CLASSIFICATION INTO TWO MULTIVARIATE NORMAL DISTRIBUTIONS WITH DIFFERENT COVARIANCE MATRICES

by

YOU-YEN YANG

B. S., National Taiwan University, 1958

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1965

Approved by:

Leslie F. Marcus

Major Professor

CONTENTS

INTRODUCTION

The development of a theory of statistical tests, as distinct from a collection of spacial examples, may be said to have begun with the introduction of the notation of types of error by Neyman and Pearson in 1928 and 1933 1,2/. Correspondingly, the initiation of a theoretical attack on the classification problem may be said to have begun when the Neyman-Pearson ideas were adapted to the discriminant function by Welch 1939 3/.

Welch only considered the problem of classifying an individual into one of two populations, say π_1 and π_2, and further restricted the problem by assuming that the probability density function of the measured quantities is completely known within each of the populations. Let $f_1(x_1, x_2, \ldots, x_p)$ denote the probability density of the observable quantities $X_1, X_2, \ldots, X_p$ in π_1 and let $f_2(x_1, x_2, \ldots, x_p)$ be the corresponding density in π_2. Welch 3/ observed that many methods of classifying an individual I into one of the two populations on the basis of the observations on $X_1, X_2, \ldots, X_p$, amount to a partition of the p-dimentional sample space of the X's into two regions, say R_1 and R_2, with the rule that I will be assigned to π_1 if the random point with coordinates $(X_1, \ldots, X_p)$ falls in R_1 and will be assigned to π_2 if $(X_1, X_2, \ldots, X_p)$ falls in R_2. The choice of a rule for classification or discrimination is thus equivalent to the choice of a partitioning of the sample space into the regions R_1 and R_2.

A discriminant function may be used for the purpose of classification in the case of two multivariete normal distributions with different mean vectors $\underline{\mu}_1$ and $\underline{\mu}_2$ and common covariance matrices Σ 4/. We shall consider the classification problem in the case of $\underline{\mu}_1 \neq \underline{\mu}_2$ and $\sum_1 \neq \sum_2$ where $\underline{\mu}_1$ and $\sum_1$ are from π_1 and $\underline{\mu}_2$ and $\sum_2$ are from π_2.

Suppose two vectors $\underline{x}$'s with p components are distributed in accordance with $N(\underline{\mu}_1, \sum_1)$ and $N(\underline{\mu}_2, \sum_2)$. The density function of the ith (i = 1, 2) distribution is

$$(1) \quad n(\underline{x} \mid \underline{\mu}_i, \textstyle\sum_i) = (2\pi)^{-p/2} \, |\textstyle\sum_i|^{-1/2} \exp[-1/2(\underline{x} - \underline{\mu}_i)'\textstyle\sum_i^{-1}(\underline{x} - \underline{\mu}_i)]$$

The best procedure for discrimination will be the likelihood ratio, i.e.,

$$L = \frac{n(\underline{x} \mid \underline{\mu}_2, \sum_2)}{n(\underline{x} \mid \underline{\mu}_1, \sum_1)} \quad .$$

An individual is classified into π_1 if L is less than a constant, say c, and into π_2 if $L \geq c$. If $\sum_1 = \sum_2$, then L is a linear function of $\underline{x}$.

In the p-variate case, $\underline{b}'\sum^{-1}\underline{b} = c$, which is an ellipsoid, where $\underline{b}$ is a vector with p-components, the shape and rotation of the ellipsoid are determined by $\sum$, and the size is determined by c. Consider two populations being centered around different points and with different patterns of scatterings. Under such a circumstance, the simplest procedure for classification will be a line or hyperplane. The following discussion will be restricted to a linear procedure.

Now, let $\underline{b} \neq 0$ be a vector of p components, and c be a scalar. An observation $\underline{x}$ is classified as from π_1, if $\underline{b}'\underline{x} < c$ and as from π_2, if $\underline{b}'\underline{x} \geq c$.

Suppose $\underline{\mu}_1 \neq \underline{\mu}_2$ and $\sum_1 \neq \sum_2$, and further assume that $\underline{x}$ is sampled from the ith ($i = 1, 2$) population then

$$E_i(\underline{b}'\underline{x}) = \underline{b}'\underline{\mu}_i$$

$$E_i[\underline{b}'\underline{x} - E_i(\underline{b}'\underline{x})]^2 = E_i(\underline{b}'\underline{x} - \underline{b}'\ \underline{\mu}_i)^2$$

$$= E_i(\underline{b}'\underline{x} - \underline{b}'\underline{\mu}_i)(\underline{b}'\underline{x} - \underline{b}'\underline{\mu}_i)'$$

