Linear Discriminant Analysis

by

Robert Harry Riffenburgh, B.S., M.S.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICS

APPROVED:

_____

Chairman, Advisory Committee

_____ _____

_____ _____

May 1957

Blacksburg, Virginia

TABLE OF CONTENTS

IV. COMPARISON OF DISCRIMINATORS AT LEVEL (c)

V. LARGE SAMPLE THEORY

TABLES

FIGURES

I. INTRODUCTION

1.1. Introduction

There is frequent need to classify an individual as having arisen from one of two populations. Discriminant analysis for the two-population case is the study of methods of classification by means of a discriminator, or scalar function of measurements of the individual. Inasmuch as the multivariate normal distribution is almost exclusively assumed in the practice of discriminant analysis, this paper specifies and restricts itself to populations having the multivariate normal distribution. Two types of discriminators have been devised by other authors, linear and quadratic.

A quadratic discriminator consists of a quadratic function of measurements of an individual; similarly a linear discriminator consists of a linear function of these same measurements. Smith (11) has shown that a second-degree discriminator is adequate for two-population classification when both populations have multivariate normal distributions.

However, the quadratic discriminator is undesirable for both practical and theoretical considerations. Practically, the quadratic discriminator involves the inversion of two matrices rather than one and introduces additional labor in

evaluating a certain constant. Theoretically, the quadratic discriminator presents difficulty in finding the probabilities of misclassification, as they involve integrating a multivariate normal density with a hyperconic section as a limit of integration; this would involve extensive tabulation. (Some simplification in the bivariate case of this integration is given in the Appendix, 10.1.)

This paper will consider linear discrimination for two populations where an individual is represented by k measurements and where each population will give rise to a distinct k-variate normal distribution for randomly picked individuals. The set of values of the k variates will be considered as defining a point in k-space.

The linear discriminator, $f(x_1,\ldots,x_k)$, is of the form

$$f(x_1,\ldots,x_k) = a_1x_1 + a_2x_2 + \ldots + a_kx_k,$$

where the a's are constants and the x's are the measurements on an individual to be classified. In matrix notation, when A is a row vector of the a's, X is the corresponding row vector of the x's, and primes denote transposition,

$$f(X) = AX' \quad .$$

Thus, the measurements of the individual are used as arguments for the discriminator, $f(X)$, and the individual is classified according to whether the resulting value of $f(X)$ is greater than or less than some fixed number, say c.

The linear discriminator $\ell(X) = c$ may be considered to be a hyperplane dividing the k-space into two parts. The vector A determines the direction of the hyperplane and the constant c determines the distance of the hyperplane from the origin; i.e. its "position".

The linear discriminator is used in practice almost exclusively. However, the linear discriminator is elsewhere considered as a special case of the quadratic which occurs when the dispersion (variance-covariance) matrices of the populations are equal. Linear discrimination has been investigated in the past as a technique only; population studies and sampling distributions have been neglected. While the assumption of equal dispersion matrices is a frequently met special case, it is by no means justified in general. Furthermore, adequate investigation has not been made using loss functions associated with misclassification and a priori information regarding the probabilities that a randomly drawn individual arose from one or the other population.

It is the purpose of this paper:

(1) to present a population study of linear discrimination with respect to certain criteria,

and (2) to investigate sampling distributions (for large samples) where practicable.

The population study will be considered on three levels:

(a) risk when loss functions and a priori probabilities of randomly drawn individuals arising from one or the other population are known;

(b) expected errors when the above-mentioned a priori probabilities are known;

(c) accuracy of discrimination disregarding loss functions and a priori probabilities, this accuracy measured in terms of probabilities of misclassification.

Suppose we have a loss function associated with misclassification of an individual; then this loss function weighted by the appropriate density and integrated over the entire sample space is called the risk. Hence, level (a) is the risk associated with a misclassification. On level (b), we have the probability that an individual will be classified as having arisen from one population, given that it arose from the other, times the probability that a randomly drawn individual arose from a particular population. This is the expectation that a misclassification will be made, which may be called expected error.

There are two reasonable criteria by which a discriminator may be judged:

(i) total risk associated with misclassification, total expected misclassification, and sum of conditional probabilities of misclassification (respectively for levels a, b, and c) for an individual to be classified;

(ii) maximum among risks of misclassification, maximum among expected misclassifications, and maximum among probabilities of misclassification (respectively for levels a, b, and c) for an individual to be classified.

The first of these was introduced on level (a) by Brown [3] and the second introduced on level (b) by Welch [12]. Welch restricted the expected errors to be equal and then minimized them, a technique which facilitates the use of criterion (ii). A paper by Lindley [8] gives some discussion of criterion (i), pointing out that in the two-populational case classification satisfying criterion (i) is classification by minimum unlikelihood. However, although "optimum" properties are readily established for quadratic discriminators, no explicit results have been derived for the risks associated with, or probabilities of, misclassification; these can be obtained explicitly in general only when the discriminator is linear. Although Welch and Brown introduced these criteria, they both left them in the most general form, not applying them to any specific distribution.

1.2. Some mathematical definitions

Note that linear discrimination is equivalent to partitioning the k-space by a hyperplane of the form $AX' = c$. We lose no generality if we use a classificatory rule such that an individual is classified into population I if $\ell(X) = AX' \leqslant c$, and into population II if $\ell(X) = AX' > c$.

