SOME APPLICATIONS OF LOEWNER'S ORDERING ON SYMMETRIC MATRICES

MINORU SIOTANI

(Received Feb. 25, 1967)

1. Introduction and summary

The method of least squares and the Gauss-Markov theorem are well known and are used widely in univariate statistical analysis. We shall consider in this paper some generalization of them to the multivariate case, based on Loewner's partial ordering [1] on symmetric matrices, which is stated below.

Let $f(\boldsymbol{x}:\boldsymbol{\theta})$ be a density function with p dimensional (column) vector variable $\boldsymbol{x}$ and s dimensional (column) vector parameter $\boldsymbol{\theta}$. Let us denote the class of unbiased estimators of $\boldsymbol{\theta}$ based on the sample of size n by $\mathcal{C}\equiv\{\boldsymbol{t}\}$ and the class of covariance matrices of the unbiased estimators by $\mathcal{L}\equiv\{\Psi_t\}$, suffix $\boldsymbol{t}$ showing that Ψ_t is the covariance matrix of $\boldsymbol{t}$. Then we have the class consisting of *concentration ellipsoids*, which are defined by H. Cramér [2], as

$$(1.1)\qquad (\boldsymbol{t}-\boldsymbol{\theta})'\Psi_t^{-1}(\boldsymbol{t}-\boldsymbol{\theta})=s+2\,,\qquad \boldsymbol{t}\in\mathcal{C}\,,\quad \Psi_t\in\mathcal{L}\,.$$

If we define Fisher's information matrix as

$$M\equiv(m_{ij}(\boldsymbol{\theta}))\,,\qquad (s\times s)$$

$$m_{ij}(\boldsymbol{\theta})=\mathcal{E}\left\{\frac{\partial\log f(\boldsymbol{x}:\boldsymbol{\theta})}{\partial\theta_i}\,\frac{\partial\log f(\boldsymbol{x}:\boldsymbol{\theta})}{\partial\theta_j}\right\}\,,$$

then the Cramér-Rao lemma ([2], p. 495) tells us that the ellipsoid

$$(1.2)\qquad n(\boldsymbol{t}-\boldsymbol{\theta})'M(\boldsymbol{t}-\boldsymbol{\theta})=s+2$$

lies entirely within the concentration ellipsoid of any $\boldsymbol{t}\in\mathcal{C}$. Hence, if there exists $\boldsymbol{t}_0$ in $\mathcal{C}$ such that its concentration ellipsoid coincides with that in (1.2), then $\boldsymbol{t}_0$ has the smallest concentration ellipsoid among all unbiased estimators of $\boldsymbol{\theta}$ and is an *efficient* estimator of $\boldsymbol{\theta}$.

Now let us express the efficiency in the above sense in terms of covariance matrices of $\boldsymbol{t}\in\mathcal{C}$.

DEFINITION 1.1. (Loewner's partial ordering). For any symmetric matrices A and B of same order, we write $A>B$ ($A\geqq B$) if and only if $A-B$ is positive definite (positive semi-definite).

This definition is equivalent to the statement

$$A \underset{(\geqq)}{>} B \Longleftrightarrow z'Az \underset{(\geqq)}{>} z'Bz \tag{1.3}$$

for all non-zero z such that $z'z=1$.

Suppose now that t_1 and t_2 are unbiased estimators of θ, and that t_1 is more efficient than t_2; namely, the concentration ellipsoid: $(t_1-\theta)'\Psi_{t_1}^{-1}(t_1-\theta)=s+2$ is contained entirely wholly within the concentration ellipsoid: $(t_2-\theta)'\Psi_{t_2}^{-1}(t_2-\theta)=s+2$. It is easy to see that this statement is exactly equivalent to the statement that

$$z'\Psi_{t_1}^{-1}z \geqq z'\Psi_{t_2}^{-1}z$$

for all non-zero z such that $z'z=1$, which is again, by (1.3), equivalent to the statement that

$$\Psi_{t_1}^{-1} \geqq \Psi_{t_2}^{-1}$$

or

$$\Psi_{t_2} \geqq \Psi_{t_1}.$$

From the similar argument, the Cramér-Rao lemma in the multivariate case can now be expressed by

$$\Psi_t \geqq \frac{1}{n} M^{-1} \tag{1.4}$$

for covariance matrix Ψ_t of any unbiased estimator t of θ.

DEFINITION 1.2. Let $\mathfrak{A}$ be a class of symmetric matrices. If there exists an A_0 such that

$$A\geqq A_0 \quad (A_0\geqq A)$$

for all A in $\mathfrak{A}$, we say that A_0 is a lower (an upper) bound of $\mathfrak{A}$. If A_0 belongs to $\mathfrak{A}$, A_0 is said to be minimum (maximum) in $\mathfrak{A}$ in the sense of Loewner's ordering.

In the minimizing or maximizing problem in multivariate analysis, it is ordinary that scalar quantities such as determinant, trace, some norms, etc. of a certain matrix of variable elements are adopted as criteria to be minimized or maximized, and the calculations are carried out separately for each criterion under investigation. They do not in general give the same solution. However, suppose, for example, that we have a class of symmetric positive definite matrices for a minimizing problem. If we can obtain the minimum matrix A_0 in the class, this is more appropriate for us than when we use a particular scalar index or when we use several indices separately, because $A \geqq A_0$ implies simultaneously the relations $|A| \geqq |A_0|$, $\mathrm{tr}\, A \geqq \mathrm{tr}\, A_0$, $ch_i(A) \geqq ch_i(A_0)$, $i=1, \cdots, p$, where the characteristic roots are arranged on descending order, and so on.

Along this line we shall consider a generalization of the Gauss-Markov theorem to the multivariate case and principal components of a vector random variable.