$$= \underline{b}'E_i(\underline{x} - \underline{\mu}_i)(x - \underline{\mu}_i)'\underline{b}$$

$$= \underline{b}'\textstyle\sum_i\underline{b}$$

The probability of misclassifying an observation when it is from π_1 is

$$P_1 = 1 - p(y_1)$$

where

$$p(y_1) = \int_{-\infty}^{y_1} n(0, 1)dx$$

and the probability of misclassifying an observation when it comes from π_2 is

$$P_2 = 1 - p(y_2)$$

where

$$p(y_2) = \int_{-\infty}^{y_2} n(0, 1)dx$$

and where

(2)
$$y_1 = \frac{c - \underline{b}'\underline{\mu}_1}{(\underline{b}'\Sigma_1\underline{b})^{1/2}}$$

$$y_2 = \frac{\underline{b}'\underline{\mu}_2 - c}{(\underline{b}'\Sigma_2\underline{b})^{1/2}}$$

It is desired to minimize P_1 and/or P_2 or to maximize y_1 and/or y_2.

Some specific problems are to find 5/,

1) a procedure that minimizes the probability of one error of classification when the other is specified.

2) the procedure that minimizes the maximum probability of error.

The above solutions will be found within the set of admissible linear procedures 5/.

ADMISSIBLE PROCEDURE

Each procedure in terms of the two probabilities of misclassification can be put into the frame-work of the general decision problem 6/. There are two possible decisions D_1 and D_2. The appropriate decision depends on the value of y_1 and y_2 which are elements of sample space R. R can be decomposed into two subspaces R_1 and R_2 so that decision D_1 is preferred if y_1 belongs to R_1 and D_2 is preferred if y_2 belongs to R_2. The probability of misclassification associated with D and y is given by the loss for misclassification m(D; y), where

if y_1 belongs to R_1, then $m(D_1; y_1) = 0$

if y_2 belongs to R_2, then $m(D_2; y_2) = 0$

An attempt is made to minimize $m(D; y)$ (≥ 0).

Definition: A procedure is admissible if there is no other procedure which is better.

In the present case, one linear procedure is better than another if the m(D; y) obtained for that procedure is at least as small as the corresponding m(D; y) for the other procedure.

Consider the y_1, y_2-plane. From (2), we obtain

$$y_1(\underline{b}'\Sigma_1\underline{b})^{1/2} = c - \underline{b}'\underline{\mu}_1$$

$$y_2(\underline{b}'\Sigma_2\underline{b})^{1/2} = \underline{b}'\underline{\mu}_2 - c$$

Solving for y_1

$$(3) \qquad y_1 = \frac{\underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{1/2}}{(\underline{b}'\Sigma_1\underline{b})^{1/2}}$$

where $\underline{d} = \underline{\mu}_2 - \underline{\mu}_1$.

For a given y_2, y_1 is a function of $\underline{b}$, i.e., $y_1 = f(\underline{b})$. From (3), y_1 is continuous everywhere in $\underline{b}$, except $\underline{b} = 0$; and y_1 is homogeneous in $\underline{b}$ of degree zero, i.e., under a scalar transformation of $\underline{b}$, y_1 is invariant. We can, therefore, restrict $\underline{b}$ to lie on an ellipse, say $\underline{b}'\Sigma\underline{b} = k$, which is closed and bounded. y_1 is continuous and has a maximum. Thus, for a given y_2, y_1 has a maximum.

We differentitate (3) with respect to y_2, we get

$$\frac{dy_1}{dy_2} = -\frac{(\underline{b}'\Sigma_2\underline{b})^{1/2}}{(\underline{b}'\Sigma_1\underline{b})^{1/2}}$$

which is a negative quantity, therefore y_1 is a decreasing function of y_2.

From (3)

$$y_1 = \frac{\underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{1/2}}{(\underline{b}'\Sigma_1\underline{b})^{1/2}}$$

$$\frac{dy_1}{d\underline{b}} = \frac{1}{\underline{b}'\Sigma_1\underline{b}}\left[(\underline{b}'\Sigma_1\underline{b})^{1/2}\left\{\underline{d} - 1/2y_2(\underline{b}'\Sigma_2\underline{b})^{-1/2}\ (2\Sigma_2\underline{b})\right\} - \left\{\underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{1/2}\right\}\right.$$

$$\left. \cdot\ 1/2(\underline{b}'\Sigma_2\underline{b})^{-1/2}\ (2\Sigma_1\underline{b})\right]$$

$$= \frac{1}{(\underline{b}'\Sigma_1\underline{b})}\left[(\underline{b}'\Sigma_1\underline{b})^{1/2}\left\{\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{-1/2}\ (\Sigma_2\underline{b})\right\} - \left\{\underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{-1/2}\right\}\right.$$

$$\left. \cdot\ (\underline{b}'\Sigma_1\underline{b})^{1/2}\ (\Sigma_1\underline{b})\right]$$

$\underline{\underline{\text{set}}}$ 0.

