NOTE ON A MULTIDIMENSIONAL LINEAR DISCRIMINANT FUNCTION

J. K. WANI AND D. G. KABE

(Received Jan. 30, 1971)

1. Introduction and summary

When p measurements $x_1, x_2, \cdots, x_p$ are available on an individual belonging to one of s ($\geqq 2$) groups, to allot the individual to one of these groups a linear discriminant function is formed, say, $y=\sum_{i=1}^{p} a_i x_i$. The coefficients a's are determined by a certain procedure, see e.g., Kendall and Stuart ([3], pp. 316–318, 44.6–44.7), so that the linear function will minimize the probability of misclassification. Hayashi [1] in his studies of quantification theory proposed an alternative method for determining the coefficients a's by maximizing the observed correlation ratio $\eta^2=\sigma_b^2/\sigma^2$, where σ^2 is the estimated variance of y and σ_b^2 is the estimated variance of y between the groups. In many situations it becomes necessary to carry on the discrimination by more than a single discriminant function, see e.g., Radcliffe [4]. Let then these simultaneous m linear discriminant functions be

$$(1) \qquad y_i=\sum_{j=1}^{p} a_{ij} x_j , \qquad i=1, 2, \cdots, m ;\ 2\leqq m\leqq p .$$

In order to determine the coefficients a_{ij}'s we consider the generalization of η^2 given by Hayashi [2]. This generalization is $\lambda=1-\sigma_w^2/\sigma^2$, where now σ^2 is the observed generalized variance of $y'=(y_1, y_2, \cdots, y_m)$ and σ_w^2 is the observed generalized variance of y within groups. We now define certain sample characteristics of y which are obtained from the sample characteristics of x. It is assumed that s samples one for each group of sizes $N_1, N_2, \cdots, N_s$ are available on the p component vector $x'=(x_1, x_2, \cdots, x_p)$.

The sample characteristics of y

$\mu_i(\nu)$: the mean of y_i within the νth group,
$\sigma_{ii}(\nu)$: the variance of y_i within νth group,
$\sigma_{ij}(\nu)$: the covariance of y_i and y_j within νth group,

μ_i : the overall mean of y_i
σ_{ii} : the overall variance of y_i
σ_{ij} : the overall covariance of y_i and y_j,
π_ν : the relative size of the νth group, $\sum_{\nu=1}^{s} \pi_\nu = 1$,

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

$\Sigma = (\sigma_{ij})$, $\Sigma_\nu = (\sigma_{ij}(\nu))$, $\Sigma_b = \left(\sum_{\nu=1}^{s} \pi_\nu[\mu_i(\nu) - \mu_i][\mu_j(\nu) - \mu_j]\right)$,

$\Sigma_w = \pi_1\Sigma_1 + \pi_2\Sigma_2 + \cdots + \pi_s\Sigma_s = \left[\sum_{\nu=1}^{s} \pi_\nu \sigma_{ij}(\nu)\right]$

$\sigma^2 = |\Sigma|$, $\sigma_w^2 = |\Sigma_w|$.

Thus we have that

$$\lambda = 1 - |\Sigma_w|/|\Sigma| , \tag{2}$$

and we may also consider the function

$$\lambda^* = |\Sigma_b|/|\Sigma| , \tag{3}$$

as a function to be maximized with respect to a_{ij}'s.

The sample characteristics of x

$m_i(\nu)$: the mean of x_i within νth group,
$\lambda_{ii}(\nu)$: the variance of x_i within the νth group,
$\lambda_{ij}(\nu)$: the covariance of x_i and x_j in νth group,
m_i : the overall mean of x_i,
λ_{ii} : the overall variance of x_i,
λ_{ij} : the overall covariance of x_i and x_j,

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

$\Lambda = (\lambda_{ij})$, $\Lambda_\nu = (\lambda_{ij}(\nu))$, $\Lambda_w = \sum_{\nu=1}^{s} \pi_\nu \Lambda_\nu$

$a_i' = (a_{i1}, a_{i2}, \cdots, a_{ip})$, $A = (a_{ij})$.

If follows that $\mu(\nu) = m(\nu)A'$, $\mu = mA'$ $\Sigma_\nu = A\Lambda_\nu A'$ $\Sigma = A\Lambda A'$, $\Sigma_w = A\Lambda_w A'$, and $\Sigma_b = ADD'A'$, where

$$D = \sum_{\nu=1}^{s} \sqrt{\pi_\nu}(m'(\nu) - m') . \tag{4}$$

Thus we have to determine A which will maximize

$$\lambda = 1 - |A\Lambda_w A'|/|A\Lambda A'| , \tag{5}$$

or that A which will maximize

$$\lambda^* = |ADD'A'|/|A\Lambda A'| . \tag{6}$$

We shall find an $m \times p$ matrix A that will maximize λ^*, the matrix A

that maximizes λ may be found on similar lines. The solution for the particular case $m=2$ is given by Uematu [6] by a very complicated procedure. We give the solution for the general case by using the following result.