2. Preliminary lemmas

The following are well known:

LEMMA 2.1. *For symmetric matrices A and B of order p, $A \geqq B$ implies $\alpha_i \geqq \beta_i$, $i=1, \cdots, p$, where $\alpha_1 \geqq \cdots \geqq \alpha_p$ and $\beta_1 \geqq \cdots \geqq \beta_p$ are characteristic roots of A and B respectively.*

As corollaries of this lemma, we have immediately

LEMMA 2.2. *$A \geqq B$ implies the following:*

(i) $|A| \geqq |B|$,
(ii) $\alpha_1 + \cdots + \alpha_k \geqq \beta_1 + \cdots + \beta_k$, $k=1, \cdots, p$
(iii) $\mathrm{tr}\, A \geqq \mathrm{tr}\, B$

LEMMA 2.3. *Let A and B be symmetric positive semi-definite, i.e., $A \geqq 0$, $B \geqq 0$. Then $A \geqq B$ implies*

$$\mathrm{tr}\, A^m \geqq \mathrm{tr}\, B^m , \qquad m=1, 2, \cdots . \tag{2.1}$$

PROOF. Let the characteristic roots of A and B be $\alpha_1 \geqq \cdots \geqq \alpha_p$ and $\beta_1 \geqq \cdots \geqq \beta_p$ respectively. Then the characteristic roots of A^m and B^m are α_i^m's and β_i^m's. Since $A \geqq B \geqq 0$, we have $\alpha_i \geqq \beta_i \geqq 0$, $i=1, \cdots, p$ by lemma 2.1 which implies $\alpha_i^m \geqq \beta_i^m$, $i=1, \cdots, p$. Hence

$$\mathrm{tr}\, A^m = \alpha_1^m + \cdots + \alpha_p^m \geqq \beta_1^m + \cdots + \beta_p^m = \mathrm{tr}\, B^m .$$

LEMMA 2.4. *Let Σ, A and B be symmetric positive semi-definite matrices and let A and B be of same order. Then*

$$\Sigma \times A \geqq \Sigma \times B \Longleftrightarrow A \geqq B . \tag{2.2}$$

PROOF. Let us denote the characteristic roots of Σ by $\lambda_1, \cdots, \lambda_p$ and those of $A-B$ by $\mu_1, \cdots, \mu_q$. Then the characteristic roots of $\Sigma \times (A-B)$ are $\lambda_i \mu_j$, $i=1, \cdots, p$; $j=1, \cdots, q$. Since λ_i's are non-negative, and since a symmetric matrix is positive semi-definite if and only if its characteristic roots are all non-negative, both sides in (2.2) hold simultaneously if and only if $\mu_i \geqq 0$, $i=1, \cdots, p$.

LEMMA 2.5. *Let* $Z'=(z_1, z_2, \cdots, z_m)$ *be* $n\times m$, V *be* $n\times n$ *and symmetric, and* $\boldsymbol{a}$ *be* $m\times 1$. *Then*

$$\frac{\partial}{\partial z'_\alpha}(\boldsymbol{a}'ZVZ'\boldsymbol{a})=2a_\alpha \boldsymbol{a}'ZV \qquad \alpha=1, \cdots, m \tag{2.3}$$

where $\dfrac{\partial}{\partial z'_\alpha}\equiv\left(\dfrac{\partial}{\partial z_{\alpha 1}}, \dfrac{\partial}{\partial z_{\alpha 2}}, \cdots, \dfrac{\partial}{\partial z_{\alpha n}}\right)$.

PROOF.

$$\begin{aligned}\frac{\partial}{\partial z_{\alpha\beta}}(\boldsymbol{a}'ZVZ'\boldsymbol{a})&=\boldsymbol{a}'\left(\frac{\partial}{\partial z_{\alpha\beta}}Z\right)VZ'\boldsymbol{a}+\boldsymbol{a}'ZV\left(\frac{\partial}{\partial z_{\alpha\beta}}Z'\right)\boldsymbol{a}\\ &=2\boldsymbol{a}'ZV\left(\frac{\partial}{\partial z_{\alpha\beta}}Z'\right)\boldsymbol{a}\\ &=2(\boldsymbol{a}'ZV)_\beta a_\alpha\end{aligned}$$

where $(\boldsymbol{a}'ZV)_\beta$ is the βth component of $\boldsymbol{a}'ZV$. (2.3) is just an expression in vector form.

Now we consider the practical procedure of the maximization or the minimization. Let $A(X)$ be a real symmetric matrix whose elements are differentiable functions of X in a domain D. Then the maximization or the minimization of $A(X)$ under conditions $r_i(X)=0$, $i=1, \cdots, r$ is accomplished by maximizing or minimizing, in the usual way, the scalar valued function

$$\phi(X)=\boldsymbol{a}'A(X)\boldsymbol{a}-\Sigma\lambda_i r_i(X) \tag{2.4}$$

for any fixed non-zero scalar vector $\boldsymbol{a}$ with $\boldsymbol{a}'\boldsymbol{a}=1$, where λ_i's are Lagrange's multipliers. If the maximizer or the minimizer X^* is determined *independently* of $\boldsymbol{a}$, and is *in* D, then $\boldsymbol{a}'A(X^*)\boldsymbol{a}$ is the maximum or the minimum for every $\boldsymbol{a}$ and hence $A(X^*)$ itself is the maximum or the minimum in the Loewner sense, because of (1.3).

3. A generalization of the Gauss-Markov theorem on the method of least squares

Let us consider the general linear model

$$\mathcal{E}(X)=\Theta A \tag{3.1}$$

where $X(p\times n\,;\ p<n)$ is an observation matrix on p-component random

vector $\boldsymbol{x}$, $\Theta(p\times m)$ a matrix of unknown parameters, and $A(m\times n)$ a design matrix which is known beforehand and of rank $r(\leqq m<n)$, i.e.,

$$A=\underset{n}{\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}}\begin{matrix} r \\ m-r \end{matrix}$$

where A_1 is a basis of A. Suppose that

$$\Phi=\Theta C$$

be a matrix to be estimated by X, where C is an $m\times q$ known matrix of coefficients, and of rank q.

