theta, \theta'\right) t(n)^{-1} \ge c_1^2 |\theta_2 - \theta_1|^2 x^2(t) \ge c_1^2 A^2 |\theta_2 - \theta_1|^2$$

and condition (4) for the process $y(t)$ holds true. Then $\forall \theta_1, \theta_2 \in [\alpha, \beta] : \mathbf{E}\eta(n, \theta_1, \theta_2) = 0$, and we can use the theorem of Dharmadhikari and Jogdeo [2] to get the inequality

$$\mathbf{E}\,(\eta(n, \theta_1, \theta_2))^4 \le R(4) t^{-3}(n) \sum_{k=1}^{t(n)} \mu_4 c_2^4 |\theta_2 - \theta_1|^4 B^4 = R(4) t^{-2}(n) \mu_4 c_2^4 |\theta_2 - \theta_1|^4 B^4.$$

Here $R(p) = \frac{1}{2}p(p-1)\max(1, 2^{p-3})\left(1 + \frac{2}{p}K_{2m}^{(p-2)/2m}\right)$, $K_{2m} = \sum_{r=1}^{m} \frac{r^{2m-1}}{(r-1)!}$. The integral value $m$ is such that $2m \le p < 2m+2$. In this case we have: $\Pi = R(4)\mu_4 c_2^4 |\theta_2 - \theta_1|^4 B^4$ and $W = \sum_{n\ge 1} n^{-2} = \pi^2/6$.

Thus we have got constants $\Pi$, $W$ from the conditions of Theorem 2. Now we have checked all the assumptions of Theorems 1 and 2 for the considered sample model.

## 6. Numerical results

The following model was used for the simulation study: $y(t) = H \cdot \exp(-\theta x(t)) + \xi(t)$. Here $\{\xi(t)\}$ is a sequence of independent random variables distributed by $N(0, 1)$, and $\{x(t)\}$ is a sequence of independent random variables evenly distributed on the interval [0,1], and $\theta^* \in [0, 1]$. The confidence coefficient $P_c$ and the required size of the confidence interval $\delta$ are fixed at 0.95 and 0.05, respectively. In this case if $r = 1$, the following is true: $\forall n \ge 1 : c(n) = 4 \cdot \pi \cdot H \cdot (n \cdot 0.3)^{-0.5}$.

The observation was stopped at random moments $\tau = \inf\{n \ge 1\, |\theta_U(n) - \theta_L(n) \le \delta\}$, where $\theta_U(n)$ is the upper bound of the confidence interval and $\theta_L(n)$ is the lower bound of the confidence interval. Thus, at the moment when the observation was stopped, the required value of the confidence interval for the parameter $\theta^*$ was achieved.

**Table 1**
Confidence intervals of the parameter $\theta^*$.

| $H$ | $\theta^*$ | $\bar{\tau}(\theta^*)$ | Average lower bound | Average upper bound |
|---|---|---|---|---|
| 20 | 0 | 8 | −0.02 | 0.02 |
| | 0.1 | 8 | 0.07 | 0.11 |
| | 0.15 | 8 | 0.11 | 0.16 |
| | 0.2 | 25 | 0.17 | 0.22 |
| | 0.25 | 56 | 0.21 | 0.26 |
| 10 | 0 | 55 | −0.03 | 0.02 |
| | 0.1 | 65 | 0.06 | 0.11 |
| | 0.15 | 100 | 0.11 | 0.15 |
| | 0.2 | 105 | 0.17 | 0.22 |
| | 0.25 | 180 | 0.23 | 0.28 |

The estimate for the average observation time $\bar{\tau}(\theta^*)$ in the sequential design has been defined as the mean time value in the realizations of the simulation experiment groups. Every experiment group simulation has been carried out for a fixed unique pair of the model parameters $(H, \theta^*)$ and consisted of 20 series. For every experiment group the lower bound of the confidence interval $(\bar{\theta}_L(\tau))$ and upper bound of the confidence interval $(\bar{\theta}_U(\tau))$ were determined. The results are presented in Table 1. As shown, $\bar{\tau}(\theta^*)$ depends on values of the parameters $H$ and $\theta^*$ to a significant degree. Parameter $H$ determines the signal to noise ratio for the observed sample. Therefore, if $\theta^*$ is fixed, greater values of the $H$ correspond to smaller values of the $\bar{\tau}(\theta^*)$ and vice versa: smaller values of the $H$ correspond to greater values of the average observation time $\bar{\tau}(\theta^*)$. This proves an assumption that might seem obvious: with fixed $P_c$ and $\theta^*$ we need a smaller volume of the sample to get the required size of the confidence interval with greater signal to noise ratio as opposed to a sample with a smaller signal to noise ratio. On the other hand, with fixed $P_c$ and $H$ the dependence between values $\theta^*$ and $\bar{\tau}(\theta^*)$ is determined by parametric sensitivity of the functional $I(n, \theta)$ in the vicinity of the point $\theta^*$. Following [10] we have: the parametric sensitivity of the functional $I(n, \theta)$ is characterized by a coefficient sensitivity $k(n, \theta)$ which is defined as a derivative of the functional $I(n, \theta)$ with respect to $\theta$.

Reasoning informally, we can see that the dependence of the value $\bar{\tau}(\theta^*)$ on $\theta^*$ is obvious enough: the greater the value of the function $k(n, \theta)$ in the vicinity of the point $\theta^*$ is, the smaller is the value of the average observation time $\bar{\tau}(\theta^*)$. In other words, the greater the sensitivity of the functional $I(n, \theta)$ to variations of the parameter $\theta$ in the vicinity of the point $\theta^*$, the smaller the sample volume needed to design the confidence interval with a given size and fixed value of the confidence coefficient $P_c$.

