

ON THE APPLICATION OF THE MAXIMUM LIKELIHOOD METHOD FOR PROCESSING RANDOM GAUSSIAN PROCESSES

Yu. S. Tumashev

1970

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-197001.76060>

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

Abstract

Full Text

UDC 519.281.2

MATHEMATICS

Yu. S. TUMASHEV

ON THE APPLICATION OF THE MAXIMUM LIKELIHOOD METHOD FOR PROCESSING RANDOM GAUSSIAN PROCESSES

(Presented by Academician B. N. Petrov, October 18, 1968)

Usually, for processing a finite number of measurements with random errors, the maximum likelihood method is used ^(1,2). In the case when the set of measured quantities can be regarded as a random vector with a Gaussian (normal) probability density distribution, the application of this method generally requires knowledge of the matrix inverse to the correlation matrix of the errors. In passing from discrete measurements to continuous ones, the random vector of measurements becomes a random Gaussian process, and the correlation matrix of measurement errors becomes a correlation function. Therefore, it is impossible to carry out a direct generalization of the maximum likelihood method for processing random Gaussian processes.

The results obtained in this paper make it possible, when processing random Gaussian processes, to find estimates of the parameters to be determined from equations analogous to the likelihood equations for processing discrete measurements.

§ 1. Let the measured quantity $y(t)$ be related to the parameters to be determined $c_1 \dots c_n$ by the equation

$$y(t) = \varphi(c, t). \quad (1)$$

The function $\varphi(c, t)$ is continuous together with its first derivatives with respect to $c_1 \dots c_n$.

Measurements of $y(t)$ are made at the instants t_i ($i = 1, \dots, N$) on the interval $[0T]$. The sample of measurements $y_{\text{meas}}(t)$ is a random N -dimensional vector. It is assumed that the correlation matrix $K(\delta y)$ is known. The application of the maximum likelihood method to the processing of $y_{\text{meas}}(t)$ leads to the solution of a nonlinear system of equations with respect to the estimated parameters c

$$\frac{\partial \varphi^T}{\partial c} K^{-1}(\delta y)(y_{\text{meas}} - \varphi) = 0. \quad (2)$$

Here

$$\frac{\partial \varphi^T}{\partial c} = \begin{bmatrix} \frac{\partial \varphi}{\partial c_1}(t_1) & \dots & \frac{\partial \varphi}{\partial c_1}(t_N) \\ \dots & \dots & \dots \\ \frac{\partial \varphi}{\partial c_n}(t_1) & \dots & \frac{\partial \varphi}{\partial c_n}(t_N) \end{bmatrix}; \quad \varphi = \begin{bmatrix} \varphi(ct_1) \\ \vdots \\ \varphi(ct_N) \end{bmatrix}; \quad y_{\text{meas}} = \begin{bmatrix} y_{\text{meas}}(t_1) \\ \vdots \\ y_{\text{meas}}(t_N) \end{bmatrix}$$

are N -dimensional vectors.

In the theory of random functions, all statistical properties of Gaussian processes are completely determined by specifying the correlation function $R(t, \tau)$. This function can always be approximated on the measurement interval with any prescribed accuracy by a set of linearly independent functions $a_i(t)$ ($i = 1, \dots, m$)

$$R(t, \tau) = \mathbf{a}^T(t) H \mathbf{a}(\tau). \quad (3)$$

Here $\mathbf{a}^T(t) = [a_1(t), \dots, a_m(t)]$; $\mathbf{a}(\tau) = \begin{bmatrix} a_1(\tau) \\ \vdots \\ a_m(\tau) \end{bmatrix}$, and H is a nonsingular constant matrix defined for the given correlation function and the basis set $a_i(t)$.

For values of the random function $y_{n3}(t)$ at discrete instants t_i ($i = 1, \dots, N$), an approximate correlation matrix of the random vector $\mathbf{y}_{n3}$ can be obtained from (3). We write the approximate correlation matrix for the sample $\mathbf{y}_{n3}$ in the form

$$K_0(\delta y) = A^T H A, \quad (4)$$

where

$$A = \begin{bmatrix} a_1(t_1) & \dots & a_1(t_N) \\ \vdots & \ddots & \vdots \\ a_m(t_1) & \dots & a_m(t_N) \end{bmatrix}.$$

The rank of the matrix A is equal to m . This follows from the linear independence of the functions $a_i(t)$ and from the fact that $t_i \neq t_j$ for $i \neq j$. As shown in (3), the rank of the matrix $K_0(\delta y)$ must not exceed m . Since $m < N$, the matrix (4) cannot be used directly in system (2).

Transform the vectors $\mathbf{y}_{n3}$ and φ according to the formulas

$$\mathbf{z}_{\text{и3}} = M\mathbf{y}_{\text{и3}}, \quad \varphi_1(c) = M\varphi(c). \quad (5)$$

The correlation matrices of the random vectors $\mathbf{y}_{\text{и3}}$ and $\mathbf{z}_{\text{и3}}$ are related by

$$K(\delta z) = MK(\delta y)M^T \quad (6)$$

and, consequently,

$$K_0(\delta z) = MA^T HAM^T. \quad (6')$$

Consider the random vector $\mathbf{z}_{\text{и3}}$ with fictitious statistical characteristics determined by the pseudocorrelation matrix according to the formula

$$K(\delta z) = K_0(\delta z) + \lambda\varepsilon, \quad (7)$$

where $K_0(\delta z)$ is determined by (6), and ε is a nonsingular matrix.