Assume that

(i) Φ is estimable (The condition for estimability is given in (3.12) or (3.13) below.),

(3.2) (ii) $\mathcal{E}[\boldsymbol{x}_i^*-\mathcal{E}(\boldsymbol{x}_i^*)][\boldsymbol{x}_j^*-\mathcal{E}(\boldsymbol{x}_j^*)]'=\sigma_{ij}H$ where $\boldsymbol{x}_i^{*\prime}=(x_{i1},\cdots,x_{in})$

is the vector consisting of n observations on the ith component of $\boldsymbol{x}$, H is an $n\times n$ symmetric positive definite matrix which is known, and σ_{ij} is the covariance between the ith and the jth components of $\boldsymbol{x}$. Under these assumptions let us consider a natural generalization of the Gauss-Markov theorem to the multivariate case.

First of all, we shall determine the best linear unbiased estimator

$$Y(p\times q)=X(p\times n)B(n\times q) \tag{3.3}$$

of Φ in the sense of the minimum covariance matrix. Rewriting ΘC as

$$\Theta C={}_p(\underset{r}{\Theta_1}\ \underset{m-r}{\Theta_2})\underset{q}{\begin{bmatrix} C_1 \\ C_2 \end{bmatrix}}\begin{matrix} r \\ m-r \end{matrix}$$

we have the conditions

$$\begin{aligned}\Phi&=\Theta C=\Theta_1 C_1+\Theta_2 C_2\\ &=\mathcal{E}(XB)\\ &=\Theta AB=\Theta_1 A_1 B+\Theta_2 A_2 B\end{aligned}$$

for all Θ, since Φ is estimable and hence

$$C=AB \quad \text{or} \quad C_1=A_1B\,,\quad C_2=A_2B\,. \tag{3.4}$$

Now let us express the covariance matrix of Y, which is to be minimized under the conditions (3.4). If $\boldsymbol{y}_i'=(y_{i1},\cdots,y_{iq})$, $i=1,\cdots,p$, then $\boldsymbol{y}_i'=\boldsymbol{x}_i^{*\prime}B$, and

$$\begin{aligned}\mathcal{E}[\boldsymbol{y}_i-\mathcal{E}(\boldsymbol{y}_i)][\boldsymbol{y}_j-\mathcal{E}(\boldsymbol{y}_j)]' &= B'\mathcal{E}[\boldsymbol{x}_i^*-\mathcal{E}(\boldsymbol{x}_i^*)][\boldsymbol{x}_i^*-\mathcal{E}(\boldsymbol{x}_i^*)]'B \\ &= \sigma_{ij}B'HB\ .\end{aligned}$$

Hence the covariance matrix $V(pq\times pq)$ of $(\boldsymbol{y}_1', \boldsymbol{y}_2', \cdots, \boldsymbol{y}_p')$ can be written in the form of the Kronecker product

$$V=\Sigma\times(B'HB) \tag{3.5}$$

where $\Sigma=(\sigma_{ij})$ is the covariance matrix of $\boldsymbol{x}$. Our problem is therefore to minimize V with respect to B under the restrictions (3.4). This minimization is, by lemma 2.4, equivalent to that of $B'HB$, because Σ is a non-variable matrix and >0. Then the actual procedure is carried out as follows: For an arbitrary non-zero vector $\boldsymbol{a}$ of q components such that $\boldsymbol{a}'\boldsymbol{a}=1$, put

$$g(B)=\boldsymbol{a}'B'HB\boldsymbol{a}-2\,\mathrm{tr}\,M'(AB-C) \tag{3.6}$$

where $M=(\lambda_{ij})=(\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_q)$, $(m\times q)$, a matrix of Lagrange's constants, and minimize $g(B)$ with respect to elements of B. By lemma 2.5, $\partial g(B)/\partial\boldsymbol{b}_\alpha'=\boldsymbol{0}'$ leads to

$$a_\alpha\boldsymbol{a}'B'H-\boldsymbol{\lambda}_\alpha' A=\boldsymbol{0}'$$

or

$$a_\alpha\boldsymbol{a}'B'=\boldsymbol{\lambda}_\alpha' AH^{-1}\ , \qquad \alpha=1, \cdots, q \tag{3.7}$$

where $\boldsymbol{b}_\alpha'=(b_{1\alpha}, \cdots, b_{n\alpha})$ and $\boldsymbol{\lambda}_\alpha'=(\lambda_{1\alpha}, \cdots, \lambda_{m\alpha})$. Multiplying both sides by A_1' from the right, we have

$$a_\alpha\boldsymbol{a}'B'A_1'=\boldsymbol{\lambda}_\alpha' AH^{-1}A_1'$$

or by the condition in (3.4)

$$a_\alpha\boldsymbol{a}'C_1'=\boldsymbol{\lambda}_\alpha' AH^{-1}A_1'\ . \tag{3.8}$$

Now we use the representation ([4], (A.3.11))

$$A=\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}=\begin{bmatrix} \tilde{T}_1 \\ T_2 \end{bmatrix}L \tag{3.9}$$

or

$$A_1=\tilde{T}_1L\ , \qquad A_2=T_2L=T_2\tilde{T}_1^{-1}A_1$$

where $\tilde{T}_1$ $(r\times r)$ is triangular and >0, T_2 is $(m-r)\times r$, and L $(r\times n)$ is row-orthogonal, i.e., $LL'=I_r$. Then (3.8) can be written as

$$a_\alpha \boldsymbol{a}'C_1' = \boldsymbol{\lambda}_\alpha' \begin{bmatrix} I_r \\ T_2\tilde{T}_1^{-1} \end{bmatrix} A_1H^{-1}A_1' .$$

Since $A_1H^{-1}A_1'$ is non-singular,

$$\boldsymbol{\lambda}_\alpha' \begin{bmatrix} I_r \\ T_2\tilde{T}_1^{-1} \end{bmatrix} = a_\alpha \boldsymbol{a}'C_1'(A_1H^{-1}A_1')^{-1} .$$