## 7. Conclusion remarks

In this paper we derived non-asymptotic confidence regions with fixed sizes for the LS estimation of the multivariate parameter. The solution is based on sequential analysis with no information about the distribution function of observations. The suggested method can be used for non-asymptotic estimation of the multivariate parameters of the wide class of nonlinear regressions. The results of the simulation study proved the efficiency of the suggested method for confidence estimation.

## Appendix

**Proof of Theorem 1.** The proof uses the following facts:

**Lemma 1** (*[3]*). *Let $\{I_n(\theta)\}_{n\geq 1}$ be a sequence of stochastic functions complying with the conditions:*

(1) *for each $\theta \in \Theta$ the sequence $\{I_n(\theta)\}_{n\geq 1}$ is consistent with a nondecreasing flow of $\sigma$-subalgebras $\{F_n\}_{n\geq 0}$;*

(2) *for each $\theta \in \Theta$ when $\theta^* \in \Theta$ is fixed the following relation holds true:*

$$P\left(\lim_{n\to\infty} I_n(\theta) = \Phi(\theta,\theta^*)\right) = 1,$$

*where $\Phi(\theta,\theta^*), \theta \in \Theta$ is a real valued function which is continuous on $\Theta$ and is such that $\Phi(\theta,\theta^*) > \Phi(\theta^*,\theta^*), \theta \neq \theta^*$.*

(3) *when $n \geq 1$ is fixed, the function $I_n(\theta), \theta^* \in \Theta$ is continuous on $\Theta$;*

(4) *for any $\delta > 0$ there exists $\gamma_0 > 0$ and a function $\mathbf{c}(\gamma), \gamma > 0, \mathbf{c}(\gamma) \to 0, \gamma \to 0$, such that for any $\theta' \in \Theta$ and any $0 < \gamma < \gamma_0$ the following relation holds true:*

$$P\left(\overline{\lim_{n\to\infty}} \sup_{\|\theta-\theta'\|<\gamma, \|\theta-\theta^*\|\geq\delta} \left(I_n(\theta) - I_n(\theta')\right) < c(\gamma)\right) = 1.$$

*Further, if $\Theta$ is compact then*

$$P\left(\lim_{n\to\infty} \left\|\theta_n - \theta^*\right\| = 0\right) = 1,$$

*where*

$$\theta_n = \arg\inf_{\theta\in\Theta} I_n(\theta),$$

*$\|\cdot\|$ is a norm of the space in which the compact $\Theta$ is embedded.*

**Lemma 2.** *Assume that conditions (1)–(5) of the considered theorem are met. Then*

$$P\left(\lim_{n\to\infty} \left\|\theta_L(n) - \theta^*\right\| = 0\right) = 1,$$
$$P\left(\lim_{n\to\infty} \left\|\theta_U(n) - \theta^*\right\| = 0\right) = 1.$$

**Proof of Lemma 2.** The representation

$$\begin{aligned} I(n,\theta) &= \sum_{k=1}^{t(n)} R_k(\theta^*,\theta) + 2\xi(k)L^{-1}(k)\left(A(k,\theta^*) - A(k,\theta)\right)^{\mathrm{T}} + \xi(k)L^{-1}(k)\xi^{\mathrm{T}}(k)\,(t(n)m)^{-1} \\ &= \sum_{k=1}^{t(n)} R_k(\theta^*,\theta)\,(t(n)m)^{-1} + 2\sum_{k=1}^{t(n)} \xi(k)L^{-1}(k)\left(A(k,\theta^*)\right. \\ &\quad \left. - A(k,\theta)\right)^{\mathrm{T}}(t(n)m)^{-1} + \sum_{k=1}^{t(n)} \left(\xi(k)L^{-1}(k)\xi^{\mathrm{T}}(k)\right)(t(n)m)^{-1} \end{aligned}$$

is obvious. The properties of the matrixes $\{L_n | n \geq 1\}$ make it easy to observe that

$$\lim_{n\to\infty} \sum_{k=0}^{t(n)} \sum_{j=1}^{m} \left(\mathbf{E}(\langle\xi(k)\rangle_j^2 l_j^{-2}(k))\right)(t(n)m)^{-1} = 1$$

and

$$\lim_{n\to 0} \sum_{k=1}^{t(n)} \xi(k)L^{-1}(k)\xi^{\mathrm{T}}(k)(t(n)m)^{-1} = 1 \quad \text{a.s.} \tag{5}$$

Consider a process $\left(z^*_{t(n)}(\theta)\right)_{t(n)\geq 1}$, where

$$z^*_{t(n)}(\theta) = \sum_{k=0}^{t(n)} 2\xi(k)L^{-1}(k)\left(A(k,\underline{x(k)},\theta^*) - A(k,\underline{x(k)},\theta)\right)^{\mathrm{T}} m^{-1}.$$

Due to the finiteness of the set $\Theta$, there exists a finite value $q = \sup_{\theta_1,\theta_2\in\Theta} \|\theta_1 - \theta_2\|$. It is obvious that for any $\theta_1, \theta_2 \in 1$ the following representation is possible:

$$\forall (t(n) \geq 1, j = \overline{1, m}) : \mathbf{E}\left(\langle \xi(t(n))\rangle_j^2 l_j^{-2} \cdot (\langle A(t(n), \theta_1)\rangle_j - \langle A(t(n), \theta_2)\rangle_j)^2\right)$$
$$= \left(\langle A(t(n), \theta_1)\rangle_j - \langle A(t(n), \theta_2)\rangle_j\right)^2 \mathbf{E}\left(\langle \xi(t(n))\rangle_j^2 l_j^{-2}\right) \leq L_0^2 \|\theta_1 - \theta_2\| \leq L_0^2 q^2.$$

