

ON ASYMPTOTICALLY EFFICIENT ESTIMATION OF REGRESSION COEFFICIENTS

MATHEMATICS

1969

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196901.66495>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

UDC 519.281.2

MATHEMATICS

Yu. A. ROZANOV, M. V. KOZLOV

ON ASYMPTOTICALLY EFFICIENT ESTIMATION OF REGRESSION COEFFICIENTS

(Presented by Academician A. N. Kolmogorov on 31 I 1969)

Let $\xi(t) = \theta(t) + \eta(t)$ be a Gaussian process with discrete time, representing the sum of a real function $\theta(t) = \alpha_1\theta_1(t) + \dots + \alpha_p\theta_p(t)$ from the linear span of known functions $\theta_1(t), \dots, \theta_p(t)$ and a real stationary process $\eta(t)$ with zero mean. It is required to estimate the unknown coefficients $\alpha_1, \dots, \alpha_p$ from observation of the process $\xi(t)$ on the interval $0 \leq t \leq n$. As is known, for the parameters $\alpha_1, \dots, \alpha_p$ there exist best unbiased estimates, coinciding with maximum-likelihood estimates (see, for example, ⁽¹⁾). However, constructing the best estimates requires knowledge of the correlation function $B(t)$ (or the spectral measure $F(d\lambda)$) of the process $\eta(t)$; moreover, the corresponding computations are connected with inversion of the matrix $\{B(t-s), t, s = 0, 1, \dots, n\}$, and this substantially complicates the practical use of the best estimates for large n . In the present work a new class of unbiased estimates for the coefficients $\alpha_1, \dots, \alpha_p$, proposed by one of the authors in ⁽²⁾, is considered. Namely, as estimates $\hat{\alpha}_1, \dots, \hat{\alpha}_p$ one proposes maximum-likelihood estimates computed under the assumption that the Gaussian stationary process $\eta(t)$ has as its spectral density an a priori chosen function $g(\lambda)$. In what follows $g(\lambda)$ will be called a pseudospectral density, and the corresponding estimates $\hat{\alpha}_1, \dots, \hat{\alpha}_p$ pseudobest. Below the asymptotic properties of pseudobest estimates with respect to the true probability distribution (corresponding to the true spectral density $f(\lambda)$) are studied.

Let us indicate explicit expressions for the pseudobest estimates. Let

$$\psi_j(\lambda) = \sum_{0 \leq s \leq n} c_j(s) e^{i\lambda s}, \quad j = 1, \dots, p,$$

be solutions of the equations

$$\int e^{-i\lambda t} \psi_j(\lambda) g(\lambda) d\lambda = \theta_j(t), \quad 0 \leq t \leq n. \quad (1)$$

Denote by $\{\psi_j^*, j = 1, \dots, p\}$ the system conjugate to $\{\psi_j, j = 1, \dots, p\}$ in the Hilbert space $L^2(g)$ of functions on the interval $[-\pi, \pi]$ with scalar product

$\langle \varphi, \psi \rangle_g = \int \varphi(\lambda) \overline{\psi(\lambda)} g(\lambda) d\lambda$. The observed random process is harmonizable ($\xi(t) = \int e^{i\lambda t} \Phi(d\lambda)$) with stochastic measure $\Phi(d\lambda) = \Psi(d\lambda) + [\alpha_1 \psi_1(\lambda) + \dots + \alpha_p \psi_p(\lambda)] \times g(\lambda) d\lambda$, where $\Psi(d\lambda)$ is the stochastic spectral measure of the process $\eta(t)$. The pseudobest estimates can be represented in the following spectral form:

$$\hat{\alpha}_j = \int \psi_j^*(\lambda) \Phi(d\lambda), \quad j = 1, \dots, p. \quad (2)$$

We note that for $g(\lambda) \equiv 1$ we obtain the classical least-squares estimates, and for $g(\lambda) \equiv f(\lambda)$, the truly best unbiased estimates.

