CLASSIFICATION OF AN OBSERVATION INTO LINEAR MANIFOLDS, THEIR CLOSED SUBSETS AND CLOSED CONVEX SUBSETS

by

R. P. Bhargava

Technical Report No. 220
November 1985

Prepared under the Auspices of
National Science Foundation
DMS 84-11411
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
CLASSIFICATION OF AN OBSERVATION INTO LINEAR
MANIFOLDS, THEIR CLOSED SUBSETS
AND CLOSED CONVEX SUBSETS

R.P. Bhargava*

Ontario Institute for Studies in Education
and University of Toronto

Key words and phrases: Classification; general linear hypothesis;
linear classification; group constellation.

ABSTRACT

In this paper, the notion of classification of an observation
(sample), into one of the given \( n \) populations is generalized
to the case when some or all of the populations into which the
new observation is to be classified may be new but related in a
simple way to the given \( n \) populations. The discussion is in
the framework of the given set of observations obeying the usual
multivariate general linear hypothesis model. The set of popula-
tions into which the new observation may be classified could be
linear manifolds of the parameter space or their closed subsets
or closed convex subsets or a combination of them or simply \( t \)
subsets of the parameter space each of which has a finite number
of elements. It is noted that in the last case a likelihood
ratio procedure can be obtained easily. Classification proce-
dures given here are based on Mahalanobis distance. Bonferoni
lower bound estimate of the probability of correctly classifying
an observation is given for the case when the covariance matrix
is known or is estimated from a large sample. Two numerical
examples relating to the classification procedures suggested here
are given.

*This work was initiated at the Department of Statistics, Stanford
University
1. INTRODUCTION

Srivastava (1967) considered the problem of classifying an observation into one of the \( n \) multivariate normal populations when the population means were linearly restricted and the common non-singular covariance matrix was unknown. In this paper, we consider the same problem except that now the problem is not of classifying an observation into one of the given \( n \) multivariate normal populations but of classifying it into one of the \( t \) linear manifolds (hereafter called manifolds) of the parameter space. The \( t \) manifolds could be pairwise disjoint or overlapping. (We call two manifolds to be disjoint if their intersection contains only the zero vector; otherwise they are overlapping.) This is considered in section 2.1. Section 2.2 considers a special case. In section 3.1 we consider a problem of classifying an observation into closed subsets of the manifolds and in section 3.2 that of classifying an observation into closed convex subsets of the manifolds. Still another problem may consist of classifying an observation into sets some of which are manifolds and some closed convex sets and still another of classifying into closed subsets each of which has a finite number of elements. These problems can be handled by our approach. If the \( n \) sets each consist of a single point, namely the mean of each of the \( n \) normal populations, we have Srivastava's (1967) problem. If we add an extra \( k \) points, we have a refinement of the original Srivastava (1967) problem. This could result in calibrating a better classification scale. It is noted that when each of the closed sets consist of a finite number of points, the likelihood classification rule can be obtained easily, although computationally, it may be cumbersome and/or unmanageable, when the total number of points in the finite sets is large.

Thus if the new observation is coming from a new population(s) which is related to the given populations in a simple manner through manifolds or their closed subsets or closed convex
subsets or finite subsets, we can obtain the classification procedure.

In growth/learning processes with linear growth in time, the new observation may belong to one of the new populations corresponding to a new time point. The study here provides the classification rule for this situation.

In section 4.1 and 4.2, we discuss two numerical examples. Example of section 4.2 relates to group constellation problem.

In section 5, Bonferoni lower bound estimate of the probability of correctly classifying an observation is given for the case when the common covariance matrix is known or is estimated from a large sample.

Lastly, we comment that our approach is applicable to Cacollos (1965) problem where the object is to classify an observation to the nearest population.

2. CLASSIFICATION OF AN OBSERVATION INTO MANIFOLDS

2.1 Problem and Solution

Let \( X(p \times n) \) be a random observation matrix whose columns are independent normal with a common unknown \( p \times p \) covariance matrix \( \Sigma, \Sigma(p \times p), \Sigma > 0 \). Let

\[
\begin{align*}
E X' &= A \xi, \quad \text{rank} \ A = m \leq n, \ p < n - m. \\
&\quad n \times p \quad n \times m \quad m \times p
\end{align*}
\]

(1)

\( A \) is known. \( \xi \) is unknown.

