THEOREM 2. Suppose that the l.f. q is invariant on G^m and the class of g.B.e.'s is not empty. Then the equivariant g.B.e. of $v(x)$ relative to the l.f. q is potentially optimal in the class Q.

Proof. It is readily verified that if $v_1(x)$ is a q-unbiased estimate, then so is $\theta v_1(\theta^{-1}x)$ for every $\theta \in V^m$. Also, if $v_1(x)$ is optimal in Q, then so is $\theta v_1(\theta^{-1}x)$; this follows from the invariance of the l.f. q. Consequently, $E_\theta q(v(x), \theta) = const$ for all $\theta \in V^m$. To complete the proof, notice that $V(x)$ is a Bayesian estimate with constant risk.

LITERATURE CITED

1. A. L. Rukhin, "Estimation of the rotation parameter on the sphere," Trudy Mat. Inst. Akad. Nauk SSSR, 29, Nauka, Leningrad (1972).
2. A. L. Rukhin, "Some statistical and probabilistic problems on groups," Trudy Mat. Inst. Akad. Nauk SSSR, 111, Nauka, Moscow (1970).
3. V. N. Nikulin, "Estimation of the rotation parameter on the circle," Dep. VINITI No. 840 (1982).
4. M. M. Postnikov, Linear Algebra and Differential Geometry [in Russian], Nauka, Moscow (1979).
5. L. B. Klebanov, "On a general definition of unbiasedness," Abstracts of International Conf. on Probability Theory and Mat. Statistics [in Russian], Vilnius (1973).

A BAYESIAN METHOD OF PARAMETER IDENTIFICATION AND PREDICTION OF STATES OF LINEAR STATIONARY DYNAMICAL SYSTEMS

N. D. Pavlov

UDC 519.21

Linear discrete dynamical systems with unknown coefficients, subjected to the action of random perturbations with unknown covariance matrix, are considered. In the first stage the probability density of the unknown parameters of the given dynamical system is found using Bayesian estimation methods. In the second stage the predicted density of the distribution of the state vector over an arbitrary number of forward steps is constructed.

Suppose the discrete linearized model of a certain object is given by the equation

$$x(k) = A x(k-1) + B u(k-1) + w(k-1)$$

in which $x(k) \in \mathcal{F}_{n\times 1}$ is the state vector of the object, $u(k) \in \mathcal{F}_{m\times 1}$ is a known control vector, $A \in \mathcal{F}_{n\times n}$ and $B \in \mathcal{F}_{n\times m}$ are matrices of constant unknown parameters, subject to identification, $w(k) \in \mathcal{F}_{n\times 1}$ is a Gaussian white sequence with characteristics

$$M[w(k)] = 0, \ \forall k, \ M[w(k) w^T(j)] = \Sigma \delta(k-j),$$

where $\delta(k-j)$ is the Kronecker symbol, and Σ is the (positive definite) covariance matrix of the perturbations. We assume that the state vector $x(k)$ is measured exactly at the moments of time $k = \overline{1, N}$. Given these observations, it is required to find the probability characteristics of the matrices A and B on the interval $k = \overline{1, N}$, and then use the latter to predict the state vector over the time interval $k = \overline{N+1, N+p}$.

To determine the probability characteristics of the matrices A and B we use the method of the maximum of the a posteriori probability [1]. By the well-known Bayes formula,

$$f[X, A, B, \Sigma] = f[X | A, B, \Sigma] f[A, B, \Sigma] = f[A, B, \Sigma | X] f[X], \tag{1}$$

where $f[\cdot]$ is the probability density function (PDF) and

$$X \triangleq [x(N), x(N-1), \ldots, x(1)] \in \mathcal{F}_{n\times N}.$$

N. E. Bauman Higher Technical Institute, Moscow.

Translated from Matematicheskie Metody Otsenivaniya i Proverki Gipotez, pp. 97-102, 1984.

0090-4104/88/4101-0838$12.50

Since

$$f[X] = \int f[X \mid A, B, \Sigma] f[A, B, \Sigma] \, dA \, dB \, d\Sigma,$$

it follows from (1) that

$$f[A, B, \Sigma \mid X] \propto f[X \mid A, B, \Sigma] f[A, B, \Sigma], \tag{2}$$

where $\propto$ stands for proportionality.

