Biometrika (1966), 53, 3 and 4, p. 339
Printed in Great Britain

Discriminant function between composite hypotheses and related problems

BY C. RADHAKRISHNA RAO
Indian Statistical Institute, Calcutta

SUMMARY

The paper deals with the problem of constructing discriminant functions when the alternative hypotheses are not simple but composite. Such a problem arises when it is intended to identify an individual as belonging to one of two sets, each set consisting of several populations mixed in unknown proportions. A general approach to this problem, using the concepts of decision theory, sufficient statistics and ancillary statistics is given. In particular, when the means of the alternative populations within a given set are linearly related and the distributions are p variate normal, the discriminant function comes out in a simpler form. It is linear when the dispersion matrices are the same for all the populations and quadratic when the dispersion matrices within sets differ. Methods of estimating the discriminant function from sample data are fully discussed. The fact that in the situations considered one has observations from a mixture of populations within a set does not create any difficulty.

1. INTRODUCTION

The discriminant function, as introduced by the late Sir Ronald Fisher, for deciding between two simple hypotheses (alternative populations) on the basis of observed data is the logarithm of the likelihood ratio for two simple hypotheses given the observations (Welch, 1939). It is known that the discriminant function so obtained provides a sufficient reduction of data for drawing inferences on the two alternative hypotheses (Smith, 1947). In some situations a discriminant function derived from two simple hypotheses may provide a sufficient reduction of data for drawing inferences on a wider set of alternatives (Rao, 1962). The importance of the latter result in practical applications of the discriminant function was demonstrated in earlier publications (Rao, 1961, 1962, 1965).

The question naturally arises as to what is a suitable discriminant function when the alternative hypotheses are not simple but composite. Such a problem was faced by Burnaby (1966) when he wanted to identify an individual as belonging to one of two *sets* of populations. Each set consisted of several populations, mixed in unknown proportions, of organisms of one kind representing different unknown stages of growth. The object was to decide to which of two kinds a given organism belongs when nothing is known about its stage of growth. The problem becomes different when some indicator of the organism's stage of growth is available (see Delany & Healy, 1964). The present paper is devoted to a general discussion of the former type of problem, and in particular to some theoretical considerations based on Burnaby's paper.

Some algebraic lemmas. We give some results in matrix algebra which are useful in discussions of the type of problems considered in this paper.

LEMMA 1*a*. *Let* B *be* $k \times p$ *matrix of rank* k *and* C *be* $(p-k) \times p$ *matrix of rank* $(p-k)$ *such that* $\mathbf{BC}' = \mathbf{0}$. *Then*

$$\mathbf{B}'(\mathbf{BB}')^{-1}\mathbf{B} + \mathbf{C}'(\mathbf{CC}')^{-1}\mathbf{C} = \mathbf{I}, \tag{1·1}$$

$$\mathbf{C}'(\mathbf{C\Lambda C}')^{-1}\mathbf{C} + \mathbf{\Lambda}^{-1}\mathbf{B}'(\mathbf{B\Lambda}^{-1}\mathbf{B}')^{-1}\mathbf{B\Lambda}^{-1} = \mathbf{\Lambda}^{-1}, \tag{1·2}$$

where $\mathbf{\Lambda}$ *is any positive definite matrix.*

The proof is easy and consists in showing that post-multiplication of (1·1) by $(\mathbf{B}' \vdots \mathbf{C}')$ and (1·2) by $(\mathbf{B}' \vdots \mathbf{\Lambda C}')$ reduce to identities. For a generalization of (1·1), see p. 60 of (Rao, 1965) and (Rao, 1966).

LEMMA 1*b*. *Let* L *and* $\boldsymbol{\delta}$ *be p-vectors and* B *and* $\mathbf{\Lambda}$ *be as defined in Lemma* 1*a*. *Then*

$$\sup_{\mathbf{L}} \frac{(\mathbf{L}'\boldsymbol{\delta})^2}{\mathbf{L}'\mathbf{\Lambda L}} \tag{1·3}$$

subject to the condition $\mathbf{L}'\mathbf{B}' = \mathbf{0}$ *is attained at*

$$\mathbf{L}^* = \mathbf{\Lambda}^{-1}(\mathbf{I} - \mathbf{B}'(\mathbf{B\Lambda}^{-1}\mathbf{B}')^{-1}\mathbf{B\Lambda}^{-1})\boldsymbol{\delta}. \tag{1·4}$$

Lemma 1*b* is a special case of Lemma 1*c*, which deals with the problem of restricted eigenvalues discussed by the author in an earlier paper (Rao, 1964). The result of Lemma 1*b* is also stated and proved by Burnaby (1966). We state Lemma 1*c* which is useful in many situations.

