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 \( x_1, x_2, \ldots, x_p \) are available on an individual belonging to one of \( s \) (\( \geq 2 \)) 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 \( a_i \)'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 \( a_i \)'s by maximizing the observed correlation ratio \( \eta^2 = \sigma_y^2/\sigma^2 \), where \( \sigma_y^2 \) is the estimated variance of \( y \) and \( \sigma^2 \) 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

\[
y_i = \sum_{j=1}^{p} a_{ij} x_j, \quad i = 1, 2, \ldots, m; \quad 2 \leq m \leq p.
\]

In order to determine the coefficients \( a_{ij} \)'s we consider the generalization of \( \eta^2 \) given by Hayashi [2]. This generalization is \( \lambda = 1 - \sigma_y^2/\sigma^2 \), where now \( \sigma^2 \) is the observed generalized variance of \( y' = (y_1, y_2, \ldots, y_m) \) and \( \sigma_y^2 \) 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 \( N_1, N_2, \ldots, N_s \) are available on the \( p \) component vector \( x' = (x_1, x_2, \ldots, x_p) \).

The sample characteristics of \( y \)

- \( \mu_i(\nu) \): the mean of \( y_i \) within the \( \nu \)th group,
- \( \sigma_{ii}(\nu) \): the variance of \( y_i \) within \( \nu \)th group,
- \( \sigma_{ij}(\nu) \): the covariance of \( y_i \) and \( y_j \) within \( \nu \)th group,
\( \mu_i : \) the overall mean of \( y_i \)
\( \sigma_{ii} : \) the overall variance of \( y_i \)
\( \sigma_{ij} : \) the overall covariance of \( y_i \) and \( y_j \),
\( \pi_\nu : \) the relative size of the \( \nu \)th group, \( \sum_{\nu=1}^s \pi_\nu = 1 \),
\( \mu(\nu) = (\mu_1(\nu), \mu_2(\nu), \cdots, \mu_p(\nu)) \), \( \mu = (\mu_1, \mu_2, \cdots, \mu_p) \),
\( \Sigma = (\sigma_{ij}) \), \( \Sigma_\nu = (\sigma_{ij}(\nu)) \), \( \Sigma_s = \left( \sum_{\nu=1}^s \pi_\nu [\mu(\nu) - \mu] [\mu(\nu) - \mu] \right) \),
\( \Sigma_w = \pi_1 \Sigma_1 + \pi_2 \Sigma_2 + \cdots + \pi_s \Sigma_s = \left[ \sum_{\nu=1}^s \pi_\nu \sigma_{ij}(\nu) \right] \),
\( \sigma^2 = |\Sigma| \), \( \sigma_w^2 = |\Sigma_w| \).

Thus we have that
\[
\lambda = 1 - |\Sigma_w|/|\Sigma|
\]
and we may also consider the function
\[
\lambda^* = |\Sigma_s|/|\Sigma|
\]
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 \( \nu \)th group,
\( \lambda_i(\nu) : \) the variance of \( x_i \) within the \( \nu \)th group,
\( \lambda_{ij}(\nu) : \) the covariance of \( x_i \) and \( x_j \) in \( \nu \)th group,
\( m_i : \) the overall mean of \( x_i \),
\( \lambda_i : \) the overall variance of \( x_i \),
\( \lambda_{ij} : \) the overall covariance of \( x_i \) and \( x_j \),
\( m(\nu) = (m_1(\nu), m_2(\nu), \cdots, m_p(\nu)) \), \( m = (m_1, m_2, \cdots, m_p) \),
\( A = (\lambda_{ij}) \), \( A_s = (\lambda_{ij}(\nu)) \), \( A_w = \sum_{\nu=1}^s \pi_\nu A_s \)
\( a' = (a_{11}, a_{12}, \cdots, a_{ip}) \), \( A = (a_{ij}) \).

If follows that \( \mu(\nu) = m(\nu)A' \), \( \mu = mA' \) \( \Sigma_\nu = AA' \) \( \Sigma = AAA' \), \( \Sigma_w = AA_wA' \), and \( \Sigma_s = ADD'A' \) where
\[
D = \sum_{\nu=1}^s \sqrt{\pi_\nu} (m(\nu) - m')
\]
Thus we have to determine \( A \) which will maximize
\[
\lambda = 1 - |AA_wA'|/|AAA'|
\]
or that \( A \) which will maximize
\[
\lambda^* = |ADD'A'|/|AAA'|
\]
We shall find an \( m \times p \) matrix \( A \) that will maximize \( \lambda^* \), the matrix \( A \)
that maximizes $\lambda$ 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

\begin{equation}
\text{Min} \sum_{v_1, \ldots, v_k} \frac{Y'_i A y_i}{y_i y_i} = \text{Min} \text{tr} \, Y A Y' = \theta_{p-k+1} + \cdots + \theta_p,
\end{equation}

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

\begin{equation}
\text{Min} \, y' A y_y, \quad \text{subject to } y'y = 1, \text{ and } H y = 0 \text{ is } \theta_{p-k},
\end{equation}

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

\begin{equation}
\text{Min} \frac{y'_i A y_i}{y_i y_i}, \quad \text{subject to } y_i y_i = 1, \; y_i y_i = 0, \cdots, \; y_i y_i = 0 \text{ is } \theta_{p-k+1}.
\end{equation}

Next we consider the $\text{Min} \, (y'_i A y_i) / y_i y_i$ subject to $y_i y_i = 1$, and $y_i y_i = 0 \cdots$, $y_i y_i = 0$ and this minimum is $\theta_{p-k+1}$ and so on, finally we consider $\text{Min} \, y'_i A y_i / y_i y_i$ subject to $y_i y_i = 1$ which is $\theta_p$. 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.,

\begin{equation}
\text{Max} \, \text{tr} \, Y A Y', \quad \text{subject to } Y Y' = I \text{ is } \theta_1 + \theta_2 + \cdots + \theta_k.
\end{equation}

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

3. Determination of the coefficients $A$

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

\begin{equation}
\lambda^* = |B A^{-1/2} D D' A^{-1/2} B' | / |B B'|.
\end{equation}

Since the maximum of (11) is sought over all $B$, we take $B B' = I$, and thus from (10) we conclude that
\[(12) \quad \text{Max } \lambda^* = \alpha_1 \alpha_2 \cdots \alpha_m ,\]

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 \(\alpha_i\). 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