Suppose that a classification of an individual, X, from population u into v involves a finite loss $W_{u,v}(X) \geqslant 0$, $u,v = I, II$; suppose that a correct classification involves zero loss, i.e. $W_{u,u}(X) = 0$. Suppose we have a priori information regarding the

probabilities that a randomly drawn individual belongs to one or the other population; say p for population I and q for population II. Designate the risk (expected loss) and density function for population u as R_u and $f_u(X)$ respectively; then

$$(1.2.1) \qquad R_I = p \int_{AX' > c} \cdots \int W_{I,II}(X) f_I(X) dX \quad ,$$

$$(1.2.2) \qquad R_{II} = q \int_{AX' \leq c} \cdots \int W_{II,I}(X) f_{II}(X) dX \quad ,$$

since $W_{I,I}(X) = W_{II,II}(X) = 0$.

We may express criterion (i) for risks as $R = R_I + R_{II}$ and criterion (ii) for risks as R_L, where R_L is the larger of R_I and R_{II}.

We may note that only in certain situations is the problem (i.e. that of finding A and c to satisfy one of the criteria) resolvable. If both the integrals R_I and R_{II} exist for all hyperplanes, then a solution exists. If one of these integrals converges while the other diverges, the solution is trivial, with all individuals being classified as having arisen from the population having the divergent integral. If both integrals diverge, the problem is not resolvable. This applies for both criteria.

Let us note some loss functions for which the problem is resolvable:

(a) $W_{u,v}(X) \propto$ k-variate normal density; $u,v = I,II,$

(b) $W_{u,v}(X) \propto \left[f_u(X)\right]^{t_u}$, where $t_u > -1$;

(c) $W_{u,v}(X) =$ constant.

(a) involves any k-variate normal density and (b) is the special case of (a) which involves the density of population u. In (b), the loss function varies inversely with the density if $t_u < 0$, and varies directly with the density if $t_u > 0$. In the first case, the central values involve lesser loss than peripheral values when misclassified, and in the latter case, the central values involve a greater loss than the peripheral values when misclassified.

Before continuing, let us note the following identities:

$$(1.2.3) \qquad \int_{AX' \leq c} \cdots \int \mathcal{N}(M, \Sigma)\,dX \equiv \int_{-\infty}^{c} N(m,\sigma^2)\,dx \equiv \int_{-\infty}^{\frac{c-m}{\sigma}} N(0,1)\,dx$$

and

$$(1.2.4) \qquad \int_{AX' > c} \cdots \int \mathcal{N}(M, \Sigma)\,dX \equiv \int_{c}^{\infty} N(m,\sigma^2)\,dx \equiv \int_{\frac{c-m}{\sigma}}^{\infty} N(0,1)\,dx$$

$$\equiv \int_{-\infty}^{\frac{m-c}{\sigma}} N(0,1)\,dx,$$

where M is the row vector of means, μ_i, Σ is the dispersion matrix with elements σ_{ij}, i, j = 1,...,k,

$$(1.2.5) \qquad m = AM' = \sum_{i=1}^{k} a_i \mu_i ,$$

$$(1.2.6) \qquad \sigma^2 = A \Sigma A' = \sum_{i,j=1}^{k} a_i \sigma_{ij} a_j ,$$

$$(1.2.7) \qquad \mathcal{N}(M, \Sigma) = (2\pi)^{-k/2} |\Sigma|^{-1} \exp\left\{-\tfrac{1}{2}(X-M)\Sigma^{-1}(X-M)'\right\} ,$$

and

$$(1.2.8) \qquad N(m, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left(\frac{X-m}{\sigma}\right)^2\right\} .$$

We shall use $\mathcal{N}(M, \Sigma)$ interchangeably as a population specification and the corresponding density function and shall specify the two populations as $\mathcal{N}(M_1, \Sigma_1)$ and $\mathcal{N}(M_2, \Sigma_2)$.

The probability of an individual arising from population I and being classified into II will be designated P_I^* ,

$$(1.2.9) \qquad P_I^* = p \int\cdots\int_{AX' > c} \mathcal{N}(M_1, \Sigma_1)dX = p \int_{-\infty}^{\frac{m_1 - c}{\sigma_1}} N(0,1)dx, \quad \text{(from 1.2.4)}$$

and the probability of an individual arising from population II and being classified into I will be designated P_{II}^* ,

$$(1.2.10) \qquad P_{II}^* = q \int\cdots\int_{AX' \leq c} \mathcal{N}(M_2, \Sigma_2)dX = q \int_{-\infty}^{\frac{c - m_2}{\sigma_2}} N(0,1)dx, \quad \text{(from 1.2.3)}$$

since the classificatory rule states that an individual is classified into population I if $f(X) = AX' \leq c$ and into population II if $f(X) = AX' > c$.

On level (b), we may express criterion (i) as $P^* = P_I^* + P_{II}^*$ and criterion (ii) as P_L^*, where P_L^* is the larger of P_I^* and P_{II}^*.

When the prior probabilities are not known, we have the third (c) level. Designate the probability of an individual being classified into population II given that it arose from population I as P_I,

$$(1.2.11) \qquad P_I = \int_{AX' > c} \cdots \int \eta(M_1, \Sigma_1)\, dX = \int_{-\infty}^{\frac{m_1 c}{\sigma_1}} N(0,1)\, dx, \quad \text{(from 1.2.4)}$$

and the probability of an individual being classified into population I given that it arose from population II as P_{II},

$$(1.2.12) \qquad P_{II} = \int_{AX' \leqslant c} \cdots \int \eta(M_2, \Sigma_2)\, dX = \int_{-\infty}^{\frac{c-m_2}{\sigma_2}} N(0,1)\, dx, \quad \text{(from 1.2.3)}$$

since the classificatory rule states that an individual is classified into population I if $\ell(X) = AX' \leqslant c$ and into population II if $\ell(X) = AX' > c$.

On the (c) level, criterion (i) may be expressed as $P = P_I + P_{II}$, and criterion (ii) as P_L, where P_L is the larger of P_I and P_{II}.