2. A useful result

We wish to prove that

$$\text{(7)} \qquad \operatorname*{Min}_{y_1,\cdots,y_k} \sum_1^k \frac{y_i' \Delta y_i}{y_i' y_i} = \operatorname*{Min}_Y \operatorname{tr} Y\Delta Y' = \theta_{p-k+1} + \cdots + \theta_p \, ,$$

where $\theta_1 > \theta_2 > \cdots > \theta_p$ are roots of Δ, Y is $k \times p$ and of rank k, such that $YY'=I$. The minimum is actually attained when y_i is proportional to a linear function of the eigenvectors of Δ. The result (7) is established by repeated application of Rao's result ([5], p. 51, If. 2.5) that

$$\text{(8)} \qquad \operatorname*{Min}_y y'\Delta y \, , \qquad \text{subject to } y'y=1, \text{ and } Hy=0 \text{ is } \theta_{p-k} \, ,$$

where H is $k \times p$ and of rank k and the minimum is sought over all H. Thus from (8) we know that

$$\text{(9)} \qquad \operatorname*{Min}_{y_1} \frac{y_1' \Delta y_1}{y_1' y_1} \, , \qquad \text{subject to } y_1'y_1=1, \; y_1'y_2=0, \cdots, y_1'y_k=0 \text{ is } \theta_{p-k+1} \, .$$

Next we consider the $\operatorname{Min}(y_2'\Delta y_2/y_2'y_2)$ subject to $y_2'y_2=1$, and $y_2'y_3=0 \cdots$, $y_2'y_k=0$ and this minimum is θ_{p-k+2} and so on, finally we consider $\operatorname{Min} y_k'\Delta y_k/y_k'y_k$ subject to $y_k'y_k=1$ which is θ_p. Adding these minima we get (7). Thus from (7) we note that minimum is given by k last roots and hence the maximum must be given by first k roots, i.e.,

$$\text{(10)} \qquad \operatorname*{Max}_Y \operatorname{tr} Y\Delta Y' \, , \qquad \text{subject to } YY'=I \text{ is } \theta_1+\theta_2+\cdots+\theta_k \, .$$

The minimum in (7) is obtained by setting $y_i = P_{p-k+1}$, $i=1, 2, \cdots, k$ and maximum in (10) is given by $y_i=P_i$, $i=1, \cdots, k$, where P_i is the eigenvector of Δ corresponding to θ_i.

3. Determination of the coefficients A

By using (6) and setting $B=A\Lambda^{1/2}$ where $\Lambda^{1/2}$ is any positive definite symmetric square root of Λ, we find that

$$\text{(11)} \qquad \lambda^* = |B\Lambda^{-1/2}DD'\Lambda^{-1/2}B'|/|BB'| \, .$$

Since the maximum of (11) is sought over all B, we take $BB'=I$, and thus from (10) we conclude that

$$\text{(12)} \qquad \operatorname{Max} \lambda^* = \alpha_1 \alpha_2 \cdots \alpha_m ,$$

where $\alpha_1 > \alpha_2 > \cdots > \alpha_p$ are roots of $DD'\Lambda^{-1}$, and the maximum is attained when the row vectors of $b_1', b_2', \cdots, b_m'$ of B are such that $b_i = P_i$, $i=1, \cdots, m$, where P_i is the eigenvector of $DD'\Lambda^{-1}$ corresponding to α_i. From B we determine A. By taking $m=2$ in (12), we get Uematu's result.

THE UNIVERSITY OF CALGARY AND SAINT MARY'S UNIVERSITY

REFERENCES

[1] Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematics-statistical point of view, *Ann. Inst. Statist. Math.*, **3**, 69-98.

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

[3] Kendall, M. G. and Stuart, A. (1966). *Advanced Theory of Statistics*, **3**, Charles Griffin, London, England.

[4] Radcliffe, J. (1966). Factorization of the residual likelihood criterion in discriminant analysis, *Proc. Camb. Phil. Soc.*, **62**, 743-751.

[5] Rao, C. R. (1965). *Linear Statistical Inference and its Applications*, Wiley, New York.

[6] Uematu, T. (1964). On a multidimensional linear discriminant function, *Ann. Inst. Statist. Math.*, **16**, 431-437.