# NOTE ON THE NUMERICAL COMPUTATION IN THE DISCRIMINATION PROBLEM

By TOSIO UEMATU

(Received Jan. 20, 1959)

1. The use of the linear discriminant function method in the discrimination problem leads to the need for solving a special type of secular equation $ANA'\mathfrak{x}=\lambda B\mathfrak{x}$, where the detailed forms of the matrices $A, N, A'$ and $B$ will be given in § 2. The usual method for solving this equation involves the computation of $B^{-1}$ and the matrix multiplications to make the product $B^{-1}ANA'$, which make the numerical computation work very troublesome when, as is often the case with practical application, the dimension of this equation is large. The need for any simple computation method stimulated Mr. H. Akaike at our Institute to study the successive approximation method. He devised a new method, which did not make use of $B^{-1}$, even for the more general type of the equation [3]. However, so far as we are concerned with the equation in the discrimination problem, the use of $B^{-1}$ is not necessarily inexpedient if we make full use of the speciality of the equation. In fact, as will be seen later, an appropriate use of $B^{-1}$ enables us for the initial equation to degenerate into the one with some smaller dimension. In many cases of practical applications, this reduced dimension is very small, while the initial dimension is large, and so much simplification is possible in the numerical computation.

2. Now we give the detailed form of the equation arising in the linear discriminant function method. Consider the classification of objects into $s(\geqq 2)$ groups and $n_i$ be the size of the $i$-th group, $i=1, \cdots, s$, $n$ being the sum of the $n_i$.

First consider the linear discriminant function $X=x_1X_1+\cdots+x_pX_p$, where the $X_i$ be the random variables which represent measurements for $p$ characteristics of object. Let $B$ be the variance-covariance matrix of $(X_1, \cdots, X_p)$ multipled by $n$, and $a_{ij}$ be the difference between $E_j(X_i)$ and $E(X_i)$, where $E_j(X_i)$ represents conditional expectation of $X_i$ over the

$j$-th group. We put $A=(a_{ij})_{\substack{i=1,\dots,p,\\ j=1,\dots,s}}$ $N=\begin{pmatrix} n_1 & & \\ & \ddots & \\ & & n_s \end{pmatrix}$, and $\mathfrak{x}=\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$, then we have to solve the following equation,

(1) $$ANA'\mathfrak{x}=\lambda B\mathfrak{x},$$

where $A'$ denotes the transpose of $A$ (see, for example, [3]). In case of $s=2$, we may put $A=\begin{pmatrix} E_1(X_1)-E_2(X_1) \\ \vdots \\ E_1(X_p)-E_2(X_p) \end{pmatrix}$, $N=s\times s$ identity matrix, and $B=$the varince-covariance matrix of $(X_1, \cdots, X_p)$ in (1) (see, for example, [2]).

Next consider the quantification problem. The linear discriminant function method applied to this problem leads to the same type of equation. Let $m$ be the number of characteristics of object which are used for discrimination, and suppose that the measurement for the $i$-th characteristic comes into one and only one of the $l_i$ mutually exclusive categories, $i=1, \cdots, m$. If we represent the $j$-th category of the $i$-th characteristic by a number $w_{ij}$, $i=1, \cdots, l_i$, we have a random variable $X_i$, $i=1, \cdots, m$. We have to determine the $w_{ij}$ in such a way that the discrimination using $X=X_1+\cdots+X_m$ attains the maximum. For that we have to solve the equation (1), where the meanings of $A, B, \mathfrak{x}$ and $p$ are given as follows, the meaning of $N$ being exactly the same as the above. Let $n_{(ij)}$ be the number of objects the $i$-th measurement of which come into the $j$-th category, $n_{(ij)}^{(k)}$ be that for the $k$-th group, and $n_{(ij)(\alpha\beta)}$ be the number of objects the $i$-th and the $\alpha$-th measuremets of which come into respectively the $j$-th and the $\beta$-th categories. We specify the order of the suffixes appropriately. Then we have

$$A=\left(\frac{n_{(ij)}^{(k)}}{n_k}-\frac{n_{(ij)}}{n}\right)_{\substack{(ij)=(11),(12),\dots,(ml_m),\\ k=1,\dots,s}} \quad B=\left(n_{(ij)(\alpha\beta)}-\frac{n_{(ij)}n_{(\alpha\beta)}}{n}\right)_{\substack{(ij)=(11)(12)\dots(ml_m)\\ (\alpha\beta)=(11)(12)\dots(ml_m)}},$$

$$\mathfrak{x}=\begin{pmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{ml_m} \end{pmatrix} \text{ and } p=l_1+\cdots+l_m.$$

