ON A MULTIDIMENSIONAL LINEAR DISCRIMINANT FUNCTION

TOSIO UEMATU

1. Suppose that we want to assign any individual, drawn from a mixed population consisting of s ($\geqq 2$) groups of individuals, to one of these groups. If we want to use for this discrimination a linear discriminant function $y=\sum_{i=1}^{p} a_i x_i$ of p measurements $x_1, x_2, \cdots, x_p$ about the individual, we need the coefficients $\{a_i\}$ of this function. These coefficients are best if they are determined such that the resulting function affords the maximum possible discrimination as the linear function. But to determine such coefficients will be impossible except for special cases.

An alternative way of determining the coefficients is the method of maximizing the correlation ratio $\eta^2=\sigma_b{}^2/\sigma^2$, where σ^2 is the variance of y, and $\sigma_b{}^2$ is the variance of y between groups. This is the way which was adopted by C. Hayashi in his quantification theory [1]. We can generalize the above method in the following way. For the discrimination we use m ($2\leqq m<p$) linear combinations

$$y_i=\sum_{j=1}^{p} a_{ij}x_j\,, \qquad i=1, 2, \cdots, m\,, \tag{1}$$

instead of a single linear discriminant function. In order to determine the coefficients $\{a_{ij}\}$ we consider some measure of discrimination, which is a generalization of η^2 in some sense or other. We determine the coefficients so as to maximize such a measure of discrimination.

2. In [2] C. Hayashi presented two measures which may be considered as generalizations of η^2. One of them is $\lambda=1-[\sigma_w^2/\sigma^2]$, where σ^2 is the generalized variance of $\boldsymbol{y}=(y_1, y_2, \cdots, y_m)$, and σ_w^2 is the generalized variance of $\boldsymbol{y}$ within groups.

Now we define:

- $\mu_i(\nu)$: the mean of y_i within the νth group,
- $\sigma_{ii}(\nu)$: the variance of y_i within the νth group,
- $\sigma_{ij}(\nu)$: the covariance between y_i and y_j within the νth group ($i\neq j$),
- μ_i : the mean of y_i,
- σ_i : the variance of y_i,

σ_{ij}: the convariance between y_i and y_j $(i\neq j)$,

π_ν : the relative size of the νth group $(\sum_{\nu=1}^{s}\pi_\nu=1)$,

$\boldsymbol{\mu}(\nu)=(\mu_1(\nu),\ \mu_2(\nu),\ \cdots,\ \mu_m(\nu)),\ \boldsymbol{\mu}=(\mu_1,\ \mu_2,\ \cdots,\ \mu_m)$,

$\boldsymbol{\Sigma}=(\sigma_{ij}),\ \boldsymbol{\Sigma}_\nu=(\sigma_{ij}(\nu)),\ \boldsymbol{\Sigma}_b=(\sum_{\nu=1}^{s}\pi_\nu(\mu_i(\nu)-\mu_i)(\mu_j(\nu)-\mu_j))$,

$\boldsymbol{\Sigma}_w=\pi_1\boldsymbol{\Sigma}_1+\pi_2\boldsymbol{\Sigma}_2+\ \cdots\ +\pi_s\boldsymbol{\Sigma}_s=(\sum_{\nu=1}^{s}\pi_\nu\sigma_{ij}(\nu))$.

Then the above σ^2 and σ_w^2 are defined respectively as $\sigma^2=|\boldsymbol{\Sigma}|$ and $\sigma_w^2=|\boldsymbol{\Sigma}_w|$. Thus we have

$$\lambda=1-[\sigma_w^2/\sigma^2]=1-[|\boldsymbol{\Sigma}_w|/|\boldsymbol{\Sigma}|]\,. \tag{2}$$

Alternatively we may choose the following measure:

$$\lambda^*=\sigma_b{}^2/\sigma^2=|\boldsymbol{\Sigma}_b|/|\boldsymbol{\Sigma}|\,, \tag{3}$$

where $\sigma_b{}^2=|\boldsymbol{\Sigma}_b|$ can be interpreted (cf. [3]) as the measure of scatter of s points $\boldsymbol{\mu}(1),\ \boldsymbol{\mu}(2),\ \cdots,\ \boldsymbol{\mu}(s)$ in an m dimensional space, multiplicity of each point being taken into consideration according to the size of the corresponding group.

