Aplikace matematiky

Gejza Wimmer

Minimum mean square error estimation

Aplikace matematiky, Vol. 24 (1979), No. 5, 382–388

Persistent URL: http://dml.cz/dmlcz/103818

MINIMUM MEAN SQUARE ERROR ESTIMATION

GEJZA WIMMER

(Received December 6, 1977)

INTRODUCTION

Repeated determination of the regression parameters e.g. in clinical practice, demographic investigations etc., indicates that the parameters could be often considered as the realizations of random variables (provided that the repeatedly obtained regression parameters are nonhomogeneous on a statistically significant level). The aim of this paper is to show one possible way of utilizing the information obtained from previous experiments. It is assumed that the information obtained from previous experiments enables us to determine the mean value and the covariance matrix of the regression parameters and of the error vector, respectively.

I. FORMULATION OF THE PROBLEM

Let us consider the statistical model of the type

$$(1) \qquad \mathbf{x}_{n,1} = \mathbf{X}_{n,k}\boldsymbol{\beta}_{k,1} + \boldsymbol{\varepsilon}_{n,1} ,$$

where the random vector $\boldsymbol{\beta} \in \mathscr{R}^k$ (k-dimensional vector space) has a priori probability distribution with mean value $E\boldsymbol{\beta} = \bar{\boldsymbol{\beta}}$ and covariance matrix $\operatorname{cov} \boldsymbol{\beta} = E(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) . (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' = \mathbf{U}_{k,k}$. The error vector $\boldsymbol{\varepsilon}_{n,1}$ is a random vector with mean value $E\boldsymbol{\varepsilon} = \mathbf{O}$ and covariance matrix $\mathbf{V}_{n,n}$. We assume that $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are uncorelated, i.e. $E(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})\, \boldsymbol{\varepsilon}' = \mathbf{O}$.

Definition I.1. *A linear estimator*

$$(2) \qquad \tilde{\boldsymbol{\beta}}_{k,1} = \mathbf{A}_{k,n}\mathbf{x}_{n,1} + \mathbf{b}_{k,1}$$

is said to be a minimum mean square error estimator of $\boldsymbol{\beta}$*, if* $\mathbf{A}$ *and* $\mathbf{b}$ *are chosen in such a way that the matrix*

(3) $$\mathbf{R}_{k,k} = E(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'$$

has the minimum norm.

Note I.1. We consider the following norm of a positive semi-definite matrix $\mathbf{A}_{k,k}$:

$$\|\mathbf{A}_{k,k}\| = \sup_{\substack{\|\mathbf{x}\|=1\\ \|\mathbf{y}\|=1}} \mathbf{x}'\mathbf{A}\mathbf{y}\,.$$

Note I.2. The matrix $\mathbf{R}$ given by (3) is called the *risk matrix.*

II. PRELIMINARY ASSERTIONS

Lemma II.1. *Let $\mathbf{A}_{n,n}$ be a regular symmetric matrix, $\mathbf{B}_{k,k}$ a symmetric matrix and $\mathbf{C}_{n,k}$ an arbitrary matrix. Then*

$$\begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{C}(\mathbf{B} - \mathbf{C}'\mathbf{A}^{-1}\mathbf{C})^{-}_{\mathrm{sym}}\,\mathbf{C}'\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{C}(\mathbf{B} - \mathbf{C}'\mathbf{A}^{-1}\mathbf{C})^{-}_{\mathrm{sym}} \\ -(\mathbf{B} - \mathbf{C}'\mathbf{A}^{-1}\mathbf{C})^{-}_{\mathrm{sym}}\,\mathbf{C}'\mathbf{A}^{-1} & (\mathbf{B} - \mathbf{C}'\mathbf{A}^{-1}\mathbf{C})^{-}_{\mathrm{sym}} \end{bmatrix}$$

is a g-inverse (see Rao-Mitra [1], p. 20) *of the matrix*

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{B} \end{bmatrix},$$

where the symbol $\mathbf{D}^-$ means a g-inverse of the matrix $\mathbf{D}$ (defined as a matrix which satisfies the condition $\mathbf{D} = \mathbf{D}\mathbf{D}^-\mathbf{D}$), $\mathbf{S}^-_{\mathrm{sym}}$ is a symmetric g-inverse of the symmetric matrix $\mathbf{S}$ and $\mathbf{D}'$ is the transpose of the matrix $\mathbf{D}$.