Multiplying by $(\underline{b}'\Sigma_1\underline{b})^{1/2}$

$$(\underline{b}'\Sigma_1\underline{b})\left[\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{-1/2}\ (\Sigma_2\underline{b})\right] - \left[\underline{b}'\underline{d} - y_2(\underline{b}\Sigma_2\underline{b})^{1/2}\right] \cdot (\Sigma_2\underline{b}) = 0$$

$$t_1'\Sigma_1 b + t_2'\Sigma_1\underline{b} = t_3'\underline{d}$$

where

$$t_1' = \underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{1/2}$$

$$t_2' = y_2(\underline{b}'\Sigma_1\underline{b})/(\underline{b}'\Sigma_2\underline{b})^{1/2}$$

$$t_3' = \underline{b}'\Sigma_1\underline{b}$$

Let

$$t_1 = t_1'/t_3' = \frac{\underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{1/2}}{(\underline{b}'\Sigma_1\underline{b})}$$

$$t_2 = t_2'/t_3' = \frac{y_2(\underline{b}'\Sigma_1\underline{b})}{(\underline{b}'\Sigma_2\underline{b})^{1/2}(\underline{b}'\Sigma_1\underline{b})}$$

$$= y_2(\underline{b}'\Sigma_2\underline{b})^{-1/2}$$

$$\therefore \qquad (t_1\Sigma_1 + t_2\Sigma_2)\underline{b} = \underline{d}$$

(4) or $$b = (t_1\Sigma_1 + t_2\Sigma_2)^{-1}\,\underline{d}$$

Notice that a point (y_1, y_2), where y_1 is maximized with respect to $\underline{b}$ for a given y_2, corresponds to an admissible procedure (this will be proved later). By definition for a given point on the ellipse $\underline{b}'\Sigma\underline{b} = k$, there exists one and only one line through the point (y_1, y_2) which is the tangent to $\underline{b}'\Sigma\underline{b} = k$. This implies the fact that y_1 has one and only one maximum with respect to $\underline{b}$ for a given y_2.

From (2) and (3)

$$c = y_1(\underline{b}'\Sigma_1\underline{b})^{1/2} + \underline{b}'\underline{\mu}$$

$$= \frac{\underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{1/2}}{(\underline{b}'\Sigma_1\underline{b})^{1/2}} \cdot (\underline{b}'\Sigma_1\underline{b})^{1/2} + \underline{b}'\underline{\mu}_1$$

$$= \underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{1/2} + \underline{b}'\underline{\mu}_1$$

$$= t_1(\underline{b}'\Sigma_1\underline{b}) + \underline{b}'\underline{\mu}_1$$

or

$$c = \underline{b}'\underline{\mu}_2 - y_2\ (\underline{b}'\Sigma_2\underline{b})^{1/2}$$

$$= \underline{b}'\underline{\mu}_2 - t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}\ (\underline{b}'\Sigma_2)^{1/2}$$

$$= \underline{b}'\underline{\mu}_2 - t_2(\underline{b}'\Sigma_2\underline{b})$$

$$(5) \qquad \therefore \qquad c = \underline{b}'\underline{\mu}_1 + t_1\ (\underline{b}'\ \Sigma_1\underline{b})$$

$$= \underline{b}'\underline{\mu}_2 - t_2(\underline{b}'\Sigma_2\underline{b})$$

Theorem 1. If a point $z_1 > 0$, $z_2 > 0$ is admissible, there exist $t_1 > 0, t_2 > 0$ such that the procedure is defined by (4) and (5) 5/.

Proof: Let the admissible procedure be defined by the vector $\underline{c}$ and scalar s. The line

(6) $$y_1 = \frac{a - \underline{q}'\underline{\mu}_1}{(\underline{q}'\Sigma_1\underline{q})^{1/2}}\ , \qquad y_2 = \frac{\underline{q}'\underline{\mu}_2 - a}{(\underline{q}'\Sigma_2\underline{q})^{1/2}}$$

with a as parameter has the point (g_1, g_2) on it.

Solving for a, we obtain

$$y_1 = \frac{\underline{q}'\underline{d} - y_2(\underline{q}'\Sigma_2\underline{q})^{1/2}}{(\underline{q}'\Sigma_1\underline{q})^{1/2}}$$

(6') $$\frac{dy_1}{dy_2} = - \frac{(\underline{q}'\Sigma_2\underline{q})^{1/2}}{(\underline{q}'\Sigma_1\underline{q})^{1/2}}$$

and since both Σ_1 and Σ_2 are positive definite, dy_1/dy_2 is negative.