LEMMA 1*c* (restricted eigenvalue problem). *Consider a symmetric* $p \times p$ *matrix* G *and a positive definite matrix* $\mathbf{\Lambda}$. *Let* $\mathbf{B}'$ *be as defined in Lemma* 1*a*.

(i) *Let* $\mathbf{L}_1, \ldots, \mathbf{L}_q$ *be p-dimensional vectors such that*

$$\left.\begin{aligned} &(a)\ \mathbf{L}_i'\mathbf{L}_i = 1,\ \mathbf{L}_i'\mathbf{L}_j = 0 \quad \text{for} \quad i \neq j, \\ &(b)\ \mathbf{L}_i'\mathbf{B}' = \mathbf{0} \quad (i = 1, \ldots, q). \end{aligned}\right\} \tag{1·5}$$

Then

$$\sup_{\mathbf{L}_1, \ldots, \mathbf{L}_q} (\mathbf{L}_1'\mathbf{GL}_1 + \ldots + \mathbf{L}_q'\mathbf{GL}_q) \tag{1·6}$$

is attained when $\mathbf{L}_i = \mathbf{R}_i$, *the ith eigenvector of* $(\mathbf{I} - \mathbf{B}'(\mathbf{BB}')^{-1}\mathbf{B})\mathbf{G}$.

(ii) *Let* $\mathbf{L}_1, \ldots, \mathbf{L}_q$ *satisfy the conditions*

$$\left.\begin{aligned} &(a)\ \mathbf{L}_i'\mathbf{\Lambda L}_i = 1,\ \mathbf{L}_i'\mathbf{\Lambda L}_j = 0 \quad \text{for} \quad i \neq j, \\ &(b)\ \mathbf{L}_i'\mathbf{B}' = \mathbf{0} \quad (i = 1, \ldots, q). \end{aligned}\right\} \tag{1·7}$$

Then

$$\sup_{\mathbf{L}_1, \ldots, \mathbf{L}_q} (\mathbf{L}_1'\mathbf{GL}_1 + \ldots + \mathbf{L}_q'\mathbf{GL}_q) \tag{1·8}$$

is attained when $\mathbf{L}_i = \mathbf{R}_i$ *the ith eigenvector of*

$$\mathbf{G} - \mathbf{B}'(\mathbf{B\Lambda}^{-1}\mathbf{B}')^{-1}\mathbf{B\Lambda}^{-1}\mathbf{G} \tag{1·9}$$

with respect to $\mathbf{\Lambda}$.

2. DISCRIMINATION BETWEEN COMPOSITE HYPOTHESES: GENERAL METHODS

Let X denote a random variable and $P(.|\theta)$ the density function of X depending on a parameter θ belonging to a set Θ. Let H_1 be the hypothesis that $\theta \in \Theta_1$ and H_2 be the hypothesis that $\theta \in \Theta_2$, where Θ_1 and Θ_2 are exclusive subsets of Θ. The problem we consider is that of choosing between the composite hypotheses H_1 and H_2 on the basis of an observed value of X. Let us discuss a few possible approaches to the problem.

Solution based on similar divisions. Let R_1 and R_2 be two exclusive regions covering the entire sample space. The regions R_1, R_2 are said to provide a similar division of the space if there exist constants e_1, e_2 such that

$$\int_{R_2} P(x|\theta)\,dx = e_1 \quad \text{for each} \quad \theta \in \Theta_1, \tag{2·1}$$

$$\int_{R_1} P(x|\theta)\,dx = e_2 \quad \text{for each} \quad \theta \in \Theta_2. \tag{2·2}$$

Let us decide to choose H_1 if $x \in R_1$ and H_2 if $x \in R_2$. In such a case the errors committed are e_1 and e_2. For determining an optimum decision rule, we consider all similar divisions and choose the one for which the magnitudes of errors are the smallest subject to a given ratio of errors, or for which a given linear compound of errors is a minimum. There are two ways of arriving at such a solution.

