General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

CR 167 813

E83-10196

AgRISTARS

SR-L3-04388
JSC-18590

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Supporting Research

December, 1982

LINEAR DISCRIMINANT ANALYSIS WITH MISALLOCATION IN TRAINING SAMPLES

(E83-10196) LINEAR DISCRIMINANT ANALYSIS WITH MISALLOCATION IN TRAINING SAMPLES (Lockheed Engineering and Management) 34 p HC A03/MF A01 CSCL 02C

N83-20318

Unclas

G3/43 00196

R. Chhikara
Lockheed Engineering and Management Services Company, Inc.

J. McKeon
OLD DOMINION UNIVERSITY
Norfolk, Virginia 23508

Earth Resources Research Division
Lyndon B. Johnson Space Center
Houston, Texas 77058

1. Report No. SR-L2-04388, JSC-18590	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Linear Discriminant Analysis with Misallocation in Training Samples	5. Report Date December 1982 6. Performing Organization Code
7. Author(s) R. Chhikara, Lockheed-EMSCO Jim McKeon, Old Dominion University	8. Performing Organization Report No. LEMSCO-19020 10. Work Unit No.
9. Performing Organization Name and Address Lockheed Engineering and Management Services Company, Inc 1830 NASA Road 1 Houston, Texas 77258	11. Contract or Grant No. NAS 9-15800 & NAS 9-16431 13. Type of Report and Period Covered Technical Report
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Technical Monitor: F. G. Hall	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Linear discriminant analysis for a two-class case is studied in the presence of misallocation in training samples. A general approach to modeling of misallocation is formulated, and the mean vectors and covariance matrices of the mixture distributions are derived. The asymptotic distribution of the discriminant boundary is obtained and the asymptotic first two moments of the two types of error rate given. Certain numerical results for the error rates are presented by considering the random and two non-random misallocation models. It is shown that when the allocation procedure for training samples is objectively formulated, the effect of misallocation on the error rates of the Bayes linear discriminant rule can almost be eliminated. If, however, this is not possible, the use of Fisher rule may be preferred over the Bayes rule.

17. Key Words (Suggested by Author(s)) Random and non-random misallocation, mixture distributions, discriminant boundary, error rates, Fisher and Bays discriminant rules, Asymptotic and sampling results.	18. Distribution Statement

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price*

*For sale by the National Technical Information Service Springfield, Virginia 22161

SR-L3-.4388
JSC-18590

LINEAR DISCRIMINANT ANALYSIS WITH MISALLOCATION IN TRAINING SAMPLES

Job Order 71-302

This report describes Error Analysis Research activities of the Supporting Research project of the AgRISTARS program.

PREPARED BY

R. S. Chhikara and Jim McKeon

APPROVED BY

R. K. Lennington, Manager
Supporting Research Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800 & NAS 9-16431

For

Earth Resources Resarch Division

Space and LIfe Sciences Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

December 1982

1. INTRODUCTION

In discriminant analysis, often a two-step procedure is followed; first, training samples are obtained to set up a discriminant rule and then, individuals are classified using the sample-based rule. However, if the criterion for assigning the training samples to their true classes is imperfect, some training samples may be misallocated. For example, this arises in discrimination of crops in an area using spectral data acquired from a satellite. The scene image of the area is analyzed to delineate crop features and training samples are assigned crop labels based on visual interpretation of their spectral observations. This can lead to mislabeling of crops for some training samples and thus, may adversely affect the performance of a discriminant rule.

Presently we study the linear discriminant analysis in the presence of misallocation in a training set. Suppose that individuals come from one of the two classes C_1 and C_2. A p-dimensional random vector $\underset{\sim}{X}$ is measured on each individual. It is assumed that $\underset{\sim}{X}$ has the multivariate normal distribution with mean $\underset{\sim}{\mu}_i$ and covariance matrix $\underset{\sim}{\Sigma}$ for C_i, i=1,2. In a training sample of n individuals, suppose n_1 are allocated into C_1 and $n_2=n-n_1$ into C_2. If α_i is the fraction of training samples from C_i that are misallocated, i=1,2, the two samples of sizes n_1 and n_2 represent mixed classes, say C_1^* and C_2^*, instead of the original classes C_1 and C_2. Let $\underset{\sim}{\bar{X}}_1^*$ and $\underset{\sim}{\bar{X}}_2^*$ and $\underset{\sim}{S}^*$ denote the sample means and the pooled sample covariance matrix, respectively. Then a random observation $\underset{\sim}{X}$ can be classified on the basis of linear discriminant function (Anderson, 1958) given by

$$\hat{\lambda}(\underset{\sim}{X}) = \hat{\beta}_0 + \hat{\underset{\sim}{\beta}}' \underset{\sim}{X} \qquad (1.1)$$

where

$$\hat{\beta}_0 = \log(n_1/n_2) - (1/2)(\bar{X}_2^* - \bar{X}_1^*)' S^{*^{-1}} (\bar{X}_2^* + \bar{X}_1^*)$$

$$\hat{\beta} = S^{*^{-1}}(\bar{X}_2^* - \bar{X}_1^*). \tag{1.2}$$

The classification procedure is to regard the observed value, X coming from C_1 or C_2 according as the discriminant value, $\hat{\lambda}(X) < 0$ or > 0, respectively. Then the error rates for the procedure are given by

$$R_1 = \text{Prob}\,\{\hat{\lambda}(X) > 0 \mid X \in C_1, \bar{X}_1^*, \bar{X}_2^*, S^*\}$$

$$R_2 = \text{Prob}\,\{\hat{\lambda}(X) < 0 \mid X \in C_2, \bar{X}_1^*, \bar{X}_2^*, S^*\} \tag{1.3}$$

and its average error rate is given by

$$R = \pi_1 R_1 + \pi_2 R_2 \tag{1.4}$$

where π_1 and π_2 are the probabilities associated with C_1 and C_2.

Assuming that training samples are randomly misallocated, Lachenbruch (1966) and McLachlan (1972) studied R_1 and R_2 for their expected values and variances. However, a random misallocation model is unrealistic, particularly if the observation X is itself used in determining the allocation. Lachenbruch (1974) suggested a non-random allocation model with two variations to it. His criterion for allocation was based on the distances of an observation from the class means. Presently, we propose an allocation model in which misallocation of a sample depends upon its observation. The random and non-random misallocation models of Lachenbruch become special cases of this new model (Section 2).

For the discriminant function in (1.1), we give the asymptotic distribution of the discriminant boundary and obtain the asymptotic mean and variance of each of the error rates, R_1, R_2, and R (Section 3). We take the same approach that was used by Efron (1975) and extend his normal discrimination results to the case of misallocated training samples. The present study can also be viewed as an extension of Sayre (1980) who gives the asymptotic distribution of R assuming correct allocation for the training samples; although we here do not explicity give the distribution. McLachlan (1972) has given the asymptotic means and variances of the error rates for random misallocation, but his derivation is limited to only one of the two misallocation rates being non-zero. Lachenbruch (1966, 1974) investigated the means and variances of R_1 and R_2 for his models using simulations. Michalek and Tripathi (1980) discussed the problem for random misallocation, but they stuoied the discrimination between the mixed classes and not between the original classes. Given in Sections 4 and 5 are certain numerical results showing the adverse effect of misallocation on the linear discriminant boundary and the associated error rates.