Proof follows easily from the definition of the g-inverse.

Lemma II.2. *Let $\mathbf{U}_{n,k}$ and $\mathbf{V}_{n,p}$ be arbitrary matrices. Then*

$$\begin{bmatrix} \mathbf{U}^+(\mathbf{I} - \mathbf{V}\mathbf{C}^+) \\ \mathbf{C}^+ \end{bmatrix},$$

where $\mathbf{C}_{n,p} = (\mathbf{I}_{n,n} - \mathbf{U}\mathbf{U}^+)\,\mathbf{V}$, is a g-inverse of the matrix $(\mathbf{U}, \mathbf{V})$. The symbol $\mathbf{U}^+$ means $\mathbf{U}^-_{l(\mathbf{M}),m(\mathbf{N})}$ for $\mathbf{M} = \mathbf{I}$, $\mathbf{N} = \mathbf{I}$. The matrix $\mathbf{U}^-_{l(\mathbf{M}),m(\mathbf{N})}$ is a matrix satisfying the equations

$$\mathbf{M}\mathbf{U}\mathbf{U}^-\mathbf{U} = \mathbf{M}\mathbf{U}\,,\quad \mathbf{N}\mathbf{U}^-\mathbf{U}\mathbf{U}^- = \mathbf{N}\mathbf{U}^-\,,$$

$$(\mathbf{U}^-\mathbf{U})'\,\mathbf{N} = \mathbf{N}\mathbf{U}^-\mathbf{U}\,,\quad (\mathbf{U}\mathbf{U}^-)'\,\mathbf{M} = \mathbf{M}\mathbf{U}\mathbf{U}^-\,,$$

where $\mathbf{M}$ and $\mathbf{N}$ are positive semi-definite or positive definite matrices.

Proof. If $\mu(\mathbf{R})$ denotes the vector space generated by the columns of the matrix $\mathbf{R}$ and $\operatorname{Ker}(\mathbf{R}) = \{\mathbf{x} : \mathbf{Rx} = \mathbf{O}\}$, then we obtain from the properties of the matrix $\mathbf{C}^+$ (see Rao-Mitra [1])

$$\mu(\mathbf{CC}^+) = \mu(\mathbf{C}) \subset \operatorname{Ker} \mathbf{U}'$$

or, equivalently,

$$\mu(\mathbf{U}) \subset \operatorname{Ker}(\mathbf{CC}^+)' = \operatorname{Ker}(\mathbf{CC}^+),$$

which implies

$$\mathbf{CC}^+\mathbf{U} = \mathbf{O}. \tag{4}$$

Further

$$\mu(\mathbf{UU}^+\mathbf{V}) \subset \mu(\mathbf{U})$$

or, equivalently,

$$\operatorname{Ker} \mathbf{U}' \subset \operatorname{Ker}(\mathbf{UU}^+\mathbf{V})';$$

hence we obtain

$$\mu(\mathbf{CC}^+) = \mu(\mathbf{C}) \subset \operatorname{Ker} \mathbf{U}' \subset \operatorname{Ker}(\mathbf{UU}^+\mathbf{V})'$$

so that

$$\mu(\mathbf{UU}^+\mathbf{V}) \subset \operatorname{Ker}(\mathbf{CC}^+)$$

and

$$\mathbf{CC}^+\mathbf{UU}^+\mathbf{V} = \mathbf{O}. \tag{5}$$

Now we can easily see that

$$(\mathbf{U}, \mathbf{V})\begin{bmatrix} \mathbf{U}^+(\mathbf{I} - \mathbf{VC}^+) \\ \mathbf{C}^+ \end{bmatrix}(\mathbf{U}, \mathbf{V}) = (\mathbf{U}, \mathbf{V}),$$

and the proof is complete.

Corollary II.1. *If $\mathbf{U}_{k,n}$ and $\mathbf{V}_{p,n}$ are arbitrary matrices, then*

$$[\mathbf{I} - (\mathbf{K}^+)' \mathbf{V}((\mathbf{U}')^+)', (\mathbf{K}^+)'],$$

where $\mathbf{K} = [\mathbf{I} - \mathbf{U}'(\mathbf{U}')^+] \mathbf{V}'$, is a g-inverse of $\begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$.