Using a sufficient statistic. Let T be a sufficient statistic, a function of X, for θ restricted to Θ_1, and let the same statistic be sufficient also for θ restricted to Θ_2. Using the well-known factorization theorem (see Lehmann, 1959, p. 49) we may write

$$P(x|\theta) = \begin{Bmatrix} P(t|\theta)\,P_1(x|t), \ \theta \in \Theta_1, \\ P(t|\theta)\,P_2(x|t), \ \theta \in \Theta_2, \end{Bmatrix} \tag{2·3}$$

where the functions $P_1(x|t)$ and $P_2(x|t)$ are independent of θ and may be interpreted as conditional densities of observations given $T = t$.

If we choose two values $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$, then the discriminant function for distinguishing between θ_1 and θ_2 is $\log\{P(x|\theta_1)/P(x|\theta_2)\}$. Using the factorizations (2·3) we have

$$\frac{P(x|\theta_1)}{P(x|\theta_2)} = \frac{P(t|\theta_1)}{P(t|\theta_2)}\frac{P_1(x|t)}{P_2(x|t)}. \tag{2·4}$$

Taking logarithms, we have

$$\log\frac{P(x|\theta_1)}{P(x|\theta_2)} = \log\frac{P(t|\theta_1)}{P(t|\theta_2)} + \log\frac{P_1(x|t)}{P_2(x|t)} \tag{2·5}$$

which provides a decomposition of the discriminant function for the simple hypotheses θ_1, θ_2 as the sum of two discriminant functions, one based on t alone for discrimination 'within' a composite hypothesis and another on the conditional distributions given t for discrimination 'between' composite hypotheses.

Using an ancillary statistic. Another method is to consider a statistic S, function of X, such that its probability density,

$$P(s|\theta) = \begin{Bmatrix} P_1(s) & \text{independent of} & \theta \in \Theta_1, \\ P_2(s) & \text{independent of} & \theta \in \Theta_2, \end{Bmatrix} \tag{2·6}$$

or, in other words, S is an ancillary statistic for $\theta \in \Theta_1$ and also for $\theta \in \Theta_2$. When $P_1(s)$ and $P_2(s)$ are different, a discriminant function for choosing between H_1 and H_2 is provided by the likelihood ratio $P_1(s)/P_2(s)$.

Method of maximum-likelihood ratio. A discriminant function which may have wide applicability is the ratio

$$\{\sup_{\theta\in\Theta_1} P(x|\theta)\}/\{\sup_{\theta\in\Theta_2} P(x|\theta)\} \tag{2·7}$$

(see Lehmann, 1959, p. 15).

It will be difficult to give a general discussion of the applicability or of the relative performances of the various suggested procedures. We shall, therefore, consider some special cases which have important applications.

3. DISCRIMINATION BETWEEN COMPOSITE HYPOTHESES: SPECIAL CASES

Let us consider the special case where $\mathbf{X}$ has a p-variate normal distribution.

Problem 1. Let H_1 and H_2 be defined as follows, where E and D stand for expectation and dispersion operators respectively

$$\left.\begin{aligned} H_1&: E(\mathbf{X}) = \boldsymbol{\alpha}_1 + \mathbf{B}'\boldsymbol{\theta}_1 \quad D(\mathbf{X}) = \boldsymbol{\Lambda}, \\ H_2&: E(\mathbf{X}) = \boldsymbol{\alpha}_2 + \mathbf{B}'\boldsymbol{\theta}_2 \quad D(\mathbf{X}) = \boldsymbol{\Lambda}, \end{aligned}\right\} \tag{3·1}$$

where $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are p-vectors $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ are k-vectors and $\mathbf{B}'$ is $p \times k$ matrix of rank k. The values of $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ and $\mathbf{B}'$ are fixed and known but those of $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ are arbitrary. Thus H_1 and H_2 are composite hypotheses.