2. MISALLOCATION MODELS

Suppose $\Delta^2=(\mu_1-\mu_2)' \Sigma^{-1} (\mu_1-\mu_2)$. By means of linear transformations, one can reduce the class structures in the canonical form (Efron 1975), where

$$\underset{\sim}{\mu}1 = \begin{bmatrix} -\Delta/2 \\ \underset{\sim}{0} \end{bmatrix}, \quad \underset{\sim}{\mu}2 = \begin{bmatrix} \Delta/2 \\ \underset{\sim}{0} \end{bmatrix}, \quad \underset{\sim}{\Sigma} = \underset{\sim}{I} \tag{2.1}$$

so that the class means $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ are aligned along the x_1-axis. Suppose allocation of an individual is made using its observation $\underset{\sim}{X}$. It is desirable

ORIGINAL PAGE IS
OF POOR QUALITY

to consider an allocation so that chance of misallocation for an individual increases as its observation deviates further away from the mean of its true class in the direction of the mean of the other class. So let the probability of misallocation of an individual from C_i into C_{3-i} be $g_i(z)$, $i = 1,2$, where $g_1(z)$ is a monotone increasing function and $g_2(z)$ is a monotone decreasing function with z to be along the x_1-axis. Suppose $f_i(z)$ is the frequency function of the first component of random vector $\underset{\sim}{X}$ for C_i and

$$\pi_i = \int_{-\infty}^{\infty} f_i(z)dz, \quad i = 1,2.$$

Define the misallocation rate α_i by

$$\alpha_i = (1/\pi_i) \int_{-\infty}^{\infty} g_i(z)f_i(z)dz, \; i = 1,2. \tag{2.2}$$

Given α_1 and α_2, the functions g_1 and g_2 can be specified differently. The random misallocation model (Lachenbruch 1966, McLachlan 1972, Michalek and Tripathi 1980) corresponds to the uniform case given by, and to be called model (a):

(a) Random Misallocation

For $\underset{\sim}{X} \in C_i$, let

$$g_i(z) = \alpha_i, \quad i = 1,2. \tag{2.3}$$

Another model, to be called model (b), is obtained by specifying g_1 and g_2 as follows:

(b) "Truncated" Model:

For $\underset{\sim}{X} \in C_1$, let

$$g_1(z) = \begin{cases} 0 & , z < a_1 \\ u & , z > a_1 \end{cases}$$

and for $\underset{\sim}{X} \in C_2$, let

$$g_2(z) = \begin{cases} u & , z < a_2 \\ 0 & , z > a_2 \end{cases} \tag{2.4}$$

where a_i is determined from (2.2). After solving it, we obtain

$$a_1 = -(\Delta/2) + Z_{1-\alpha_1}/u$$

$$a_2 = \Delta/2 + Z_{\alpha_2}/u$$

where Z_γ denotes the γ-percentage point of the standard normal distribution.

If we assume $u=1$ and $a_1=a_2=0$, then one obtains the complete separation model of Lachenbruch (1974). His other non-random model can be obtained by taking the a_i as percentage points of the chi square distribution with p degrees of freedom.

Though models (a) and (b) are easy to implement and hence, these are appealing, they may not be always suitable. Instead, it may perhaps be more appropriate to let the probability of misallocation increase as the observed value deviates away from the mean of its true class. One such model can be defined as follows:

(c) Exponential Model:

For $\underset{\sim}{X} \in C_1$, let

$$g_1(z) = \begin{cases} 0 , & z < -\Delta/2 \\ 1-\exp(-k_1[z+\Delta/2]^2/2), & z > -\Delta/2 \end{cases}$$

and for $\underset{\sim}{X} \in C_2$, let

$$g_2(z) = \begin{cases} 1-\exp(-k_2[z-\Delta/2]^2/2), & z < \Delta/2 \\ 0 \quad , & z > \Delta/2 \end{cases} \tag{2.5}$$

where k_i is determined from (2.2). It easily follows that

$$k_i = (1-2\alpha_i)^{-2} - 1.$$

In practice, the misallocation rates α_i will be subject to sampling variation. Hence, these rates are being considered as random variables.

In Appendix A, we derive the mean vectors and the covariance matrices of the mixture distributions of C_1^* and C_2^*, and in section 3, we give the discriminant analysis for arbitrary functions g_1 and g_2 as defined earlier. For numerical computations presented in sections 4 and 5, we consider the special cases, models (a), (b) and (c), and compare the performances of the discriminant rule associated with the discriminant function in (1.1) for these models.

3. DISCRIMINANT BOUNDARY AND ERROR RATES

When the parameters are known, the discriminant rule is: classify $\underset{\sim}{X}$ into C_1 if $\lambda(\underset{\sim}{X}) < 0$ and into C_2, otherwise, where

$$\lambda(\underset{\sim}{X}) = \beta_0 + \beta' X \tag{3.1}$$

$$\beta_0 = \log(\pi_1^*/\pi_2^*) - (\mu_{21}^{*2} - \mu_{11}^{*2})/2\,(1 + \xi)$$

$$\beta_1 = (\mu_{21}^* - \mu_{11}^*)/(1 + \xi)$$

$\beta_j = 0 \quad , j = 2, \ldots, p$

and π_1^*, π_2^*, μ_{11}^*, μ_{21}^* and ξ as defined in Appendix A.

As discussed by Efron (1975), the "Optimum" boundary, $\lambda(\underset{\sim}{X})=0$, is the (p-1)-dimensional plane orthogonal to x_1-axis and intersecting it at

$$\tau = -\beta_0/\beta_1. \tag{3.2}$$

For large sample size n, the sample-based boundary, $\hat{\lambda}(\underset{\sim}{X})=0$, is the plane intersecting the x_1-axis at $\hat{\tau} = \tau + d\tau$ with normal vector at an angle $d\varepsilon$ from the x_1-axis, where $d\tau$ and $d\varepsilon$ represent small deviations from 0. With no loss of generality, suppose $\tau>0$. Then the distances of $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ from the optimum boundary are

$$D_1 = \Delta/2 + \tau, \quad D_2 = \Delta/2 - \tau, \tag{3.3}$$

and those from the sample-based boundary are

$$d_1 = (D_1+d\tau)\cos(d\varepsilon), \; d_2=(D_2-d\tau)\cos(d\varepsilon). \tag{3.4}$$

Refer to Efron(1975) for a pictorial description of the two-discriminant boundaries and other related details.

The error rates can now be written in terms of these distances:

$$R_1^0 = \Phi(-D_1) \quad , \quad R_2^0 = \Phi(-D_2) \tag{3.5}$$

for the "optimum" boundary, and

$$R_1 = \Phi(-d), \quad R_2 = \Phi(-d_2) \tag{3.6}$$

for the sample-based boundary, where Φ stands for the standard normal cdf. Let ϕ denote the density function of standard normal. Then, ignoring higher than second order differential terms, we have (Efron, 1975)

$$R_1 \approx R_1^0 - \phi(D_1)d\tau + (D_1/2)\ \phi(D_1)[(d\tau)^2 + (d\epsilon)^2]$$

$$R_2 = R_2^0 + \phi(D_2)d\tau + (D_2/2)\ \phi(D_2)[(d\tau)^2 + (d\epsilon)^2] \tag{3.7}$$

where

$$d\tau = -(d\beta_0 + \tau d\beta_1)/\beta_1$$

$$(d\epsilon)^2 = [(d\beta_2)^2 + (d\beta_3)^2 + \ldots. + (d\beta_p)^2]/\beta_1^2 \tag{3.8}$$

with $d\beta_j = (\hat{\beta}_j - \beta_j)$ denoting the error in the estimate $\hat{\beta}_j$, $j = 0,1,2,\ldots,p$, given in (1.2). We denote $d\underset{\sim}{\beta}^{(1)} = (d\beta_1, d\beta_2, \ldots, d\beta_p)'$.

Since n is large, one may assume that $\sqrt{n}(d\beta_0, d\underset{\sim}{\beta}^{(1)})'$ has a limiting normal distribution with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{V}_{\hat{\beta}}$. In Appendix B, we obtain $\underset{\sim}{V}_{\hat{\beta}}$ and write it in the form,

$$\underset{\sim}{V}_{\hat{\beta}} = \begin{bmatrix} \sigma_{00} & \sigma_{01} & \underset{\sim}{0}' \\ \sigma_{01} & \sigma_{11} & \underset{\sim}{0}' \\ \underset{\sim}{0} & \underset{\sim}{0} & \sigma_{22}\underset{\sim}{I} \end{bmatrix}$$

with quantities σ_{00}, σ_{01}, σ_{11} and σ_{22} expressed in terms of basic input parameters, π_1, π_2, α_1, α_2 and Δ, among others.