Let

\[
C = (C_1, C_2, \ldots, C_t), \ C: m \times s, \ C_i: m \times s_i, \ \Sigma_{i=1}^{t} s_i = s.
\]

(2)

Let rank \( C_i = s_i \), rank \( C = m \leq s \). \( C \) is given. The column spaces of \( C_i \) and \( C \) are respectively isomorphic to the manifolds \( M_i \) and \( M \) (of linear parametric functions) given by

\[
M_i = \{ \lambda_i' C_i \xi: \lambda_i \in \mathbb{R} \}_{s_i}
\]

(3)
\[ M = \{ \xi' C' \xi : \xi \in R_k \} , \]

where \( R_k \) denotes Euclidean \( k \)-dimensional space.

Let \( \xi_0 \) \((p \times 1)\) be a new observation from \( N(\mu_0, \Sigma) \). We say \( \xi_0 \) belongs to \( M_i \) iff \( \mu_0' \in M_i \). Suppose it is known that \( \xi_0 \) belongs to at least one \( M_i, i=1, \ldots, t \). (When \( s > m_0 \), \( \xi_0 \) can belong to more than one \( M_i \)). We want to classify \( \xi_0 \) as belonging to (at least) one \( M_i, i=1, \ldots, t \).

Let \( \hat{\xi} \) denote the m.l.e. of \( \xi \) and \( V \) the Wishart estimate of \( \Sigma \) based on \( n-m \) d.f. Then

\[
\hat{\xi} = (A'A)^{-1} A'X' \text{ and } (n-m) V = (X' - A\hat{\xi})' (X' - A\hat{\xi}). \tag{5}
\]

Let \( \xi' \) be a \( s_i \times 1 \) real vector. Then

\[
\xi_0' - \xi' C_i \hat{\xi} = x_0' - \xi_0' B_i', \quad B_i = XA(A'A)^{-1} C_i . \tag{6}
\]

Define

\[
D^2(M_i) = \min_{\xi_0' \in R_{s_i}} (\xi_0' - B_i \xi_0')', V^{-1}(\xi_0' - B_i \xi_0') \cdot \tag{7}
\]

where

\[
\hat{\xi}_0 = (B_i V^{-1} B_i)^{-1} B_i V^{-1} \xi_0 . \tag{8}
\]

We call \( \hat{\xi}_0 \), the least square estimate of \( \xi_0 \). Sometimes \( \xi_0 \) and its estimate \( \hat{\xi}_0 \) may be very meaningful. \( D^2(M_i) \) in (7) can also be expressed as

\[
D^2(M_i) = \xi_0' V^{-1} \xi_0 - \xi_0' V^{-1} B_i (B_i V^{-1} B_i)^{-1} B_i V^{-1} \xi_0 . \tag{9}
\]

\( D^2(M_i) \) is the estimated Mahalanobis distance of \( \xi_0 \) from \( M_i \).
Classification Rule

Classify \( \chi_o \) as belonging to \( M_j \) iff

\[
D^2(M_j) = \min_{1 \leq i \leq t} D^2_i.
\] (10)

If minimum occurs for more than one \( j \), then any such \( j \) is acceptable.

Estimate of \( E \chi_o \) (\( = \mu_o \))

\[
E \chi_o' = \hat{\mu}_o' c_j' \xi. \text{ Hence } \hat{\mu}_o, \text{ the estimate of } \mu_o \text{ is given by}
\]

\[
\hat{\mu}_o = \hat{\xi}' c_j \hat{\mu}_j = X' A (A'A)^{-1} C_j (B_j' V^{-1} B_j)^{-1} B_j' V^{-1} \chi_o.
\] (11)

Remark: The above approach also holds when normality assumptions do not hold but the distributions are such that the Mahalanobis distance is meaningful. In those cases \( \hat{\xi} \) and \( V \) given in (5), are the least square estimates.

2.2 Special Case.

Let

\[
E X' = A \xi, A' = (a_{i1}, a_{i2}, \ldots, a_{im}).
\] (12)

Without loss of generality, assume \( a_{i1}, a_{i2}, \ldots, a_{im} \) to be linearly independent. Let \( \hat{\mu}_i' \xi = \mu_i', i = 1, \ldots, m \). Let \( \mu_1, \mu_2, \ldots, \mu_m \) respectively relate to \( m \) time points \( t_1 < t_2 < \ldots < t_m \). Let \( C_i = (a_{i1}, a_{i+1}) i = 1, \ldots, m-1 \). Then \( M_i \) is the linear space spanned by \( \mu_{i1}, \mu_{i1+1} \). We denote

\[
M_i = \langle \mu_i, \mu_{i1+1} \rangle, i = 1, \ldots, m-1.
\] (13)