From a practical point of view, the most interesting case is when pseudospectral density is chosen in the form

$$g(\lambda) = \frac{1}{2\pi} |Q(e^{i\lambda})|^{-2}, \quad Q(z) = \sum_{0 \leq k \leq m} g_k z^k, \quad (3)$$

where the polynomial (with real coefficients) $Q(z)$ has no zeros in the disk $|z| \leq 1$. Let Θ_n be the $(n+1) \times p$ -matrix whose columns are the values of the functions $\theta_1(t), \dots, \theta_p(t)$, $0 \leq t \leq n$. Introduce the shift operator Δ ($\Delta\theta(t) = \theta(t+1)$) and put

$$Q(\Delta) = \sum_{0 \leq k \leq m} q_k \Delta^k, \quad Q(\Delta^{-1}) = \sum_{0 \leq k \leq m} q_k \Delta^{-k}.$$

Then

$$\hat{\alpha}^{(n)} = \{\Theta_n^* [Q(\Delta)Q(\Delta^{-1})\Theta_n]\}^{-1} [Q(\Delta)Q(\Delta^{-1})\Theta_n] \xi^{(n)}, \quad (4)$$

where $\hat{\alpha}^{(n)}$ denotes the vector $(\hat{\alpha}_1, \dots, \hat{\alpha}_p)$, $\xi^{(n)} = (\xi(0), \xi(1), \dots, \xi(n))$, and the operator $Q(\Delta)Q(\Delta^{-1})$ is applied to the functions $\theta_j(t)$, previously continued from the interval $0 \leq t \leq n$ to the whole axis by formulas (1). We note that for $n \geq m$ the indicated continuation is easily carried out by means of linear recurrence relations:

$$\theta(-j) = -\frac{1}{q_0} \sum_{1 \leq k \leq m} q_k \theta(k-j),$$

$$\theta(n+j) = -\frac{1}{q_0} \sum_{1 \leq k \leq m} q_k \theta(n+j-k), \quad j > 0.$$

From expression (4) it is easy to see that the number of operations necessary for computing $\hat{\alpha}^{(n)}$ (for fixed n) is a linear function of n . If, however, the

observations are made sequentially, then it is expedient at each computational step to keep in memory the $p \times p$ -matrix $\Theta_n^*[Q(\Delta)Q(\Delta^{-1})\Theta_n]$ and the p -vector $[Q(\Delta)Q(\Delta^{-1})\Theta_n]\xi^{(n)}$. In this case, increasing the observation interval by 1 leads to additional computations of bounded complexity (depending only on p), and as a result the number of operations of such a computational process is also a linear function of n .

We shall assume that the true spectral density $f(\lambda)$ and the chosen pseudospectral density $g(\lambda)$ are continuous and positive, and that the functions $\theta_1(t), \dots, \theta_p(t)$ satisfy the conditions:

$$\sum_{0 \leq t \leq n} |\theta_j(t)|^2 \rightarrow \infty, \quad n \rightarrow \infty, \quad j = 1, \dots, p; \quad (5)$$

$$\max_{0 \leq t \leq n} |\theta_j(t)| = o \left(\sum_{0 \leq t \leq n} |\theta_j(t)|^2 \right)^{1/2}, \quad (6)$$

for all s there exists the limit

$$\lim_{n \rightarrow \infty} \sum_{0 \leq t \leq n} \theta_k(t+s)\theta_j(t) / \left(\sum_{0 \leq t \leq n} |\theta_k(t)|^2 \sum_{0 \leq t \leq n} |\theta_j(t)|^2 \right)^{1/2} = R_{kj}(s),$$

$$\det\{R_{kj}(0)\} > 0. \quad (7)$$

As is known, the matrix function $R(s)$ admits the spectral representation $R(s) = \int e^{i\lambda s} H^0(d\lambda)$ with a positive definite matrix measure $H^0(d\lambda) = \{H_{kj}^0(d\lambda)\}$ (usually called the regression measure, see (1)). Put $H(d\lambda) = \frac{1}{g(\lambda)} H^0(d\lambda)$, $H = \int H(d\lambda)$. In addition, introduce the normalizing diagonal matrix D_n with elements

$$\left(\sum_{0 \leq t \leq n} |\theta_j(t)|^2 \right)^{1/2}$$

on the main diagonal.