In case of $s=2$ we may put $A=\begin{pmatrix} \dfrac{n_{(11)}^{(1)}}{n_1}-\dfrac{n_{(11)}^{(2)}}{n_1} \\ \vdots \\ \dfrac{n_{(mlm)}^{(1)}}{n_1}-\dfrac{n_{(mlm)}^{(2)}}{n_2} \end{pmatrix}$,

$B=\left(\dfrac{n_{(ij)(\alpha\beta)}}{n}-\dfrac{n_{(ij)}n_{(\alpha\beta)}}{n^2}\right)_{\substack{(ij)=(11)(12)\dots(ml_m)\\ (\alpha\beta)=(11)(12)\dots(ml_m)}}$ and $N=s\times s$ identity matrix [1].

Now denote by $\mathfrak{a}_i$ the $i$-th column vector of $A$, $i=1, \cdots, s$.

Then we have for $s \geqq 3$

$$n_1\mathfrak{a}_1 + \cdots + n_s\mathfrak{a}_s = 0 \tag{2}$$

It follows that the rank $r$ of $A$ is less than $s$. In practical applications we may consider that $r=s-1$ and $\mathfrak{a}_1, \cdots, \mathfrak{a}_{s-1}$ are linearly independent, and we assume this hereafter. Also we assume that $B$ is non-singular.

3. In studying the computation method of (1), the author noticed the degeneration of (1) into an equation with smaller dimension and the possibility of simplification in the numerical computation by making use of the fact. Mr. H. Akaike pointed out to the author that this fact was contained in the results of §10, [2], in which the discrimination problem using the linear discriminant function was derived as a special case of the canonical correlation analysis. Though in his paper he did not attach importance to the fact of degeneration of (1), it seems very useful in numerical computation of (1).

The degeneration of (1) can be derived from elementary consideration too. It is easily seen that any eigenvector $\mathfrak{x}$ belonging to any non-zero eigenvalue $\lambda$ of $B^{-1}ANA'$ can be expressed as a linear combination of vectors $B^{-1}\mathfrak{a}_1, \cdots, B^{-1}\mathfrak{a}_{s-1}$. Let this expression be

$$\mathfrak{x} = c_1 B^{-1}\mathfrak{a}_1 + \cdots + c_{s-1}B^{-1}\mathfrak{a}_{s-1} \tag{3}$$

From (2), (1) can be written as follows:

$$(\mathfrak{a}_1 \cdots \mathfrak{a}_{s-1}) M \begin{pmatrix} \mathfrak{a}'_1 \\ \vdots \\ \mathfrak{a}'_{s-1} \end{pmatrix} \mathfrak{x} = \lambda B \mathfrak{x} \tag{4}$$

where

$$M = \begin{pmatrix} n_1 & & \\ & \ddots & \\ & & n_{s-1} \end{pmatrix} + \frac{1}{n_s} \begin{pmatrix} n_1 \\ \vdots \\ n_{s-1} \end{pmatrix} (n_1 \cdots n_{s-1}), \tag{5}$$

and $\mathfrak{a}'_i$ denotes the transpose of $\mathfrak{a}_i$.

From (3) and (4), we get the following equivalent to (1):

$$M \begin{pmatrix} \mathfrak{a}'_1 \\ \vdots \\ \mathfrak{a}'_{s-1} \end{pmatrix} B^{-1}(\mathfrak{a}_1 \cdots \mathfrak{a}_{s-1})\mathfrak{c} = \lambda \mathfrak{c}, \tag{6}$$

where $\mathfrak{c}=\begin{pmatrix} c_1 \\ \vdots \\ c_{s-1} \end{pmatrix}$, and

(7) $$\mathfrak{x}=B^{-1}(\mathfrak{a}_1\cdots\mathfrak{a}_{s-1})\mathfrak{c}$$

Thus we have the following computation procedure:

(i) Solve $s-1$ linear equations $B\mathfrak{y}_1=\mathfrak{a}_1, \cdots, B\mathfrak{y}_{s-1}=\mathfrak{a}_{s-1}$ to get $\mathfrak{y}_1, \cdots \mathfrak{y}_{s-1}$.

(ii) Compute $\begin{pmatrix} \mathfrak{y}_1' \\ \vdots \\ \mathfrak{y}_{s-1}' \end{pmatrix}(\mathfrak{y}_1 \cdots \mathfrak{y}_{s-1})$ and multiply $M$ from left of it to get $H$.