III. SOLUTION OF THE PROBLEM

In the sequel we shall need the following matrices:

$$\mathbf{L}_{k,k+n}(\mathbf{A}) = (\mathbf{I}_{k,k} - \mathbf{A}_{k,n}\mathbf{X}_{n,k}, A_{k,n}), \tag{6}$$
$$\mathbf{K}_{k+n,k} = (\mathbf{I}, \mathbf{X}')'$$

and

$$\boldsymbol{K}^{\ddagger} = ((K')^{-}_{m(\boldsymbol{M})})' , \tag{7}$$

where

$$\boldsymbol{M}_{k+n,k+n} = \begin{bmatrix} \boldsymbol{U}_{k,k} & \boldsymbol{O}_{k,n} \\ \boldsymbol{O}_{n,k} & \boldsymbol{V}_{n,n} \end{bmatrix} . \tag{8}$$

The symbol $\boldsymbol{T}^{-}_{m(\boldsymbol{M})}$ denotes a g-inverse of the matrix $\boldsymbol{T}$ satisfying the equation $\boldsymbol{T}\boldsymbol{T}^{-}\boldsymbol{T} = \boldsymbol{T}$ and $(\boldsymbol{T}^{-}\boldsymbol{T})' \boldsymbol{M} = \boldsymbol{M}\boldsymbol{T}^{-}\boldsymbol{T}$, where $\boldsymbol{M}$ is a positive semi-definite or definite matrix.

Lemma III.1. *Vector $\boldsymbol{b}_0$ minimizes the norm of the risk matrix* (3) *if and only if*

$$\boldsymbol{b}_0 = (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\bar{\boldsymbol{\beta}} . \tag{9}$$

Proof. We have

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \boldsymbol{\beta} = \boldsymbol{A}(\boldsymbol{X}\boldsymbol{\beta} + \varepsilon) + \boldsymbol{b} - \boldsymbol{\beta} =$$
$$= \boldsymbol{A}\varepsilon + \boldsymbol{b} - (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\boldsymbol{\beta} = (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \boldsymbol{A}\varepsilon + (\boldsymbol{b} - (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\bar{\boldsymbol{\beta}}) .$$

This implies

$$\boldsymbol{R} = E(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\,(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})' = (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\boldsymbol{U}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})' + \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}' + (\boldsymbol{b} - (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\bar{\boldsymbol{\beta}})\,(\boldsymbol{b} - (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\bar{\boldsymbol{\beta}})' . \tag{10}$$

Consequently, we see that $\boldsymbol{R}$ has the minimal norm iff $\boldsymbol{b} = \boldsymbol{b}_0 = (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\bar{\boldsymbol{\beta}}$. The proof is complete.

Lemma III.2. *The relation*

$$\min \{\|(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\boldsymbol{U}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})' + \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}'\| : \boldsymbol{A} \text{ is } k \times n \text{ matrix}\} = \|(\boldsymbol{I} - \tilde{\boldsymbol{A}}\boldsymbol{X})\,\boldsymbol{U}(\boldsymbol{I} - \tilde{\boldsymbol{A}}\boldsymbol{X})' + \tilde{\boldsymbol{A}}\boldsymbol{V}\tilde{\boldsymbol{A}}'\| . \tag{11}$$

holds if and only if $\boldsymbol{L}(\tilde{\boldsymbol{A}}) = \boldsymbol{K}^{\ddagger}$, where $\boldsymbol{L}(\tilde{\boldsymbol{A}})$ and $\boldsymbol{K}^{\ddagger}$ are given by (6) *and* (7), *respectively.*

Proof. We have

$$\boldsymbol{L}(\boldsymbol{A})\,\boldsymbol{K} = \boldsymbol{I}_{k,k} \tag{12}$$

and

$$(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})\,\boldsymbol{U}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{X})' + \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}' = \boldsymbol{L}(\boldsymbol{A})\,\boldsymbol{M}\boldsymbol{L}'(\boldsymbol{A}) , \tag{13}$$

where $\boldsymbol{M}$ is given in (8).