II. DIRECTION OF THE HYPERPLANE

2.1 The vector A for level (a) of (1.1), criterion (i)

Consider now the parameters of the hyperplane. If we wish to find a vector A and a c for criterion (i), we wish to find these parameters such that $R = R_I + R_{II}$ is minimized. From (1.2.1) and (1.2.2),

$$(2.1.1) \qquad R = p \int_{AX' > c} \cdots \int W_{I,II}(X) f_I(X) dX + q \int_{AX' \leqslant c} \cdots \int W_{II,I}(X) f_I(X) dX.$$

No solution can be given for general loss functions. The special cases of loss functions mentioned in (1.2) will be considered later.

2.2. The vector A for level (b) of (1.1), criterion (i)

We wish to obtain A and c which minimize the sum of expected errors, $P^* = P_I^* + P_{II}^*$. From (1.2.9) and (1.2.10),

$$(2.2.1) \qquad P^* = p \int_{-\infty}^{\frac{m_1 - c}{\sigma_1}} N(0,1)dx + q \int_{-\infty}^{\frac{c - m_2}{\sigma_2}} N(0,1)dx \ .$$

Differentiating P* with respect to c and a_i, i = 1,...,k,

then setting the derivatives equal to zero, we have:

$$(2.2.2)\qquad \frac{\partial P^*}{\partial c} = -\frac{P}{\sigma_1\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{m_1-c}{\sigma_1}\right)^2\right\} + \frac{q}{\sigma_2\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{c-m_2}{\sigma_2}\right)^2\right\} = 0;$$

$$(2.2.3)\qquad \frac{\partial P^*}{\partial a_i} = \frac{P}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{m_1-c}{\sigma_1}\right)^2\right\} \frac{\partial}{\partial a_i}\left(\frac{m_1-c}{\sigma_1}\right) + \frac{q}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{c-m_2}{\sigma_2}\right)^2\right\} \frac{\partial}{\partial a_i}\left(\frac{c-m_2}{\sigma_2}\right) = 0.$$

But from (2.2.2),

$$(2.2.4)\qquad \frac{P}{\sigma_1} \exp\left\{-\tfrac{1}{2}\left(\frac{m_1-c}{\sigma_1}\right)^2\right\} = \frac{q}{\sigma_2} \exp\left\{-\tfrac{1}{2}\left(\frac{c-m_2}{\sigma_2}\right)^2\right\} \neq 0.$$

The equality of (2.2.4) permits removal of a non-zero factor from (2.2.3) yielding

$$\sigma_1 \frac{\partial}{\partial a_i}\left(\frac{m_1-c}{\sigma_1}\right) + \sigma_2 \frac{\partial}{\partial a_i}\left(\frac{c-m_2}{\sigma_2}\right) = 0,$$

or

$$\sigma_1 \frac{\partial}{\partial a_i}\left(\frac{m_1-c}{\sigma_1}\right) = \sigma_2 \frac{\partial}{\partial a_i}\left(\frac{m_2-c}{\sigma_2}\right).$$

But $m_1 = \sum_{i=1}^{k} a_i \mu_{iI}$, $m_2 = \sum_{i=1}^{k} a_i \mu_{iII}$, $\sigma_1^2 = \sum_{i,j=1}^{k} a_i \sigma_{ijI} a_j$, and $\sigma_2^2 = \sum_{i,j=1}^{k} a_i \sigma_{ijII} a_j$; thus

$$(2.2.5)\qquad \frac{\partial m_1}{\partial a_i} - \frac{m_1 - c}{2\sigma_1^2} \frac{\partial \sigma_1^2}{\partial a_i} = \frac{\partial m_2}{\partial a_i} - \frac{m_2 - c}{2\sigma_2^2} \frac{\partial \sigma_2^2}{\partial a_i},$$

or

$$(2.2.6)\qquad \mu_{iI} - \frac{m_1 - c}{\sigma_1^2} \sum_{j=1}^{k} a_j \sigma_{ijI} = \mu_{iII} - \frac{m_2 - c}{\sigma_2^2} \sum_{j=1}^{k} a_j \sigma_{ijII}.$$

If we take this set of k equations and express the set in matrix notation,

$$(2.2.7)\qquad M_1 - \frac{m_1 - c}{\sigma_1^2} A\Sigma_1 = M_2 - \frac{m_2 - c}{\sigma_2^2} A\Sigma_2.$$

This equation is not readily solvable for the a's which occur in A, m_1, m_2, σ_1^2, and σ_2^2.

Consider the special case of proportionate matrices in which

(2.2.8) $$\Sigma_2 = \alpha^2 \Sigma_1 , \quad \alpha \geq 1.$$

Then equation (2.2.7) reduces to

(2.2.9) $$M_1 - M_2 = \frac{m_1 - c}{\sigma_1^2} A\Sigma_1 - \frac{m_2 - c}{\sigma_1^2 \alpha^2} \alpha^2 A \Sigma_1 \quad .$$

If $M_1 - M_2$ be designated as D and Σ_1 as Σ,

$$D = \frac{m_1 - m_2}{\sigma^2} A\Sigma \quad .$$

But $m_1 = AM_1'$, $m_2 = AM_2'$, and $\sigma^2 = A\Sigma A'$. Then

$$D = \frac{AD' \cdot A\Sigma}{A\Sigma A'} \quad .$$

Let

(2.2.10) $$A = D\Sigma^{-1} \quad .$$

Then

$$D = \frac{D\Sigma^{-1}D' \cdot D\Sigma^{-1}\Sigma}{D\Sigma^{-1}\Sigma\Sigma^{-1}D'} = \frac{D\Sigma^{-1}D' \cdot D}{D\Sigma^{-1}D'} = D \quad .$$

Note that A is independent of c. The evaluation of c, then, will be considered in chapter III.