The question from which \( M_i, \chi_o \) comes can be answered using the above approach. Here each \( \mu_i \) is a 2-dimensional vector. In growth/learning processes \( \mu_1, \mu_2, \ldots, \mu_m \) may refer to \( m \) states.
of growth/learning, which are ordered. Let \( \beta_1 = (\beta_{11}, \beta_{12})' \). For 
\[ \beta_{11} + \beta_{12} = \lambda \]
and \( \beta_{11} > 0, \beta_{12} > 0 \), the exterior state if one of the \( \beta_{11}, \beta_{12} \) is positive and the other negative. If \( \beta_{12} > 1, \beta_{11} < 0 \), we say that the exterior state is closer to \( \beta_{i+i} \). Similar considerations apply for the exterior state closer to \( \beta_{1i} \). In the next two sections, we continue discussing this special case.

3.1 Classification into Closed Subsets of the Manifolds

Let

\[ U_i = (1-k_i) \beta_i + k_i \beta_{i+1} : k_i \in \mathbb{R}, i=1, \ldots, m-1 \]  \hspace{1cm} (14)

Define

\[ D^2(U_i) = \min_{k_i \in \mathbb{R}} (\chi_0 - (1-k_i) \beta_i - k_i \beta_{i+1})' V^{-1} (\chi_0 - (1-k_i) \beta_i - k_i \beta_{i+1}) \]

\[ = (\chi_0 - (1-k_i) \beta_i - k_i \beta_{i+1})' V^{-1} (\chi_0 - (1-k_i) \beta_i - k_i \beta_{i+1}) \]  \hspace{1cm} (15)

where

\[ k_i = \frac{\beta_{i+1} - \beta_i}{\beta_{i+1} - \beta_i}' V^{-1} ((\lambda_{i1} - 1) \beta_i + \lambda_{i2} \beta_{i+1}) \]

\[ \hat{\beta}_{i+1} = \hat{\beta}_i + \frac{\hat{\mu}_{i+1}}{\hat{\beta}_{i+1}' \beta_{i+1}} \]

and \( \hat{\beta}_i \) is defined by (8) in which 
\[ B_i = X A (A'A)^{-1} (\hat{\mu}_i, \hat{\beta}_{i+1}) \]

Classification Rule: Classify \( \chi_0 \) as belonging to or closer to \( U_i \) iff

\[ D^2(U_i) = \min_{1 \leq i \leq m-1} D^2(U_i) \]  \hspace{1cm} (17)

The estimate of \( E \chi_0 \) when it is given that it belongs to one of the \( U_i \)'s is given by 
\[ (1-k_j) \hat{\mu}_j + k_j \hat{\mu}_{i+1} \]

Remark: If \( U_i \) consists of a finite number of points, the above approach is available. Also the likelihood ratio procedure is
available. Computationally, however, ours as well as the likelihood ratio approach will become cumbersome, when the total number of points is large.

3.2 Classification into Closed Convex Subsets of the Manifolds

Let

\[ S_i = \{ \lambda_{11}^*, \lambda_{12}^*: \lambda_{11}^* > 0, \lambda_{12}^* > 0, \lambda_{11}^* + \lambda_{12}^* = 1, i = 1, \ldots, m-1 \} \]  \hspace{1em} (18)

Define

\[ D^2(S_i) = \min_{\lambda_{11}^+, \lambda_{12}^+} (x_0 - \lambda_{11}^+ \hat{u}_{i1} - \lambda_{12}^+ \hat{u}_{i1+1})' V^{-1} (x_0 - \lambda_{11}^+ \hat{u}_{i1} - \lambda_{12}^+ \hat{u}_{i1+1}) \]

\[ = (x_0 - \hat{\lambda}_{i1}^+ \hat{u}_{i1} - \hat{\lambda}_{i1+1}^+ \hat{u}_{i1+1})' V^{-1} (x_0 - \hat{\lambda}_{i1}^+ \hat{u}_{i1} - \hat{\lambda}_{i1+1}^+ \hat{u}_{i1+1}) \]  \hspace{1em} (19)

where \( \hat{u}_{i1} \) and \( \hat{u}_{i1+1} \) are as given in section 3.1. We call \( \hat{\lambda}_{i1}^+ \) and \( \hat{\lambda}_{i1+1}^+ \) the least square estimates of \( \lambda_{11}^+ \) and \( \lambda_{12}^+ \) corresponding to the convex set \( S_i \). They are as follows:

\[ \hat{\lambda}_{i1}^+ = 1 - k_i, \hat{\lambda}_{i2}^+ = k_i, \text{ if } 0 < k_i < 1, \]

\[ \hat{\lambda}_{i1}^+ = 1, \hat{\lambda}_{i2}^+ = 0, \text{ if } k_i < 0 \]

\[ \hat{\lambda}_{i1}^+ = 0, \hat{\lambda}_{i2}^+ = 1, \text{ if } k_i > 1 \]  \hspace{1em} (20)