It follows from (3.8) that

$$\sigma_\tau^2 = E[(d\tau)^2]$$
$$= (\sigma_{00} + 2\tau\sigma_{01} + \tau^2 \sigma_{11})/\beta_1^2. \qquad (3.10)$$

Suppose we define

$$d\omega_j = d\beta_j/\beta_1, \; j=2,3,\ldots p.$$

Then its variance is

$$\sigma_{\omega_j}^2 = \sigma_{22}/\beta_1^2, \quad j=2,3,\ldots,p. \qquad (3.11)$$

Next, $\sqrt{n}(d\tau, d\underset{\sim}{\omega})$ has a limiting normal distribution with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{T}\underset{\sim}{V}_{\hat{\beta}}\underset{\sim}{T}'$, where

$$\underset{\sim}{T} = (1/\beta_1) \begin{bmatrix} 1 & \tau & \underset{\sim}{0}' \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{I} \end{bmatrix}.$$

The covariance matrix may be written as

$$\begin{bmatrix} \sigma_\tau^2 & \underset{\sim}{0}' \\ \underset{\sim}{0} & \sigma_\omega^2 \underset{\sim}{I} \end{bmatrix}$$

where $\sigma_\omega^2 = \sigma_{22}/\beta_1^2$.

Since $(d\varepsilon)^2 = \sum_{j=2}^{p} (d\omega_j)^2$ and $n(d\omega_j)^2/\sigma_\omega^2 \sim \chi_1^2$, $j=2,3,\ldots,p$, $n(d\varepsilon)^2/\sigma_\omega^2 \sim \chi_{k-1}^2$. Furthermore, $n(d\tau)^2/\sigma_\tau^2 \sim \chi^2_1$.

(The symbol ~ should read "asymptotically distributed as".)

From (3.7) and the above distributional results, the asymptotic moments of the error rates can now be easily obtained. Since $(d\tau)^2$ and $(d\varepsilon)^2$ are asymptotically uncorrelated and

$$E[(d\tau)^2] = \sigma_\tau^{\,2}/n, \quad E[(d\varepsilon)^2] = (p-1)\,\sigma_\omega^{\,2}/n$$

and

$$V[(d\tau)^2] = 2\sigma_\tau^{\,4}/n^2, \quad V[(d\varepsilon)^2] = 2(p-1)\sigma_\omega^{\,4}/n^2,$$

the asymptotic means of R_1 and R_2, ignoring second and higher order terms, are given by

$$\begin{aligned} E[R_1] &= R_1^0 + (D_1/2n)\,\phi(D_1)\,[\sigma_\tau^{\,2} + (p-1)\sigma_\omega^{\,2}] \\ E[R_2] &= R_2^0 + (D_2/2n)\,\phi(D_2)\,[\sigma_\tau^{\,2} + (p-1)\sigma_\omega^{\,2}] \end{aligned} \tag{3.12}$$

For the asymptotic second order moments, ignoring third and higher order terms, we have the variances and covariances of R_1 and R_2 as follows:

$$\begin{aligned} V[R_1] &= (1/n)\phi^2(D_1)\,\{\sigma_\tau^{\,2} + (D_1^2/2n)[\sigma_\tau^{\,4} + (p-1)\sigma_\omega^{\,4}]\} \\ V[R_2] &= (1/n)\phi^2(D_2)\,\{\sigma_\tau^{\,2} + (D_2^2/2n)[\sigma_\tau^{\,4} + (p-1)\sigma_\omega^{\,4}]\} \\ Cov[R_1, R_2] &= (1/n)\phi(D_1)\phi(D_2)\{-\sigma_\tau^{\,2} + (D_1D_2/2n) \\ &\qquad [\sigma_\tau^{\,4} + (p-1)\sigma_\omega^{\,4}]\}, \end{aligned} \tag{3.13}$$

where $\sigma_\tau^{\,2}$ and $\sigma_\omega^{\,2}$ are functions of elements of β_0, β_1 and $\underset{\sim}{V}_{\hat{\beta}}$.

Clearly, $E[R_i]$ approaches R_i^0, i=1,2, as n becomes infinite.

For the average error rate, we have

$$\begin{aligned} E[R] &= R^0 + (1/2n)\,[\pi_1 D_1\,\phi(D_1) + \pi_2 D_2\,\phi(D_2)]\,[\sigma_\tau^{\,2} + (p-1)\,\sigma_\omega^{\,2}] \\ V[R] &= \pi_1^2\,V[R_1] + \pi_2^2\,V[R_2] + 2\,\pi_1\,\pi_2\,Cov(R_1, R_2), \end{aligned} \tag{3.14}$$

where $V[R_1]$, $V[R_2]$ and $Cov[R_1, R_2]$ are as given in (3.13).

4. NUMERICAL RESULTS

Computations were made to evaluate the asymptotic covariance matrix $\underset{\sim}{V}_{\hat{\beta}}$ for following cases of input parameters:

$\pi_1 = .5, .7$

$\Delta = 2, 4$

$\alpha_1 = 0, .1, .2, .3, .4$ and $\alpha_2 = 0$

This was done for all three misallocation models discussed in section 2. We specified u =.5 in model (b), equation (2.4), so that there is a fifty-fifty chance of misallocation for an observation that falls beyond a threshold point. Based on these computations, we obtained τ, σ_τ^2, σ_ω^2 and the means and variances of the error rates given in equations (3.12), (3.13) and (3.14). Table 1 lists the values of τ, σ_τ^2 and σ_ω^2. From these numerical results, we find that σ_τ^2 increases as α_1 increases from 0 to .4, except there is a slight decrease when Δ=2, π_1=.7 and model (c) for misallocation. The results for σ_ω^2 are mixed; it is constant in the case of misallocation model (a) and it decreases as α_1 increases for models (b) and (c), provided Δ=2. When Δ=4, it first decreases and then increases.

The values of σ_τ^2 and σ_ω^2 are considerably higher for model (a) than for other two models. This is an expected result because the boundary is subject to higher variability under random mixing in training samples. Next, the rate of increase in σ_τ^2 as a function of α_1 is higher for Δ=4 than for Δ=2. Again, this is expected since a higher rate of misallocation in training samples will lead to a larger change in the variance of a mixture distribution when C_1 and C_2 are more separated and, hence, causing a large increase in σ_τ^2.