It is easy to observe that $\sum_{k\geq 1} L_0^2 q^2 k^{-2} < \infty$. Therefore, for any $\theta_1, \theta_2 \in \Theta$

$$\overset{m}{\underset{j=1}{\forall}} \sum_{k\geq 1} k^{-2} \mathbf{E}\left(\frac{\langle \xi(k)\rangle_j^2}{l_j^2(k)} \left(\langle A(k, \theta_1)\rangle_j - \langle A(k, \theta_2)\rangle_j\right)^2\right) < \infty.$$

By strong law of large numbers, we have:

$$\forall \theta, \quad \theta^* \in \Theta : P_{\theta^*}\left(\lim_{n\to\infty} t(n)^{-1} z^*_{t(n)}(\theta) = 0\right) = 1. \tag{6}$$

Taking into account (5), (6) and condition 5 of the theorem being proved,

$$\forall \theta, \quad \theta^* \in \Theta : P_{\theta^*}\left(\lim_{n\to\infty} I(n, \theta) = R(\theta^*, \theta) + 1\right) = 1.$$

Hence, condition 2 of Lemma 1 is met. It follows from condition 3 and the Rademachers Theorem [9] that the function $\langle A(k, \theta)\rangle_j, k \geq 1, j = \overline{1, m}$ is continuous on $\Theta$ with a probability 1. Therefore, using the theorem on continuity of a composite function, we conclude that for any finite $n \geq 1$ the function $I(n, \theta)$ is continuous on $\Theta$ and condition 3 of Lemma 1 is met. Let us set

$$\mathbf{G}(n, \theta_1, \theta_2) = \sum_{k=1}^{t(n)} R_k(\theta_1, \theta_2)(t(n)m)^{-1}, \quad n \geq 1.$$

Further, taking into account condition 3 of the considered theorem,

$$\overset{m}{\underset{j=1}{\forall}} \left|\frac{\partial \langle A(k, \theta)\rangle_j}{\partial \langle \theta\rangle_j}\right| \leq L_0 \sup_j l_j^2(k) \leq KL_0$$

and

$$\begin{aligned}
\overset{m}{\underset{j=1}{\forall}} \frac{\partial G(n, \theta^*, \theta)}{\partial \langle \theta\rangle_j} &= 2\sum_{k=0}^{t(n)} \sum_{j=1}^{m} \frac{\partial \langle A(k, \theta)\rangle_j}{\partial \langle \theta\rangle_j} \left(\frac{\langle A(k, \theta^*)\rangle_j}{l_j^2(k) t(n) m} - \frac{\langle A(k, \theta)\rangle_j}{l_j^2(k) t(n) m}\right) \\
&\leq 2(t(n)m)^{-1} \sum_{k=0}^{t(n)} \sum_{j=1}^{m} \frac{\partial \langle A(k, \theta)\rangle_j}{\partial \langle \theta\rangle_j} \left|\frac{\langle A(k, \theta^*)\rangle_j}{l_j^2(k)} - \frac{\langle A(k, \theta)\rangle_j}{l_j^2(k)}\right| \\
&\leq 2(t(n)m)^{-1} \sum_{k=0}^{t(n)} \sum_{j=1}^{m} \left|\frac{\partial \langle A(k, \theta)\rangle_j}{\partial \langle \theta\rangle_j}\right| \cdot \left|\frac{\langle A(k, \theta^*)\rangle_j}{l_j^2(k)} - \frac{\langle A(k, \theta)\rangle_j}{l_j^2(k)}\right| \\
&\leq 2KL_0^2 \left\|\theta^* - \theta\right\| \leq 2KL_0^2 q.
\end{aligned}$$

For any $\theta_1, \theta_2 \in \Theta$ when $\theta^*$ is fixed, applying Taylor's expansion of the function $\mathbf{G}(\mathrm{n}, \theta, \theta^*)$ at the point $\theta_1$ we have:

$$\mathbf{G}\left(n, \theta^*, \theta_1\right) = \mathbf{G}\left(n, \theta^*, \theta_2\right) + D(n, \theta_1, \theta_2). \tag{7}$$

where $D(n, \theta_1, \theta_2)$ is a remainder term and

$$|D(n, \theta_1, \theta_2)| \leq 2KL_0^2 q \sum_{j=1}^{m} \left|\langle \theta_1\rangle_j - \langle \theta_2\rangle_j\right|.$$

Further, (7) allows us to conclude that

$$\forall\left(\theta_1,\theta_2\in\Theta\right):\left|\mathbf{G}\left(n,\theta^*,\theta_1\right)-\mathbf{G}\left(n,\theta^*,\theta_2\right)\right|\le 2KL_0^2q\sum_{j=1}^{m}\left|\langle\theta_1\rangle_j-\langle\theta_2\rangle_j\right|.$$

Applying Jensen's inequality

$$\left(\sum_{j=1}^{m}\left|\langle\theta_1\rangle_j-\langle\theta_2\rangle_j\right|m^{-1}\right)^2\le\sum_{j=1}^{m}\left|\langle\theta_1\rangle_j-\langle\theta_2\rangle_j\right|^2m^{-1},$$

and therefore

$$\sum_{j=1}^{m}\left|\langle\theta_1\rangle_j-\langle\theta_2\rangle_j\right|\le\|\theta_1-\theta_2\|\,m^{0.5}.$$

The final conclusion is

$$\forall\left(\theta^*,\theta_1,\theta_2\in\Theta,n\ge 1\right):\left|\mathbf{G}\left(n,\theta^*,\theta_1\right)-\mathbf{G}\left(n,\theta^*,\theta_2\right)\right|\le 2KL_0^2q\,\|\theta_1-\theta_2\|\,m^{0.5}. \tag{8}$$

Consider the following events:

$$\omega_n\in F_n:\left|B(n,\theta^*,\theta_1,\theta_2)+2\sum_{k=0}^{t(n)}\xi(k)L^{-1}(k)\left(A(k,\theta_2)-A(k,\theta_1)\right)^{\mathrm{T}}\left(t(n)m\right)^{-1}\right|\le L_1\|\theta_1-\theta_2\|,$$

$$\tilde{\omega}_n\in F_n:\left|B(n,\theta^*,\theta_1,\theta_2)\right|\le L_1\|\theta_1-\theta_2\|,$$

where $B(n,\theta^*,\theta_1,\theta_2)=\left|\mathbf{G}\left(n,\theta^*,\theta_1\right)-\mathbf{G}\left(n,\theta^*,\theta_2\right)\right|,\ L_1=2KL_0^2qm^{0.5}$.