Hence, there exist positive numbers t_1 and t_2 and k such that the line

$$y_1 = \frac{\underline{q}'\underline{d} - y_2(\underline{q}'\Sigma_2\underline{q})^{1/2}}{(\underline{q}'\Sigma_1\underline{q})^{1/2}}$$

is tangent to the ellipse

(7) $$\frac{y_1^2}{t_1} + \frac{y_2^2}{t_2} = k$$

at the point (g_1, g_2). The slope of the line tangent to the ellipse at a point (y_1', y_2') is

$$-\frac{y_1' t_2}{y_2' t_1}$$

Consider a line defined by

(8) $$y_1 = \frac{c - \underline{b}'\underline{\mu}_1}{(\underline{b}'\Sigma_1\underline{b})^{1/2}} \qquad\qquad y_2 = \frac{\underline{b}'\underline{\mu}_2 - c}{(\underline{b}'\Sigma_2\underline{b})^{1/2}}$$

where $\underline{b}$ is a coefficient vector and c is a scalar which may be so chosen that the line is the tangent to (7) at point (y_1', y_2').

From (7)

$$\frac{y_1(c - \underline{b}'\underline{\mu}_1)}{t_1(\underline{b}'\Sigma_1\underline{b})^{1/2}} + \frac{y_2(\underline{b}'\underline{\mu}_2 - c)}{t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}} = k$$

$$\frac{(c - \underline{b}'\underline{\mu}_1)}{t_1(\underline{b}'\Sigma_1\underline{b})^{1/2}} \cdot \frac{dy_1}{dy_2} + \frac{(\underline{b}'\underline{\mu}_2 - c)}{t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}} = 0$$

From (6')

$$-\frac{(\underline{b}'\Sigma_2\underline{b})^{1/2}}{(\underline{b}'\Sigma_1\underline{b})^{1/2}} \cdot \frac{(c - \underline{b}'\underline{\mu}_1)}{t_1(\underline{b}'\Sigma_2\underline{b})^{1/2}} + \frac{(\underline{b}'\underline{\mu}_2 - c)}{t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}} = 0$$

$$c(t_2\underline{b}'\Sigma_2\underline{b} + t_1\underline{b}'\Sigma_1\underline{b}) = t_2(\underline{b}'\underline{\mu}_1)(\underline{b}'\Sigma_2\underline{b}) + t_2(\underline{b}'\underline{\mu}_2) \cdot (\underline{b}'\Sigma_1\underline{b})$$

(9)
$$c = \frac{t_1(\underline{b}'\underline{\mu}_2)(\underline{b}'\Sigma_1\underline{b}) + t_2(\underline{b}'\underline{\mu}_1)(\underline{b}'\Sigma_2\underline{b})}{\underline{b}'(t_1\Sigma_1 + t_2\,\Sigma_2)\underline{b}}$$

Substituting (9) into (8), we obtain

$$y_1 = \frac{t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}\underline{b}'\underline{d}}{\underline{b}'(t_1\Sigma_1 + t_2\Sigma_2)\underline{b}}$$

(10)

$$y_2 = \frac{t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}\underline{b}'\underline{d}}{\underline{b}'(t_1\Sigma_1 + t_2\Sigma_2)\underline{b}}$$

Substituting (10) into (7), we get

$$\frac{\left[\dfrac{t_1(\underline{b}'\Sigma_1\underline{b})^{1/2}\underline{b}'\underline{d}}{\underline{b}'(t_1\Sigma_1 + t_2\Sigma_2)\underline{b}}\right]^2}{t_1} + \frac{\left[\dfrac{t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}\ \underline{b}'\underline{d}}{\underline{b}'(t_1\Sigma_1 + t_2\Sigma_1)\underline{b}}\right]^2}{t_2} = k$$

(11)

$$k = \frac{t_1(\underline{b}'\Sigma_1\underline{b})(\underline{b}'\underline{d})^2 + t_2(\underline{b}'\Sigma_2\underline{b})(\underline{b}'\underline{d})^2}{[\underline{b}'(t_1\Sigma_1 + t_2\Sigma_2)b]^2}$$

Therefore, the point (y_1', y_2') is on the ellipse with a constant

(12)

$$\frac{(y_1')^2}{t_1} + \frac{(y_2')^2}{t_2} = \frac{(\underline{b}'\underline{d})^2}{\underline{b}'(t_1\Sigma_1 + t_2\Sigma_2)\underline{b}}$$