Theorem 1. For the correlation matrix $\{\sigma_{kj}^{(n)}\}$ of the pseudobest of the estimates $\hat{\alpha}_1^{(n)}, \dots, \hat{\alpha}_p^{(n)}$ the relation holds

$$\lim_{n \rightarrow \infty} D_n \{\sigma_{kj}^{(n)}\} D_n = 2\pi H^{-1} \int \frac{f(\lambda)}{g(\lambda)} H(d\lambda) H^{-1}. \quad (8)$$

Theorem 1 makes it possible to formulate conditions for the asymptotic efficiency of pseudo-best estimates, similarly to how this is done in (1) for the case of least-squares estimates ($g(\lambda) \equiv 1$). Namely, equating the asymptotic expressions (8) for the correlation matrices of the best ($g(\lambda) \equiv f(\lambda)$) and pseudo-best estimates, we arrive at the following theorem:

Theorem 2. *For the asymptotic efficiency of pseudo-best estimates corresponding to the pseudospectral density $g(\lambda)$, it is necessary and sufficient that the matrix function $\frac{f(\lambda)}{g(\lambda)}E$ (where E is the identity $p \times p$ matrix) be constant almost everywhere with respect to the matrix measure*

$$\tilde{H}(d\lambda) = H^{-1/2}H(d\lambda)H^{-1/2},$$

i.e.

$$\left[\frac{f(\lambda)}{g(\lambda)}E - C \right] \tilde{H}(d\lambda) = 0$$

for some constant matrix $C = \{c_{kj}\}$.

It is easy to see that, with a proper choice of $g(\lambda)$, the estimates $\hat{\alpha}_1^{(n)}, \dots, \hat{\alpha}_p^{(n)}$ are asymptotically efficient, whereas ordinary least-squares estimates ($g(\lambda) \equiv 1$), generally speaking, do not possess this property. We emphasize that in the practically most important case, when the measure $H^0(d\lambda)$ is concentrated at a finite number of points (for example, when $\theta_j(t) = P_j(t) \int e^{i\lambda t} m_j(d\lambda)$, where $P_j(t)$ are polynomials and $m_j(t)$ are finite measures with a finite number of jumps), in order to obtain asymptotically efficient estimates, one may take as the pseudospectral density a function $g(\lambda)$ of type (3).

We give below the proof of Theorem 1.

Lemma 1. *If the matrix measure*

$$H^{(n)}(d\lambda) = \left\{ \frac{\psi_k(\lambda)\overline{\psi_j(\lambda)}}{\|\psi_k\|_g\|\psi_j\|_g} g(\lambda) d\lambda \right\} \quad (\|\psi\|_g^2 = \langle \psi, \psi \rangle_g) \quad (9)$$

converges weakly as $n \rightarrow \infty$ to some measure $\tilde{H}(d\lambda)$, then

$$\lim_{n \rightarrow \infty} \tilde{D}_n \{\sigma_{kj}^{(n)}\} \tilde{D}_n = \tilde{H}^{-1} \int \frac{f(\lambda)}{g(\lambda)} \tilde{H}(d\lambda) \tilde{H}^{-1},$$

where $\tilde{H} = \int \tilde{H}(d\lambda)$, and \tilde{D}_n is the diagonal matrix with elements $\|\psi_j\|_g$ on the main diagonal.

The proof follows easily from the following relation for the correlation matrix $\{\sigma_{kj}^{(n)}\}$ of the pseudo-best estimates:

$$\{\sigma_{kj}^{(n)}\} = \{\langle \psi_k, \psi_j \rangle_f\} = \{\langle \psi_k, \psi_j \rangle_g\}^{-1} \{\langle \psi_k, \psi_j \rangle_f\} \{\langle \psi_k, \psi_j \rangle_g\}^{-1}.$$

Lemma 2. For $g(\lambda)$ of the form (3), the matrix measure (9) converges weakly to

$$\{\tilde{H}_{kj}(d\lambda)\} = \{H_{kk}^{-1/2} H_{kj}(d\lambda) H_{jj}^{-1/2}\}.$$

For the proof one must use the fact that the solutions of equations (1) can be represented in the form (see (3))

$$\psi_j^{(n)}(\lambda) = \tilde{\theta}_j^{(n)}(\lambda) |Q(e^{i\lambda})|^2 + R_j^{(n)}(\lambda),$$

where

$$\tilde{\theta}_j^{(n)}(\lambda) = \sum_{0 \leq t \leq n} e^{i\lambda t} \theta_j(t), \quad \max |R_j^{(n)}(\lambda)| = o \left(\sum_{0 \leq t \leq n} |\theta_j(t)|^2 \right)^{1/2}.$$

From Lemmas 1 and 2 follows the validity of Theorem 1 for functions $g(\lambda)$ having the special form (3). To obtain the assertion of the theorem in the general case, it suffices to show that the matrix measure (9) for an arbitrary continuous positive $g(\lambda)$ converges weakly. The latter result is derived by uniformly approximating $g(\lambda)$ from below by functions of the form (3).