Concerning the prior PDF $f[A,B,\Sigma]$, we assume that the entries of the matrices A, B, and Σ are independently distributed, i.e.,

$$f[A, B, \Sigma] = f[A] \, f[B] \, f[\Sigma].$$

It is natural to assume that we have no information whatsoever on the values of the unknown matrices; then they may be chosen as follows:

$$f[A] = const, \quad f[B] = const, \quad f[\Sigma] = |\Sigma|^{-\frac{n+1}{2}}.$$

The application of the given singular distributions in Bayesian analysis is justified in the literature [2-4]. Since $w(k)$ is a Markov sequence, it can be readily verified (see [5]) that $x(k)$, too, is a Markov sequence, i.e.,

$$f[X \mid A, B, \Sigma] = f[X(0)] \, f[X(1) \mid X(0), A, B, \Sigma] \cdot \ldots \cdot f[X(N) \mid X(N-1), A, B, \Sigma],$$

where each conditional PDF is a normal PDF. Using this remark we can reexpress (2) as

$$f[A, B, \Sigma \mid X] \propto |\Sigma|^{-\frac{n+N+1}{2}} \cdot exp\left\{-\frac{1}{2}\sum_{k=1}^{N} Y(k)\Sigma^{-1}Y^T(k)\right\}; \; Y(k) = X(k) - AX(k-1) - Bu(k-1). \tag{3}$$

To simplify the ensuing analysis, we introduce the notations

$$X_1 = [x(N-1), \ldots, x(0)] \in \mathcal{F}_{n \times N}, \quad u_1 = [u(N-1), \ldots, u(0)] \in \mathcal{F}_{m \times N},$$
$$u = \begin{bmatrix} X_1 \\ u_1 \end{bmatrix} \in \mathcal{F}_{(n+m) \times N}, \quad C = [A, B] \in \mathcal{F}_{n \times (n+m)}. \tag{4}$$

We can now rewrite (3) in the form

$$f[A, B, \Sigma \mid X] \propto |\Sigma|^{-\frac{n+N+1}{2}} exp\left\{-\frac{1}{2} tr (X - Cu)(X - Cu)^T \Sigma^{-1}\right\}, \tag{5}$$

where $tr(\cdot)$ denotes the trace of matrices. The expression under the exponential in (5) can be transformed as follows

$$(X - Cu)(X - Cu)^T = (X - \hat{C}u)(X - \hat{C}u)^T + (C - \hat{C}) u u^T (C - \hat{C})^T, \tag{6}$$

where $\hat{C} = Xu^T(uu^T)^{-1}$ is the least-square estimate of the matrix C. We remark that the first term in (6) is proportional to the sampling covariance matrix Σ of the perturbations, which permits us to put (5) in the form

$$f[A, B, \Sigma \mid X] \propto |\Sigma|^{-\frac{n+N+1}{2}} exp\left\{-\frac{1}{2} tr \, S\Sigma^{-1} - \frac{1}{2} tr (C - \hat{C}) u u^T (C - \hat{C})^T \Sigma^{-1}\right\}. \tag{7}$$

From this expression we can derive the sought-for marginal PDF through integration over all the entries of the matrix Σ:

$$f[A, B \mid X] \propto \int |\Sigma|^{-\frac{n+N+1}{2}} exp\left\{-\frac{1}{2} tr (S + (C - \hat{C}) u u^T (C - \hat{C})^T)\Sigma^{-1}\right\} d\Sigma.$$

To evaluate this integral we use properties of the inverse Wishart distribution [6], whose PDF is given by the formula

$$f[\Sigma \mid H, \nu, m] \propto \frac{|H|^{\nu/2}}{|\Sigma|^{\frac{\nu+m+1}{2}}} exp\left\{-\frac{1}{2} tr \, \Sigma^{-1} H\right\}. \tag{8}$$