**Classification Rule:** Classify \( x_0 \) as belonging to or closer to \( S_j \) iff

\[ D^2(S_j) = \min_{1 < i < m-1} D^2(S_i) \]  \hspace{1em} (21)

The estimate of \( E x_0 \), when it is given that it belongs to one of \( S_i \)'s is given by \( \hat{x}_{j1}^+ \hat{u}_{j1} + \hat{x}_{j2}^+ \hat{u}_{j1+1} \)
4.1 Numerical Illustrative Example 1

The data for this example is taken from Srivastava (1983). He gives the 3-dimensional sample means for the three species, concinna, Heikertlingeri and chaetocnema heptapotamica each based on 10 observations and the $3 \times 3$ common sample covariance matrix based on 27 d.f. Each observation $\chi_i$ consisted of three measurements $x_1, x_2, x_3$ where

- $x_1 =$ width of the first joint of the first tarsus (in microns)
- $x_2 =$ same for the second joint and
- $x_3 =$ maximal width of the aedegus (in microns).

$x_i$ was the $i$th component of the vector $\chi_i$, $i = 1, 2, 3$. The sample means for the above three species were respectively,

$$\hat{\mu}_1' = (179.1, 127.4, 50.5)', \quad \hat{\mu}_2' = (208.2, 119.8, 48.9)'$$
$$\hat{\mu}_3' = (136.8, 122.8, 50.7)'$$

and the sample covariance matrix was

$$V = \begin{bmatrix}
198.67 & 94.13 & 32.82 \\
94.13 & 63.24 & 18.08 \\
32.82 & 18.08 & 8.72
\end{bmatrix}$$  \hspace{1cm} (23)

We construct the new observation $\chi_o$, as

$$\chi_o' = \frac{1}{3} (\hat{\mu}_1 + \hat{\mu}_2 + \hat{\mu}_3)' = (174.70, 123.33, 50.03),$$  \hspace{1cm} (24)

and pretend that it is coming from a new population.

Let $\mu_4 = \mu_1$ and

$$M_1 = <\mu_1, \mu_2>, M_2 = <\mu_2, \mu_3>, \text{ and } M_3 = <\mu_3, \mu_4>,$$  \hspace{1cm} (25)

$$U_1 = \{(1 - k_1)\mu_1 + k_1 \mu_2 : k_1 \in \mathbb{R}_1\},$$

$$U_2 = \{(1 - k_2)\mu_2 + k_2 \mu_3 : k_2 \in \mathbb{R}_1\},$$

$$U_3 = \{(1 - k_3)\mu_3 + k_3 \mu_4 : k_3 \in \mathbb{R}_1\},$$  \hspace{1cm} (26)
\[ S_1 = \{ z_{11}^* + z_{12}^* : z_{11}^* \geq 0; z_{12}^* \geq 0, z_{11}^* + z_{12}^* = 1 \} , \]
\[ S_2 = \{ z_{21}^* + z_{22}^* : z_{21}^* \geq 0, z_{22}^* \geq 0, z_{21}^* + z_{22}^* = 1 \} , \]
\[ S_3 = \{ z_{31}^* + z_{32}^* : z_{31}^* \geq 0, z_{32}^* \geq 0, z_{31}^* + z_{32}^* = 1 \} . \]

\[ M_1, M_2, M_3 \] are manifolds, \( U_1, U_2, U_3 \) are lines joining two points out of \( \lambda_1^*, \lambda_2^* \) and \( \lambda_3^* \) and \( S_1, S_2, S_3 \) are convex subsets of \( U_1, U_2, U_3 \). \( S_i \subset U_i \subset M_i, i = 1, 2, 3 \).