Let $g(\lambda)$ and $g'(\lambda)$ be continuous, positive, $0 < m \leq g'(\lambda) \leq g(\lambda) \leq M$, and let

$$\psi_j(\lambda) = \sum c_j(s) e^{i\lambda s}, \quad \psi'_j(\lambda) = \sum c'_j(s) e^{i\lambda s}$$

be the corresponding solutions of equations (1), i.e., for $0 \leq t \leq n$,

$$\int e^{-i\lambda t} \psi'_j(\lambda) g'(\lambda) d\lambda = \int e^{-i\lambda t} \psi_j(\lambda) g(\lambda) d\lambda \quad (= \theta_j(t)) \quad (10)$$

or

$$\sum c'_j(s) b'(t-s) = \sum c_j(s) b(t-s) = \theta_j(t), \quad (11)$$

where $2\pi b(t)$, $2\pi b'(t)$ are the Fourier coefficients of the functions $g(\lambda)$, $g'(\lambda)$.

To estimate the discrepancy of the matrix measures (9) corresponding to $g(\lambda)$ and $g'(\lambda)$, consider

$$\int e^{i\lambda t} [\psi_j(\lambda) \overline{\psi_k(\lambda)} g(\lambda) - \psi'_j(\lambda) \overline{\psi'_k(\lambda)} g'(\lambda)] d\lambda. \quad (12)$$

Adding and subtracting from (12) the expressions $\psi_j(\lambda)\overline{\psi'_k(\lambda)}g(\lambda)$ and $\psi_j(\lambda)\overline{\psi'_k(\lambda)}g'(\lambda)$, we split the integral (12) into three parts, each of which is estimated separately using Cauchy's inequality and the relations (10); the latter are applied after replacing $\psi'_k(\lambda)$ (or $\psi_j(\lambda)$) by their representation as linear combinations of $e^{i\lambda s}$. As a result, the modulus of the integral (12) is bounded above by

$$M \left(\sum^* |c_j(s)|^2 \right)^{1/2} (\|\psi_k\| + \|\psi'_k\|) + M \left(\sum^* |c'_k(s)|^2 \right)^{1/2} (\|\psi_j\| + \|\psi'_j\|) + \left(1 + \frac{1}{\sqrt{2\pi}} \right) \max |g(\lambda) - g'(\lambda)| (\|\psi_k\| + \|\psi'_k\|) (\|\psi_j\| + \|\psi'_j\|), \quad (13)$$

where the summation \sum^* is over all s such that $0 \leq s \leq t-1$, $n-t+1 \leq s \leq n$, and

$$\|\psi\|^2 = \int |\psi(\lambda)|^2 d\lambda.$$

Passing in the right-hand equality (11) to the matrix inverse to $\{b(t-s)\}$, and using Cauchy's inequality, it is easy to show that

$$\left(\sum |c_k(s) - c'_k(s)|^2 \right)^{1/2} \leq \frac{1}{m^2} [g(\lambda) - g'(\lambda)] \left(\sum |c_k(s)|^2 \right)^{1/2}. \quad (14)$$

It is somewhat more difficult to prove that, for each fixed t ,

$$c_j(t), \quad c_j(n-t) = o\left(\|\tilde{\theta}_j^{(n)}\|\right). \quad (15)$$

Together with (14) and (15), estimate (13) makes it possible to establish the existence of the corresponding limit for the matrix measure (9).

Mathematical Institute named after V. A. Steklov
Academy of Sciences of the USSR
Moscow

Moscow State University
named after M. V. Lomonosov

Received
9 I 1969

CITED LITERATURE

1. U. Grenander, M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, N. Y., 1957.
2. Yu. A. Rozanov, II International Symposium on Multivariate Analysis, Dayton, 1968.

3. V. F. Pisarenko, Yu. A. Rozanov, *Problems of Information Transmission*, vol. 14, 1963.

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.