Thus, $A = D\Sigma^{-1}$ is a set of roots for equation (2.2.3).

Recall that the a's are direction numbers; therefore the ratios of the various a's is all that is required. Any A_1 proportional to A will also solve equations (2.2.3), provided that c is adjusted correspondingly.

It is shown by Smith [10] that these roots minimize P when $\alpha = 1$; it can be shown that, when linear discrimination is appropriate (c.f. 4.2), these roots minimize P when $\alpha \neq 1$.

Minimizing $P_I + P_{II} = P$ on level (c) is a similar procedure and yields an identical result.

If $\Sigma_1 \neq \Sigma_2$, we suggest

$$(2.2.11) \qquad A = D(p\Sigma_1^{-1} + q\Sigma_2^{-1})$$

as a reasonable approximation. If the optimum (i) hyperplane intersects the line connecting the centroids at some point between the centroids, say X_o, then,

(1) Among all hyperplanes through X_o, the ones minimizing the probabilities of misclassification for either population considered separately are the hyperplanes tangential to the equidensity surfaces through X_o (c.f. 2.5.2) which are $D\Sigma_1^{-1}(X - X_o)' = 0$, $D\Sigma_2^{-1}(X - X_o)' = 0$. If $D\Sigma_1^{-1} \propto D\Sigma_2^{-1}$, a solution to (2.2.7) exists, as shown below. (2.2.7) gives

$$M_1 - M_2 = \frac{AM_1' - c}{A\Sigma_1 A'} A\Sigma_1 - \frac{AM_2' - c}{A\Sigma_2 A'} A\Sigma_2 .$$

Suppose $A = D\Sigma_1^{-1} = \frac{1}{\rho} D\Sigma_2^{-1}$; $D\Sigma_2^{-1} = \rho D\Sigma_1^{-1}$. Then

$$D = \frac{D\Sigma_1^{-1}M_1' - c}{D\Sigma_1^{-1}\Sigma_1\Sigma_1^{-1}D'} D\Sigma_1^{-1}\Sigma_1 - \frac{D\Sigma_1^{-1}M_2' - c}{D\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}D'} D\Sigma_1^{-1}\Sigma_2$$

$$= \frac{D\Sigma_1^{-1}M_1' - c}{D\Sigma_1^{-1}D'} D - \frac{D\Sigma_1^{-1}M_2' - c}{\rho D\Sigma_1^{-1}D'} \rho D$$

$$= \frac{D\Sigma_1^{-1}(M_1' - M_2')}{D\Sigma_1^{-1}D'} D = D .$$

(2) Provided $D\Sigma_1^{-1}$, $D\Sigma_2^{-1}$ are not excessively disproportionate, the above hyperplanes are approximately the same; then it seems reasonable that the hyperplane satisfying criterion (i) would be within the smaller of the two supplementary

angles between the two hyperplanes. Thus (2.2.11) may be considered as an approximation to the minimizing values of A.

2.3. The vector A for level (c) of (1.1), criterion (i)

$P = P_I + P_{II}$ is identical with $P^* = P_I^* + P_{II}^*$ if p and q are replaced by 1's. The A in 2.2 is independent of p. Thus, the solution for proportionate matrices is the same as in 2.2, viz. $A = D\Sigma^{-1}$, Σ either Σ_1 or Σ_2. Similarly when the matrices are not proportionate, there is no simple solution and we suggest the use of

$$A = D(\Sigma_1^{-1} + \Sigma_2^{-1}) \tag{2.3.1}$$

corresponding to the previous suggestions.

2.4. The vector A for level (a) of (1.1), criterion (ii)

On the basis of criterion (ii) we wish to find a set of a's to minimize R_L. As in 2.1, no solution can be given for general loss functions. The special cases of loss functions mentioned in (1.2) will be considered later.

2.5. The vector A for level (c) of (1.1), criterion (ii)

Consider the two populations in k-space with centroids M_1 and M_2, and the line connecting the centroids. Any discriminator of hyperplane form intersects this line (take

a parallel line as intersecting it at infinity). Let the point of intersection of the hyperplane and this line have co-ordinates X_o . X_o may or may not lie between the centroids. If X_o does not lie between the centroids, one probability of misclassification will be greater than one-half and we may immediately discard this case. If X_o lies between the centroids, each probability of misclassification will be less than one-half, provided the classificatory rule is equivalent to classifying an observed individual into the population which has its centroid on the same side of the hyperplane as the observed individual.

For criterion (ii) we wish to minimize P_L, the larger of P_I and P_{II} .

Now the equidensity surfaces of the populations passing through X_o are

$$(2.5.1) \qquad (X - M_1)\Sigma_1^{-1}(X - M_1)' = (X_o - M_1)\Sigma_1^{-1}(X_o - M_1)'$$

and

$$(2.5.2) \qquad (X - M_2)\Sigma_2^{-1}(X - M_2)' = (X_o - M_2)\Sigma_2^{-1}(X_o - M_2)' .$$

It has been shown by Clunies-Ross [4] that, if $\Sigma_1 \propto \Sigma_2$, among the family of hyperplanes passing through X_o, the hyperplane minimizing both P_I and P_{II} (and therefore both $pP_I = P_I^*$ and $qP_{II} = P_{II}^*$) will be the common hyperplane tangent to the surfaces at X_o .