We examine three questions, namely, whether (i) \( \chi_0 \) comes from (or is closer to) \( M_1 \) or \( M_2 \) or \( M_3 \); (2) \( \chi_0 \) comes from (or is closer to) \( U_1 \) or \( U_2 \) or \( U_3 \) and (iii) \( \chi_0 \) comes from (or is closer to) \( S_1 \) or \( S_2 \) or \( S_3 \). Since \( \chi_0 \) is a \( 3 \times 1 \) vector and \( M_i \) are 2-dimensional manifolds, \( \chi_0 \) is likely to be very close to at least one of the \( M_i \)'s and examining question (i) may not be very meaningful. We do it for the sake of completeness.

Let

\[ B_1 = (\hat{\eta}_1, \hat{\eta}_2), B_2 = (\hat{\eta}_2, \hat{\eta}_3), B_3 = (\hat{\eta}_3, \hat{\eta}_4), \hat{\eta}_4 = \hat{\eta}_1 , \]
\[ \hat{\chi}_i = (B_i^T V^{-1} B_i)^{-1} B_i^T V^{-1} \chi_0 , \quad i = 1, 2, 3 . \]

Then

\[ \hat{\chi}_1 = (0.9732, 0.01103)', \quad \hat{\chi}_2 = (0.5052, 0.5030)' \]
\[ \hat{\chi}_3 = (0.02070, 0.9606)' . \]

Using (7), we obtain

\[ D^2(M_1) = 0.2991 , \quad D^2(M_2) = 0.0802 , \quad D^2(M_3) = 0.2901 . \]

The minimum \( D^2(M_i) \) is \( D^2(M_2) \). Among \( M_1, M_2 \) or \( M_3 \), \( \chi_0 \) is closest to \( M_2 \).

The estimate of \( \hat{\chi}_0 \) corresponding to \( M_1 \) is \( B_1 \hat{\chi}_1 \). The best one corresponding to \( M_2 \) is \( B_2 \hat{\chi}_2 = 0.5052 \hat{\chi}_2 + 0.5030 \hat{\chi}_3 \).
Using (16), we obtain
\[ \hat{k}_1 = 0.0306, \hat{k}_2 = 0.5031, \hat{k}_3 = 0.9935 \]  \hspace{1cm} (32)

From (15), we obtain
\[ D^2(U_1) = 0.3681, D^2(U_2) = 0.0997, D^2(U_3) = 0.4013. \]  \hspace{1cm} (33)

The minimum \( D^2(U_1) \) is \( D^2(U_2) \).

The estimate of \( \mu \) corresponding to \( U_i \) is given by

\( (1 - \hat{k}_i) \hat{\mu}_1 + \hat{k}_i \hat{\mu}_{i+1} \). The best one corresponding to \( U_2 \) is

\( 0.4969 \hat{\mu}_2 + 0.5031 \hat{\mu}_3 \).

From (20) and (19) respectively, we obtain
\[ \hat{\lambda}^*_{11} = 1 - \hat{k}_1, \hat{\lambda}^*_{12} = \hat{k}_1; \hat{\lambda}^*_{21} = 1 - \hat{k}_2, \hat{\lambda}^*_{22} = \hat{k}_2; \hat{\lambda}^*_{31} = 1 - \hat{k}_3, \hat{\lambda}^*_{32} = \hat{k}_3, \]  \hspace{1cm} (34)

and
\[ D^2(S_1) = D^2(U_1) = 0.3681, D^2(S_2) = D^2(U_2) = 0.0997, \]
\[ D^2(S_3) = D^2(U_3) = 0.4013. \]  \hspace{1cm} (35)

The minimum \( D^2(S_1) \) is \( D^2(S_2) \).

The estimate of \( \mu \) corresponding to \( S_i \) is given by

\( \hat{\lambda}^*_{i1} \hat{\mu}_1 + \hat{\lambda}^*_{i2} \hat{\mu}_{i+1} \). The best one corresponding to \( S_2 \) is given by

\( 0.4969 \hat{\mu}_2 + 0.5031 \hat{\mu}_3 \).