Taking (8) and (6) into account for any $\theta^*,\theta_1,\theta_2\in\Theta$,

$$P_{\theta^*}\left(\overline{\lim_{n\to\infty}}\,\omega_n=\overline{\lim_{n\to\infty}}\,\tilde{\omega}_n\right)=1,$$

i.e. for any $\gamma>0$ we can use the following representation:

$$P_{\theta^*}\left(\overline{\lim_{n\to\infty}}\sup_{\|\theta_1-\theta_2\|<\gamma}\left|I(n,\theta_1)-I(n,\theta_2)\right|=\overline{\lim_{n\to\infty}}\sup_{\|\theta_1-\theta_2\|<\gamma}\left|B(n,\theta^*,\theta_1,\theta_2)\right.\right.$$
$$\left.\left.+\ 2\sum_{k=0}^{t(n)}\xi(k)L^{-1}(k)(A(k,\theta_2)-A(k,\theta_1))^{\mathrm{T}}/(t(n)m)\right|<L_1\gamma\right)=1. \tag{9}$$

Given (9), condition 4 of Lemma 1 is met with the function $\mathbf{c}(\gamma)=L_1\gamma$. Therefore, for the sequence $\{I(n,\theta)|n\ge 1\}$ all the conditions of Lemma 1 are met; so

$$P_{\theta^*}\left(\lim_{n\to\infty}\left\|\theta_n-\theta^*\right\|=0\right)=1. \tag{10}$$

Further,

$$\forall\theta\in\Theta:P_\theta\left(\lim_{n\to\infty}c(n)=0\right)=1. \tag{11}$$

Consider an arbitrary sequence $\left\{\bar{\theta}_n|n\ge 1\right\}$ of the elements of the compact $\Theta$ such that

$$n\ge 1:\left(\bar{\theta}_n\in\Xi(n),\bar{\theta}_n\ne\theta_n\right).$$

For the sequence $\left\{\bar{\theta}_n|n\ge 1\right\}$ there exists a real positive sequence $\varepsilon=\{\varepsilon_n|n\ge 1\}$ which can be described as follows:

$$n\ge 1:\bar{\theta}_n(\varepsilon_n)=\arg\inf_{\theta\in\Theta}\left|\Phi(n,\theta,\theta_n)-\varepsilon_n\right|.$$

It is obvious that

$$n \geq 1 : \exists \theta^* \in \Xi(n) : P_{\theta^*}\left(\varepsilon_n \leq c(n)\right) = 1. \tag{12}$$

Let us check whether conditions 2 and 4 of Lemma 1 are met for the sequence of functionals $\{|\Phi(n,\theta,\theta_n) - \varepsilon_n|\}_{n\geq 1}$. (It is obvious that the rest of the conditions are met.)

Taking into account (11) and (12),

$$n \geq 1, \quad \theta^* \in \Theta : P_{\theta^*}\left(\lim_{n\to\infty} \varepsilon_n = 0\right) = 1.$$

Besides,

$$P_{\theta^*}\left(\lim_{n\to\infty} |\Phi(n,\theta,\theta_n) - \varepsilon_n| = \left|R(\theta,\theta^*) - R(\theta^*,\theta^*)\right| = \left|R(\theta,\theta^*)\right|\right) = 1,$$

where the function $|R(\theta,\theta^*)|$ has the only minimum on $\Theta$ at $\theta = \theta^*$ and is continuous. Therefore, condition 2 of Lemma 1 is met. For any $\theta_1, \theta_2 \in \Theta$

$$||\Phi(n,\theta_1,\theta_n) - \varepsilon_n| - |\Phi(n,\theta_2,\theta_n) - \varepsilon_n|| \leq |I(n,\theta_1) - I(n,\theta_2)|.$$

Taking (8) into account, for any $\theta_1, \theta_2 \in \Theta$

$$\forall\,(n \geq 1, \gamma > 0) : P_{\theta^*}\left(\overline{\lim_{n\to\infty}} \sup_{\|\theta_1-\theta_2\|<\gamma} ||\Phi(n,\theta_1,\theta_n) - \varepsilon_n| - |\Phi(n,\theta_2,\theta_n) - \varepsilon_n|| \leq L_1\gamma\right) = 1.$$

In this case for the sequence $\{|\Phi(n,\theta,\theta_n) - \varepsilon_n|\}_{n\geq 1}$ condition 4 of Lemma 1 is met with the function $c(\gamma) = L_1\gamma$ based on Lemma 1.

$$P_{\theta^*}\left(\lim_{n\to\infty} \left\|\bar{\theta}_n(\varepsilon_n) - \theta^*\right\| = 0\right) = 1.$$

Due to the arbitrary choice of the sequence $\left\{\bar{\theta}_n(\varepsilon_n)\,|n \geq 1\right\}$

$$\mathop{\forall}_{j=1}^{m} P_{\theta^*}\left(\lim_{n\to\infty} \left|\langle\theta_L(n)\rangle_j - \langle\theta^*\rangle_j\right| = 0\right) = 1,$$

$$\mathop{\forall}_{j=1}^{m} P_{\theta^*}\left(\lim_{n\to\infty} \left|\langle\theta_U(n)\rangle_j - \langle\theta^*\rangle_j\right| = 0\right) = 1.$$

Hence, Lemma 2 is proven. ■

From the strong consistency of the sequences $\{\theta_L(n)\,|n \geq 1\}$ and $\{\theta_U(n)\,|n \geq 1\}$ it follows that $\tau$ is an end point which means that the second assertion of the considered theorem holds true.