This common tangent is

$$(2.5.3) \qquad (X_o - M_u)\Sigma^{-1}(X - M_u)' = (X_o - M_u)\Sigma^{-1}(X_o - M_u)'$$

The equation for the tangent may be written

$$(2.5.4) \qquad (X_o - M_u)\Sigma^{-1}(X - X_o)' = 0 ,$$

or

$$(2.5.5) \qquad A(X - X_o)' = 0 ,$$

where $A = (X_o - M_u)\Sigma^{-1} \propto D\Sigma^{-1}$, since X_o is on the line connecting the centroids.

Thus the A for minimizing P_L is, for $\Sigma_1 \propto \Sigma_2$, the same A as before.

This technique is inapplicable in general when the dispersion matrices are not proportional. Although it can be shown that when the matrices are not proportional the desired A for minimizing P_L differs from that for minimizing $P_I + P_{II}$, we nevertheless (provided the departure from proportionality is reasonably small) suggest that the same A (2.3.1) should be used.

2.6 The vector A for level (b) of (1.1), criterion (ii)

It was mentioned in the last section that P_I^* and P_{II}^* may in this case be considered in the same way as was P_I and P_{II}. Provided again that the matrices are proportional, the same results appear. We suggest that the corresponding A (2.2.11) be used when the matrices are not proportionate.

2.7. The vector A for special cases of $W_{I,II}(X)$ and $W_{II,I}(X)$, both criteria

If the loss functions are any of the special cases mentioned in Section (1.2), i.e. proportional to a multivariate normal

or a constant, so that

$$W_{I,II}(X) = b_I \mathcal{N}(B_1, V_1) \text{ or } = b_I \,, \tag{2.7.1}$$

$$W_{II,I}(X) = b_{II} \mathcal{N}(B_2, V_2) \text{ or } = b_{II} \,, \tag{2.7.2}$$

then

$$W_{I,II}(X) f_I(X) = (2\pi)^{-k} b_I \, |V_1|^{-1} |\Sigma_1|^{-1} \tag{2.7.3}$$

$$\exp\left\{-\tfrac{1}{2}\left[(X-M_1)\Sigma_1^{-1}(X-M_1)' - (X-B_1)V_1^{-1}(X-B_1)'\right]\right\}$$

$$= (2\pi)^{-k/2} b_I^* \left|\Sigma_1^*\right|^{-1} \exp\left\{-\tfrac{1}{2}(X-M_1^*)\Sigma_1^{*-1}(X-M_1^*)'\right\} \,, \text{say,}$$

which defines M_1^*, Σ_1^*, and b_I^* implicitly and uniquely.

Similarly,

$$W_{II,I}(X) f_{II}(X) = (2\pi)^{-k/2} b_{II}^* \left|\Sigma_2^*\right|^{-1} \exp\left\{-\tfrac{1}{2}(X-M_2^*)\Sigma_2^{*-1}(X-M_2^*)'\right\} \tag{2.7.4}$$

which defines M_2^*, Σ_2^*, and b_{II}^* implicitly and uniquely.

For both criteria, if $\Sigma_1^* \propto \Sigma_2^*$, corresponding to (2.2.10),

$$A = (M_2^* - M_1^*)\Sigma_1^{*-1} \,; \tag{2.7.5}$$

if $\Sigma_1^* \not\propto \Sigma_2^*$, corresponding to (2.2.11), then

$$A = (M_2^* - M_1^*)\Sigma^{*-1} \,, \tag{2.7.6}$$

where

$$\Sigma^{*-1} = b_1^* p \Sigma_1^{*-1} + b_2^* q \Sigma_2^{*-1} \,. \tag{2.7.7}$$

III. LOCATION OF THE HYPERPLANE

3.1 Evaluation of c for level (a) of (1.1), criterion (i)

Take the definitions of the several parameters and functions as in 1.2. The a's may now be considered as constants since they are independent of c.

Under criterion (i), we wish to minimize $R = R_I + R_{II}$ and solve for c.

$$\frac{\partial R}{\partial c} = \frac{\partial R_I}{\partial c} + \frac{\partial R_{II}}{\partial c} . \tag{3.1.1}$$

This may not be solved for general loss functions. The previously mentioned special cases will be considered later.

3.2. Evaluation of c for levels (b) and (c) of (1.1), criterion (i)

We lose no generality if we let $m_1 < m_2$. The classificatory rule remains as before.

Here we wish to minimize $P^* = P_I^* + P_{II}^*$.

$$\frac{\partial P^*}{\partial c} = \frac{\partial P_I^*}{\partial c} + \frac{\partial P_{II}^*}{\partial c} . \tag{3.2.1}$$

From (2.2.2), we have

$$\frac{\partial P^*}{\partial c} = -\frac{p}{\sigma_1\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{m_1-c}{\sigma_1}\right)^2\right\} + \frac{q}{\sigma_2\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{c-m_2}{\sigma_2}\right)^2\right\} = 0 . \tag{3.2.2}$$

Hence

$$\ln \frac{p\sigma_2}{q\sigma_1} - \tfrac{1}{2}\left(\frac{m_1-c}{\sigma_1}\right)^2 + \tfrac{1}{2}\left(\frac{c-m_2}{\sigma_2}\right)^2 = 0 . \tag{3.2.3}$$

Expand the squared terms and collect coefficients of like powers of c; then

$$(3.2.4) \qquad c^2\left(\frac{1}{2\sigma_2^2}-\frac{1}{2\sigma_1^2}\right)+c\left(\frac{m_1}{\sigma_1^2}-\frac{m_2}{\sigma_2^2}\right)+\left(\frac{m_2^2}{2\sigma_2^2}-\frac{m_1^2}{2\sigma_1^2}+\ln\frac{p\sigma_2}{q\sigma_1}\right)=0,$$

which yields

$$(3.2.5) \qquad c(\min)=\frac{1}{\sigma_1^2-\sigma_2^2}\left[\sigma_1^2 m_2-\sigma_2^2 m_1 \pm \sigma_1\sigma_2\sqrt{(m_2-m_1)^2+2(\sigma_2^2-\sigma_1^2)\ln\frac{p\sigma_2}{q\sigma_1}}\right]$$

(3.2.4) has two possible roots. There are three ways for this to occur:

(1) when there are no roots,

(2) when no roots fall between the centroid, and

(3) when one and only one root falls between the centroids.