\( \hat{\mu}_\infty \) seems to lie midway between \( \hat{\mu}_2 \) and \( \hat{\mu}_3 \). Conclusions like this in medical and biological sciences and educational studies may be very enlightening. The state of cancer disease after treatment could recede to lie between the less advanced and the (previously attained) more advanced stage of the disease, showing effectiveness of the treatment. In biological sciences, a new species could be classed to lie between two known species. In education, a child's ability in mathematics could be classed between grades 5 and 6, perhaps two-thirds closer to grade 6.
4.2 Example 2

The data for this example is taken from Rao (1952) where he considers group constellation for the 12 castes and tribes of the United Provinces, India, on the basis of 9 characters. The 9 characters were: (1) Head length ($x_1$), (2) Head Breadth ($x_2$), (3) Biszygomatic Breadth ($x_3$), (4) Nasal height ($x_4$), (5) Nasal Breadth ($x_5$), (6) Nasal Depth ($x_6$), (7) Stature ($x_7$), (8) Sitting Height ($x_8$) and (9) Frontal Breadth ($x_9$). For the ease of calculations of $D^2$, Rao (1952) transformed $x_1, \ldots, x_9$ into uncorrelated variables $y_1, \ldots, y_9$, each having unit variance. This enables computing $D^2$ corresponding to any two groups as the sum of squares of the differences in the mean values of $y_1, \ldots, y_9$. For convenience, we also use $y$ measurements for calculating our $D^2$. Let $\bar{y}_i = (y_1, \ldots, y_9)'$ and $\mu_i$ denote the sample mean $y$ for the caste $A_i, i=1,2,3$ and let $\bar{x}_o$ denote the sample mean of $\bar{x}_o$ for $A_4$. From Rao (1952),

\[
\begin{align*}
\hat{\mu}_1' &= -0.143 \ 0.061 -0.053 -0.342 -0.289 0.005 -0.235, 0.462 -0.015 \\
\hat{\mu}_2' &= 0.071 -0.041 -0.032 -0.298 -0.075 -0.205, -0.286 0.081 -0.016 \\
\hat{\mu}_3' &= -0.107 -0.237 0.009 -0.387 -0.004 -0.321 -0.196 -0.020 0.064 \\
\hat{\mu}_4' &= 0.066 -0.399 -0.112 -0.414 0.142 -0.258 -0.360 -0.530 -0.012.
\end{align*}
\]

(36)

Rao (1952) classes the four Artisans $A_1, A_2, A_3$ and $A_4$ as a cluster and further remarks that they appear to be linearly arranged in the order $A_1, A_2, A_3$ and $A_4$ with $A_4$ being most distant from $A_1$.

Let $M_1, U_1, S_1, D^2(M_1), D^2(U_1), D^2(S_1), i=1,2,3$ be as defined in section 4.1. Then for this example,

\[
\begin{align*}
\hat{\lambda}_1 &= (-1.08968, 2.08904)' \ , \ \\
\hat{\lambda}_2 &= (-0.07748, 1.11916)' \ , \\
\hat{\lambda}_3 &= (1.43713, -0.71890) \\
D^2(M_1) &= 0.2249, \ D^2(M_2) = 0.4238, \ D^2(M_3) = 0.2235
\end{align*}
\]
\[ \hat{k}_1 = 2.0899, \quad \hat{k}_2 = 1.1220, \quad \hat{k}_3 = -0.6165 \]
\[ D^2(U_1) = 0.2302, \quad D^2(U_2) = 0.4304, \quad D^2(U_3) = 0.2523 \] (37)
\[ \hat{k}^*_{11} = 0.0, \quad \hat{k}^*_{12} = 1.0, \quad \hat{k}^*_{21} = 0.0, \quad \hat{k}^*_{22} = 1.0, \quad \hat{k}^*_{31} = 1.0, \quad \hat{k}^*_{32} = 0.0 \]
\[ D^2(S_1) = 0.58, \quad D^2(S_2) = 0.4321, \quad D^2(S_3) = 0.4321 \]

Thus \( x_0 \) seems to lie on the line joining \( \hat{\mu}_{11} \) and \( \hat{\mu}_{12} \) and is an exterior point closer to \( \hat{\mu}_{12} \). Also \( x_0 \) seems to lie on the line joining \( \hat{\mu}_{11} \) and \( \hat{\mu}_{13} \) and is an exterior point closer to \( \hat{\mu}_{13} \). (From Rao (1952), we know that \( \hat{\mu}_{12} \) and \( \hat{\mu}_{13} \) are close, hence the above remarks make sense.)

Let us now denote \( x_0 \) as \( \hat{\mu}_4 \) and choose the closed set \( U \) as the line joining \( \hat{\mu}_{11} \) and \( \hat{\mu}_4 \) (where \( \mu_4 = E \hat{\mu}_4 \)) and ask for the best estimates of \( \hat{\mu}_{12} \) and \( \hat{\mu}_{13} \) in \( U \).