Using (8), we have

$$f[A,B|X] \propto |S+(C-\hat{C})UU^T(C-\hat{C})^T|^{-\frac{N}{2}}. \tag{9}$$

The PDF (9) is referred to in the literature as the matrix or generalized Student t-PDF. Using the properties of this distribution we can obtain the PDF's for the matrices A and B. They have the form

$$f[A|X] \propto |S+(A-\hat{A})F_1(A-\hat{A})^T|^{-\frac{N-m}{2}},$$
$$f[B|X] \propto |S+(B-\hat{B})F_2(B-\hat{B})^T|^{-\frac{N-n}{2}},$$

i.e., they are matrix Student t-PDF's with

$$F_1 = X_1 R X_1^T, \quad F_2 = U_1 P U_1^T$$
$$P = I - X_1^T(X_1X_1^T)^{-1}X_1, \quad R = I - U_1^T(U_1U_1^T)^{-1}U_1,$$
$$\hat{A} = X[I-PU_1^T(U_1PU_1^T)^{-1}U_1]X_1^T(X_1X_1^T)^{-1} = XRX_1^T(X_1RX_1^T)^{-1},$$
$$\hat{B} = X[I-RX_1^T(X_1RX_1^T)^{-1}X_1]U_1^T(U_1U_1^T)^{-1} = XPU_1^T(U_1PU_1^T)^{-1}.$$

To derive the predicted PDF of the state vector for the moments of time $k=\overline{N+1, N+p}$ we assume that the observations are generated by the following model:

$$x(k) = Ax(k-1)+Bu(k-1)+w(k-1), \quad k=\overline{N+1, N+p}. \tag{10}$$

We denote

$$X_2=[x(N+p),\dots,x(N+1)] \in \mathcal{F}_{n\times p}, \quad X_{21}=[x(N+p-1),\dots,x(N)] \in \mathcal{F}_{n\times p},$$

$$U_{21}=[u(N+p-1),\dots,u(N)] \in \mathcal{F}_{m\times p}, \quad W=[w(N+p-1),\dots,w(N)] \in \mathcal{F}_{n\times p},$$

$$U_2 = \begin{bmatrix} X_{21} \\ U_{21} \end{bmatrix} \in \mathcal{F}_{(n+m)\times p}.$$

Then (10) can be rewritten as $X_2 = CU_2 + W$. Concerning the perturbation vector $w(k)$, $k=\overline{N+1, N+p}$ we assume that it has the same characteristics as $w(k)$, $k=\overline{1,N}$. Then by Bayes' theorem

$$f[X_2,X,A,B,\Sigma] = f[X_2,A,B,\Sigma|X]f[X] = f[X_2|X_1,A,B,\Sigma]f[A,B,\Sigma|X]\,f[X]$$

or, equivalently,

$$f[X_2,A,B,\Sigma|X] \propto f[X_2|X,A,B,\Sigma]\,f[A,B,\Sigma|X]. \tag{11}$$

Integrating expression (11) with respect to C and Σ, we obtain the predicted PDF for the matrix X_2, where $f[X_2|X,C,\Sigma]$, the a posteriori PDF of X_2 given C,Σ is a normal PDF:

$$\begin{gathered} f[X_2|X] = \int f[X_2|X,A,B,\Sigma]f[A,B,\Sigma|X]\,dA\,dB\,d\Sigma, \\ f[X_2|X,A,B,\Sigma] \propto |\Sigma|^{-\frac{p}{2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}(X_2-CU_2)(X_2-CU_2)^T\Sigma^{-1}\right\}, \end{gathered} \tag{12}$$

The PDF $f[A,B,\Sigma|X]$ is defined by (9). Next, using a property of the inverse Wishart distribution, we integrate expression (12) with respect to all the entries of Σ , which gives

$$|(X_2-CU_2)(X_2-CU_2)^T+(X-CU)(X-CU)^T|^{-\frac{p+N}{2}}. \tag{13}$$

To integrate expression (13) with respect to C, we isolate in (13) a perfect square in C:

$$(X_2-CU_2)(X_2-CU_2)^T+(X-CU)(X-CU)^T = X_2X_2^T+XX^T+(C-\tilde{C})M(C-\tilde{C})^T-\tilde{C}M\tilde{C}^T,$$

in which $\tilde{C}=[XU^T+X_2U_2^T]M^{-1}$ with $M=[UU^T+U_2U_2^T]$. Using a property of the matrix Student t-PDF we get

$$f[X_2|X] \propto |XX^T+X_2X_2^T-\tilde{C}M\tilde{C}^T|^{-\frac{1}{2}(p+N-m-n)}. \tag{14}$$