If a root should fall at a centroid, this may be considered as a limiting case of situation (2). Situation (1) is trivial; all individuals are classified into one population. In situation (2), linear discrimination is not very helpful; quadratic discrimination is indicated. In these situations, probably (depending on parameters) there is no discrimination which will be much improvement over classification of all individuals into one population. Thus, situation (3) will be considered in this paper. If $p = q = \frac{1}{2}$, situations (1), (2) cannot occur.

When a root falls between the centroids, this is the root which minimizes $P^* = P_I^* + P_{II}^*$, and is therefore the root desired

for discrimination by critorion (i). The other root maximizes an expected error and therefore will not be used. If $\sigma_2 < \sigma_1$, the positive root is the root desired; if $\sigma_2 > \sigma_1$, the negative root is the root desired.

On level (c) of (1.1), c(min) is identical with (3.2.5) except for the absence of p and q.

It may be noted that the classificatory rule could be reversed from its present statement, but that would make $P_I > \frac{1}{2}$ and $P_{II} > \frac{1}{2}$, and will therefore not be considered further.

It should be noted that if $\sigma_1 = \sigma_2$, c(min) reduces to

$$\frac{m_1 + m_2}{2} = c(m), \tag{3.2.6}$$

designated as c(m) since it depends only on m_1 and m_2. This c(m) is the c introduced for samples by Barnard [2] and Fisher [6], and currently used in all linear discriminant analysis.

When $\sigma_1 = \sigma_2$, c(min) is the root between the centroids. The other root goes to infinity.

3.3. Evaluation of c for level (a) of (1.1), criterion (ii)

If we locate the hyperplane such that $R_I = R_{II}$, we have a c optimum in sense (ii), since R_I is monotonically decreasing and R_{II} is monotonically increasing in c. In this case

$$p \int_{AX' > c} \cdots \int W_{I,II}(X) f_I(X) dX = q \int_{AX' \leq c} \cdots \int W_{II,I}(X) f_{II}(X) dX. \tag{3.3.1}$$

An explicit c may not be found in general. The previously mentioned special cases will be considered later.

3.4. Evaluation of c for levels (b) and (c) of (1.1), criterion (ii)

P_I^* and P_{II}^* are also monotonic, decreasing and increasing respectively, in c and therefore if the hyperplane is located such that $P_I^* = P_{II}^*$, we will obtain the c for criterion (ii). Here

$$(3.4.1) \qquad p \int_{-\infty}^{\frac{m_1 - c}{\sigma_1}} N(0,1)dx = q \int_{-\infty}^{\frac{c - m_2}{\sigma_2}} N(0,1)dx \ .$$

Again an explicit result for c may not be found in general, since the integrals have not been evaluated explicitly. If $p = q$, we have the integrals identical except for upper limits of integration, and equation (3.4.1) reduces to

$$(3.4.2) \qquad \frac{m_1 - c}{\sigma_1} = \frac{c - m_2}{\sigma_2} \ .$$

Designate the c satisfying (3.4.2) as $c(\sigma)$ since it places c proportionally to the standard deviations; then

$$(3.4.3) \qquad c(\sigma) = \frac{m_1 \sigma_2 + m_2 \sigma_1}{\sigma_1 + \sigma_2} \ .$$

As in the case of c(min), if $\sigma_1 = \sigma_2$, $c(\sigma)$ reduces to

$$c(m) = \frac{m_1 + m_2}{2} \ .$$

On the (c) level, $c(\sigma)$ is identical with (3.4.3).

3.5. Evaluation of c for special cases of loss functions

If loss functions should be any of the special cases mentioned in Section (1.2), they lead to the parameters Σ_1^*, Σ_2^*, M_1^*, M_2^*, and A as defined implicitly in Section (2.7). Then c(min), $c(\bar{\sigma})$, and c(m) are obtained from (3.2.5), (3.4.3), and (3.2.6) respectively by the following substitutions:

(3.5.1) $\bar{\sigma}_u^{*2} = A\Sigma_u^* A'$ for $\bar{\sigma}_u^2$, $u = 1,2$,

(3.5.2) $m_u^* = AM_u^{*\prime}$ for m_u, $u = 1,2$,

(3.5.3) $b_1^* p$ for p,

and

(3.5.4) $b_2^* q$ for q .

IV. COMPARISON OF DISCRIMINATORS AT LEVEL (c)

4.1 Introduction

Under certain circumstances linear discrimination does not yield good results; an example of this is the situation in which the centroids of the two populations are the same. Any description of the conditions necessary for linear discrimination to be able to lead to reasonable results must be, to some extent, arbitrary. Generally the situations in which linear discrimination may be rejected are typified by c(min) not contained in (m_1, m_2) when the A optimum under criterion (i) is used. Consequently, as mentioned in (3.2), we shall consider that linear discrimination is appropriate only when c(min) is contained in (m_1, m_2).

There are two criteria by which discriminators may be judged on level (c) of (1.1), namely

(i) $P = P_I + P_{II}$

(ii) $P_L = \max(P_I, P_{II})$

where P_I is the probability of an individual being classified into population II given it arose from population I and P_{II} is the probability of an individual being classified into I given it arose from II.