We are going to consider the problem of determining the coefficients $\{a_{ij}\}$ so as to maximize λ or λ^*.

3. We define

$m_i(\nu)$: the mean of x_i within the νth group,

$\lambda_{ii}(\nu)$: the variance of x_i within the νth group,

$\lambda_{ij}(\nu)$: the covariance between x_i and x_j within the νth group $(i\neq j)$,

m_i : the mean of x_i,

λ_{ii} : the variance of x_i,

λ_{ij} : the covariance between x_i and x_j $(i\neq j)$,

$\boldsymbol{m}(\nu)=(m_1(\nu),\ m_2(\nu),\ \cdots,\ m_p(\nu)),\ \boldsymbol{m}=(m_1,\ m_2,\ \cdots,\ m_p)$,

$\boldsymbol{\Lambda}=(\lambda_{ij}),\ \boldsymbol{\Lambda}_\nu=(\lambda_{ij}(\nu)),\ \boldsymbol{\Lambda}_w=\sum_{\nu=1}^{s}\pi_\nu\boldsymbol{\Lambda}_\nu$,

$$\boldsymbol{a}_i=(a_{i1},\ a_{i2},\ \cdots,\ a_{ip}),\quad \boldsymbol{A}=(a_{ij})=\begin{pmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\\ \vdots\\ \boldsymbol{a}_m\end{pmatrix}.$$

Then we have $\boldsymbol{\mu}(\nu)=\boldsymbol{m}(\nu)\boldsymbol{A}'$, $\boldsymbol{\mu}=\boldsymbol{m}\boldsymbol{A}'$, $\boldsymbol{\Sigma}_\nu=\boldsymbol{A}\boldsymbol{\Lambda}_\nu\boldsymbol{A}'$, and $\boldsymbol{\Sigma}=\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}'$, where $\boldsymbol{A}'$ means the transpose of $\boldsymbol{A}$. From those relations we can easily see that

$$\boldsymbol{\Sigma}_w=\boldsymbol{A}\boldsymbol{\Lambda}_w\boldsymbol{A}'\,, \tag{4}$$

$$\boldsymbol{\Sigma}_b=\boldsymbol{A}\boldsymbol{D}\boldsymbol{D}'\boldsymbol{A}', \tag{5}$$

where

$$(6)\qquad \boldsymbol{D}=(\boldsymbol{m}'(1)-\boldsymbol{m}',\ \boldsymbol{m}'(2)-\boldsymbol{m}',\ \cdots,\ \boldsymbol{m}'(s)-\boldsymbol{m}')\begin{pmatrix} \sqrt{\pi_1} & & & \\ & \sqrt{\pi_2} & & \\ & & \ddots & \\ & & & \sqrt{\pi_s} \end{pmatrix},$$

and we can represent λ and λ^* as follows:

$$(7)\qquad \lambda=1-[|\boldsymbol{A}\boldsymbol{\varLambda}_w\boldsymbol{A}'|\,/\,|\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'|],$$

and

$$(8)\qquad \lambda^*=|\boldsymbol{A}\boldsymbol{D}\boldsymbol{D}'\boldsymbol{A}'|\,/\,|\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'|.$$

We seek the maximum of λ or λ^*. Let $\boldsymbol{U}$ denote the matrix $\boldsymbol{\varLambda}_w$ or $\boldsymbol{D}\boldsymbol{D}'$ according as we consider λ or λ^*. Then our problem is to minimize or maximize $|\boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'|\,/\,|\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'|$, the elements $\{a_{ij}\}$ of $\boldsymbol{A}$ being variables which may vary under the restriction $|\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'|\neq 0$.

Now we can find a non-singular matrix $\boldsymbol{P}$ depending on $\boldsymbol{A}$ so that $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'\boldsymbol{P}=\boldsymbol{I}$ (unit matrix) and $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'\boldsymbol{P}$ has a diagonal form. But $|\boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'|/|\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'|=|\boldsymbol{P}'\boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'\boldsymbol{P}|\,/\,|\boldsymbol{P}'\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'\boldsymbol{P}|$. Therefore we can conclude that, when we want to minimize or maximize $|\boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'|/|\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'|$, we may do it under the conditions that $\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'=\boldsymbol{I}$ and $\boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'$ has a diagonal form.