The estimate for \( \hat{\mu}_{12} \) in \( U \) is
\[ \hat{\mu}_{11} + \hat{k}^* (\hat{\mu}_4 - \hat{\mu}_{11}) \] (38)
where
\[ \hat{k}^*_2 = \frac{\hat{\mu}_{11} - \hat{\mu}_{14} - \hat{\mu}_{12} + \hat{\mu}_{14}}{\hat{\mu}_{11} - \hat{\mu}_{14} + \hat{\mu}_{12}} = 0.3795 \] (39)

The estimate for \( \hat{\mu}_{13} \) in \( U \) is
\[ \hat{\mu}_{11} + \hat{k}^* (\hat{\mu}_4 - \hat{\mu}_{11}) \] (40)
where
\[ \hat{k}^*_3 = \frac{\hat{\mu}_{13} - \hat{\mu}_{12} - \hat{\mu}_{11} + \hat{\mu}_{14}}{\hat{\mu}_{13} - \hat{\mu}_{12} + \hat{\mu}_{11}} = 0.5234 \] (41)

The values of \( \hat{k}^*_2 \) and \( \hat{k}^*_3 \), confirm Rao's (1952) remarkable comments "...that \( A_1, A_2, A_3, A_4 \) would appear to be linearly
arranged in the order $A_1, A_2, A_3$ and $A_4$ with $A_4$ being most distant from $A_1$.

5. **Bonferoni Lower Bound of Probability of Correctly Classifying an Observation into One of the $t$ Given Normal Populations When the Given Set of Observations Follow a Multivariate General Linear Hypothesis Model and the Common Covariance Matrix $\Sigma$ Is Known**

We assume that our given observations $X(p \times n)$ obey a multivariate general linear hypothesis model as specified in (1) and $\Sigma$ is known. Let $X_0 (p \times 1)$ be a new observation and let it be given that it belongs to one of the $t$ populations $N(\theta_i, \Sigma)$, $i = 1, \ldots, t$, where $\theta_i' = C_i \xi$. $C_i : s_i \times 1$, $C_i : m \times s_i$ are known, $\xi : m \times p$ is unknown. We shall also call $N(\theta_i, \Sigma)$ as population $P_i$.

Let

$$D^2(\xi_i) = (X_0 - B_i \xi_i)'\Sigma^{-1}(X_0 - B_i \xi_i), \quad i = 1, \ldots, t,$$

(42)

where $B_i$ is given by (6).

Our classification rule is, classify $X_0$ under $P_j$ iff

$$D^2(\xi_j) = \min_{1 \leq i \leq t} D^2(\xi_i).$$

(43)

The estimate of the probability of correctly classifying $X_0$ as belonging to $P_i$ (under our procedure), when it does come from $P_i$ is difficult. We obtain an estimate of its lower bound and call it Bonferoni lower bound. In the following, $V(\xi_0)$ will denote the covariance matrix of the random vector $\xi_0$.

Let

$$E = X_0, \Sigma^{-\frac{1}{2}}X_0 = \chi_0^*, \Sigma^{-\frac{1}{2}}X = \chi^*, A' = (\xi_1, \ldots, \xi_n)$$

$$A(A'A)^{-1}C_i \xi_i = (r_{i1}, \ldots, r_{in})', \quad i = 1, \ldots, t.$$

(44)

Then

$$V(\Sigma^{-\frac{1}{2}}X_0 - \Sigma^{-\frac{1}{2}}B_i \xi_i) = V(\chi_0^* - \chi^* \xi_i) = (1 + \xi_i')\Sigma \xi_i).$$

(45)
Let

\[ \chi_{ij} = (k_{ij1})^{-1}[(x^*_{ij1} - x^*_{ij}) - (x^*_{ij1} - x^*_{ij})], \quad k_{ij1} = \sum_{q=1}^{n} (r_{i1q} - r_{i1q})^2 \]

\[ \chi_{ij} = (k_{ij2})^{-1}[(x^*_{ij1} + x^*_{ij}) - (x^*_{ij1} + x^*_{ij})], \quad k_{ij2} = \sum_{q=1}^{n} (r_{i1q} + r_{i1q})^2 \]

for \( i \neq j = 1, \ldots, t \)