ORIGINAL PAGE IS
OF POOR QUALITY

1. Values o. τ and Variances σ_τ^2 and σ_ω^2 Associated With the Sample-Based Boundary

	(α_1, α_2)	$\pi_1 = .5$ Misallocation Model (a)	(b)	(c)	$\pi_1 = .7$ Misallocation Model (a)	(b)	(c)
				(i) $\Delta = 2$			
τ	(0, 0)	0	0	0	.424	.424	.424
	(.1, 0)	-.221	-.192	-.191	.214	.092	.089
	(.2, 0)	-.491	-.398	-.375	-.074	-.167	-.186
	(.3, 0)	-.819	-.649	-.565	-.463	-.443	-.435
	(.4, 0)	-1.218	-1.001	-.776	-.976	-.815	-.681
σ_τ^2	(0, 0)	1.000	1.000	1.000	1.360	1.360	1.360
	(.1, 0)	2.157	1.1[illegible]6	1.130	1.929	1.308	1.235
	(.2, 0)	4.327	1.541	1.184	3.475	1.717	1.216
	(.3, 0)	8.248	2.473	1.211	7.088	2.542	1.133
	(.4, 0)	15.549	5.373	1.296	15.564	5.178	1.072
σ_ω^2	(0, 0)	2.000	2.000	2.000	2.190	2.190	2.190
	(.1, 0)	2.000	1.068	1.051	2.190	.845	.824
	(.2, 0)	2.000	.747	.533	2.190	.488	.286
	(.3, 0)	2.000	.644	.248	2.190	.387	.074
	(.4, 0)	2.000	.773	.098	2.190	.515	.005
				(ii) $\Delta = 4$			
τ	(0, 0)	0	0	0	.212	.212	.212
	(.1, 0)	-.277	-.257	-.257	-.065	-.145	-.147
	(.2, 0)	-.617	-.553	-.535	-.413	-.477	-.493
	(.3, 0)	-1.034	-.916	-.847	-.874	-.860	-.853
	(.4, 0)	-1.546	-1.398	-1.207	-1.483	-1.373	-1.248
σ_τ^2	(0, 0)	1.000	1.000	1.000	1.193	1.193	1.193
	(.1, 0)	2.741	1.497	1.480	2.236	1.751	1.703
	(.2, 0)	5.821	2.653	2.012	4.558	2.756	2.182
	(.3, 0)	10.948	4.976	2.642	9.464	4.765	2.623
	(.4, 0)	19.324	10.391	3.494	18.998	9.885	3.125
σ_ω^2	(0, 0)	1.250	1.193	1.193	1.298	1.193	1.193
	(.1, 0)	1.250	.511	.497	1.298	.324	.309
	(.2, 0)	1.250	.266	.122	1.298	.094	.009
	(.3, 0)	1.250	.194	.002	1.298	.044	.077
	(.4, 0)	1.250	.285	.055	1.298	.109	.343

If we consider the complete separation model, i.e., u=1, or the other non-random model of Lachenbruch (1974), the mixture distributions will have smaller variances than the original distributions have. As such, the variance σ_τ^2 may be smaller as compared to the case of no misallocation allowed in samples. In turn, this may lead to smaller values for the expected error rates, as it was observed by Lachenbruch in his sampling study. His study was, of course, restricted to the linear discriminant function without the term of $\log \pi_1^*/\pi_2^*$ or its estimate $\log n_1/n_2$ as may be the case with respect to the discriminant boundary, optimum or sample-based.

In Table 2, we present the asymptotic expected values and standard deviations (SD) of R_1, R_2, and R corresponding to $\pi_1=.5$, $\Delta=2$, $p=2$ and α_1 and α_2 as considered in Table 1. Similar results can be easily computed for the other cases by making use of the values of τ, σ_τ^2 and σ_ω^2 from Table 1. It is seen that $E[R_1]$ and $SD[R_1]$ increase, whereas $E[R_2]$ and $SD[R_2]$ decrease as α_1 increases. When $\alpha_1>0$ and $\alpha_2=0$, $\pi_1^*/\pi_2^* < \pi_1/\pi_2=1$ and $\alpha_1-\alpha_2>0$ and hence, the discriminant boundary shifts away from μ_{21} in the direction of μ_{11} as α_1 increases, causing the error rate to increase for C_1 and to decrease for C_2. For the average error rate, $E[R]$ and $SD[R]$ increase as the misallocation rate α_1 increases. Thus, there is an adverse effect on the average error rate R due to misallocation of samples from one class to another.

In limit, $E[R_i]=R_i^0$, $i=1, 2$, and $E[R]=R^0$ as n becomes infinite. The values of R_1^0, R_2^0, and R^0 obtained for $n=\infty$ are also given in Table 2. The corresponding standard deviations are, of course, zero.

2. Asymptotic Means and Standard Deviations of R_1, R_2 and R (π_1=.5, Δ=2, p=2)

	(α_1 α_2)	n=100 Misallocation Model (a)	(b)	(c)	n= ∞ Misallocation Model (a)	(b)	(c)
$E[R_1]$	(0, 0)	.162	.162	.162	.159	.159	.159
	(.1, 0)	.223	.212	.212	.218	.210	.209
	(.2, 0)	.311	.276	.268	.305	.273	.266
	(.3, 0)	.432	.365	.333	.428	.363	.332
	(.4, 0)	.594	.500	.412	.586	.500	.411
$E[R_2]$	(0, 0)	.162	.162	.162	.159	.159	.159
	(.1, 0)	.116	.119	.119	.111	.117	.117
	(.2, 0)	.074	.084	.086	.068	.081	.085
	(.3, 0)	.042	.052	.060	.034	.050	.059
	(.4, 0)	.020	.026	.039	.013	.023	.038
E[R]	(0, 0)	.162	.162	.162	.159	.159	.159
	(.1, 0)	.169	.166	.166	.165	.163	.163
	(.2, 0)	.193	.180	.177	.187	.177	.175
	(.3, 0)	.237	.208	.197	.231	.206	.195
	(.4, 0)	.307	.263	.225	.300	.262	.225
$SD[R_1]$	(0, 0)	.025	.025	.025			
	(.1, 0)	.044	.031	.031			
	(.2, 0)	.073	.041	.036			
	(.3, 0)	.113	.059	.040			
	(.4, 0)	.154	.092	.044			
$SD[R_2]$	(0, 0)	.025	.025	.025			
	(.1, 0)	.028	.021	.021			
	(.2, 0)	.028	.019	.017			
	(.3, 0)	.023	.016	.013			
	(.4, 0)	.016	.013	.009			
SD[R]	(0, 0)	.004	.004	.004			
	(.1, 0)	.010	.006	.006			
	(.2, 0)	.024	.012	.010			
	(.3, 0)	.046	.022	.014			
	(.4, 0)	.070	.040	.017			

5. SMALL SAMPLE RESULTS

Because of complex algebric expressions involved in the evaluation of $\hat{V}_{\beta}$, we conducted a Monte Carlo sampling experiment to check the accuracy of asymptotic results as well as to study the error rates when the training sample size is small. Normal random numbers were generated using the technique of Box and Muller (1958). The simulation study was limited to p=2, Δ =2, 4, and n=20, 50, 100. The numbers of training samples from C_1 and C_2 were taken to be proportional to their a-priori probabilities. Though there were many other cases, we have chosen to give here the results for the case of π_1=.69, α_1=.087, α_2=.226 (this is equivalent to π_1^*=.7, α_1^*=.1 and α_2^*=.2 in terms of mixed classes), Δ=2. Table 3 presents the means and standard deviations of R_1 and R_2 for n=20, 50, 100 obtained from the sampling experiment as well as from the theoretical results given in (3.12) and (3.13).

Besides misallocation models (a), (b), and (c), we also consider the case of no misallocation in training samples, i.e., $\alpha_1=\alpha_2=0$. This is listed as model (o) in Table 3. Based on these and other results, we find a good agreement between the sampling and asymptotic results. When n=100, the two sets of values of $E[R_1]$, $E[R_2]$, $SD[R_1]$ and $SD[R_2]$ agree at least up to second decimal place. Moreover, the agreement holds quite well even for small sample size of n=20.

A comparison between the results for model (o) and of other three models shows that misallocations under models (b) and (c) lead to about the same results that are obtained with no misallocation in training samples. The actual error rates are considerably biased and have much larger variances with random misallocation.

3. The Means and Standard Deviations of R_1 and R_2

(π_1=.69, α_1=.087 α_2=.226, Δ=2, p=2)

Parameter	Sampling				Asymptotic			
	Misallocation Model			No Misallo-cation	Misallocation Model			No Misallo-cation
	(a)	(b)	(c)	(o)	(a)	(b)	(c)	(o)
				(i) n=100				
$E[R_1]$	.044	.081	.090	.078	.046	.082	.086	.081
$E[R_2]$	.434	.286	.267	.291	.416	.280	.268	.286
$SD[R_1]$	.023	.021	.022	.017	.027	.021	.019	.017
$SD[R_2]$	.118	.048	.047	.042	.117	.047	.040	.040
				(ii) n=50				
$E[R_1]$	.057	.085	.087	.084	.055	.084	.088	.085
$E[R_2]$	.423	.291	.276	.289	.421	.282	.269	.289
$SD[R_1]$	.036	.029	.025	.025	.040	.030	.027	.025
$SD[R_2]$	.143	.072	.051	.054	.166	.067	.057	.056
				(iii) n=20				
$E[R_1]$	.083	.095	.096	.089	.079	.091	.094	.096
$E[R_2]$	.482	.323	.295	.323	.435	.289	.275	.299
$SD[R_1]$	.112	.060	.043	.044	.074	.049	.044	.042
$SD[R_2]$	.237	.137	.115	.101	.264	.106	.090	.090

So, if an allocation procedure for training samples is formulated based on the concept underlying models (b) and (c), the effect of misallocation on the linear discriminant analysis for two classes can be minimized.