The right hand side of (12) is maximized with respect to $\underline{b}$, where $\underline{b}$ was given in (4). This procedure must be admissible. For, if there were a vector $\underline{b}$ such that (12) was larger than k, the point (g_1, g_2) would be within the ellipse with constant (11) (Fig. 1). This contradicts the fact

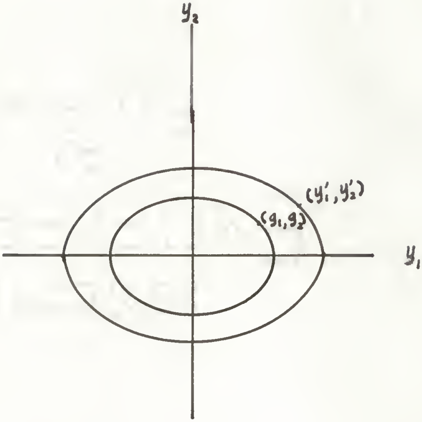

Fig. 1

that the procedure is admissible, for (z_1, z_2) would be nearer the origin than the tangent at (y_1', y_2'); then some points on this tangent line (corresponding to procedures with vector $\underline{b}$ and scalar c) would be better. This proves the assertion.

Since Σ_1 and Σ_2 are positive definite, and $t_1, t_2 > 0$, $t_1\Sigma_1 + t_2\Sigma_2$ is also positive definite. The right hand side of (12) is homogeneous of degree zero in $\underline{b}$, therefore for any scalar transformation of $\underline{b}$, the value of k is invariant.

When $\underline{b}$ is given by (4)

$$\underline{b} = (t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d}$$

$$d = (t_1\Sigma_1 + t_2\Sigma_2)\underline{b}$$

$$\underline{d}'\underline{b} = \underline{b}'\underline{d} = \underline{d}'(t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d}$$

$$\begin{aligned}\underline{b}'\underline{d} = \underline{d}'\underline{b} &= \underline{d}'(t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d} \\ &= \underline{b}'(t_1\Sigma_1 + t_2\Sigma_2)\underline{b}\end{aligned}$$

Then (10) reduces to

$$\begin{aligned} y_1 &= \frac{t_1(\underline{b}'\Sigma_1\underline{b})\underline{b}'\underline{d}}{\underline{b}'(t_1\Sigma_1 + t_2\Sigma_2)\underline{b}} \\ &= t_1(\underline{b}'\Sigma_1\underline{b})^{1/2} \\ y_2 &= t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}\end{aligned} \tag{13}$$

For simplicity, we normalize $t_1 + t_2 = 1$ if $t_1, t_2 > 0$, $t_1 - t_2 = 1$, if $t_1 > 0$, $t_2 < 0$ and $t_2 - t_1 = 1$ when $t_1 < 0$, $t_2 > 0$.

Lemma: Any positive definite symmetric matrix A can be expressed as the product of a non-singular matrix and its transpose. Conversely, every such product is a positive definite symmetric matrix 7/.

Proof: If A is nxn and positive definite, then the rank of A is n. This means that the congruent cannonical form of A is I. Therefore, there exists a non-singular matrix P such that

$$P'AP = I$$

$$A = (P')^{-1}I(P)^{-1}$$

$$= (P^{-1})'(P^{-1})$$

$$= Q'Q$$

where $Q = P^{-1}$. Conversely, if Q is non-singular, then Q'Q is non-singular. This implies that Q'Q is congruently similar to (δ_{ij}) which is the Kronecker delta.

$$(Q'Q)' = Q'Q$$

implies that Q'Q is symmetric. Therefore, Q'Q is positive definite. (Q.E.D.)

We know that $\sum_1$ and $\sum_2$ are positive definite. There exists a non-singular matrix P such that

$$P'\textstyle\sum_2 P = I$$

$$\textstyle\sum_2 = N'N$$

where $N = P^{-1}$. And there exists an orthogonal matrix R such that

$$R'(P'\textstyle\sum_1 P)R = M$$

where

$$M = (\delta_{ij}\, m_i);\ m_i \geq m_{i+1},\ j, i = 1, \ldots, p.$$

Notice that

$$R'(P'\Sigma_2 P)R = R'IR$$
$$= R'R$$
$$\Sigma_2 = I$$

and

$$P'\Sigma_2 P = I$$

$$(PR)'\Sigma_2(PR) = P'\Sigma_2 P$$

implies that

$$PR = P$$

$$(PR)'\Sigma_2(PR) = N'\Sigma_2 N$$

Define

$$\underline{d} = N'\lambda$$

From (13)