For example, each composite hypothesis may correspond to populations representing various stages of growth of an organism. The mean of any character X_i (the ith component of $\mathbf{X}$) for an organism with age t may be written $E(X_i) = \alpha_i + \beta_i t$, where β_i is the regression coefficient on time. The regression coefficients β_i are taken to be the same for the two sets of populations but α_i may be different. The problem is to identify an organism as belonging to one of two sets of populations when the age of the organism is not known.

Considering the general case of (3·1) it is easy to verify that the statistic $\mathbf{B}\boldsymbol{\Lambda}^{-1}\mathbf{X}$ is sufficient for θ_1 and also for θ_2. Applying the formula (2·3) we find the discriminant function between the composite hypotheses H_1 and H_2 to be

$$(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)'(\boldsymbol{\Lambda}^{-1} - \boldsymbol{\Lambda}^{-1}\mathbf{B}'(\mathbf{B}\boldsymbol{\Lambda}^{-1}\mathbf{B}')^{-1}\mathbf{B}\boldsymbol{\Lambda}^{-1})\mathbf{X} \tag{3·2}$$

which depends only on $(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)$ and is independent of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ as was to be expected.

To apply the method of ancillary statistics, let us consider the statistic $\mathbf{CX}$ where $\mathbf{C}$ is a $(p-k) \times p$ matrix of rank $(p-k)$ such that $\mathbf{BC}' = \mathbf{0}$. Then

$$\left.\begin{aligned} E(\mathbf{CX}|H_1) &= \mathbf{C}\boldsymbol{\alpha}_1, \quad D(\mathbf{CX}|H_1) = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}', \\ E(\mathbf{CX}|H_2) &= \mathbf{C}\boldsymbol{\alpha}_2, \quad D(\mathbf{CX}|H_2) = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}', \end{aligned}\right\} \tag{3·3}$$

under the hypotheses H_1 and H_2 respectively. Thus $\mathbf{CX}$ is ancillary under the alternatives in H_1 and also in H_2, and in terms of $\mathbf{CX}$ the problem is reduced to discrimination between two simple hypotheses. The discriminant function based on $\mathbf{CX}$ is

$$(\mathbf{C}\boldsymbol{\alpha}_1 - \mathbf{C}\boldsymbol{\alpha}_2)'(\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}')^{-1}\mathbf{CX} = (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)'(\mathbf{C}'(\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}')^{-1}\mathbf{C})\mathbf{X}. \tag{3·4}$$

It follows from the identity (1·2) of Lemma 1*a* that (3·2) and (3·4) are the same.

It is easily shown that the method of maximum-likelihood ratio as defined in (2·5) also yields the same discriminant function.

Problem 2. In problem 1, the dispersion matrices under the two hypotheses were the same. Let us now consider the alternative composite hypotheses

$$\left.\begin{aligned} H_1&: E(\mathbf{X}) = \boldsymbol{\alpha}_1 + \mathbf{B}'\boldsymbol{\theta}_1,\ D(\mathbf{X}) = \boldsymbol{\Lambda}_1, \\ H_2&: E(\mathbf{X}) = \boldsymbol{\alpha}_2 + \mathbf{B}'\boldsymbol{\theta}_2,\ D(\mathbf{X}) = \boldsymbol{\Lambda}_2, \end{aligned}\right\} \tag{3·5}$$

where $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ are arbitrary as in problem 1.

It is easily seen that $\mathbf{B}\Lambda_1^{-1}\mathbf{X}$ is sufficient for $\boldsymbol{\theta}_1$ while $\mathbf{B}\Lambda_2^{-1}\mathbf{X}$ is sufficient for $\boldsymbol{\theta}_2$. Since the two sufficient statistics are not the same, the method of conditional distributions cannot be applied, unless one considers the statistic $(\mathbf{B}\Lambda_1^{-1}\mathbf{X}, \mathbf{B}\Lambda_2^{-1}\mathbf{X})$ as jointly sufficient for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. But such a statistic is too wide.