In [4] C. Hayashi treated some multidimensional measure in connection with a quantification problem, and sought the solution such that the measure attains its maximum under some orthogonality conditions. We remark that by this procedure the unconditioned maximum of the measure is attained in this case, too.

4. In what follows we consider the case $m=2$ only. We assume that $\boldsymbol{\varLambda}$ and $\boldsymbol{\varLambda}_w$ are non-singular. This will be a natural assumption in the usual application of the linear discriminant function. In the quantification problem ([1], [2]), however, $\boldsymbol{\varLambda}$ and $\boldsymbol{\varLambda}_w$ always appear as singular matrices. But in this case there is a non-singular matrix $\boldsymbol{Q}$ for which

$$\boldsymbol{Q}\boldsymbol{U}\boldsymbol{Q}'=\begin{pmatrix} \boldsymbol{H} & \vdots & \boldsymbol{O} \\ \cdots & \vdots & \cdots \\ \boldsymbol{O} & \vdots & \boldsymbol{O} \end{pmatrix},\quad \boldsymbol{Q}\boldsymbol{\varLambda}\boldsymbol{Q}'=\begin{pmatrix} \boldsymbol{F} & \vdots & \boldsymbol{O} \\ \cdots & \vdots & \cdots \\ \boldsymbol{O} & \vdots & \boldsymbol{O} \end{pmatrix},$$

where $\boldsymbol{H}$ is some positive semi-definite matrix of order $p-t$ for a certain $t\,(>0)$ and $\boldsymbol{F}$ is some positive definite matrix of order $p-t$. Moreover in the case of $\boldsymbol{U}=\boldsymbol{\varLambda}_w$, $\boldsymbol{H}$ is positive definite. Consequently in the quantification problem we may consider $|\boldsymbol{A}\boldsymbol{H}\boldsymbol{A}'|\,/\,|\boldsymbol{A}\boldsymbol{F}\boldsymbol{A}'|$ instead of $|\boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'|/|\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'|$. From this consideration we may assume that $\boldsymbol{\varLambda}$ and $\boldsymbol{\varLambda}_w$ are non-singular, without loss of generality. From the remark in the preceding section our problem in the case $m=2$ is to minimize or maximize $\tau=|\boldsymbol{A}\boldsymbol{U}\boldsymbol{A}'|/|\boldsymbol{A}\boldsymbol{\varLambda}\boldsymbol{A}'|$

under the conditions

$$a_1 A a_1' = 1 , \tag{9}$$

$$a_2 A a_2' = 1 , \tag{10}$$

$$a_1 A a_2' = 0 , \tag{11}$$

$$a_1 U a_2' = 0 . \tag{12}$$

We remark that under these conditions τ becomes

$$\tau = (a_1 U a_1')(a_2 U a_2') . \tag{13}$$

From the assumption that A and A_w are non-singular, it follows that if $aAa'=1$ then $aUa'\neq 0$, in the case $U=A_w$.

In the case $U=DD'$ our objective is to maximize τ. But $\tau\geqq 0$, and if $a_1Ua_1'=0$ or $a_2Ua_2'=0$, then $\tau=0$. Hence we may consider only such a_1, a_2 that $a_1Ua_1\neq 0$ and $a_2Ua_2\neq 0$ except the trivial case of $\max \tau=0$. From these considerations it is justified to impose the additional condition

$$a_1 U a_1' \neq 0 \text{ and } a_2 U a_2' \neq 0 . \tag{14}$$

Next we solve the problem of minimizing or maximizing (13) under the conditions (9), (10), (11), (12), and (14). Introducing Lagrangian parameters l_1, l_2, l_3 and l_4 we obtain the following equations, as necessary conditions to minimize or maximize (13) under the above conditions:

$$(a_2 U a_2') a_1 U = l_1 a_2 U + l_2 a_2 A + l_3 a_1 A , \tag{15}$$

$$(a_1 U a_1') a_2 U = l_1 a_1 U + l_2 a_1 A + l_4 a_2 A . \tag{16}$$

These equations together with (9), (10), (11), and (12) give

$$(a_2 U a_2')(a_1 U a_1') = l_3 = l_4 . \tag{17}$$

Hence (15) and (16) become

$$(a_2 U a_2') a_1 U = l_1 a_2 U + l_2 a_2 A + l_3 a_1 A , \tag{18}$$

$$(a_1 U a_1') a_2 U = l_1 a_1 U + l_2 a_1 A + l_3 a_2 A . \tag{19}$$

We have to solve the system of equations (9), (10), (11), (12), (14), (18) and (19). This solution is obtained under some additional assumptions which will be stated later.