$$y_1 = t_1(\underline{b}'\Sigma_1\underline{b})^{1/2}$$

$$y_1 = t_1[((t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d})'\Sigma_1(t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d})]^{1/2}$$

$$= t_1[((t_1N'MN + t_2N'N)^{-1}N'\underline{\lambda})'N'MN((t_1N'MN + t_2N'N)^{-1}N'\underline{\lambda})]^{1/2}$$

$$= t_1[\underline{\lambda}'(t_1M + t_2I)^{-1}M(t_1M + t_2I)^{-1}\underline{\lambda}]^{1/2}$$

$$(14) \qquad = t_1[\sum_{j=1}^{p}(\lambda_j^2 m_j/(t_1m_j + t_2)^2]^{1/2}$$

For the same reason

$$y_2 = t_2[\sum_{j=1}^{p} \lambda_j^2/(t_1 m_j + t_2)^2]^{1/2} \tag{15}$$

From (14)

$$y_1^2 = t_1^2 \sum_{j=1}^{p} (\lambda_j^2/(t_1 m_j + t_2)^2)$$

$$\frac{dy_1^2}{dt_1} = \sum_{j=1}^{p} \frac{2t_1\lambda_j^2 m_j(t_1 m_j + t_2)^2 - 2t_1 m_j^2 \lambda_j^2 (t_1 m_j + t_2)}{(t_1 m_j + t_2)^4}$$

$$= 2t_1 t_2 \sum_{j=1}^{p} [\lambda_j^3 m_j/(t_1 m_j + t_2)^3] \tag{16}$$

The quantity is positive, since $t_1, t_2 > 0$, this means that y_1^2 is an increasing function of t_1. The same argument shows that y_2^2 is a decreasing function of t_2, i.e.,

$$\frac{dy_2^2}{dt_1} = -2 \sum_{j=1}^{p} [t_2^2 r_j^2 m_j/(t_1 m_j + t_2)^3] \tag{17}$$

Theorem 2. The procedure defined by $\underline{b} = (t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d}$ and c given by (5) for any t_1 and t_2 such that $t_1\Sigma_1 + t_2\Sigma_2$ is positive definite is admissible 3/.

Proof:

i) $t_1 + t_2 = 1$, i.e., let $t_1 > 0$, $t_2 > 0$

If the procedure defined by $\underline{b} = (t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d}$ were not admissible, there would be an admissible procedure which would, by theorem 1, be defined by $\underline{g} = (\tau_1\Sigma_1 + \tau_2\Sigma_2)^{-1}\underline{d}$ for some $\tau_1 + \tau_2 = 1$. However, (16), (17) indicate that y_1^2 and y_2^2 are increasing and decreasing functions of t_1, respectively; one of the coordinates corresponding to τ_1 would have to be less than one of coordinates corresponding to t_1. We can conclude that the procedure defined by $\underline{g} = (\tau_1\Sigma_1 + \tau_2\Sigma_2)^{-1}\underline{d}$ is not better than the procedure originally defined by $\underline{b}$, and this contradicts the assumption. Therefore, the original procedure is admissible.

ii) $t_1 = 0$, $t_2 = 1$

If $t_1 = 0$, from (13), $y_1 = 0$ and $y_2 = t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}$. But from (3)

$$y_1 = \frac{\underline{b}'\underline{d} - y_2(\underline{b}'\Sigma_2\underline{b})^{1/2}}{(\underline{b}'\Sigma_1\underline{b})^{1/2}}$$

since $y_1 = 0$, we get

$$y_2 = \frac{\underline{b}'\underline{d}}{(\underline{b}'\Sigma_2\underline{b})^{1/2}}$$

From (13)

$$\underline{b}'\Sigma_2\underline{b} = \underline{b}'\underline{d}$$

$$\underline{b} = \Sigma_2^{-1}\underline{d}$$

This is a form for $\underline{b}$ which maximizes y_2. Therefore, the procedure is admissible. In the case of $t_2 = 0$, $t_1 = 1$, the result will be similar to the above argument.

Remark: When y_2 is maximized with respect to $\underline{b}$, we obtain $\underline{b} = (t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d}$. In the case (ii), $t_1 = 0$, $t_2 = 1$,

$$\underline{b} = \Sigma_2^{-1}\underline{d} .$$

iii) $t_1 > 0$ and $t_2 < 0$, i.e., let $t_1 - t_2 = 1$.