6. CONCLUDING REMARKS

In practice, $\zeta^*=\log \pi_1^*/\pi_2^*$ or its estimate, as may be the case, is not included in the discriminant boundary. This leads to what is sometimes referred to as the Fisher classification rule. Otherwise, it may be called the Bayes classification rule. To study the difference in the error rates caused by the exclusion of $\zeta^*=\log n_1/n_2$ from the discriminant function as given in (1.1), we obtained the means and standard deviations of R_1 and R_2 for each rule. The results are presented in Table 4 for the case of $\pi_1=.69$, $\alpha_1=.087$, $\alpha_2=.226$ and $n=100$. Results are also given for the case of $\pi_1=.69$, $\alpha_1=0, \alpha_2=0$. Since simulation and asymptotic results are almost same when $n=100$, either of two sets of results can be considered. We have listed in Table 4 the results obtained by the Monte Carlo method.

A comparison between the results of misallocation models (a), (b), (c), and those of no misallocation model (o) shows that the means and standard deviations of R_1 and R_2, and hence, of R, are less affected due to misallocation in the case of Fisher rule than for the Bayes rule, particularly when misallocation is random. This difference is more in the case of $\Delta=4$. Since $\pi_1^*=.7$, and $\pi_1=.69$, π_1^*/π_2^* is approximately equal to π_1/π_2. So the ratio n_1/n_2 can be considered an equally good estimate of π_1/π_2, and thus, hardly introduces any additional shift in the discriminant boundary, otherwise obtained from the correctly allocated samples. However, when the two ratios, π_1^*/π_2^*

ORIGINAL PAGE IS OF POOR QUALITY

4. The Means and Standard Deviations of R_1 and R_2 for Fisher and Bayes Classification Rules

(π_1=.69, α_1=.087, α_2=.226, p=2, n=100)

Parameter	Fisher: Misallocation Model (a)	(b)	(c)	Fisher: No Misallocation (o)	Bayes: Misallocation Model (a)	(b)	(c)	Bayes: No Misallocation (o)
				(i) Δ=2				
$E[R_1]$	.189	.150	.147	.158	.044	.081	.090	.078
$E[R_2]$	.148	.175	.179	.166	.434	.286	.267	.291
$SD[R_1]$	.034	.023	.024	.027	.023	.021	.022	.017
$SD[R_2]$	.031	.026	.028	.027	.118	.048	.047	.042
				(ii) Δ=4				
$E[R_1]$	.039	.028	.026	.024	.007	.010	.011	.015
$E[R_2]$	.019	.023	.025	.024	.112	.061	.057	.038
$SD[R_1]$	.015	.008	.008	.006	.006	.005	.005	.004
$SD[R_2]$	.008	.007	.008	.006	.066	.021	.020	.011

and π_1/π_2 are not the same, the shift due to the inclusion of log n_1/n_2 in the discriminant function may become considerable and hence, it may cause higher bias as well as higher variance for an error rate. Thus, unless the allocation procedure for the training samples is objectively formulated as reflected in our models (b) and (c), the use of Fisher rule may be preferred over the Bayes rule because of its robustness property.

APPENDIX A

Mixture Distributions of C_1* and C_2*

We obtain parameters of the two mixture distributions by expressing these in terms of $\underset{\sim}{\mu}_1$, $\underset{\sim}{\mu}_2$, $\underset{\sim}{\Sigma}$, α_1 and α_2. First we obtain these parameters by considering the orginal class structure and then give these parameters for the case of canonical form.

Without loss of generality, let $\underset{\sim}{\mu}_1$ and $\underset{\sim}{\mu}_2$ be aligned along the x_1-axis and the conditional means in other dimensions, given x_1, be

$$\mu_{ij|x_1} = \mu_{ij} + \gamma_j \, (x_1 - \mu_{i1}), \; j = 2, 3, \cdots, p \tag{A-1}$$

for $\underset{\sim}{X} \in C_i$, i=1, 2. Suppose σ^2 denotes the common variance of the two distributions for X_1, the first component of random vector $\underset{\sim}{X}$. Let $\underset{\sim}{\mu}_i^*$ and $\underset{\sim}{\Sigma}_i^*$ denote the mean vector and covariance matrix for C_i^*, i = 1, 2. The frequency function of X_1 for C_i^* can be written as

$$f_i^*(z) = [1 - g_i(z)] \, f_i(z) + g_{3-i}(z) \, f_{3-i}(z) \tag{A-2}$$

where $g_i(z)$ and $f_i(z)$, i = 1, 2, are as defined in section 2.

Then the probability associated with C_i^* is

$$\begin{aligned} \pi_i^* &= \int_{-\infty}^{\infty} f_i^*(z) \, dz \\ &= (1 - \alpha_i) \, \pi_i + \alpha_{3-i}\pi_{3-i}, \; i = 1, 2 \end{aligned} \tag{A-3}$$

and $\pi_1^* + \pi_2^* = 1$.

Define

$$m_i = \frac{1}{\pi_i\alpha_i} \int_{-\infty}^{\infty} \left(\frac{z - \mu_{i1}}{\sigma}\right) g_i(z) f_i(z) dz$$

ORIGINAL PAGE IS
OF POOR QUALITY

and (A-4)

$$V_i = \frac{1}{\pi_i \alpha_i} \int_{-\infty}^{\infty} \left(\frac{z - \mu_{i1}}{\sigma}\right)^2 g_i(z) f_i(z) dz, \quad i = 1, 2.$$

Now the elements of μ_i^*, $i = 1, 2$, are obtained as follows:

For

$$\mu_{11}^* = \frac{1}{\pi_1^*} \int_{-\infty}^{\infty} z f_1^*(z) dz$$

it follows from (2.2) and (A-2) to (A-4) that

$$\pi_1^* \mu_{11}^* = \pi_1 \mu_{11} - \pi_1 \alpha_1 (\mu_{11} + m_1 \sigma) + \pi_2 \alpha_2 (\mu_{21} + m_2 \sigma)$$

$$= \pi_1^* \mu_{11} + \pi_2 \alpha_2 (\mu_{21} - \mu_{11}) + (\pi_2 \alpha_2 m_2 - \pi_1 \alpha_1 m_1) \sigma.$$

Similarly

$$\pi_2^* \mu_{21}^* = \pi_2^* \mu_{21} - \pi_1 \alpha_1 (\mu_{21} - \mu_{11}) - (\pi_2 \alpha_2 m_2 - \pi_1 \alpha_1 m_1) \sigma.$$

For $j = 2, 3, \cdots, p$, we have

$$\pi_1^* \mu_{ij}^* = \int_{-\infty}^{\infty} \mu_{ij|z} [1 - g_1(z)] f_1(z) dz + \int_{-\infty}^{\infty} \mu_{2j|z} g_2(z) f_2(z) dz.$$

Making substitutions from (A-1) and simplifying it, we get

$$\pi_1^* \mu_{1j}^* = \pi_1^* \mu_{1j} + \pi_2 \alpha_2 (\mu_{2j} - \mu_{1j}) + \gamma_j (\pi_2 \alpha_2 m_2 - \pi_1 \alpha_1 m_1) \sigma.$$