First we see that the only solution of l_1, l_2 is $l_1=l_2=0$. From (18), (10), (11), and (12) it is obvious that if $l_1=0$ then $l_2=0$. Now suppose that there is a set of solutions of the system of equations concerned such that $l_1\neq 0$. Consider any such a set of solutions a_1, a_2, l_1, l_2, l_3. From (18), (19), (9), (10), (11), and (12) we have

(20) $$\begin{cases} l_1 \boldsymbol{a}_2 \boldsymbol{U} \boldsymbol{a}_2' + l_2 = 0 , \\ l_1 \boldsymbol{a}_1 \boldsymbol{U} \boldsymbol{a}_1' + l_2 = 0 . \end{cases}$$

Noticing that $l_1 \neq 0$, we have $\boldsymbol{a}_1 \boldsymbol{U} \boldsymbol{a}_1' = \boldsymbol{a}_2 \boldsymbol{U} \boldsymbol{a}_2'$. Then from (17) we get

(21) $$\boldsymbol{a}_1 \boldsymbol{U} \boldsymbol{a}_1' = \boldsymbol{a}_2 \boldsymbol{U} \boldsymbol{a}_2' = \sqrt{l_3} .$$

(20) and (21) give

(22) $$l_2 = -l_1 \sqrt{l_3} .$$

From (18), (19), and (21) we have

$$\sqrt{l_3}(\boldsymbol{a}_1 + \boldsymbol{a}_2)\boldsymbol{U} = l_1(\boldsymbol{a}_1 + \boldsymbol{a}_2)\boldsymbol{U} + l_2(\boldsymbol{a}_1 + \boldsymbol{a}_2)\boldsymbol{A} + l_3(\boldsymbol{a}_1 + \boldsymbol{a}_2)\boldsymbol{A} ,$$
$$\sqrt{l_3}(\boldsymbol{a}_1 - \boldsymbol{a}_2)\boldsymbol{U} = l_1(\boldsymbol{a}_2 - \boldsymbol{a}_1)\boldsymbol{U} + l_2(\boldsymbol{a}_2 - \boldsymbol{a}_1)\boldsymbol{A} + l_3(\boldsymbol{a}_1 - \boldsymbol{a}_2)\boldsymbol{A} .$$

These relations together with (22) give

(23) $$(\sqrt{l_3} - l_1)(\boldsymbol{a}_1 + \boldsymbol{a}_2)\boldsymbol{U} = \sqrt{l_3}(\sqrt{l_3} - l_1)(\boldsymbol{a}_1 + \boldsymbol{a}_2)\boldsymbol{A}$$

(24) $$(\sqrt{l_3} + l_1)(\boldsymbol{a}_1 - \boldsymbol{a}_2)\boldsymbol{U} = \sqrt{l_3}(\sqrt{l_3} + l_1)(\boldsymbol{a}_1 - \boldsymbol{a}_2)\boldsymbol{A}$$

where $\boldsymbol{a}_1 \pm \boldsymbol{a}_2 \neq 0$ because of (9), (10), and (11). If $\sqrt{l_3} = l_1$, then $\sqrt{l_3} + l_1 \neq 0$, so from (24) we conclude that $\sqrt{l_3}$ is an eigenvalue of the characteristic equation

(25) $$\boldsymbol{a}\boldsymbol{U} = \gamma \boldsymbol{a}\boldsymbol{A} .$$

In the case $\sqrt{l_3} \neq l_1$ it is obvious from (23) that $\sqrt{l_3}$ is an eigenvalue of (25). Thus, when $l_1 \neq 0$, $\sqrt{l_3}$ is always a non-zero eigenvalue of (25). On the other hand $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ satisfy (21) and (9), (10). Therefore, we can conclude that $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are characteristic vectors of (25) belonging to the same non-zero eigenvalue $\gamma = \sqrt{l_3}$.