Thus, if we want to compare two linear discriminators, it can happen either that one has both criteria less than or equal to

those of the other or that the above does not occur. If the former holds, then the discriminator with the smaller criteria may be said to be better than the other. This is true whether the discrimination is linear or not.

The linear discriminators are functions of A and c. The A's for both criteria were the same (where exact results were obtained) and therefore do not effect comparisons. Further, the A's are analogies of those used by other authors in sample studies. Consequently, we shall not consider discriminators using other A's .

However, the c's obtained for the criteria are different so that comparisons of discriminators using different c's will be useful. The comparisons will be referred to as comparisons of the c's, although they involve A, since A will be the same for the different discriminators. The restriction to level (c), together with the constant A, enables us to keep the number of parameters down to two for comparisons of the discriminators $AX' = c(\min)$, $AX' = c(\sigma)$, and $AX' = c(m)$, which involve $c(\min)$, $c(\sigma)$, and $c(m)$, respectively.

$c(\min)$ and $c(\sigma)$ are the c's derived for the two criteria; both reduce to $c(m)$ in the special case of equal dispersion matrices. Further, $c(m)$ is the population analog of the c used in practice and is easier to compute than are $c(\min)$ and $c(\sigma)$. Thus, although the same comparisons could be made using other functions for c, they do not appear as worthwhile as comparisons using $c(\min)$, $c(\sigma)$, and $c(m)$. Since $c(\min)$ and $c(\sigma)$ each satisfy a criterion, the comparisons will be to find for what ranges of the two essential parameters does $c(m)$

lead to both a smaller sum of conditional probabilities of misclassification than does $c(\sigma)$ and a smaller maximum conditional probability of misclassification than does c(min). If this never occurs when a linear discriminator is appropriate, then we may say that c(m) is of no use since one of c(min), $c(\sigma)$ have both criteria smaller than those of c(m).

We lose no generality if we let $m_2 > m_1$, and $\sigma_2 > \sigma_1$. The designation of the population having the larger standard deviation as population II is arbitrary. We may then multiply by ± 1, whichever is necessary to obtain $m_2 > m_1$.

4.2 Definition of parameters

It was mentioned that the functions in our comparisons reduce to two parameters; these parameters are defined as:

$$K = \frac{m_2 - m_1}{\sigma_1 + \sigma_2} \tag{4.2.1}$$

and

$$\alpha = \frac{\sigma_2}{\sigma_1} \tag{4.2.2}$$

We may restrict $K > 0$ and $\alpha \geq 1$. When results in K and α are tabulated, the tables are symmetric K, -K and $\log \alpha$, $-\log \alpha$.

It was noted in Sections (3.2) and (4.1) that only when one root of c(min) lies between the centroids is a linear discriminator appropriate; otherwise a quadratic discriminator is appropriate.

Thus when

(4.2.3) $\quad c(\min) < m_1 ,$

quadratic discrimination is appropriate and linear discrimination will not be considered. (4.2.3) may be written $m_1 - c(\min) > 0$ and expanded to obtain

$$\frac{\sigma_1^2 (m_1-m_2) + \sigma_1\sigma_2 \sqrt{(m_2-m_1)^2 + 2(\sigma_2^2-\sigma_1^2)\ln \sigma_2/\sigma_1}}{\sigma_1^2 - \sigma_2^2} > 0$$

$$\frac{\sigma_1^2}{\sigma_2-\sigma_1}\left(\frac{m_2-m_1}{\sigma_1+\sigma_2}\right) + \frac{(\sigma_1+\sigma_2)\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2}\sqrt{\left(\frac{m_2-m_1}{\sigma_1+\sigma_2}\right)^2 + 2\,\frac{\sigma_2-\sigma_1}{\sigma_2+\sigma_1}\ln \sigma_2/\sigma_1} > 0$$

$$K - \alpha\sqrt{K^2 + 2\,\frac{\alpha-1}{\alpha+1}\ln\alpha} > 0$$

$$K^2 - \alpha^2\left(K^2 + 2\,\frac{\alpha-1}{\alpha+1}\ln\alpha\right) > 0$$

(4.2.4) $\quad \alpha^2(\alpha+1)^{-2}\ln\alpha^2 - K^2 > 0$

Fig. 1 exhibits the region in (K,α) for which this inequality holds.