Consequently, from (3.7)

$$\begin{aligned} a_\alpha \boldsymbol{a}'B' &= \boldsymbol{\lambda}_\alpha' \begin{bmatrix} I_r \\ T_2\tilde{T}_1^{-1} \end{bmatrix} A_1H^{-1} \\ &= a_\alpha \boldsymbol{a}'C_1'(A_1H^{-1}A_1')^{-1}A_1H^{-1} \end{aligned}$$

for every non-zero $\boldsymbol{a}$ and hence we have as the solution

$$B_0 = H^{-1}A_1'(A_1H^{-1}A_1')^{-1}C_1 . \tag{3.10}$$

Thus we finally have the best linear unbiased estimator of Φ,

$$Y_0 = XB_0 = XH^{-1}A_1'(A_1H^{-1}A_1')^{-1}C_1 \tag{3.11}$$

with the minimum covariance matrix

$$V_0 = \Sigma \times (B_0'HB_0) = \Sigma \times [C_1'(A_1H^{-1}A_1')^{-1}C_1] \tag{3.12}$$

which is positive definite.

It should be noted that the condition for Φ to be estimable is now easily obtained from (3.4), (3.9) and (3.10) as

$$\begin{aligned} C_2 &= A_2B_0 = A_2H^{-1}A_1'(A_1H^{-1}A_1')^{-1}C_1 \qquad (3.13) \\ &= T_2\tilde{T}_1^{-1}C_1 \\ &= A_2A_1'(A_1A_1')^{-1}C_1 . \qquad (3.14) \end{aligned}$$

Next we show that the solution (3.11) can be obtained from the minimization of the weighted residual covariance matrix

$$S = (X - \Theta A)H^{-1}(X - \Theta A)' \tag{3.15}$$

under the variation of Θ. To do so, let us reduce (3.15) to the case of full rank:

$$\Theta A = (\Theta_1 + \Theta_2 T_2\tilde{T}_1^{-1})A_1 ,$$

$$\Phi = \Theta C = (\Theta_1 + \Theta_2 T_2\tilde{T}_1^{-1})C_1 .$$

If we put

$$\xi(p\times r)=\Theta_1+\Theta_2 T_2\tilde{T}_1^{-1}$$

we can write

$$S=(X-\xi A_1)H^{-1}(X-\xi A_1)' ,$$

$$\Phi=\xi C_1 .$$

It is now required to minimize S with respect to ξ. For any fixed non-zero $\boldsymbol{a}(p\times 1)$, $\partial(\boldsymbol{a}'S\boldsymbol{a})/\partial\boldsymbol{\xi}_\alpha=\boldsymbol{0}'$ gives the equation

$$a_\alpha \boldsymbol{a}'(X-\xi A_1)H^{-1}A_1'=\boldsymbol{0}' .$$

Hence, denoting the minimizer by ξ_0, we have

$$\xi_0 A_1 H^{-1}A_1'=XH^{-1}A_1'$$

or

(3.16) $$\xi_0=XH^{-1}A_1'(A_1H^{-1}A_1')^{-1} .$$

Consequently $\Phi_0=\xi_0 C_1$ coincides with Y_0 in (3.11); in other words, the best linear unbiased estimator of the estimable function $\Phi=\Theta C$ can be given by calculating ξ_0 $(=\Theta_{10}+\Theta_{20}T_2\tilde{T}_1^{-1})$ which makes S minimum and by forming $\xi_0 C_1$ $(=\Theta_0 C)$. So, the minimum of S is

(3.17) $$S_0=(X-\xi_0A_1)H^{-1}(X-\xi_0A_1)'=(X-\Theta_0A)H^{-1}(X-\Theta_0A)' .$$

Let us next show that an analogous relation as in the univariate case of the Gauss-Markov theorem, i.e.,

(3.18) $$\mathcal{E}(S_0)=(n-r)\Sigma$$

holds in our multivariate case. First we notice that

(3.19) $$\begin{aligned}S&=\{(X-\xi_0A_1)+(\xi_0-\xi)A_1\}H^{-1}\{(X-\xi_0A_1)+(\xi_0-\xi)A_1\}'\\ &=S_0+(\xi_0-\xi)A_1H^{-1}A_1'(\xi_0-\xi)'\end{aligned}$$

since

$$(X-\xi_0A_1)H^{-1}A_1'(\xi_0-\xi)=0 .$$

It is easy to prove that S, S_0 and $(\xi_0-\xi)A_1H^{-1}A_1'(\xi_0-\xi)'$ can be expressed as follows:

$$S=(X-\xi A_1)H^{-1}(X-\xi A_1)' ,$$

$$S_0=(X-\xi A_1)H^{-1/2}\{I_n-H^{-1/2}A_1'(A_1H^{-1}A_1')^{-1}A_1H^{-1/2}\}H^{-1/2}(X-\xi A_1)' ,$$

$$\begin{aligned}&(\xi_0-\xi)A_1H^{-1}A_1'(\xi_0-\xi)'\\ &\quad=(X-\xi A_1)H^{-1/2}\{H^{-1/2}A_1'(A_1H^{-1}A_1')^{-1}A_1H^{-1/2}\}(X-\xi A_1)' .\end{aligned}$$