But the method of ancillary statistics is applicable since the statistic $\mathbf{CX}$, where $\mathbf{C}$ is as defined in (3·3), is ancillary under both the hypotheses. The distributions under H_1 and H_2 are specified by

$$\left.\begin{aligned} E(\mathbf{CX}|H_1) &= \mathbf{C}\boldsymbol{\alpha}_1, \quad D(\mathbf{CX}|H_1) = \mathbf{C}\Lambda_1\mathbf{C}',\\ E(\mathbf{CX}|H_2) &= \mathbf{C}\boldsymbol{\alpha}_2, \quad D(\mathbf{CX}|H_2) = \mathbf{C}\Lambda_2\mathbf{C}'.\end{aligned}\right\} \tag{3·6}$$

Taking the logarithm of the likelihood ratio we have the discriminant function, $Q(\mathbf{X})$ equal to

$$\mathbf{X}'\mathbf{C}'\{(\mathbf{C}\Lambda_1\mathbf{C}')^{-1} - (\mathbf{C}\Lambda_2\mathbf{C}')^{-1}\}\mathbf{CX} - 2\{\boldsymbol{\alpha}_1'\mathbf{C}'(\mathbf{C}\Lambda_1\mathbf{C}')^{-1} - \boldsymbol{\alpha}_2'\mathbf{C}'(\mathbf{C}\Lambda_2\mathbf{C}')^{-1}\}\mathbf{CX} \tag{3·7}$$

which is quadratic in $\mathbf{X}$. Using the identity (1·2)

$$\mathbf{C}'(\mathbf{C}\Lambda_i\mathbf{C}')^{-1}\mathbf{C} = \Lambda_i^{-1} - \Lambda_i^{-1}\mathbf{B}'(\mathbf{B}\Lambda_i^{-1}\mathbf{B}')^{-1}\mathbf{B}\Lambda_i^{-1} \tag{3·8}$$

we can write (3·8) in terms of $\mathbf{B}$ only. It may be verified that the method of maximum likelihood ratio also provides the same quadratic discriminant function.

Problem 3. (Discrimination between several composite hypotheses.) Let us consider several composite hypotheses $H_1, \ldots, H_k$ such that

$$E(\mathbf{X}|H_i) = \boldsymbol{\alpha}_i + \mathbf{B}'\boldsymbol{\theta}_i \quad (\boldsymbol{\theta}_i \text{ arbitrary}),$$
$$D(\mathbf{X}|H_i) = \Lambda_i \quad (i = 1, \ldots, k).$$

The general theory of §2 for determining decision rules independent of $\boldsymbol{\theta}_i$ applies. In terms of $\mathbf{Y} = \mathbf{CX}$, the common ancillary statistic for each composite hypothesis, the problem reduces to the discrimination of k simple hypotheses such that

$$E(\mathbf{Y}|H_i) = \mathbf{C}\boldsymbol{\alpha}_i, \quad D(\mathbf{Y}|H_i) = \mathbf{C}\Lambda_i\mathbf{C}' \quad (i = 1, \ldots, k).$$

The solution in such a case follows on standard lines (Lindley, 1953; Rao, 1948, 1952, 1965; Wald, 1950, etc.)

It has been pointed out by a referee that the decision rules obtained in this section also follow from invariance theory by observing that $\mathbf{CX}$, where $\mathbf{BC}' = \mathbf{0}$ and $\mathbf{C}$ is of rank $(p-k)$, is a maximal invariant under the transformations $\mathbf{X} \rightarrow \mathbf{X} + \mathbf{B}'\mathbf{t}$ for variable $\mathbf{t}$.

4. Properties of the linear and quadratic discriminant functions

We shall characterize the discriminant functions obtained in §3 in a series of lemmas, without considering the actual form of the distribution of $\mathbf{X}$.