We formulate for the pencil of matrices (7) a lemma, without giving its proof.

Lemma. Let K_0 be a singular matrix, and let the matrix ε be nonsingular. There exists a δ -neighborhood of the point $\lambda = 0$ in which, except for the point itself, all matrices of the pencil (7) are nonsingular.

Without loss of generality, one may assume that $\varepsilon \equiv E$. Consequently, the pseudocorrelation matrix (7) can be used in the likelihood equations.

§ 2. After substituting (7) into them and performing some transformations, the likelihood equations are expanded in a series in λ . In the resulting expansions the limiting transition λ to zero is carried out.

Introduce the matrix X by the formula

$$\frac{\partial \varphi_1^T}{\partial c} K^{-1}(\delta z) = X. \quad (8)$$

Substituting the expression for $K(\delta z)$ from (7) and (6') into (8), we obtain

$$XMA^T HAM^T + \lambda X = \frac{\partial \varphi_1^T}{\partial c}. \quad (9)$$

The linear system of equations (9) is solved with respect to X by a method analogous to the method for solving Fredholm integral equations of the second kind with degenerate kernels. As a result we obtain

$$X = \frac{1}{\lambda} \frac{\partial \varphi_1^T}{\partial c} - \frac{1}{\lambda} \frac{\partial \varphi_1^T}{\partial c} M A^T (H A M^T M A^T + \lambda E)^{-1} H^{-1} A M^T. \quad (10)$$

Using (5), expand (10) in a series in λ . The likelihood equations are represented in the form

$$\frac{1}{\lambda} \frac{\partial \varphi^T}{\partial c} [M^T M - M^T M A^T (A M^T M A^T)^{-1} A M^T M + \lambda M^T M A^T (A M^T M A^T)^{-1} \times \times H^{-1} (A M^T M A^T)^{-1} A M^T M$$

Consider a specific form of the nonsingular transformation M

$$M = A^T (A A^T)^{-1} A + \lambda E. \quad (12)$$

Here λ is a parameter identical with the parameter λ in (11). The likelihood equations (11), using (12), shall be represented in the form of a series in λ , restricting ourselves to the linear term of the expansion.

$$\frac{\partial \varphi^T}{\partial c} \{A^T (A A^T)^{-1} H^{-1} (A A^T)^{-1} A + \lambda [E - A^T (A A^T)^{-1} A - A^T (A A^T)^{-1} H^{-1} (A A^T)^{-1} H^{-1} (A A^T)^{-1} A]\} (\mathbf{y} -$$

Three cases are possible for the approximate representation of the correlation function in the form (3). The first case is when the number of terms in the expansion m is greater than the number of parameters being determined n ; the second case is $m < n$, and the third case is $m = n$.

§ 3. First case $m > n$.

Since $N \gg m$ and the functions $a_i(t)$ are linearly independent, the rank of the matrix

$$\frac{\partial \varphi^T}{\partial c} A^T (A A^T)^{-1}$$

is equal to the number of parameters being determined. Therefore the Jacobian of the transformation for c , determined by system (13) at $\lambda = 0$, is different from zero, i.e., the condition holds

$$\det \frac{\partial \varphi^T}{\partial c} A^T (A A^T)^{-1} H^{-1} (A A^T)^{-1} A \frac{\partial \varphi}{\partial c} \neq 0.$$

Consequently, the likelihood equations for this case take the form

$$\frac{\partial \varphi^T}{\partial c} A^T (AA^T)^{-1} H^{-1} (AA^T)^{-1} A (\mathbf{y} - \varphi) = 0. \quad (14)$$

The correlation matrix of the parameters estimated by formula (14) has the form

$$K(\delta c) = \left[\frac{\partial \varphi^T}{\partial c} A^T (AA^T)^{-1} H^{-1} (AA^T)^{-1} A \frac{\partial \varphi}{\partial c} \right]^{-1}. \quad (15)$$

In equations (14) and (15) a direct limiting transition to continuous measurement is possible.

Second case $m < n$. Since $N \gg m$, the rank of the matrix

$$\frac{\partial \varphi^T}{\partial c} A^T (AA^T)^{-1}$$

is equal to m , and the Jacobian of the transformation for c , determined by the first term of the expansion of system (13) in λ , is equal to zero, i.e.,

$$\det \left[\frac{\partial \varphi^T}{\partial c} A^T (AA^T)^{-1} H^{-1} (AA^T)^{-1} A \frac{\partial \varphi}{\partial c} \right] = 0.$$

We shall now show that the Jacobian of the transformation for c , determined by system (13), is different from zero. For this it is sufficient to show the nonsingularity of the matrix

$$E - A^T (AA^T)^{-1} A - A^T (AA^T)^{-1} H^{-1} (AA^T)^{-1} H^{-1} (AA^T)^{-1} A.$$

Indeed, by direct verification it is easy to show that the matrix

$$E - A^T (AA^T)^{-1} A - A^T H A A^T H A$$

is, for

$$E - A^T (AA^T)^{-1} A - A^T (AA^T)^{-1} H^{-1} (AA^T)^{-1} H^{-1} (AA^T)^{-1} A$$

the inverse.