Now we assume that all of the non-zero eigenvalues of (25) are simple roots, and denote them by $\{\gamma_i\}$. Moreover, we assume that the rank of $\boldsymbol{D}$ is not less than two, so (25) has at least two non-zero eigenvalues whether $\boldsymbol{U} = \boldsymbol{D}\boldsymbol{D}'$ or $\boldsymbol{U} = \boldsymbol{A}_w$.

The above simple root assumption implies that for $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, $\boldsymbol{a}_1 \boldsymbol{A} \boldsymbol{a}_2' \neq 0$ which contradicts (11). Thus the only solution of l_1, l_2 is $l_1 = l_2 = 0$. Therefore, (18) and (19) become

(18)′ $$\boldsymbol{a}_1 \boldsymbol{U} = \frac{l_3}{(\boldsymbol{a}_2 \boldsymbol{U} \boldsymbol{a}_2')} \boldsymbol{a}_1 \boldsymbol{A} ,$$

(19)′ $$\boldsymbol{a}_2 \boldsymbol{U} = \frac{l_3}{(\boldsymbol{a}_1 \boldsymbol{U} \boldsymbol{a}_1')} \boldsymbol{a}_2 \boldsymbol{A} .$$

Now from (18)′ and (19)′ we have

(26) $$\frac{l_3}{(a_2 U a_2')} = \gamma_i ,$$

(27) $$\frac{l_3}{(a_1 U a_1')} = \gamma_j ,$$

for any pair of non-zero eigenvalues γ_i, γ_j of (25). Here γ_i and γ_j must be different, for if $\gamma_i = \gamma_j$ it follows from the simple root assumption that for the corresponding solution $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, $\boldsymbol{a}_1 \boldsymbol{\Lambda} \boldsymbol{a}_2' \neq 0$, contradicting (11). Taking all combinations of different γ_i, γ_j and the corresponding characteristic vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ which are normalized so as to satisfy (9) and (10), and putting

(28) $$l_3 = \gamma_i \gamma_j ,$$

we get all of the sets of solutions of the system of equations concerned.

Now the minimum of τ in the case $\boldsymbol{U} = \boldsymbol{\Lambda}_w$ and the maximum of τ in the case $\boldsymbol{U} = \boldsymbol{D}\boldsymbol{D}'$ are easily obtained. Noting $\tau = l_3$ from (13) and (17), we get the minimum or maximum of τ as follows, under the assumptions previously stated. Let $\gamma_1 < \gamma_2 < \cdots < \gamma_{k-1} < \gamma_k$ ($k \geqq 2$ from our assumptions) be all of the non-zero eigenvalues of (25), where k equals to p or $s-1$ according as $\boldsymbol{U} = \boldsymbol{\Lambda}_w$ or $\boldsymbol{U} = \boldsymbol{D}\boldsymbol{D}'$ (in the quantification problem p is to be substituted by $p-t$). Then $\min \tau = \gamma_1 \gamma_2$, or $\max \tau = \gamma_{k-1} \gamma_k$. Hence $\max \lambda = 1 - \gamma_1 \gamma_2$ or $\max \lambda^* = \gamma_{k-1} \gamma_k$. Coefficients $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ of (1) in the case $m=2$ are determined as the corresponding characteristic vectors of (25) to γ_1, γ_2 or γ_{k-1}, γ_k. We remark that, when we use λ^*, $s \geqq 3$ is necessary.

5. In the quantifying method in [2], the response patterns are not added, but are used in the original dimension in order to make the classification more effective. In the application of the present method to quantification, however, the response patterns are added. Instead, this addition is carried out in more than one way and a multidimensional discriminator is obtained, in order to make the classification more effective. It is easily seen that, under the same dimension and the use of λ, the present method can give larger value of λ, hence is more effective.

The author wishes to acknowledge very helpful suggestions of Dr. C. Hayashi and Mr. M. Isida.

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," *Ann. Inst. Stat. Math.*, 3 (1952), 69-98.

[2] C. Hayashi, "Multidimensional quantification — with the applications to analysis of social phenomena," *Ann. Inst. Stat. Math.*, 5 (1954), 120-143.

[3] S. S. Wilks, "Multidimensional statistical scatter," *Contribution to Probability and Statistics*, Stanford University Press, 1960.

[4] C. Hayashi, "Theory of quantification and its examples (VI)," (in Japanese), *Proc. Inst. Stat. Math.*, 9 (1961), 29-35.