(iii) Solve the equation $H\mathfrak{c}=\lambda\mathfrak{c}$ to get $\lambda$ and $\mathfrak{c}$.

(iv) Compute $\mathfrak{x}=c_1\mathfrak{y}_1+\cdots+c_{s-1}\mathfrak{y}_{s-1}$.

$\lambda$ and $\mathfrak{x}$, thus obtained, are the desired solutions of (1)

If only the maximum eigenvalue is required, as is usually the case in discrimination problem, and if we use the elimination method, the following derivation of $\begin{pmatrix} \mathfrak{a}_1' \\ \vdots \\ \mathfrak{a}_{s-1}' \end{pmatrix}B^{-1}(\mathfrak{a}_1\cdots\mathfrak{a}_{s-1})$ will be rather simpler. $B^{-1}$ is equal to $Q'Q$, where $Q$ is the matrix which expresses the elimination of all elements under the diagonal of $B$. We compute $\mathfrak{a}_i^*=Q\mathfrak{a}_i$, $i=1, \cdots, s-1$, and $Q$. Then $\begin{pmatrix} \mathfrak{a}_1' \\ \vdots \\ \mathfrak{a}_{s-1}' \end{pmatrix}B^{-1}(\mathfrak{a}_1 \cdots \mathfrak{a}_{s-1})=\begin{pmatrix} \mathfrak{a}_1^{*\prime} \\ \vdots \\ \mathfrak{a}_{s-1}^{*\prime} \end{pmatrix}(\mathfrak{a}_1^*\cdots\mathfrak{a}_{s-1}^*)$, and $\mathfrak{x}=Q'(\mathfrak{a}_1^*\cdots\mathfrak{a}_{s-1}^*\mathfrak{c})$. In practical computation work the following procedure may be taken: first write down $B, \mathfrak{a}_1, \cdots, \mathfrak{a}_{s-1}$ and $I$, $I$ being the $p\times p$ identity matrix, and then determine $Q$ by making use $B$, and at the same time perform the operation $Q$ on $\mathfrak{a}_1, \cdots, \mathfrak{a}_{s-1}$ and $I$ to get $\mathfrak{a}_1^*, \cdots, \mathfrak{a}_{s-1}^*$ and $Q$.

Now compare our procedure with the one, which uses $B^{-1}ANA'$. Multiplication $\begin{pmatrix} \mathfrak{a}_1' \\ \vdots \\ \mathfrak{a}_{s-1}' \end{pmatrix}(\mathfrak{y}_1, \cdots, \mathfrak{y}_{s-1})$ in the former corresponds to multiplication $ANA'$ in the latter, and the troubles of computation are roughly of $s-1$ to $p$. Moreover multiplication $B^{-1}(\mathfrak{a}_1 \cdots \mathfrak{a}_{s-1})$ in the former corresponds to multiplication $B^{-1}(ANA')$ in the latter, and the troubles of computation will nearly be of one to two if $p$ is very large compared to $s$, as is often the case with practical applications. As, in usual, $s$ is small, the trouble for solving $B^{-1}ANA'\mathfrak{x}=\lambda\mathfrak{x}$ cannot bear comparison with the trouble for solving (6). These comparisons show that our computation precedure is useful. In particular if we need all of the non-zero eigenvalues and corresponding eigenvectors in case of $s=3$, our method will be very useful. This will remain true even for $s=4$.

The programming for the numerical computation using the above procedure, by a FACOM-128 automatic relay computer is available for any $p$ and $s$ satisfying the relation $p+s\leqq 58$.

The author is grateful to Professor C. Hayashi and Mr. H. Akaike for helpful advices and suggestions in connection with the work. Thanks are also to Miss S. Takakura, who undertook the programming for the numerical computation by FACOM-128.

THE INSTITUTE OF STATISTICAL MATHEMATICS

## REFERENCES

[1] C. Hayashi ' On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view," *Annals of* the *Institute of Statistical Mathematics*, Vol. III (1952), pp. 69–98.

[2] C. R. Rao, "Advanced Statistical Methods in Biometric Research," John Wiley and Sons, Inc., New York (1952), pp. 287-289.

[3] H. Akaike, "On a computation method for eigenvalue problems and its application to statistical analysis," *Annals of the Institute of Statistical Mathematics*,, Vol. X (1958), pp. 1-20.