Consider

$$\text{(18)} \qquad \frac{y_1^2}{t_1} + \frac{y_2^2}{t_2} = k, \qquad k > 0$$

which is a hyperbola which cuts y_1-axis at $\pm(t_1k)^{1/2}$. We are interested in the right hand branch. From (13), $y_1 > 0$ and $y_2 > 0$, substituting (2) into (18), we get

$$\frac{\dfrac{(c - \underline{b}'\underline{\mu}_1)^2}{\underline{b}'\Sigma_1\underline{b}}}{t_1} + \frac{\dfrac{(\underline{b}'\underline{\mu}_2 - c)^2}{\underline{b}'\Sigma_2\underline{b}}}{t_2} = k$$

$$\text{(19)} \qquad \frac{(c - \underline{b}'\underline{\mu}_1)^2}{t_1(\underline{b}'\Sigma_1\underline{b})} + \frac{(\underline{b}'\underline{\mu}_2 - c)^2}{t_2(\underline{b}'\Sigma_2\underline{b})} = k$$

As we have shown, the maximum value for (19) with respect to c for given $\underline{b}$ is exactly the same form as given in (5) and (6). The maximum of (19) is given by $\underline{b} = (t_1\Sigma_1 + t_2\Sigma_2)^{-1}\underline{d}$. By theorem 1, it follows that the procedure is admissible. We note that in the case of $t_1 < 0$, $t_2 > 0$ is similar to the case proved above.

USE OF ADMISSIBLE PROCEDURES

i) Minimization of one probability of misclassification for a specified probability of the other 5/.

For given $y_2 > 0$, i.e., the probability of misclassification when sampling from π_2, we want to maximize y_1, i.e., we want to minimize the probability of misclassification when sampling from π_1.

a) $t_1 + t_2 = 1$

If maximum $y_1 \geq 0$, we want to find $t_2 = 1 - t_1$, so $y_2 = t_2(\underline{b}'\Sigma_1\underline{b})^{1/2}$, where $\underline{b} = (t_1\Sigma_1 + t_2\Sigma_2)^{-1}\ \underline{d}$. The solution can be approximated by trial and error, since y_2 is an increasing function of t_2, until $t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}$ agrees closely with the desired y_2.

b) $t_1 - t_2 = 1$

y_2 is a decreasing function of t_2 (≤ 1). $y_2 = (\underline{d}'\Sigma_2\underline{d})^{1/2}$ when $t_2 = 1$. If the given $y_2 > (\underline{d}'\Sigma_2^{-1}\underline{d})^{1/2}$, then $y_1 < 0$, and we search for a value of t_2 so that the given $y_2 = t_2(\underline{b}'\Sigma_2\underline{b})^{1/2}$.

ii) The minimax procedure 8/.

This procedure involves finding $\underline{b}$ and t to satisfy

(20) $$[t\Sigma_1 + (1 - t)\Sigma_2]\underline{b} = \underline{d}$$

(21) $$\underline{b}'[t^2\Sigma_1 - (1 - t)^2\Sigma_2]\ \underline{b} = 0, \qquad 0 < t < 1.$$

Consider (21) may be taken as

$$\underline{b}'(\Sigma_1 - v\Sigma_2)\underline{b} = 0$$

where $v = [(1 - t)/t]^2$.

There is a non-singular matrix N such that $\Sigma_2 = N'N$ and $\Sigma_1 = N'MN$, where Λ is a diagonal matrix with λ_j $(j = 1, 2, \ldots, p)$ roots of $|\Sigma_1 - \lambda\Sigma_2| = 0$. On substitution of these values in (21) the quadratic form reduces to

$$(\underline{b}^*)' \ (\Lambda - vI)(\underline{b}^*) = 0,$$

where $\underline{b}^* = N\underline{b}$, which is equivalent to

$$\sum_{j=1}^{p} (\lambda_j - v)(b_j^*)^2 = 0.$$

This will not have a non-zero solution for $\underline{b}^*$, if the factors $(\lambda_j - v)$ are all positive or are all negative. This means, for minimax solution, v will have to lie between the largest and the smallest roots of

$$|\Sigma_2 - \lambda\Sigma_2| = 0.$$

REFERENCES

1. Neyman, J., and Pearson, E. S., (1933a). On the problem of the most efficient tests of statistical hypothesis. Phil. Trans., A, 231, 289-337.

2. __________, (1933b). The testing of statistical hypothesas in relation to probabilities a priori. Proc. Camb. Phil. Soc., 29, 492-510.

3. Welch, B., (1939). Note on discriminant functions. Biom., 31, 218-220.

4. Anderson, T. W., (1958). An introduction to multivariata statistical analysis. John Wiley, New York.

5. Andarson, T. W. and Bahadur, R. R., (1962). Classification in two multivariate normal distributions with different covariance matrices. Ann. Math. Stat. 33, 420-431.