Similarly,

$$\pi_2^* \mu_{2j}^* = \pi_2^* \mu_{2j} - \pi_1 \alpha_1 (\mu_{2j} - \mu_{1j}) - \gamma_j (\pi_2 \alpha_2 m_2 - \pi_1 \alpha_1 m_1) \sigma.$$

Let

$$\delta_j = \mu_{2j} - \mu_{1j},\ j = 1, 2, \cdots, p$$

$$t = \pi_2\alpha_2 m_2 - \pi_1\alpha_1 m_1 \qquad \text{(A-5)}$$

$$\alpha_1^* = \pi_2\alpha_2/\pi_1^* \text{ and } \alpha_2^* = \pi_1\alpha_1/\pi_2^*.$$

Then

$$\mu_{1j}^* = \mu_{1j} + \alpha_1^*\,\delta_j + \gamma_j t\ \sigma/\pi_1^*$$

$$\mu_{2j}^* = \mu_{2j} - \alpha_2^*\,\delta_j - \gamma_j t\ \sigma/\pi_2^* \qquad \text{(A-6)}$$

$$j=1, \cdots, p$$

where $\gamma_1 = 1$. Another form of (A-6) that will be used in the derivation of covariance matrices Σ_1^* and Σ_2^* is:

$$\mu_{1j}^* = \mu_{2j} - (1 - \alpha_1^*)\,\delta_j + \gamma_j t\ \sigma/\pi_1^*$$

$$\qquad \text{(A-7)}$$

$$\mu_{2j}^* = \mu_{1j} + (1 - \alpha_2^*)\,\delta_j - \gamma_j t\ \sigma/\pi_2^*.$$

Next, the covariance matrix for C_1^*,

$$\underset{\sim}{\Sigma}_1^* = E_{\underset{\sim}{X}}[(\underset{\sim}{X} - \underset{\sim}{\mu}_1^*)(\underset{\sim}{X} - \underset{\sim}{\mu}_1^*)']$$

can be written as

$$\pi_1^* \Sigma_1^* = \int_{-\infty}^{\infty} E_{X|z}[(X - \mu_1^*)(X - \mu_1^*)' | z] f_1^*(z) dz$$

$$= \int_{-\infty}^{\infty} \left[\Sigma_{X|z} + (\mu_{1|z} - \mu_1^*)(\mu_{1|z} - \mu_1^*)'\right][1 - g_1(z)] f_1(z) dz$$

$$+ \int_{-\infty}^{\infty} \left[\Sigma_{X|z} + (\mu_{2|z} - \mu_1^*)(\mu_{2|z} - \mu_1^*)'\right] g_2(z) f_2(z) dz \qquad \text{(A-8)}$$

where $\mu_{1|z}$ and $\Sigma_{X|z}$ are the conditional mean vector and covariance matrix of X given z. This easily follows from the conditional expectation argument. The elements of $\mu_{i|z}$, i = 1, 2, are given in (A-1) with x_1 replaced by z. Letting

$$\pi_1 \Sigma^0 = \int_{-\infty}^{\infty} \Sigma_{X|z} f_1(z) dz$$

$$\pi_1 \alpha_1 \Sigma^0 = \int_{-\infty}^{\infty} \Sigma_{X|z} g_1(z) f_1(z) dz$$

and making substitutions from (A-1), (A-6) and (A-7) in (A-8), it can be shown that

$$\Sigma_1^* = \Sigma^0 + \alpha_1^*(1 - \alpha_1^*)\, \delta\delta' + (\pi_1 + x_1)\, \gamma\gamma' \sigma^2/\pi_1^*$$
$$+ \psi_1(\delta\gamma' + \gamma\delta')\sigma/\pi_1^* \qquad \text{(A-9)}$$

where

$$x_1 = \pi_1(t|\pi_1^*)^2 - \pi_1\alpha_1[V_1 + (t|\pi_1^*)^2 - 2\, m_1(t|\pi_1^*)]$$
$$+ \pi_2\alpha_2[V_2 + (t/\pi_1^*)^2 - 2\, m_1(t/\pi_1^*)]$$
$$\psi_1 = \alpha_1^*[(1 - \alpha_1^*)\, t + \pi_1\alpha_1 m_1] + (1 - \alpha_1^*)[-\alpha_1^* t + \pi_1\alpha_1 m_1].$$

Similarily, the covariance matrix $\underset{\sim}{\Sigma}_2^*$ can be obtained as

$$\underset{\sim}{\Sigma}_2^* = \underset{\sim}{\Sigma}^0 + \alpha_2^*(1 - \alpha_2^*)\underset{\sim}{\delta}\underset{\sim}{\delta}' + (\pi_2 + \chi_2)\underset{\sim}{\gamma}\underset{\sim}{\gamma}'\sigma^2/\pi_2^* + \psi_2(\underset{\sim}{\delta}\underset{\sim}{\gamma}' + \underset{\sim}{\gamma}\underset{\sim}{\delta}')\sigma/\pi_2^* \qquad \text{(A-10)}$$

where

$$\chi_2 = \pi_2(t/\pi_2^*)^2 - \pi_2\alpha_2 [V_2 + (t/\pi_2^*)^2 + 2 m_2(t/\pi_2^*)] + \pi_1\alpha_1[V_1 + (t/\pi_2^*)^2 + 2 m_1(t/\pi_2^*)]$$

$$\psi_2 = \alpha_2^*[(1 - \alpha_2^*) t - \pi_2\alpha_2 m_2] - (1 - \alpha_2^*) [\alpha_2^* t + \pi_1\alpha_1 m_1].$$

In the discriminant function, we use the pooled covariance matrix which is an estimate of the weighted covariance matrix, $\underset{\sim}{\Sigma}^* = \pi_1^*\underset{\sim}{\Sigma}_1^* + \pi_2^*\underset{\sim}{\Sigma}_2^*$, which is given by

$$\underset{\sim}{\Sigma}^* = \underset{\sim}{\Sigma} + \eta\underset{\sim}{\delta}\underset{\sim}{\delta}' + \chi\underset{\sim}{\gamma}\underset{\sim}{\gamma}'\sigma^2 + \psi (\underset{\sim}{\delta}\underset{\sim}{\gamma}' + \underset{\sim}{\gamma}\underset{\sim}{\delta}')\sigma \qquad \text{(A-11)}$$

where

$$\eta = \alpha_1^* (1 - \alpha_1^*) \pi_1^* + \alpha_2^* (1 - \alpha_2^*) \pi_2^*$$

$$\chi = \chi_1 + \chi_2$$

$$\psi = \psi_1 + \psi_2.$$

with $\underset{\sim}{\delta}$, t and α^*'s are as defined in (A-5). In obtaining (A-11), we have made use of the fact that $\underset{\sim}{\Sigma}^0 + \underset{\sim}{\gamma}\,\underset{\sim}{\gamma}'\sigma^2 = \underset{\sim}{\Sigma}$.

In the case of canonical form, the mean vectors for C_i^*, i = 1, 2, and the weighted covariance matrix are

$$\underset{\sim}{\mu}_1^* = [-(1 - 2\,\alpha_1^*)(\Delta/2) + t/\pi_1^*]\,\underset{\sim}{e}_1$$

$$\underset{\sim}{\mu}_2^* = [(1 - 2\,\alpha_2^*)(\Delta/2) - t/\pi_2^*]\,\underset{\sim}{e}_1$$

$$\underset{\sim}{\Sigma}^* = I + \xi \underset{\sim}{e}_1 \underset{\sim}{e}_1'$$

where

$$\underset{\sim}{e}_1' = (1, 0, \cdots, 0)$$

$$\xi = \eta\Delta^2 + \chi + 2\,\Delta\,\psi.$$

These expressions are obtained from (A-6) and (A-11) by recognizing that $\delta_1 = \Delta$, $\gamma_1 = 1$ and $\sigma^2 = 1$, $\underset{\sim}{\Sigma} = \underset{\sim}{I}$, and $\delta_j = 0$ and $\gamma_j = 0$, $j = 2, 3, \cdots, p$.