When an assertion $x$ implies an assertion $y$, we will use the following notation: $x \Rightarrow y$. If an event $\omega(1)$ leads to an event $\omega(2)$, we will write $\omega(1) \subset \omega(2)$. Further, for any $\theta_1, \theta_2 \in \Theta$, $n \geq 1$ consider a value

$$\eta(n,\theta_1,\theta_2) = \sum_{k=1}^{t(n)} 2\xi(k)L^{-1}(k)\left(A(k,\theta_2) - A(k,\theta_1)\right)^{\mathrm{T}}\left(t(n)m\right)^{-1}.$$

It is obvious that

$$\begin{aligned}
\mathbf{E}\left(\eta^2(n,\theta_1,\theta_2)\right) &= 4\mathbf{E}\left(\sum_{k=1}^{t(n)}\sum_{j=1}^{m}\left(\frac{\langle\xi(k)\rangle_j^2}{l_j^2(k)}\right)\left(\frac{\langle A(k,\theta_1)\rangle_j - \langle A(k,\theta_2)\rangle_j}{l_j(k)t(n)m}\right)^2\right) \\
&= 4\left(\sum_{k=1}^{t(n)}\sum_{j=1}^{m}\mathbf{E}\left(\frac{\langle\xi(k)\rangle_j^2}{l_j^2(k)}\right)\mathbf{E}\left(\frac{\langle A(k,\theta_1)\rangle_j - \langle A(k,\theta_2)\rangle_j}{l_j(k)t(n)m}\right)^2\right) \\
&= 4\sum_{k=1}^{t(n)}\mathbf{E}\left(R_k(\theta_1,\theta_2)\right)\left(t(n)m\right)^{-2}.
\end{aligned}$$

Further, taking into account condition 3 of the considered theorem,

$$4\sum_{k=1}^{t(n)} \mathbf{E}\left(R_k(\theta_1,\theta_2)\right)(t(n)m)^{-2} \le 4L_0^2 q^2 (t(n)m^2)^{-1}.$$

For the sake of brevity, the following shorthand notations will be used:

$$\varepsilon(n,\bullet) = \mathbf{E}_{\theta^*}\left(\eta(n,\theta_n,\theta)\right), \qquad \chi(n,\bullet) = \eta(n,\theta_n,\theta) - \varepsilon(n,\bullet), \quad \theta\in\Theta.$$

Further, for any $\theta \in \Theta$

$$\begin{aligned} &P_{\theta^*}\left(|\eta(n,\theta_n,\theta)| \le c(n)\right) = P_{\theta^*}\left(|\chi(n,\cdot)+\varepsilon(n,\cdot)| \le c(n)\right) \\ &\quad \ge P_{\theta^*}\left(|\chi(n,\bullet)| + \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)| \le c(n)\right) = P_{\theta^*}\left(|\chi(n,\bullet)| \le c(n) - \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\right) \end{aligned}$$

Further,

$$\forall\theta\in\Theta : \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|^2 \le 4L_0^2q^2(t(n)m^2)^{-1}.$$

However, according to Lyapunov's inequality

$$\forall\theta\in\Theta : \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)| \le \left(\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|^2\right)^{-0.5};$$

so

$$\forall\theta\in\Theta : \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)| \le \left(4L_0^2q^2(t(n)m^2)^{-1}\right)^{0.5} = 2L_0^q\left(m\sqrt{t(n)}\right)^{-1}.$$

Further,

$$(t(n))^{-0.5} = \left(\left[n^{2+r}\right]+1\right)^{-0.5} \le \left[n^{2+r}\right]^{-0.5} \le n^{-r/2}, \quad n>0.$$

That is why

$$\forall\theta\in\Theta : \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)| \le 2L_0^q\,(t(n))^{-0.5}\,m^{-1} \le c(n)\,(1-P_c)^{0.5}\left(\pi^2/6\right)^{-0.5}/2.$$

It is obvious that $(1-P_c)^{0.5} < 1$и$\left(\pi^2/6\right)^{-0.5} < 1$; so

$$\forall\theta\in\Theta : \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)| \le c(n)/2$$

and for $t = c(n)/2 - \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)| > 0$ we have

$$\begin{aligned} &P_{\theta^*}\left(|\eta(n,\theta_n,\theta)| \le c(n)\right) \ge P_{\theta^*}\left(|\chi(n,\bullet) \le c(n)| - \mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|\right) \\ &\quad = P_{\theta^*}\left(|\chi(n,\bullet)| \le c(n)/2+t\right) \ge P_{\theta^*}\left(|\chi(n,\bullet)| \le c(n)/2\right). \end{aligned}$$

Further,

$$\forall\theta\in\Theta : P_{\theta^*}\left(|\eta(n,\theta_n,\theta)| > c(n)\right) < P_{\theta^*}\left(|\chi(n,\bullet)| > c(n)/2\right).$$