6. Mood, A. M. and Graybill, F. A., (1963). Introduction to tha thaory of statistics. McGraw-Hill, New York.

7. Fuller, L. E., (1965). Theory of matrices. (Unpublished note)

8. Banerjee, K. S. and Marcus, L. F., (1965). Bounds in a matrix classification procedure. Accepted for publication in Biometrika 1965.

9. Jolicoeur, P. and Mosimann, J. E., (1960). Size and shape variation in the painted turtla. Growth, 24, 338-354.

ACKNOWLEDGEMENT

The writer wishes to express his thanks to his Major Professor, Dr. Leslie F. Marcus for encouraging him in his academic years and for helping him in the preparation of this report. He is also deeply indebted to Dr. K. S. Banerjee for carefull reading of the manuscript and many valuable suggestions and comments.

APPENDIX

Numerical Example

In order to illustrate the minimax procedure developed in this report, a numerical example is given below. The data for this example are from Jolicoeur and Mosimann 9/.

Three characters of carapace dimentions of painted turtles (*Chrysemys picta marginata*) were measured in mm.

24 Males			24 Females		
Length	Width	Height	Length	Width	Height
93	74	37	98	81	38
94	78	35	103	84	38
96	80	35	103	86	42
101	84	39	105	86	40
102	85	38	109	88	44
103	81	37	123	92	50
104	83	39	123	95	46
106	83	39	133	99	51
107	82	38	133	102	51
112	89	40	133	102	51
113	88	40	134	100	48
114	86	40	136	102	49
116	90	43	137	98	51
117	90	41	138	99	51
117	91	41	141	105	53
119	93	41	147	108	57
120	89	40	149	107	55
120	93	44	153	107	56
121	95	42	155	115	63
125	93	45	155	117	60
127	96	45	158	115	62
128	95	45	159	118	63
131	95	46	162	124	61
135	106	47	177	132	67

Table 1 Three characters of carapace dimentions of painted turtles.

CLASSIFICATION INTO TWO MULTIVARIATE NORMAL DISTRIBUTIONS WITH DIFFERENT COVARIANCE MATRICES

by

YOU-YEN YANG

B. S., National Taiwan University, 1958

ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1965

The classification of vector random variables as coming from one of two multivariate normal distributions leads to linear likelihood procedure when the covariance matrices are equal. It is possible to find a linear admissible procedure which will minimize the probability of misclassification of observations from one of two multivariate normal distributions with different mean vectors $(\underline{\mu}_1, \underline{\mu}_2)$ and covariance matrices (Σ_1, Σ_2). Two approaches have been discussed in this report; 1) minimization of one probability of misclassification for a specified probability of the order, 2) minimax procedure. The former approach minimizes the probability of misclassification when sampling from π_1; and the latter finds a value of t which ranges between the maximal and minimal characteristic root of $\Sigma_1 - \lambda\Sigma_2 = 0$, such that $\underline{b}'[t^2\Sigma_1 - (1 - t)^2\Sigma_2]\underline{b} = 0$. An example, using carapace dimensions of painted turtles, has been cited to illustrate the application of the procedures.

The mean vectors and covariance matrices for the three characters of the male and female painted turtles are:

$$\underline{\mu}_1' = (113.38 \quad 88.29 \quad 40.71), \qquad \underline{\mu}_2' = (136.00 \quad 102.58 \quad 51.96)$$

$$\Sigma_1 = \begin{bmatrix} 138.77 & 79.15 & 37.38 \\ 79.15 & 50.04 & 21.65 \\ 37.38 & 21.65 & 11.26 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 451.39 & 271.17 & 168.70 \\ 271.17 & 171.73 & 103.29 \\ 168.70 & 103.29 & 66.65 \end{bmatrix}$$

Then to find $|\Sigma_1\Sigma_2^{-1} - \lambda I| = 0$, i.e.,

$$\lambda^3 - 1.69321 \ \lambda^2 + 0.75937 \ \lambda - 0.06303 = 0$$

Solving for λ, we obtain three roots as follows:

$$\lambda = 0.10710, \ 0.58427, \ 1.00784$$

But

$$\lambda = \frac{1-t}{t}^{\,2}$$

$$\therefore \quad t = \frac{1}{1 + \sqrt{\lambda}}$$

Substituting $\lambda^{1/2} = (1.00784)^{1/2}$, $(0.10710)^{1/2}$ into the above equation, we get $t = 0.49990$ and 0.75343, i.e., the value of t will fall between 0.75343 and 0.49990.

t was computed using an iterative scheme to satisfy (20) & (21). The solution was $t = 0.73286 \pm (\frac{1}{2})^{14}$ where 14 is number of iterations. This gives $b' = (0.37820 \quad 0.08406 \quad 1.62367)$.