Therefore, if we put

$$U=(X-\xi A_1)H^{-1/2}\ ,\qquad E=H^{-1/2}A_1'(A_1H^{-1}A_1')^{-1}A_1H^{-1/2}\ ,$$

then (3.19) can be rewritten as

$$UU'=U(I_n-E)U'+UEU'\ . \tag{3.20}$$

The following are easily checked:

$$E^2=E\ ,\qquad (I_n-E)^2=I_n-E\ ,\qquad E(I_n-E)=(I_n-E)E=0\ ,$$

namely, E and I_n-E are orthogonal idempotent matrices of rank r and $(n-r)$, respectively. Hence there exists an orthogonal matrix which transforms (3.20) into

$$ZZ'=Z_1Z_1'+Z_2Z_2'$$

where Z_1 is $p\times(n-r)$, Z_2 is $p\times r$ and each column of $Z=(Z_1Z_2)$ has zero mean vector and covariance matrix Σ, and has no correlation with others: in fact,

$$Z=UL=(X-\xi A_1)H^{-1/2}L$$

where L is an orthogonal matrix. Let $z_i^{*\prime}=(z_{i1}, z_{i2}, \cdots, z_{in})$, and $\boldsymbol{\xi}_i^{*\prime}=(\xi_{i1}, \xi_{i2}, \cdots, \xi_{ir})$. Then

$$\begin{aligned}\mathrm{Cov}\,(z_i^*, z_j^*)&=\mathcal{E}(z_i^* z_j^{*\prime})\\ &=L'H^{-1/2}\mathcal{E}(\boldsymbol{x}_i^*-A_1\boldsymbol{\xi}_i^*)(\boldsymbol{x}_j^*-A_1\boldsymbol{\xi}_j^*)'H^{-1/2}L\\ &=\sigma_{ij}L'H^{-1/2}HH^{-1/2}L\\ &=\sigma_{ij}I_n\ ,\end{aligned}$$

which proves the assertion. From this fact, we have

$$\mathcal{E}(S_0)=\mathcal{E}(Z_1Z_1')=(n-r)\Sigma\ .$$

Summarizing the above argument, we have

THEOREM 3.1. (*Generalized Gauss-Markov Theorem*). *For the set-up* (3.1) *and under* (3.2), *the best linear unbiased estimator of an estimable matrix function $\Phi=\Theta C$ can be given by $\Phi_0=\Theta_0 C$ where Θ_0 is a matrix such that at $\Theta=\Theta_0$*

$$S=(X-\Theta A)H^{-1}(X-\Theta A)'$$

becomes minimum under the variation of Θ, that is,

$$\Phi_0=XH^{-1}A_1'(A_1H^{-1}A_1')^{-1}C_1\ .$$

If we denote the minimum of S by S_0, then

$$\mathcal{E}(S_0)=(n-r)\Sigma \ .$$

Remark. It is a direct consequence from lemmas 2.2 and 2.3 that the solution in the theorem minimizes the various scalar measures stated in the lemmas for V or S.

Just the same as the Gauss-Markov theorem for the univariate case, this theorem has various applications to the estimation in the multivariate regression problem including the analysis of dispersion, in the sampling theory from a finite multivariate population, etc.

4. An application of theorem 3.1

We start from the multivariate linear model (3.1), i.e.,

$$\mathcal{E}(X)=\Theta A \ . \tag{3.1}$$

Let

$$X'=(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \cdots, \boldsymbol{x}_p^*) \ , \qquad \boldsymbol{x}_i^{*\prime}=(x_{i1}, x_{i2}, \cdots, x_{in})$$
$$\Theta'=(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \cdots, \boldsymbol{\theta}_p^*) \ , \qquad \boldsymbol{\theta}_i^{*\prime}=(\theta_{i1}, \theta_{i2}, \cdots, \theta_{im}) \ .$$

Then the model (3.1) for the ith variable reduces to

$$\mathcal{E}(\boldsymbol{x}_i^*)=A'\boldsymbol{\theta}_i^* \ ,$$
$$\mathrm{Cov}\,(\boldsymbol{x}_i^*, \boldsymbol{x}_j^*)=\sigma_{ij}H \ , \qquad i, j=1, \cdots, p$$

where H is, as before, an $n\times n$ symmetric positive definite matrix which is known. Now let us consider the estimation of

$$\phi=\boldsymbol{c}_1'\boldsymbol{\theta}_1^*+\boldsymbol{c}_2'\boldsymbol{\theta}_2^*+ \cdots +\boldsymbol{c}_p'\boldsymbol{\theta}_p^* \ , \tag{4.1}$$

where $\boldsymbol{c}_j'\boldsymbol{\theta}_j^*$, $(j=1, \cdots, p)$ are estimable. J. N. Srivastava [5] proved that, if we obtain the best linear unbiased estimator $\boldsymbol{c}_j'\hat{\boldsymbol{\theta}}_j^*$ of $\boldsymbol{c}_j'\boldsymbol{\theta}_j^*$ for each j by the usual Gauss-Markov theorem, then

$$y_0=\boldsymbol{c}_1'\hat{\boldsymbol{\theta}}_1^*+ \cdots +\boldsymbol{c}_p'\hat{\boldsymbol{\theta}}_p^* \tag{4.2}$$

is the best linear unbiased estimator of ϕ in the class of linear unbiased estimators of ϕ:

$$y=\boldsymbol{f}_1'\boldsymbol{x}_1^*+ \cdots +\boldsymbol{f}_p'\boldsymbol{x}_p^* \ , \tag{4.3}$$

i.e., $\mathrm{Var}\,(y)>\mathrm{Var}\,(y_0)$ for any y.

The following is another proof of it by theorem 3.1.