After a number of transformations expression (14) can be reduced to the form

$$f[x_2|x] \propto |S + (x_2 - \hat{C}u_2)D(x_2 - \hat{C}u_2)^T|^{-\frac{1}{2}(p+N-m-n)}$$

where $D = I - u_2^T M^{-1} u_2$ and $\hat{C} = X u^T (uu^T)^{-1}$. We see that the predicted PDF $f[x_2|x]$ is also a matrix Student t-PDF. We thus found the predicted PDF of the state vector using the solution of the identification problem for the matrices A and B.

LITERATURE CITED

1. E. Seth and C. White, Optimal Control Systems [Russian translation], Radio i Svyaz, Moscow (1982).
2. A. Zellner, Bayesian Methods in Econometrics [Russian translation], Statistika, Moscow (1980).
3. D. R. Cox and D. V. Hinkley, Theoretical Statistics, Chapman and Hall, London (1974).
4. M. H. de Groot, Optimal Statistical Decisions, McGraw-Hill, New York (1970).
5. J. S. Meditch, Stochastic Optimal Linear Estimation and Control, McGraw-Hill, New York (1969).
6. N. Johnson and S. Kotz, Distributions of Statistics. Continuous Multivariate Distributions (1972).

ESTIMATION OF FUNCTIONALS IN AN A PRIORI DENSITY

M. Ya. Penskaya

UDC 519.2

The problem of constructing estimates for the density of an a priori distribution is considered in papers [1-4]. In [1] the problem of estimating linear functionals in an unknown a priori distribution is solved in the discrete case. Here we are concerned with the estimation of linear and quadratic functionals of the form

$$\Phi(g) = \int_{\Theta} \varphi(\theta) g(\theta) \mu(d\theta), \tag{1}$$

and

$$\Omega(g) = \int_{\Theta} \omega(\theta) g^2(\theta) \mu(d\theta) \tag{2}$$

in an unknown a priori density.

One of the methods for obtaining the desired estimates is to estimate first $g(\theta)$ (see [1-4]) and then substitute the result in formulas (1) and (2). However, this method is not always effective since the estimation of $g(\theta)$ is rather tedious and the estimates are biased. We consider here a different approach which permits at once to estimate the functionals (1) and (2) and, in the case of $\Phi(g)$, allows us to obtain an unbiased estimate with dispersion of order N^{-1}.

Suppose that on the measurable space $(\mathcal{X}, \mathcal{A})$ there is given a family of random variables $p(x|\theta), x \in \mathcal{X}$ by means of conditional probability densities $X\{\mathcal{P}_\theta^X, \theta \in \Theta\}$. Let $\nu(dx)$ denote the measure on $\mathcal{A}$. The form of $p(x|\theta)$ is known, and the parameter θ is a random variable assuming values in a measurable space $(\Theta, \mathcal{B})$, $\mu(d\theta)$ is a given measure on $\mathcal{B}$. We let G denote the unknown a priori distribution of the quantity θ and by $g(\theta) = \frac{dG}{d\mu}(\theta)$ the probability density with respect to the given measure μ. In accordance with the a priori distribution $g(\theta)$ and the conditional distribution $p(x|\theta)$ one obtains N independent samples

$$(x_{i1}, x_{i2}, \ldots, x_{ik_i}), \quad i = \overline{1, N}, \tag{3}$$

in which x_{ij}, $i = \overline{1, N}$, $j = \overline{1, k_i}$, are known, and θ_i, $i = \overline{1, N}$, are not known. It is required to estimate the functionals (1) and (2) from the observations x_{ij}.

Perm University. Translated from Statisticheskie Metody Otsenivaniya i Proverki Gipotez, pp. 102-110, 1984.