Multiply system (13) from the left by the nonsingular matrix

$$\frac{1}{\lambda} E - \frac{1}{\lambda} \frac{\partial \varphi^T}{\partial c} A^T \left(A \frac{\partial \varphi}{\partial c} R \frac{\partial \varphi^T}{\partial c} \right)^{-1} A \frac{\partial \varphi}{\partial c} R^{-1} + \frac{\partial \varphi^T}{\partial c} A^T \left(A \frac{\partial \varphi}{\partial c} R^{-1} \frac{\partial \varphi^T}{\partial c} A^T \right)^{-1} A A^T H A A^T \left(A \frac{\partial \varphi}{\partial c} R^{-1} \frac{\partial \varphi^T}{\partial c} A^T \right)$$

where

$$R = \frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} - \frac{\partial \varphi^T}{\partial c} A^T (AA^T)^{-1} \frac{\partial \varphi}{\partial c}.$$

Then the likelihood equations take the form

$$\left\{ \frac{\partial \varphi^T}{\partial c} [E - A^T (AA^T)^{-1} A] + \frac{\partial \varphi^T}{\partial c} A^T \left[A \frac{\partial \varphi}{\partial c} \left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right)^{-1} \frac{\partial \varphi^T}{\partial c} A^T \right]^{-1} \right\} \times \times A \left[E - \frac{\partial \varphi}{\partial c} \left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right)^{-1} \frac{\partial \varphi^T}{\partial c} \right]$$

System (16) has a unique solution for c . For this it is sufficient to show that the matrix

$$\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} - \frac{\partial \varphi^T}{\partial c} A^T (AA^T)^{-1} A \frac{\partial \varphi}{\partial c}$$

has an inverse. Indeed, such a matrix is

$$\left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right)^{-1} + \left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right)^{-1} \times \times \frac{\partial \varphi}{\partial c} A^T \left[AA^T - A \frac{\partial \varphi}{\partial c} \left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right)^{-1} \frac{\partial \varphi^T}{\partial c} A^T \right]^{-1} A \frac{\partial \varphi}{\partial c} \left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right)^{-1}.$$

The correlation matrix of the parameters c estimated by formula (16) can be represented in the form

$$K(\delta c) = B^T AA^T H AA^T B, \quad (17)$$

where

$$B = \left[A \frac{\partial \varphi}{\partial c} \left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right)^{-1} \frac{\partial \varphi^T}{\partial c} A^T \right]^{-1} A \frac{\partial \varphi}{\partial c} \left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right)^{-1}.$$

In equations (16) and (17) a passage to a continuous measurement is also possible.

Third case $m = n$. In this case the matrix $\frac{\partial \varphi^T}{\partial c} A^T$ is square and nonsingular. Its rank is equal to the rank of the matrix AA^T . Therefore the likelihood equations take the form

$$A(y_{\text{meas}} - \varphi) = 0. \quad (18)$$

The correlation matrix of the parameters estimated by formula (18) has the form

$$K(\delta c) = \left(A \frac{\partial \varphi}{\partial c} \right)^{-1} A A^T H A A^T \left(\frac{\partial \varphi^T}{\partial c} A^T \right)^{-1}. \quad (19)$$

In formulas (18), (19) a passage to a continuous measurement is possible. In this case the following limiting expressions will hold for the matrices occurring in formulas (14), (15), (16), (17), (18), (19):

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} A \frac{\partial \varphi}{\partial c} \Delta t &= \int_0^t a(t) \frac{\partial \varphi}{\partial c}(c, t) dt, \\ \lim_{\Delta t \rightarrow 0} A A^T \Delta t &= \int_0^t a(t) a^T(t) dt, \\ \lim_{\Delta t \rightarrow 0} \left(\frac{\partial \varphi^T}{\partial c} \frac{\partial \varphi}{\partial c} \right) \Delta t &= \int_0^t \frac{\partial \psi^T}{\partial c}(c, t) \frac{\partial \varphi}{\partial c}(c, t) dt, \quad (20) \\ \lim_{\Delta t \rightarrow 0} A(y_{\text{meas}} - \varphi) \Delta t &= \int_0^t a(t) [y_{\text{meas}}(t) - \varphi(c, t)] dt. \end{aligned}$$

Substituting equalities (20) into systems (14), (15), (16), (17), (18), (19), we obtain the likelihood equations for processing Gaussian processes and the estimates of the correlation matrices of the parameters being determined.

Received
2 X 1968

CITED LITERATURE

1. Yu. V. Linnik, *The Method of Least Squares and the Foundations of the Theory of Observation Processing*, Moscow, 1962.
2. H. Cramér, *Mathematical Methods of Statistics*, IL, 1948.
3. F. R. Gantmakher, *The Theory of Matrices*, "Nauka," 1966.

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

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