Chebyshev's inequality allows us to say that

$$\begin{aligned} &P_{\theta^*}\left(|\chi(n,\bullet)| > c(n)/2\right) \le 4c^{-2}(n)\mathbf{E}_{\theta^*}\left(\chi(n,\bullet)\right)^2 \\ &\quad \le 4c^{-2}(n)\mathbf{E}_{\theta^*}|\eta(n,\theta_n,\theta)|^2 \le 16c^{-2}(n)L_0^2q^2(t(n)m^2)^{-1} \end{aligned}$$

and

$$\forall\theta\in\Theta : P_{\theta^*}\left(|\eta(n,\theta_n,\theta)| > c(n)\right) \le 16c^{-2}(n)L_0^2q^2(t(n)m^2)^{-1}. \tag{13}$$

The following representation is possible:

$$\Phi(n,\theta,\theta_n) = \mathbf{G}\left(n,\theta^*,\theta\right) - \mathbf{G}\left(n,\theta^*,\theta_n\right) + \eta(n,\theta_n,\theta) \ge 0, n\ge 1, \quad \theta\in\Theta.$$

If $\theta=\theta^*$ then $\mathbf{G}(n,\theta^*,\theta) - \mathbf{G}(n,\theta^*,\theta_n) \le 0$ and the parameter $\theta^*$ does not belong to the set $\Xi(n)$ only if $\eta(n,\theta_n,\theta^*)$ is such that

$$I(n,\theta^*) - I(n,\theta_n) = \mathbf{G}\left(n,\theta^*,\theta^*\right) - \mathbf{G}\left(n,\theta^*,\theta_n\right) + \eta(n,\theta_n,\theta^*) > c(n). \tag{14}$$

Consider a situation when $|\eta(n,\theta_n,\theta^*)| \le c(n)$. In this case it is obvious that $I(n,\theta^*) - I(n,\theta_n) \le c(n)$ because $\operatorname{sign}(I(n,\theta^*) - I(n,\theta_n)) = \operatorname{sign}(\mathbf{G}(n,\theta^*,\theta^*) - \mathbf{G}(n,\theta^*,\theta_n)) = -1$. In other words, if $|\eta(n,\theta_n,\theta^*)| \le c(n)$ then $\theta^* \in \Xi(n)$ and on the set $\Xi(n)$ the following conclusions hold true:

$$\forall n \ge 1 : \left(\left|\eta(n,\theta_n,\theta^*)\right| \le c(n)\right) \Rightarrow \left(\theta^* \in \Xi(n)\right) \Rightarrow \left(\theta^* \in \Xi^*(n)\right). \tag{15}$$

Using (15),

$$\forall n \ge 1 : P\left(\left|\eta(n,\theta_n,\theta^*)\right| \le c(n)\right) \le P\left(\theta^* \in \Xi^*(n)\right).$$

Taking (13) into account,

$$\forall\left(n \ge 1, \theta^* \in \Theta\right) : P\left(\theta^* \notin \Xi^*(n)\right) \le P\left(\left|\eta(n,\theta_n,\theta^*)\right| > c(n)\right) \le 16c^{-2}(n)L_0^2q^2(t(n)m^2)^{-1}.$$

Consider now the probability of the event $\omega_\tau : \theta^* \notin \Xi^*(\tau)$ not happening.

$$\begin{aligned} P\left(\theta^* \notin \Xi^*(\tau)\right) &= \sum_{n\ge1} P\left(\theta^* \notin \Xi^*(n), \tau = n\right) \le \sum_{n\ge1} P\left(\theta^* \notin \Xi^*(n)\right) \\ &\le \sum_{n\ge1} 16c^{-2}(n)L_0^2q^2(t(n)m^2)^{-1} \le \left(16L_0^2q^2m^2\right)\sum_{n\ge1} c^{-2}(n)t^{-1}(n) \\ &= (1-P_c)\left(\frac{\pi^2}{6}\right)^{-1}\sum_{n\ge1} n^r t^{-1}(n) = (1-P_c)\left(\frac{\pi^2}{6}\right)^{-1}\sum_{n\ge1} n^r\left(\left[n^{2+r}\right]+1\right)^{-1} \\ &\le (1-P_c)\left(\frac{\pi^2}{6}\right)^{-1}\sum_{n\ge1} n^r n^{-2-r} = 1-P_c. \end{aligned}$$

Here we have taken into account that $\sum_{n\ge1} n^{-2} = \frac{\pi^2}{6}$. Hence,

$$P\left(\theta^* \in \Xi^*(\tau)\right) \ge P_c,$$

and we have proved the first statement of Theorem 1. It is obvious that the rectangular parallelepiped $\Xi^*(\tau)$ is a confidence parallelepiped for the parameter $\theta^*$. It complies with the requirements formulated in the statement of the problem and is fully determined by the vectors $\theta_L(n), \theta_U(n)$. The implementation of the suggested sequential design $(\gamma(n), \tau)$ is aimed at calculation of these vectors. ■

**Proof of Theorem 2.** Taking into account that $\mathbf{G}(n,\theta^*,\theta^*) = 0, n \ge 1$,

$$\begin{aligned} n \ge 1 : I(n,\theta_U(n)) - I(n,\theta^*) &= \mathbf{G}\left(n,\theta^*,\theta_U(n)\right) + \mathbf{G}\left(n,\theta^*,\theta^*\right) + \eta(n,\theta^*,\theta_U(n)) \\ &= \mathbf{G}\left(n,\theta^*,\theta_U(n)\right) + \eta(n,\theta^*,\theta_U(n)) \ge 0. \end{aligned}$$

In a similar manner,

$$\forall n \ge 1 : I(n,\theta_L(n)) - I(n,\theta^*) = \mathbf{G}\left(n,\theta^*,\theta_L(n)\right) + \eta(n,\theta^*,\theta_U(n)) \ge 0.$$