Let

$$C(m\times p)=(\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_p)\ , \qquad F(n\times p)=(\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_p)\ ,$$
$$\hat{\Theta}'(m\times p)=(\hat{\boldsymbol{\theta}}_1^*, \hat{\boldsymbol{\theta}}_2^*, \cdots, \hat{\boldsymbol{\theta}}_p^*)\ .$$

Then we can write (4.1) and (4.3) as

$$\phi=\operatorname{tr}(\Theta C)\equiv\operatorname{tr}\Phi\ , \tag{4.1$'$}$$

$$y=\operatorname{tr}(XF)\equiv\operatorname{tr}(Y) \tag{4.3$'$}$$

where $\phi_{ij}=\boldsymbol{c}_j'\boldsymbol{\theta}_i^*$ and $y_{ij}=\boldsymbol{f}_j'\boldsymbol{x}_i^*$, $i, j=1, \cdots, p$. If we denote the $p^2\times p^2$ covariance matrix of Y by V, i.e., $V=\Sigma\times(F'HF)$, and if t^* is a vector having 1 in the positions corresponding to those of y_{ii}, $i=1, \cdots, p$ in the arrangement of Y in a row and 0 elsewhere, i.e.,

$$\boldsymbol{t}^{*\prime}=(\overset{y_{11}}{1}\,0\cdots 0\,\overset{y_{22}}{1}\,0\cdots 0\,\overset{y_{33}}{1}\,0\cdots 0\,\overset{y_{pp}}{1})\ ,$$

then we have

$$\operatorname{Var}(y)=\boldsymbol{t}^{*\prime}V\boldsymbol{t}^*\ . \tag{4.4}$$

Now notice that $\Phi=\Theta C$ is estimable and consider the best linear unbiased estimator $Y_0=XF_0$ in the generalized sense (theorem 3.1 in the case that $p=q$). Denote the covariance matrix of Y_0 by V_0, and determined F_0 so that $\boldsymbol{a}'V\boldsymbol{a}\geqq\boldsymbol{a}'V_0\boldsymbol{a}$ for all non-zero normalized $\boldsymbol{a}$ of p^2 components and for all Y. Then we have in particular

$$\operatorname{Var}(y)=p(\boldsymbol{t}^{*\prime}/\sqrt{p})V(\boldsymbol{t}^*/\sqrt{p})\geqq p(\boldsymbol{t}^*/\sqrt{p})V_0(\boldsymbol{t}^*/\sqrt{p})=\operatorname{Var}(y_0)$$

where $y_0=\operatorname{tr}XF_0=\boldsymbol{x}_1^*\boldsymbol{f}_{10}+\cdots+\boldsymbol{x}_p^*\boldsymbol{f}_{p0}$. This y_0 is our solution, i.e., the best linear unbiased estimator for ϕ and coincides with y_0 in (4.2) as is easily seen that $\boldsymbol{x}_i^{*\prime}\boldsymbol{f}_{i0}=\boldsymbol{c}_i'\hat{\boldsymbol{\theta}}^*$.

5. Principal components of a vector random variable

Let $\boldsymbol{x}$ be a p-component random vector with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Consider a transformation

$$\boldsymbol{y}=T\boldsymbol{x} \tag{5.1}$$

where T is a $q\times p$ matrix ($q\leqq p$), and the joint covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\begin{bmatrix} \Sigma & \Sigma T' \\ T\Sigma & T\Sigma T' \end{bmatrix}\ . \tag{5.2}$$

The predictive efficiency of $\boldsymbol{y}$ for $\boldsymbol{x}$ depends on the residual covariance matrix of $\boldsymbol{x}$ after subtracting its best linear predictor $\boldsymbol{y}=T\boldsymbol{x}$:

$$\Phi=\Sigma-\Sigma T'(T\Sigma T')^{-1}T\Sigma \,. \tag{5.3}$$

C. R. Rao [3] considered two measures:

$$\text{(i)}\quad \operatorname{tr}\Phi\,, \qquad \text{(ii)}\quad \|\Phi\|=(\operatorname{tr}\Phi^2)^{1/2} \tag{5.4}$$

as statistics for determining T, and showed that both measures when minimized lead to the same T, the matrix of the first q characteristic vectors of Σ. Let us now treat the problem according to our method based on Loewner's ordering. To do so, we prove the

LEMMA 5.1. *Let $\boldsymbol{x}$ be a p-component column vector and $\Sigma(p\times p)$ be a given symmetric positive definite matrix. Then the maximizer $\boldsymbol{x}_0$ of*

$$\frac{\Sigma\boldsymbol{x}\boldsymbol{x}'\Sigma}{\boldsymbol{x}'\Sigma\boldsymbol{x}} \tag{5.5}$$

is the characteristic vector corresponding to the largest characteristic root λ_1 of Σ, i.e.,

$$\frac{\Sigma\boldsymbol{x}_0\boldsymbol{x}_0'\Sigma}{\boldsymbol{x}_0'\Sigma\boldsymbol{x}_0}\geqq\frac{\Sigma\boldsymbol{x}\boldsymbol{x}'\Sigma}{\boldsymbol{x}'\Sigma\boldsymbol{x}} \tag{5.6}$$

for all $\boldsymbol{x}$.

PROOF. Without loss of generality, we can assume that

$$\boldsymbol{x}'\Sigma\boldsymbol{x}=1\,. \tag{5.7}$$

We need to maximize $\boldsymbol{a}'\Sigma\boldsymbol{x}\boldsymbol{x}'\Sigma\boldsymbol{a}$ for any fixed non-zero $\boldsymbol{a}(p\times 1)$ such that $\boldsymbol{a}'\boldsymbol{a}=1$ under the condition (5.7). Put

$$\phi(\boldsymbol{x})=\boldsymbol{a}'\Sigma\boldsymbol{x}\boldsymbol{x}'\Sigma\boldsymbol{a}-\lambda(\boldsymbol{x}'\Sigma\boldsymbol{x}-1) \tag{5.8}$$

where λ is the Lagrange multiplyer. Then $\partial\phi(\boldsymbol{x})/\partial\boldsymbol{x}=0$ gives

$$\Sigma\boldsymbol{a}\boldsymbol{a}'\Sigma\boldsymbol{x}-\lambda\Sigma\boldsymbol{x}=\boldsymbol{0} \tag{5.9}$$

or

$$\boldsymbol{a}\boldsymbol{a}'\Sigma\boldsymbol{x}-\lambda\boldsymbol{x}=\boldsymbol{0}$$

since $\Sigma>0$. If we multiply $\boldsymbol{a}'$ from the left, we have

$$\boldsymbol{a}'(\Sigma\boldsymbol{x}-\lambda\boldsymbol{x})=0$$

holding for every $\boldsymbol{a}$. Hence

$$(\Sigma-\lambda I)\boldsymbol{x}=\boldsymbol{0}\,, \tag{5.10}$$

that is, the solution $\boldsymbol{x}_0$ is a characteristic vector of Σ corresponding to a characteristic root λ_0 of Σ. Since, from (5.7) and (5.9),

$$\lambda_0=\boldsymbol{x}_0'\Sigma\boldsymbol{a}\boldsymbol{a}'\Sigma\boldsymbol{x}_0=\boldsymbol{a}'\Sigma\boldsymbol{x}_0\boldsymbol{x}_0'\Sigma\boldsymbol{a} , \tag{5.11}$$

λ_0 should be the largest characteristic root λ_1 and hence $\boldsymbol{x}_0$ is corresponding to λ_1.