Lemma 4*a*. *Let*

$$\left.\begin{aligned} E(\mathbf{X}|H_1) &= \boldsymbol{\mu}_1 = \boldsymbol{\alpha}_1 + \mathbf{B}'\boldsymbol{\theta}_1,\ D(\mathbf{X}|H_1) = \Lambda,\\ E(\mathbf{X}|H_2) &= \boldsymbol{\mu}_2 = \boldsymbol{\alpha}_2 + \mathbf{B}'\boldsymbol{\theta}_2,\ D(\mathbf{X}|H_2) = \Lambda,\end{aligned}\right\} \tag{4·1}$$

where $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \mathbf{B}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ *are as defined in* (3·1). *Further let* $\mathbf{L}$ *be a* p*-vector such that* $\mathbf{BL} = \mathbf{0}$, *and* $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. *Then*

$$\sup_{\mathbf{L}} \frac{(\mathbf{L}'\boldsymbol{\delta})^2}{\mathbf{L}'\Lambda\mathbf{L}}, \quad \textit{subject to } \mathbf{BL} = \mathbf{0} \tag{4·2}$$

is attained at

$$\mathbf{L}^* = \Lambda^{-1} - \Lambda^{-1}\mathbf{B}'(\mathbf{B}\Lambda^{-1}\mathbf{B}')^{-1}\mathbf{B}\Lambda^{-1}. \tag{4·3}$$

The result is the same as that of Lemma 1*b* and provides an optimum property of the compounding vector in the discriminant function (3·2).

The function $(L^*)'X$ may be called a restricted linear discriminant function whether X has a normal distribution or not, drawing an analogy with the linear discriminant function of Fisher obtained by maximizing the ratio $(L'\delta)^2/L'\Lambda L$ without any restriction on L.

LEMMA 4*b*. *Let* $\phi = \theta_1 - \theta_2$, *in which case* $\delta = \alpha_1 - \alpha_2 + B'\phi$. *Then*

$$\max_{L} \min_{\phi} \frac{(L'\delta)^2}{L'\Lambda L} \tag{4·4}$$

is attained at

$$L^* = \Lambda^{-1} - \Lambda^{-1}B'(B\Lambda^{-1}B')^{-1}B\Lambda^{-1}. \tag{4·5}$$

Let us observe that

$$\min (L'\delta)^2 = 0 \quad \text{if} \quad BL \neq 0$$
$$= (L'\delta)^2 \quad \text{if} \quad BL = 0.$$

Hence

$$\max_{L} \min_{\phi} \frac{(L'\delta)^2}{L'\Lambda L} = \max_{BL=0} \frac{(L'\delta)^2}{L'\Lambda L} = \frac{(L^{*\prime}\delta)^2}{L^{*\prime}\Lambda L^*},$$

where L^* is as defined in (4·3).

Lemma 4*b* provides another characterization of the discriminant function (3·2). First, we consider a linear function $L'X$ and its discriminatory power (as measured by $(L'\delta)^2/L'\Lambda L$) between two alternative simple hypotheses one from each of the composite hypotheses H_1 and H_2. Then we choose L in such a way that the minimum power with respect to all possible alternatives is as high as possible.

LEMMA 4*c*. *Let* $\delta = \alpha_1 - \alpha_2 + B'\phi$ *as before and* $\Lambda = \Lambda_0 + B'DB$ *where* D *is arbitrary. Then* L^* *as determined in Lemma* 4*a or* 4*b is independent of* ϕ *and* D.

According to Lemma 4*a*, L^* is obtained by considering the class of vectors L such that $BL = 0$ and maximizing $(L'\delta)^2/L'\Lambda L$. Now for such L

$$\left.\begin{aligned} L'\delta &= L'(\alpha_1 - \alpha_2), \\ L'\Lambda L &= L'(\Lambda_0 + B'DB)L = L'\Lambda_0 L. \end{aligned}\right\} \tag{4·6}$$

Therefore the problem is the same as that of maximizing $\{L'(\alpha_1 - \alpha_2)\}^2/\{L'\Lambda_0 L\}$ subject to the condition $BL = 0$. Then the solution is

$$L^* = (\Lambda_0^{-1} - \Lambda_0^{-1}B'(B\Lambda_0^{-1}B')^{-1}B\Lambda_0^{-1})(\alpha_1 - \alpha_2) \tag{4·7}$$

which is obviously independent of ϕ and D.