Let us set $H = 2\,\|\theta^* - \theta_U(n)\|\,\|\theta^* - \theta_L(n)\|$. Consider the following events:

$$\begin{aligned} &\omega_U(0,n) : \left\{I(n,\theta^*) - I(n,\theta_U(n)) \le g(\delta^*/2)^2\right\} \\ &\omega_U(1,n) : \left\{\mathbf{G}\left(n,\theta^*,\theta_U(n)\right) = c(n) + \eta(n,\theta^*,\theta_n) + s(n,\theta^*,\theta_U(n)) \le g(\delta^*/2)^2\right\} \\ &\omega_L(0,n) : \left\{I(n,\theta^*) - I(n,\theta_L(n)) \le g(\delta^*/2)^2\right\} \\ &\omega_L(1,n) : \left\{\mathbf{G}\left(n,\theta^*,\theta_L(n)\right) = c(n) + \eta(n,\theta_L(n),\theta^*) + s(n,\theta^*,\theta_L(n)) \le g(\delta^*/2)^2\right\} \\ &\omega_U(4,n) : \left\{\left\|\theta^* - \theta_U(n)\right\|^2 \le (\delta^*/2)^2\right\} \\ &\omega_L(4,n) : \left\{\left\|\theta^* - \theta_L(n)\right\|^2 \le (\delta^*/2)^2\right\} \\ &\omega(*,n) : \left\{H \le (\delta^*)^2/2\right\}. \end{aligned}$$

Here

$$s(n,\theta^*,\theta_U(n)) = c(n) - \left(\mathbf{G}\left(n,\theta^*,\theta_U(n)\right) + \eta(n,\theta^*,\theta_U(n))\right),$$
$$s(n,\theta^*,\theta_L(n)) = c(n) - \left(\mathbf{G}\left(n,\theta^*,\theta_L(n)\right) + \eta(n,\theta^*,\theta_L(n))\right).$$

Since

$$\left(\mathbf{G}\left(n,\theta^*,\theta_U(n)\right) + \eta(n,\theta^*,\theta_U(n))\right) > 0 \quad \text{and} \quad \left(\mathbf{G}\left(n,\theta^*,\theta_L(n)\right) + \eta(n,\theta^*,\theta_L(n))\right) > 0,$$

we have

$$s(n,\theta^*,\theta_U(n)) < c(n), \qquad s(n,\theta^*,\theta_L(n)) < c(n). \tag{16}$$

Because of $\|\theta^* - \theta_U(n)\|^2 + \|\theta^* - \theta_L(n)\|^2 \ge \|\theta_U(n) - \theta_L(n)\|^2 - H$ the following implications are true:

$$\begin{aligned}
&\left(\left\|\theta^* - \theta_L(n)\right\|^2 \le (\delta^*/2)^2\right) \,\&\, \left(\left\|\theta^* - \theta_U(n)\right\|^2 \le (\delta^*/2)^2\right) \\
&\Rightarrow \left(\left\|\theta^* - \theta_U(n)\right\|^2 + \left\|\theta^* - \theta_L(n)\right\|^2 \le (\delta^*)^2/2\right), \\
&\left(H \le (\delta^*)^2/2\right) \\
&\Rightarrow \left(\left\|\theta^* - \theta_U(n)\right\|^2 + \left\|\theta^* - \theta_L(n)\right\|^2 \ge \|\theta_U(n) - \theta_L(n)\|^2 - H \ge \|\theta_U(n) - \theta_L(n)\|^2 - (\delta^*)^2/2\right) \\
&\Rightarrow \left\{\left(\left\|\theta^* - \theta_U(n)\right\|^2 + \left\|\theta^* - \theta_L(n)\right\|^2 \le (\delta^*)^2/2\right)\right. \\
&\left.\Rightarrow \left(\|\theta_U(n) - \theta_L(n)\|^2 - (\delta^*)^2/2 \le (\delta^*)^2/2\right)\right\} \\
&\Rightarrow \left(\|\theta_U(n) - \theta_L(n)\|^2 \le (\delta^*)^2\right) \Rightarrow \left(\|\theta_U(n) - \theta_L(n)\| \le \delta^*\right).
\end{aligned}$$

Further taking (4) into account,

$$\omega_U(1,n) \subset \omega_U(4,n), \qquad \omega_L(1,n) \subset \omega_L(4,n). \tag{17}$$

Using the triangle inequality and taking (17) into account,

$$\begin{aligned}
&\omega_U(1,n)\cdot\omega_L(1,n) \subset \omega_U(4,n)\cdot\omega_L(4,n) \subset \omega(*,n) \\
&\quad \subset \omega(5,n) : \left\{\left\|\theta^* - \theta_U(n)\right\|^2 + \left\|\theta^* - \theta_L(n)\right\|^2 \le (\delta^*)^2/2\right\} \\
&\quad \subset \omega(6,n) : \left\{\|\theta_U(n) - \theta_L(n)\| \le \delta^*\right\} \subset \omega(7,n) : \{\tau \le n\}.
\end{aligned}$$

Assume the following notation for the events:

$$\begin{aligned}
&\omega_U(8,n) : \left\{\eta(n,\theta^*,\theta_U(n)) + s(n,\theta^*,\theta_U(n)) \le g(\delta^*/2)^2 - c(n)\right\} \\
&\omega_L(8,n) : \left\{\eta(n,\theta^*,\theta_L(n)) + s(n,\theta^*,\theta_L(n)) \le g(\delta^*/2)^2 - c(n)\right\} \\
&\omega_U(9,n) : \left\{\left|\eta(n,\theta^*,\theta_U(n))\right| + c(n) \le g(\delta^*/2)^2 - c(n)\right\} \\
&\omega_L(9,n) : \left\{\left|\eta(n,\theta^*,\theta_L(n))\right| + c(n) \le g(\delta^*/2)^2 - c(n)\right\} \\
&\bar{\omega}_U(9,n) : \left\{\left|\eta(n,\theta^*,\theta_U(n))\right| > g(\delta^*/2)^2 - 2c(n)\right\} \\
&\bar{\omega}_L(9,n) : \left\{\left|\eta(n,\theta^*,\theta_L(n))\right| > g(\delta^*/2)^2 - 2c(n)\right\}.
\end{aligned}$$