APPENDIX B

Derivation of $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}$

Let $\underset{\sim}{\sigma}^{(1)} = \underset{\sim}{\gamma}\sigma$ and $\underset{\sim}{\sigma}^{(2)} = (\sigma_{22}, \sigma_{23}, \cdots, \sigma_{pp})$, where $\sigma_{22}, \sigma_{23}, \cdots, \sigma_{pp}$ are the elements of the upper triangular matrix of $\underset{\sim}{\Sigma}$ with its first row excluded. Suppose $\underset{\sim}{\sigma}^{*(1)}$ is the first row of $\underset{\sim}{\Sigma}^{*-1}$ and $\underset{\sim}{\sigma}^{*(2)}$ is the vector of elements of the upper triangular matrix, less $\underset{\sim}{\sigma}^{*(1)}$ of Σ^{*-1}. In the determination of $\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}}$, there is no need to consider $\underset{\sim}{\sigma}^{(2)}$ and $\sigma^{*(2)}$; e.g., refer to Lemma 2 in Efron (1975). Suppose $\zeta = \log \pi_1/\pi_2$, $\zeta^* = \log \pi_1^*/\pi_2^*$, and

$$
\begin{aligned}
\underset{\sim}{\theta}' &= (\zeta, \underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2, \underset{\sim}{\sigma}^{(1)}) \\
\underset{\sim}{\alpha}' &= (\alpha_1, \alpha_2, \alpha_1 m_1, \alpha_2 m_2) \\
\underset{\sim}{\psi} &= \begin{pmatrix} \underset{\sim}{\theta} \\ \underset{\sim}{\alpha} \end{pmatrix}
\end{aligned}
\tag{B-1}
$$

and

$$
\underset{\sim}{\theta}^{*\prime} = (\zeta^*, \underset{\sim}{\mu}_1^*, \underset{\sim}{\mu}_2^*, \underset{\sim}{\sigma}^{*(1)}). \tag{B-2}
$$

Then by the δ-method (Rao, 1973), we have

$$
\underset{\sim}{V}_{\hat{\underset{\sim}{\beta}}} = \left(\frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\theta}^*}\right)\left(\frac{\partial \underset{\sim}{\theta}^*}{\partial \underset{\sim}{\psi}}\right) \underset{\sim}{V}_{\hat{\underset{\sim}{\psi}}} \left(\frac{\partial \underset{\sim}{\theta}^*}{\partial \underset{\sim}{\psi}}\right)' \left(\frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\theta}^*}\right)' \tag{B-3}
$$

where

$$
\underset{\sim}{V}_{\hat{\underset{\sim}{\psi}}} = \begin{bmatrix} \underset{\sim}{V}_{\hat{\underset{\sim}{\theta}}} & V_{\hat{\underset{\sim}{\theta}}\hat{\underset{\sim}{\alpha}}} \\ V_{\hat{\underset{\sim}{\theta}}\hat{\underset{\sim}{\alpha}}} & V_{\hat{\underset{\sim}{\alpha}}} \end{bmatrix}.
$$

The elements of $V_{\hat{\psi}}$ can be obtained by evaluating the asymptotic variances for the maximum likelihood estimates of ψ. Restricting ourselves to the case of canonical form, we have the following asymptotic variances of $\hat{\mu}_i$ and $\hat{\sigma}^{(1)}$,

$$n\ V[\hat{\mu}_i] = \frac{1}{\pi_i} I,\ i = 1,\ 2$$

$$n\ V[\hat{\sigma}^{(1)}] = I + E_{11} \qquad (B\text{-}4)$$

and their asympotic covariance zero, where $E_{11} = e_1 e_1'$. Determination of $V_{\hat{\alpha}}$ and $V_{\hat{\theta}\hat{\alpha}}$ would require the misallocation model to be specified. We skip the specifics and sketch the main steps involved in obtaining these matrices.

Define the random variable y by

$$y = \begin{cases} 0, & \text{Sample observation } X \text{ is correctly allocated} \\ 1, & \text{Sample observation } X \text{ is misallocated} \end{cases}$$

If $X \epsilon C_i$, then it can be seen from (2.2) and (A-4) that $E[y] = \alpha_i$, $V[y] = \alpha_i\ (1 - \alpha_i)$, $E[yz] = \alpha_i m_i$ and $E[(yz)^2] = E[yz^2] = \alpha_i \nu_2$, (say).

So the asymptotic elements of $V_{\hat{\alpha}}$ are given by

$$n\ V[\hat{\alpha}_i] = V\ [y] = \alpha_i (1 - \alpha_i)$$

$$n\ V[\widehat{\alpha_i m_i}] = V[yz] = \alpha_i \nu_2 - (\alpha_i m_i)^2 \qquad (B\text{-}5)$$

$$n\ \mathrm{Cov}\ [\hat{\alpha}_i, \widehat{\alpha_i m_i}] = \mathrm{Cov}\ [y,\ yz] = \alpha_i (1 - \alpha_i) m_i$$

$$i = 1,\ 2.$$

Noting that these variables are independent for C_1 and C_2, all elements of $V_{\hat{\underset{\sim}{\alpha}}}$ are obtained in (B-5). Next, $V_{\hat{\underset{\sim}{\theta}}\hat{\underset{\sim}{\alpha}}}$ may be derived by the use of δ-method.

Denote $\underset{\sim}{\alpha} = \underset{\sim}{\alpha}(\underset{\sim}{\theta})$. Then $d\underset{\sim}{\alpha} = \left(\frac{\partial \underset{\sim}{\alpha}}{\partial \underset{\sim}{\theta}}\right) d\underset{\sim}{\theta}$ and $d\underset{\sim}{\theta}(d\underset{\sim}{\alpha})' = d\underset{\sim}{\theta}(d\underset{\sim}{\theta})' \left(\frac{\partial \underset{\sim}{\alpha}}{\partial \underset{\sim}{\theta}}\right)$.

Thus

$$V_{\hat{\underset{\sim}{\theta}}\hat{\underset{\sim}{\alpha}}} = E[d\underset{\sim}{\theta}(d\underset{\sim}{\alpha})'] = V_{\hat{\underset{\sim}{\theta}}} \left(\frac{\partial \underset{\sim}{\alpha}}{\partial \underset{\sim}{\theta}}\right)' \tag{B-6}$$

It can be shown that

$$\left.\begin{array}{lll}
\dfrac{\partial \alpha_1}{\partial \mu_{11}} = \alpha_1 m_1, & \dfrac{\partial \alpha_1}{\partial \mu_{21}} = 0, & \dfrac{\partial \alpha_1}{\partial \sigma} = \alpha_1 (m_1^{(2)} - 1) \\[2ex]
\dfrac{\partial \alpha_1 m_1}{\partial \mu_{11}} = \alpha_1 m_1^{(2)}, & \dfrac{\partial \alpha_1 m_1}{\partial \mu_{21}} = 0, & \dfrac{\partial \alpha_1 m_1}{\partial \sigma} = \alpha_1 (m_1^{(3)} - m_1) \\[2ex]
\dfrac{\partial \alpha_2}{\partial \mu_{11}} = 0, & \dfrac{\partial \alpha_2}{\partial \mu_{21}} = \alpha_2 m_2, & \dfrac{\partial \alpha_2}{\partial \sigma} = \alpha_2 (m_2^{(2)} - 1) \\[2ex]
\dfrac{\partial \alpha_2 m_2}{\mu_{11}} = 0, & \dfrac{\partial \alpha_2 m_2}{\partial \mu_{21}} = \alpha_2 m_2^{(2)}, & \dfrac{\partial \alpha_2 m_2}{\partial \sigma} = \alpha_2 (m_2^{(3)} - m_2)
\end{array}\right\} \tag{B-7}$$

where

$$\alpha_i m_i^{(r)} = \int_{-\infty}^{\infty} z^r g_i(z) \phi(z) dz$$

which can be easily evaluated by specifying $g_i(z)$, $i = 1, 2$.