Using (12) and (13) we obtain

$$(14)\qquad (I - AX)\,U(I - AX)' + AVA' = L(A)\,ML'(A) =$$
$$= L(A)\,(KK^{\ddagger} + (I - KK^{\ddagger}))\,M(KK^{\ddagger} + (I - KK^{\ddagger}))'\,L'(A) =$$
$$= L(A)\,KK^{\ddagger}M(KK^{\ddagger})'\,L'(A) + L(A)\,(I - KK^{\ddagger})\,M(I - KK^{\ddagger})'\,L'(A) =$$
$$= K^{\ddagger}M(K^{\ddagger})' + (L(A) - K^{\ddagger})\,M(L(A) - K^{\ddagger})',$$

because of the relations

$$KK^{\ddagger}M(I - KK^{\ddagger})' = (I - KK^{\ddagger})\,M(KK^{\ddagger})' = O$$

(see the definition of $K^{\ddagger}$).

From (14), it is seen that (11) is true if and only if $L(\tilde{A}) = K^{\ddagger}$. The proof is complete.

Corollary III.1. *The matrix* $\tilde{A}$ *from* (11) *is a solution of the equation*

$$(15)\qquad L = (I - AX, A) = ((I, X')^{-}_{m(M)})' = K^{\ddagger}.$$

Lemma III.3. *If*

$$E_{k,k} = (I + U)^{-1} + (I + U)^{-1}\,X'(XX' + V - X(I + U)^{-1}\,X')^{-}_{\text{sym}}\,X(I + U)^{-1},$$
$$B_{k,n} = -(I + U)^{-1}\,X'(XX' + V - X(I + U)^{-1}\,X')^{-}_{\text{sym}},$$
$$C_{n,n} = (XX' + V - X(I + U)^{-1}\,X')^{-}_{\text{sym}},$$
$$D_{k,k} = E + X'B' + BX + X'CX =$$
$$= (I + U)^{-1} + [((I + U)^{-1} - I)\,X'(XX' + V - X(I + U)^{-1}\,X')^{-}_{\text{sym}}\,.$$
$$.\,X((I + U)^{-1} - I)],$$

then

$$((I, X')^{-}_{m(M)})' = ((D^{+})'\,(E + BX)', (D^{+})'\,(B' + CX)').$$

Proof. A special g-inverse $((I, X')^{-}_{m(M)})'$ is the matrix $((I, X')^{-}_{l(I),m(M)})'$. From the properties of $((I, X')^{-}_{l(I),m(M)})'$ (see Rao-Mitra [1] p. 54) and from Lemma II.1 we have

$$(16)\qquad (I, X')^{-}_{l(I),m(M)} =$$
$$= \left[\begin{bmatrix} U & O \\ O & V \end{bmatrix} + \begin{bmatrix} I \\ X \end{bmatrix}(I, X')\right]^{-}\begin{bmatrix} I \\ X \end{bmatrix}\left[\begin{bmatrix} I \\ X \end{bmatrix}(I, X')\left[\begin{bmatrix} U & O \\ O & V \end{bmatrix} +\right.\right.$$
$$\left.\left. + \begin{bmatrix} I \\ X \end{bmatrix}(I, X')\right]^{-}\begin{bmatrix} I \\ X \end{bmatrix}\right]^{-}\begin{bmatrix} I \\ X \end{bmatrix} =$$

$$= \begin{bmatrix} I + U & X' \\ X & XX' + V \end{bmatrix}^{-} \begin{bmatrix} I \\ X \end{bmatrix} \left[\begin{bmatrix} I & X' \\ X & XX' \end{bmatrix} \begin{bmatrix} U + I & X' \\ X & XX' + V \end{bmatrix}^{-} \begin{bmatrix} I \\ X \end{bmatrix} \right]^{-} \begin{bmatrix} I \\ X \end{bmatrix} =$$
$$= \begin{bmatrix} E & B \\ B' & C \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} \left[\begin{bmatrix} I & X' \\ X & XX' \end{bmatrix} \begin{bmatrix} E & B \\ B' & C \end{bmatrix} \begin{bmatrix} I \\ X \end{bmatrix} \right]^{-} \begin{bmatrix} I \\ X \end{bmatrix} =$$
$$= \begin{bmatrix} E + BX \\ B' + CX \end{bmatrix} \begin{bmatrix} D \\ XD \end{bmatrix}^{-} \begin{bmatrix} I \\ X \end{bmatrix}.$$

It follows from Corollary II.1 that

$$\begin{bmatrix} D \\ XD \end{bmatrix}^{-} = \left((I - (K^+)' XD)((D')^+)', (K^+)' \right),$$

where $K = (I - D'(D')^+) D'X' = O$. Consequently, we can take $K^+ = O$.