Lemma 4*c* provides an important result for practical applications. It enables us to construct the discriminant function knowing only the means and dispersion matrices of arbitrary mixtures of populations defined by each composite hypothesis. Thus if θ_1 has *a priori* mean $\bar{\theta}_1$ and dispersion matrix D_1 and θ_2 has *a priori* mean $\bar{\theta}_2$ and dispersion matrix D_2, the mixture of populations under H_1 has mean and dispersion matrix equal to

$$\mu_1 = \alpha_1 + B'\bar{\theta}_1, \quad D(X) = \Lambda_0 + B'D_1B, \tag{4·8}$$

where Λ_0 is the dispersion matrix of X for given θ_1. Similarly, the mixture of populations under H_2 has mean and dispersion matrix equal to

$$\mu_2 = \alpha_2 + B'\bar{\theta}_2, \quad D(X) = \Lambda_0 + B'D_2B. \tag{4·9}$$

Let us consider

$$\left.\begin{aligned}\boldsymbol{\delta} &= \boldsymbol{\mu}_1-\boldsymbol{\mu}_2 = \boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2+\mathbf{B}'(\bar{\boldsymbol{\theta}}_1-\bar{\boldsymbol{\theta}}_2),\\ \boldsymbol{\Lambda} &= w_1 D(\mathbf{X}|H_1)+w_2 D(\mathbf{X}|H_2) = \boldsymbol{\Lambda}_0+\mathbf{B}'(w_1\mathbf{D}_1+w_2\mathbf{D}_2)\,\mathbf{B},\end{aligned}\right\} \tag{4·10}$$

where w_1 and w_2 are arbitrary weights such that $w_1+w_2 = 1$. The substitution of $\boldsymbol{\delta}$ and $\boldsymbol{\Lambda}$ in the formula (4·3) for $\mathbf{L}^*$ gives us a result which is independent of $\bar{\boldsymbol{\theta}}_1$, $\bar{\boldsymbol{\theta}}_2$, $\mathbf{D}_1$ and $\mathbf{D}_2$. The importance of the result arises because in the type of practical situations we are considering we are likely to have only estimates of means and dispersion matrices of (unknown) mixtures of populations in the two groups which we are trying to discriminate.

Lemma 4*d*. *Let* $\boldsymbol{\mu}_1 = \boldsymbol{\alpha}_1+\mathbf{B}'\boldsymbol{\theta}_1$, $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_{01}+\mathbf{B}'\mathbf{D}_1\mathbf{B}$ *and* $\boldsymbol{\mu}_2 = \boldsymbol{\alpha}_2+\mathbf{B}'\boldsymbol{\theta}_2$, $\boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}_{02}+\mathbf{B}'\mathbf{D}_2\mathbf{B}$. *Then the quadratic discriminant function*, $Q(\mathbf{X})$ *as determined in* (3·7), *is independent of* $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\mathbf{D}_1$ *and* $\mathbf{D}_2$.

The result is established by verification. Similar results are true for the decision rules arising out of problem 3 of § 3.

References

Burnaby, T. P. (1966). Growth invariant discriminant functions and generalized distances. *Biometrics* 22, 96–110.

Delany, M. J. & Healy, M. J. R. (1964). Variations in the long tailed field-mouse, etc., II Simultaneous examination of all characters. *Proc. Roy. Soc.* B 161, 200–7.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. New York: John Wiley.

Lindley, D. V. (1953). Statistical inference. *J. R. Statist. Soc.* B 11, 30–76.

Rao, C. R. (1948). Tests of significance in multivariate analysis. *Biometrika* 35, 58–79.

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: John Wiley.

Rao, C. R. (1961). Some observations on multivariate statistical methods in anthropological research. *Bull. Int. Stat. Inst.* 38, 99–109.

Rao, C. R. (1962). Use of discriminant and allied functions in multivariate analysis. *Sankhyā* 22, 317–38.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā* 26, 329–58.

Rao, C. R. (1965). *Linear Statistical Inference and its Applications*. New York: John Wiley.

Rao, C. R. (1966). Generalized inverse for matrices and its application in mathematical statistics. *Research Papers in statistics* (Neyman Festschrift volume) edited by F. N. David. Wiley, London. 263–79.

Smith, C. A. B. (1947). Some examples of discrimination. *Ann. Eugen., Lond.* 18, 272–82.

Wald, A. (1950). *Statistical Decision Functions*. New York: John Wiley.

Welch, B. L. (1939). Note on discriminant functions. *Biometrika* 31, 213.