Though the matrices $\frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\theta}^*}$ and $\frac{\partial \underset{\sim}{\theta}^*}{\partial \underset{\sim}{\psi}}$ are somewhat complex, their derivations are fairly straight forward. These are as follows:

$$\frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\theta}^*} = \begin{bmatrix} \frac{\partial \beta_0}{\partial \underset{\sim}{\theta}^*} \\ \frac{\partial \beta^{(1)}}{\partial \underset{\sim}{\theta}^*} \end{bmatrix} = \begin{bmatrix} 1 & \frac{\mu_{11}^*}{1+\xi}\underset{\sim}{e}_1' & -\frac{\mu_{21}^*}{1+\xi}\underset{\sim}{e}_1' & -\frac{1}{2}(\mu_{21}^{*2} - \mu_{11}^{*2})\,\underset{\sim}{e}_1' \\ \underset{\sim}{0} & -\underset{\sim}{I} + \frac{\xi}{1+\xi}\underset{\sim}{E}_{11} & \underset{\sim}{I} - \frac{\xi}{1+\xi}\underset{\sim}{E}_{11} & (\mu_{21}^* - \mu_{11}^*)\underset{\sim}{I} \end{bmatrix} \quad \text{(B-8)}$$

$$\frac{\partial \underset{\sim}{\theta}^*}{\partial \underset{\sim}{\psi}} = \begin{bmatrix} \frac{\partial \underset{\sim}{\theta}^*}{\partial \underset{\sim}{\theta}} \\ \frac{\partial \underset{\sim}{\theta}}{\partial \underset{\sim}{\alpha}} \end{bmatrix}. \quad \text{(B-9)}$$

where

$$\frac{\partial \underset{\sim}{\theta}^*}{\partial \underset{\sim}{\theta}} = \begin{bmatrix} \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & (1-\alpha_1)\underset{\sim}{I} & \alpha_1\underset{\sim}{I} & -(t/\pi_1^*)\underset{\sim}{I} \\ \underset{\sim}{0} & \alpha_2\underset{\sim}{I} & (1-\alpha_2)\underset{\sim}{I} & (t/\pi_2^*)\underset{\sim}{I} \\ \underset{\sim}{0} & \frac{\partial \underset{\sim}{\sigma}^{*(1)}}{\partial \underset{\sim}{\mu}_1} & \frac{\partial \underset{\sim}{\sigma}^{*(1)}}{\partial \underset{\sim}{\mu}_2} & \frac{\partial \underset{\sim}{\sigma}^{*(1)}}{\partial \underset{\sim}{\sigma}^{(1)}} \end{bmatrix}$$

with

$$\frac{\partial \underset{\sim}{g}^{*(1)}}{\partial \underset{\sim}{\mu}_1} = \frac{2(n\Delta + \psi)}{(1+\xi)^2}\underset{\sim}{E}_{11} + \frac{(n\Delta + \psi)}{1+\xi}(\underset{\sim}{I} - \underset{\sim}{E}_{11})$$

$$\frac{\partial \underset{\sim}{g}^{*(1)}}{\partial \underset{\sim}{\mu}_2} = - \frac{\partial \underset{\sim}{g}^{*(1)}}{\partial \underset{\sim}{\mu}_1}$$

$$\frac{\partial \underset{\sim}{g}^{*(1)}}{\partial \underset{\sim}{g}^{(1)}} = - \frac{2(1 + \chi + \psi\Delta)}{(1+\xi)^2}\underset{\sim}{E}_{11} - \frac{(1 + \chi + \psi\Delta)}{1+\xi}(\underset{\sim}{I} - \underset{\sim}{E}_{11})$$

and

$$\frac{\partial \underset{\sim}{\theta}^*}{\partial \underset{\sim}{\alpha}} = \begin{bmatrix} \frac{\pi_1}{\pi_1^*\pi_2^*} & -\frac{\pi_2}{\pi_1^*\pi_2^*} & 0 & 0 \\ \frac{\pi_1}{\pi_1^*}(\frac{\Delta}{2} + \mu_{11}^*)\underset{\sim}{e}_1 & \frac{\pi_2}{\pi_1^*}(\frac{\Delta}{2} - \mu_{11}^*)\underset{\sim}{e}_1 & -\frac{\pi_1}{\pi_1^*}\underset{\sim}{e}_1 & \frac{\pi_2}{\pi_1^*}\underset{\sim}{e}_1 \\ -\frac{\pi_1}{\pi_2^*}(\frac{\Delta}{2} + \mu_{21}^*)\underset{\sim}{e}_1 & -\frac{\pi_2}{\pi_2^*}(\frac{\Delta}{2} - \mu_{21}^*)\underset{\sim}{e}_1 & \frac{\pi_1}{\pi_2^*}\underset{\sim}{e}_1 & -\frac{\pi_2}{\pi_2^*}\underset{\sim}{e}_1 \\ -\frac{\pi_1}{(1+\xi)^2} & -\frac{\pi_2}{(1+\xi)^2} & \frac{2\pi_1}{(1+\xi)^2} & -\frac{2\pi_2}{(1+\xi)^2} \\ \times\frac{\partial\xi}{\partial M_1}\underset{\sim}{e}_1 & \times\frac{\partial\xi}{\partial M_2}\underset{\sim}{e}_1 & \times(\mu_{21}^* - \mu_{11}^*)\underset{\sim}{e}_1 & \times(\mu_{21}^* - \mu_{11}^*)\underset{\sim}{e}_1 \end{bmatrix}$$

with

$$\frac{\partial\xi}{\partial M_1} = \Delta^2(1 - 2\alpha_2 - \alpha_1^2 + \alpha_2^2) + 2\,\Delta t(\alpha_1\pi_2^* + (1-\alpha_2)\pi_1^*)/\pi_1^*\pi_2^*$$

$$+ t^2(\pi_1^{*2} - \pi_2^{*2})/\pi_1^{*2}\pi_2^{*2}$$

$$\frac{\partial\xi}{\partial M_2} = \Delta^2(1 - 2\alpha_1 + \alpha_1^2 - \alpha_2^2) + 2\,\Delta t(\alpha_2\pi_1^* + (1-\alpha_1)\,\pi_2^*)/\pi_1^*\pi_2^*$$

$$+ t^2(\pi_2^{*2} - \pi_1^{*2})/\pi_1^{*2}\pi_2^{*2}.$$

REFERENCES

Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis, New York: John Wiley & Sons.

Box, G. E. P. and Muller, M. A. (1958), "A Note on the Generation of Random Normal Deviates," Annals of Mathematical Statistics, 29, 610-611.

Efron, Bradley (1975), "The Efficiency of Logistic Regression compared to Normal Discriminant Analysis," Jorunal of the American Statistical Association, 70, 892-898.

Lachenbruch, Peter A. (1966), "Discriminant Analysis When the Initial Samples are Misclassified," Technometrics, 8, 657-662.

______________ (1974), "Discriminant Analysis When the Initial Samples are Misclassified II: Non-Random Misclassification Models", Technometrics, 16, 419-424.

McLachlan, G. J. (1972), "Asymptotic Results for Discriminant Analysis When the Initial Samples are Misclassified, "Technometrics, 14, 415-422.

Michalek, Joel, E. and Tripathi, Ram C. (1980), "The Effect of Errors in Diagnosis and Measurement on the Estimation of the Probability of an Event," Journal of the American Statistical Association, 75, 713-721.

O'Neill, Terence J. (1980), "The General Distribution of the Error Rate of a Classification Procedure With Application to Logistic Regression Discrimination," Journal of the American Statistical Association, 75, 154-160.

Rao, C. R. (1973). Linear Statistical Inference and Applications, New York: John Wiley & Sons.

Sayre, James W. (1980), "The Distribution of the Actual Error Rates in Linear Discriminant Analysis," Journal of American Statistical Association, 75, 201-205.