Using Boole's inequality and (16),

$$\begin{aligned}
&\forall n > \rho : P_{\theta^*}\left(\omega_U(1,n)\cdot\omega_L(1,n)\right) = P_{\theta^*}\left(\omega_U(8,n)\cdot\omega_L(8,n)\right) \ge P_{\theta^*}\left(\omega_U(9,n)\cdot\omega_L(9,n)\right) \\
&\quad \ge 1 - \left(P_{\theta^*}\left(\bar{\omega}_U(8,n)\right) + P_{\theta^*}\left(\bar{\omega}_L(8,n)\right)\right) \ge 1 - \left(P_{\theta^*}\left(\left|\eta(n,\theta^*,\theta_U(n))\right| > z\right)\right. \\
&\qquad \left. + P_{\theta^*}\left(\left|\eta(n,\theta^*,\theta_L(n))\right| > z\right)\right).
\end{aligned} \tag{18}$$

Taking the condition of the considered theorem into account:

$$\forall \left(n \geq 1; \theta, \theta' \in \Theta\right) : \mathbf{E}_{\theta^*} \left(\eta^4(n, \theta, \theta')\right) \leq \mathbf{E}_{\theta^*} \left(\sup_{\theta_1, \theta_2 \in \Theta} (\eta^4(n, \theta, \theta'))\right) \leq \Pi n^{-p}.$$

Further, using Chebyshev's inequality

$$\forall \left(n \geq 1; \theta, \theta' \in \Theta\right) : P_{\theta^*} \left(\left|\eta(n, \theta, \theta')\right| > z\right) \leq \mathbf{E}_{\theta^*} \left(\eta^4(n, \theta, \theta')\right) z^{-4} \leq \Pi n^{-p} z^{-4}. \tag{19}$$

It follows from (18) that for any $\alpha \in ]0, 1[$

$$\left(P_{\theta^*} \left(\omega_U(1, n) \cdot \omega_L(1, n)\right) \geq \alpha\right) \Rightarrow \left(P_{\theta^*} \left(\tau \leq n\right) \geq \alpha\right) \Rightarrow \left(P_{\theta^*} \left(\tau > n\right) < 1 - \alpha\right), \quad n \geq 1.$$

In this case, using (18) and (19) when $n > \rho$

$$P_{\theta^*} \left(\tau > n\right) < P_{\theta^*} \left(\left|\eta(n, \theta^*, \theta_U(n))\right| > z\right) + P_{\theta^*} \left(\left|\eta(n, \theta^*, \theta_L(n))\right| > z\right) < 2\Pi n^{-p} z^{-4}.$$

And taking the condition of the theorem into account, we can say that

$$\mathbf{E}_{\theta^*} \tau = \sum_{n \geq 1} P_{\theta^*} \left(\tau > n\right) \leq \rho + \sum_{n \geq 1} 2\Pi n^{-p} z^{-4} \leq \rho + 2\Pi W z^{-4}.$$

Hence, the second statement of the theorem is proved. It also follows that

$$\sum_{n \geq 1} P_{\theta^*} \left(\tau > n\right) < \infty.$$

Using the Borel–Cantelli Lemma, we may conclude that

$$\mathbf{P}_{\theta^*} \left(\tau < \infty\right) = 1$$

and the first statement of the theorem is proven. ■

## References


[1] T.W. Anderson, J. Taylor, Strong consistency of least squares estimation in dynamic models, The Annals of Statistics 7 (1979) 484–489.

[2] S.W. Dharmadhikari, K. Jogdeo, Bounds on moments of sums of random variables, Annals of Mathematical Statistics 40 (1969) 1506–1509.

[3] A.Ya. Dorogovtsev, The Theory of Estimates of the Parameters of Random Processes, Vyshcha shkola, Kiev, 1982 (in Russian).

[4] W. Feller, An Introduction to Probability Theory and Its Applications, 3rd ed., vol. 1, Wiley, New York, 1968, pp. 243–245. (The strong law of large numbers, Section 10.7).

[5] I. Hu, Strong consistency of Bayes estimates in stochastic regression models, Journal of Multivariate Analysis 57 (1996) 215–227.

[6] R.I. Jennrich, Asymptotic properties of non-linear least squares estimators, Annals of Mathematical Statistics 40 (1969) 633–643.

[7] T.L. Lai, Asymptotic properties of nonlinear least squares estimates in stochastic regression models, The Annals of Statistics 22 (4) (1994) 1917–1930.

[8] L. Ljung, System Identification—Theory for User, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 1999.

[9] H. Rademacher, Uber partielle und totale Dierenzierbarkeit von Funktionen mehrerer Variabeln und uber die Transformation der Doppelintegrale, Mathematische Annalen 79 (1919) 340–359.

[10] E. Rosenwasser, R. Yusupov, Sensitivity of Automatic Control Systems, CRC Press, 2000.

[11] K. Skouras, Strong consistency in nonlinear stochastic regression models, The Annals of Statistics 28 (3) (2000) 871–879.

[12] A.V. Timofeev, Non-asymptotic solution of confidence-estimation parameter task of a non-linear regression by means of sequential analysis, Problem of Control and Information Theory 20 (5) (1991) 341–351.

[13] A.V. Timofeev, Non-asymptotic confidence estimation of nonlinear regression parameters: A sequential analysis approach, Automation and Remote Control 58 (1) (1997) 1611–1616 (10).

[14] C.F. Wu, Asymptotic theory of nonlinear least squares estimation, The Annals of Statistics 9 (1981) 501–513.