Then

\[ V \left[ \begin{array}{c} \nu_{ij} \\ \chi_{ij} \end{array} \right] = V \begin{pmatrix} \nu_{ij} \\ \chi_{ij} \end{pmatrix} = \begin{pmatrix} I_p & \rho_{ij} I_p \\ \rho_{ij} I_p & I_p \end{pmatrix}, \quad \rho_{ij} = \sum_{q=1}^{n} (r_{i1q} - r_{i1q})^2 / k_{ij1} k_{ij2}, \]

\[ E \nu_{ij} = k_{ij1}^{-1} \sum_{q=1}^{n} (r_{i1q} - r_{i1q}) (\Sigma_{i1q}^{-1} \xi), \quad i \neq j = 1, \ldots, t, \]

\[ E \chi_{ij} = k_{ij2}^{-1} [2 \Sigma_{i1q}^{-1} \xi - \Sigma_{i1q}^{-1} (r_{i1q} + r_{i1q}) (\Sigma_{i1q}^{-1} \xi)], \quad i \neq j = 1, \ldots, t. \]

Note \( E \nu_{ij} \) and \( E \chi_{ij} \) are not known since \( \xi \) is not known but estimates of \( E \nu_{ij} \), \( E \chi_{ij} \) are available by replacing \( \xi \) in (47) and (48) by \( \hat{\xi} \) given in (5). Note \( \xi_i \) and \( \rho_{ij} \) are known,

\( i = 1, \ldots, t \).

Let

\[ e_{ij} = \Pr \{ D^2(\nu_{ij}) > D^2(\chi_{ij}) \mid \nu_{i1} = \nu_{i1} \}
\]

\[ = \Pr \{ k_{ij1} k_{ij2} \nu_{ij}^* \chi_{ij} > 0 \mid \nu_{i1} = \nu_{i1} \}
\]

\[ = \Pr \left( \frac{(\nu_{ij}^* + \nu_{ij}) \nu_{ij}^* + \nu_{ij}}{(\nu_{ij}^* - \nu_{ij}) \nu_{ij}^* + \nu_{ij}} > 1 \mid \nu_{i1} = \nu_{i1} \}
\]

\[ = \Pr \{ F_{pp}(\lambda_{ij1}, \lambda_{ij2} > (1 - \rho_{ij})/(1 + \rho_{ij}) \}
\]

where \( F_{pp}(\lambda_{ij1}, \lambda_{ij2}) \) has a doubly non-central \( F \)-distribution with \( p, p \) d.f. and non-centrality parameters \( \lambda_{ij1}, \lambda_{ij2} \) given by

\[ \lambda_{ij1} = (2(1 + \rho_{ij})^{-1} E(\nu_{ij} + \nu_{ij})^* E(\nu_{ij} + \nu_{ij}), \]

\[ \lambda_{ij2} = (2(1 - \rho_{ij})^{-1} E(\nu_{ij} - \nu_{ij})^* E(\nu_{ij} - \nu_{ij}). \]

\[ \]
E(\(\gamma_{ij} + \lambda_{ij}\)) are obtainable from (47) and (48). \(\lambda_{ij1}\) and \(\lambda_{ij2}\) are not known since they involve \(\xi\), which is unknown but their estimates \(\hat{\lambda}_{ij1}, \hat{\lambda}_{ij2}\) can be obtained by replacing \(\xi\) by \(\hat{\xi}\).

Note \(e_{ii} = 0\). Let \(e_{i} = \sum_{j=1}^{t} e_{ij}\). Then

\[
\Pr \{x_0 \in N(\bar{\theta}_{i1}, \Sigma) | \theta_i = \hat{\theta}_{i1} \} > 1 - e_{i}^*.
\]  

(52)

We call \(1 - e_{i}^*\), the Bonferoni lower bound. Let

\[
e_{ij} = \Pr \{F \geq \frac{\hat{\lambda}_{ij1} - \hat{\lambda}_{ij2}}{(1 - \rho_{ij})/(1 + \rho_{ij})}, \hat{\lambda}_{ij} = 0, \hat{\lambda}_{i} = \sum_{j=1}^{t} \hat{\lambda}_{ij}\}.
\]

(53)

Then \(1 - \hat{e}_{i}\) is the estimated lower bound. It can be computed by using tables of doubly noncentral F distributions. The lower bound might be close to the desired probability and hence a nice lower bound if (a) the means of \(\bar{\theta}_i\)'s are mutually far apart in terms of Mahalanobis distance and/or (b) the classification is to be done for a sample of size k, k large, and the classification procedure is based on the sample mean and/or (c) t is small. In many other cases, it may not be a nice bound.

**Remark.** If \(\Sigma\) is not known but its estimate \(V\) is based on a large number of d.f., then in the above \(\Sigma\) can be replaced by \(V\) to obtain an estimate of \(1 - e_{i}\).

**REFERENCES**