LEMMA 5.2. *Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be two-p-component column vectors and $\Sigma(p\times p)$ a symmetric positive definite matrix. Then*

$$\max_{R_1}\frac{\Sigma\boldsymbol{x}_1\boldsymbol{x}_1'\Sigma}{\boldsymbol{x}_1'\Sigma\boldsymbol{x}_1}+\max_{R_2}\frac{\Sigma\boldsymbol{x}_2\boldsymbol{x}_2'\Sigma}{\boldsymbol{x}_2'\Sigma\boldsymbol{x}_2}=\frac{\Sigma\boldsymbol{x}_{10}\boldsymbol{x}_{10}'\Sigma}{\boldsymbol{x}_{10}'\Sigma\boldsymbol{x}_{10}}+\frac{\Sigma\boldsymbol{x}_{20}\boldsymbol{x}_{20}'\Sigma}{\boldsymbol{x}_{20}'\Sigma\boldsymbol{x}_{20}} , \tag{5.12}$$

where R_1 is the p-dimensional space of $\boldsymbol{x}_1$, $R_2=\{\boldsymbol{x}_2 : \boldsymbol{x}_{10}'\boldsymbol{x}_2=0,\ \boldsymbol{x}_2\neq 0\}$, and $\boldsymbol{x}_{10}$, $\boldsymbol{x}_{20}$ are the characteristic vectors of Σ associated with the largest and second largest characteristic roots λ_1, λ_2 of Σ, respectively.

PROOF. By lemma 5.1,

$$\max_{R_1}\frac{\Sigma\boldsymbol{x}_1\boldsymbol{x}_1'\Sigma}{\boldsymbol{x}_1'\Sigma\boldsymbol{x}_1}=\frac{\Sigma\boldsymbol{x}_{10}\boldsymbol{x}_{10}'\Sigma}{\boldsymbol{x}_{10}'\Sigma\boldsymbol{x}_{10}} .$$

To evaluate the second term in the left hand side of (5.12), let us consider, for any fixed non-zero $\boldsymbol{a}(p\times 1)$ with $\boldsymbol{a}'\boldsymbol{a}=1$,

$$\psi(\boldsymbol{x}_2)=\boldsymbol{a}'\Sigma\boldsymbol{x}_2\boldsymbol{x}_2'\Sigma\boldsymbol{a}-\mu(\boldsymbol{x}_2'\Sigma\boldsymbol{x}_2-1)-2\nu(\boldsymbol{x}_{10}'\boldsymbol{x}_2) , \tag{5.13}$$

where μ and ν are Lagrange's multipliers. Differentiating $\psi(\boldsymbol{x}_2)$ with respect to $\boldsymbol{x}_2$ and equating it to zero give the equation

$$\Sigma\boldsymbol{a}\boldsymbol{a}'\Sigma\boldsymbol{x}_2-\mu\Sigma\boldsymbol{x}_2-\nu\boldsymbol{x}_{10}=\boldsymbol{0} . \tag{5.14}$$

From the fact that $\Sigma\boldsymbol{x}_{10}=\lambda_1\boldsymbol{x}_{10}$, we have

$$\boldsymbol{a}\boldsymbol{a}'\Sigma\boldsymbol{x}_2-\mu\boldsymbol{x}_2-(\nu/\lambda_1)\boldsymbol{x}_{10}=\boldsymbol{0}$$

or

$$\boldsymbol{a}'[\Sigma\boldsymbol{x}_2-\mu\boldsymbol{x}_2-(\nu/\lambda_1)\boldsymbol{x}_{10}]=0 ,$$

for every $\boldsymbol{a}$. Hence

$$\Sigma\boldsymbol{x}_2-\mu\boldsymbol{x}_2-(\nu/\lambda_1)\boldsymbol{x}_{10}=\boldsymbol{0} .$$

Since, in R_2, $\boldsymbol{x}_{10}'\boldsymbol{x}_2=(1/\lambda_1)\boldsymbol{x}_{10}'\Sigma\boldsymbol{x}_2=0$, it is easy to see that $\nu=0$. Thus the desired maximizer $\boldsymbol{x}_{20}$ satisfies

$$(\Sigma-\mu I)\boldsymbol{x}_{20}=\boldsymbol{0}$$

$$\mu=\boldsymbol{a}'\Sigma\boldsymbol{x}_{20}\boldsymbol{x}_{20}'\Sigma\boldsymbol{a} .$$

Consequently $\boldsymbol{x}_{20}$ should be the characteristic vector associated with the second largest characteristic root λ_2 of Σ.