So we have

$$\begin{bmatrix} D \\ XD \end{bmatrix}^{-} = (D^+, O) \tag{17}$$

and from (16) we conclude

$$((I, X')^{-}_{m(M)})' = ((D^+)'(E + BX)', (D^+)'(B' + CX)'). \tag{18}$$

The lemma is proved.

Corollary III.2. *The matrix*

$$\tilde{A} = (D^+)'(B' + CX)'$$

is a solution of (15).

Proof.

$$I - \tilde{A}X = I - (D^+)'(B' + CX)' X = (DD^+)' - (D^+)'(B + X'C') X =$$
$$= (D^+)'(E' + BX + X'B' + X'C'X - BX - X'C'X) = (D^+)'(E + BX)'$$

and $\tilde{A}$ is a solution of (15) according to Lemma III.3. The proof is complete.

Theorem III.1. *In the statistical model* (1), *the vector*

$$\tilde{\beta}_{k,1} = \tilde{A}x + b_0 \tag{19}$$

is the minimum mean square error estimator of β. *The risk matrix with the minimum norm is given by*

$$R = Z'UZ + \tilde{A}V\tilde{A}',$$

where

$$Z_{k,k} =$$
$$= \{(I + U)^{-1} + (I + U)^{-1} X'(XX' + V - X(I + U)^{-1} X')^{-}_{\text{sym}} X((I + U)^{-1} - I)\}.$$

$$. [(\boldsymbol{I} + \boldsymbol{U})^{-1} + ((\boldsymbol{I} + \boldsymbol{U})^{-1} - \boldsymbol{I})\, \boldsymbol{X}'(\boldsymbol{X}\boldsymbol{X}' + \boldsymbol{V} - \boldsymbol{X}(\boldsymbol{I} + \boldsymbol{U})^{-1} \boldsymbol{X}')^{-}_{\mathrm{sym}} \,.$$
$$. \boldsymbol{X}((\boldsymbol{I} + \boldsymbol{U})^{-1} - \boldsymbol{I})]^{+} \,.$$

and

$$\boldsymbol{b}_0 = (\boldsymbol{I} - \tilde{\boldsymbol{A}}\boldsymbol{X})\, \bar{\boldsymbol{\beta}} \,.$$

Proof. Due to Lemmas III.1, III.2 and Corollaries III.1 and III.2, the vector $\tilde{\boldsymbol{\beta}}_{k,1}$ (cf. (19)) is the minimum mean square error estimator of $\boldsymbol{\beta}$. The risk matrix with the minimum norm is easily obtained from (10), (11), (13) and (18). The proof is complete.

Corollary III.3. *If the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are positive definite then*

$$\tilde{\boldsymbol{A}} = \boldsymbol{U}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}' + \boldsymbol{V})^{-1} \,.$$

The above corollary has been obtained by Chipman [2]. Hence our Theorem III.1 is a generalization of Chipman's result for singular matrices $\boldsymbol{U}$ and $\boldsymbol{V}$. In practically relevant situations, the new information concerning the parameters $\boldsymbol{\beta}$ and ε usually causes a restriction of their respective ranges. Consequently, the occurence of singular matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ is quite natural in such problems.

References

[1] *C. R. Rao, S. K. Mitra:* Generalized inverse of matrices and its applications. John Wiley, New York 1971.

[2] *J. S. Chipman:* On least squares with insufficient observations. J. Amer. Statist. Assoc. 59 (1964), 1078—1111.

Súhrn

ODHAD S NAJMENŠÍM RIZIKOM

GEJZA WIMMER

V mnohých prípadoch regresnej analýzy môžeme parametre regresnej závislosti považovať za náhodné premenné. Dôležitým a užitočným sa tu ukazuje odhad s najmenším rizikom. V práci sa našiel explicitný tvar odhadu v prípade, že variačné obory náhodného parametra aj chybového vektora nevyplňujú celý parametrický resp. chybový priestor.

Author's address: RNDr. *Gejza Wimmer,* Ústav merania a meracej techniky SAV, Patrónka-Dúbravská cesta, 885 27 Bratislava.