Now back to the problem of minimizing Φ or equivalently maximizing $\Sigma T'(T\Sigma T')^{-1}T\Sigma$. If we denote q column vectors of $T'(p\times q)$ by $\boldsymbol{t}_1, \cdots, \boldsymbol{t}_q$, we may assume without any loss of generality that

$$(5.15) \qquad \boldsymbol{t}_i'\Sigma\boldsymbol{t}_j=0\ , \qquad (i\neq j)$$

and the function of T to be minimized can be expressed as

$$(5.16) \qquad \frac{\Sigma\boldsymbol{t}_1\boldsymbol{t}_1'\Sigma}{\boldsymbol{t}_1'\Sigma\boldsymbol{t}_1}+\cdots+\frac{\Sigma\boldsymbol{t}_q\boldsymbol{t}_q'\Sigma}{\boldsymbol{t}_q'\Sigma\boldsymbol{t}_q}$$

with the conditions (5.15). This can be carried out as follows:

$$(5.17) \qquad \max_{R}\left[\frac{\Sigma\boldsymbol{t}_1\boldsymbol{t}_1'\Sigma}{\boldsymbol{t}_1'\Sigma\boldsymbol{t}_1}+\cdots+\frac{\Sigma\boldsymbol{t}_q\boldsymbol{t}_q'\Sigma}{\boldsymbol{t}_q'\Sigma\boldsymbol{t}_q}\right]$$
$$=\max_{R_1}\frac{\Sigma\boldsymbol{t}_1\boldsymbol{t}_1'\Sigma}{\boldsymbol{t}_1'\Sigma\boldsymbol{t}_1}+\max_{R_2}\frac{\Sigma\boldsymbol{t}_2\boldsymbol{t}_2'\Sigma}{\boldsymbol{t}_2'\Sigma\boldsymbol{t}_2}+\cdots+\max_{R_q}\frac{\Sigma\boldsymbol{t}_q\boldsymbol{t}_q'\Sigma}{\boldsymbol{t}_q'\Sigma\boldsymbol{t}_q}$$

where R is the whole region of $(\boldsymbol{t}_1, \cdots, \boldsymbol{t}_q)$ satisfying $\boldsymbol{t}_i'\Sigma\boldsymbol{t}_j=0$, $(i\neq j)$, R_1 the p-dimensional space of $\boldsymbol{t}_1$, and

$$R_j=\{\boldsymbol{t}_j:\ \boldsymbol{t}_{\alpha 0}'\Sigma\boldsymbol{t}_j=0,\ \alpha=1, \cdots, j-1,\ \boldsymbol{t}_j\neq 0\}\ , \qquad j=2, \cdots, q\ ,$$

where $\boldsymbol{t}_{\alpha 0}$ is the maximizer of the αth term of (5.17). By lemma 5.1, maximum of the 1st term in the right hand side of (5.17) is attained by the characteristic vector $\boldsymbol{t}_{10}$ associated with the largest characteristic root λ_1 of Σ. The maximum of the 2nd term can be achieved, according to lemma 5.2, by the characteristic vector $\boldsymbol{t}_{20}$ associated with the 2nd largest characteristic root λ_2 of Σ. Doing in this way until the qth term in (5.17), the maximizer of (5.16) is found to be

$$(5.18) \qquad T_0'=(\boldsymbol{t}_{10}, \cdots, \boldsymbol{t}_{q0})$$

where $\boldsymbol{t}_{i0}$ is the characteristic vector corresponding to the ith largest root λ_i of Σ.

The transformed variables, $\boldsymbol{t}_{10}'\boldsymbol{x}, \cdots, \boldsymbol{t}_{q0}'\boldsymbol{x}$ are well known as the first q principal components of the random variable $\boldsymbol{x}$.

6. Acknowledgement

I wish to express my gratitude to Professor J. Ogawa for the useful discussions. I also wish to thank Mrs. Y. Hayakawa for her help in the preparation of this paper.

NIHON UNIVERSITY

References

[1] E. F. Beckenback and R. Bellman, *Inequalities*, Springer-Verlag, Berlin, 1965.
[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[3] C. R. Rao, "The use and interpretation of principal component analysis in applied research," *Sankhyā*, 26 (1964), 329-358.
[4] S. N. Roy, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1957.
[5] J. N. Srivastava, "A multivariate extension of the Gauss-Markov theorem," *Ann. Inst. Statist. Math.*, 17 (1965), 63-66.

CANCELLATION OF

SECTION 5 OF "SOME APPLICATIONS OF LOEWNER'S ORDERING ON SYMMETRIC MATRICES"

MINORU SIOTANI

Soon after the above titled paper was published (this Annals, Vol. 19, No. 2 (1967), 245–259), the author found that there was a big mistake, that is, Lemmas 5.1 and 5.2 do not hold, so the whole of Section 5 should be canceled.

In fact, for any vector with $\boldsymbol{a}\boldsymbol{a}'=1$, we have

$$\begin{aligned}\frac{\boldsymbol{a}'\Sigma\boldsymbol{x}\boldsymbol{x}'\Sigma\boldsymbol{a}}{\boldsymbol{x}'\Sigma\boldsymbol{x}} &= \frac{\boldsymbol{a}'\Sigma^{1/2}\boldsymbol{y}\boldsymbol{y}'\Sigma^{1/2}\boldsymbol{a}}{\boldsymbol{y}'\boldsymbol{y}} \qquad (\Sigma^{1/2}\boldsymbol{x}=\boldsymbol{y}) \\ &= \frac{\boldsymbol{y}'\Sigma^{1/2}\boldsymbol{a}\boldsymbol{a}'\Sigma^{1/2}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} \\ &\leqq \max ch(\Sigma^{1/2}\boldsymbol{a}\boldsymbol{a}'\Sigma^{1/2}) = ch(\boldsymbol{a}'\Sigma\boldsymbol{a}) = \boldsymbol{a}'\Sigma\boldsymbol{a} \leqq \lambda_1\end{aligned}$$

where ch denotes a characteristic root and λ_1 is the largest characteristic root of Σ. The equality in the first inequality is attained by $\boldsymbol{y}=\Sigma^{1/2}\boldsymbol{a}$ or $\boldsymbol{x}=\boldsymbol{a}$ for fixed $\boldsymbol{a}$, which depends on $\boldsymbol{a}$, and the equality in the second inequality can not be attained generally except when $\boldsymbol{a}$ is equal to the characteristic vector corresponding to λ_1.