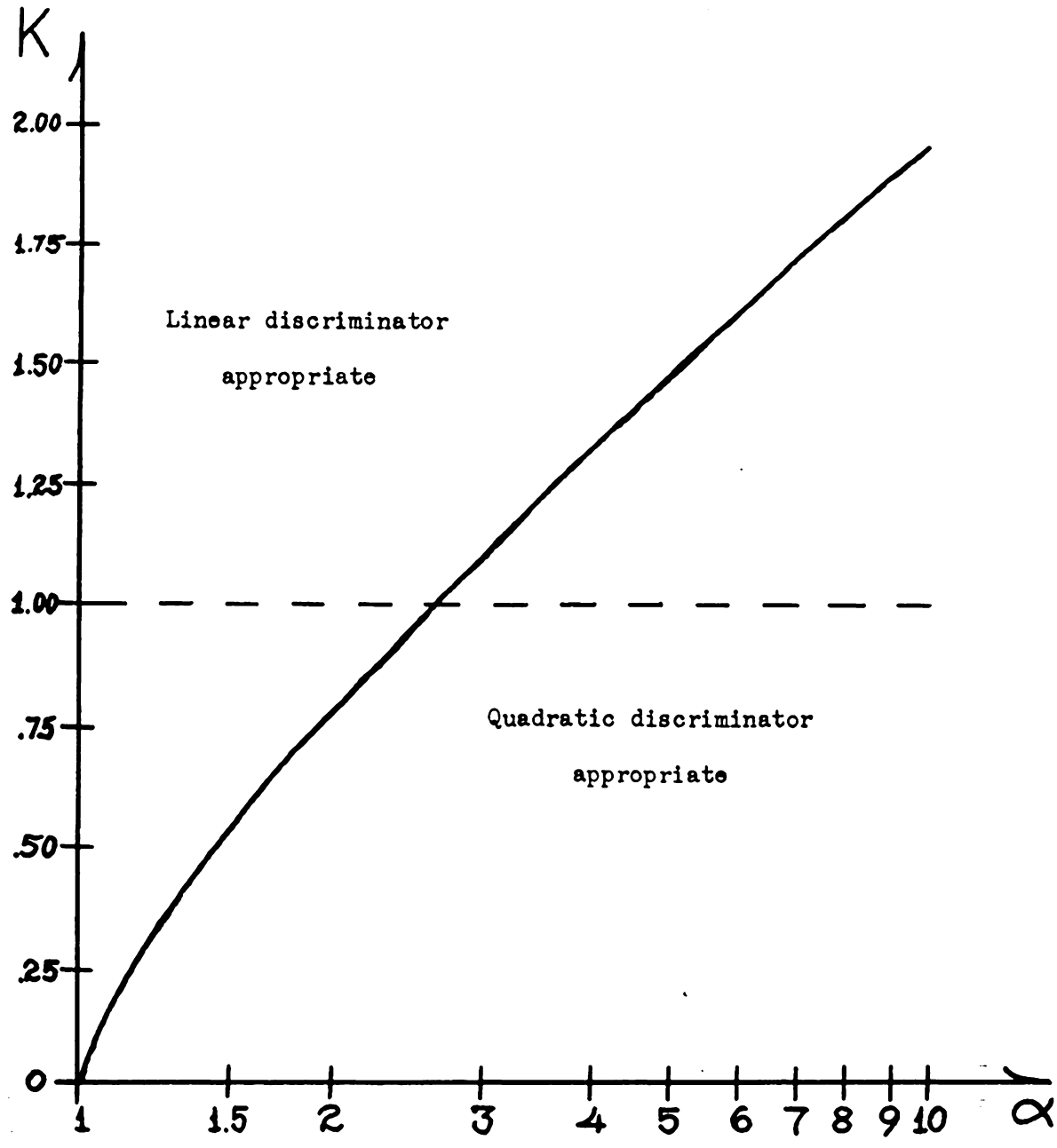


FIG. 1

$$K^2 = \alpha^2(\alpha + 1)^{-2} \ln \alpha^2$$

4.3 Sum of conditional probabilities of misclassification using c(m), c(σ)

Denote the sum of conditional probabilities of misclassification using c(m), c(σ) by P(m), P(σ) respectively.

Then

$$P(m) = \int_{-\infty}^{\frac{m_1-c(m)}{\sigma_1}} N(0,1)dx + \int_{-\infty}^{\frac{c(m)-m_2}{\sigma_2}} N(0,1)dx \quad \text{(from 1.2.11 and 1.2.12)} \tag{4.3.1}$$

and

$$P(\sigma) = \int_{-\infty}^{\frac{m_1-c(\sigma)}{\sigma_1}} N(0,1)dx + \int_{-\infty}^{\frac{c(\sigma)-m_2}{\sigma_2}} N(0,1)dx \quad \text{(from 1.2.11 and 1.2.12)} \tag{4.3.2}$$

Now

$$c(m) = \frac{m_1+m_2}{2}, \quad \text{(from 3.2.6)}$$

$$c(\sigma) = \frac{m_1\sigma_2 + m_2\sigma_1}{\sigma_1 + \sigma_2}, \quad \text{(from 3.4.3)}$$

$$K = \frac{m_2-m_1}{\sigma_1+\sigma_2}, \quad \text{(from 4.2.1)}$$

and

$$= \frac{\sigma_2}{\sigma_1}. \quad \text{(from 4.2.2)}$$

Hence

$$\frac{m_1 - c(m)}{\sigma_1} = \frac{m_1 - m_2}{2\sigma_1} = \frac{m_1 - m_2}{\sigma_1 + \sigma_2}\left(\frac{\sigma_1 + \sigma_2}{2\sigma_1}\right) = \frac{-K(\alpha + 1)}{2}.$$

Similarly,

$$(4.3.4) \qquad \frac{c(m) - m_2}{\sigma_2} = \frac{-K(\alpha + 1)}{2\alpha} \quad .$$

Also

$$(4.3.5) \qquad \frac{m_1 - c(\sigma)}{\sigma_1} = \frac{m_1 - m_2}{\sigma_1 + \sigma_2} = -K \quad .$$

Similarly

$$(4.3.6) \qquad \frac{c(\sigma) - m_2}{\sigma_2} = -K \quad .$$

Substituting (4.3.3) and (4.3.4) into (4.3.1) and (4.3.5) and (4.3.6) into (4.3.2), we obtain

$$(4.3.7) \qquad P(m) = \int_{-\infty}^{\frac{-K(\alpha+1)}{2}} N(0,1)dx + \int_{-\infty}^{\frac{-K(\alpha+1)}{2\alpha}} N(0,1)dx$$

and

$$(4.3.8) \qquad P(\sigma) = 2\int_{-\infty}^{-K} N(0,1)dx \quad .$$

Now when

$$(4.3.9) \qquad P(m) - P(\sigma) > 0$$

$c(\sigma)$ yields a lesser sum of conditional probabilities of misclassification than does $c(m)$. When the inequality is reversed, $c(m)$ yields a lesser sum of conditional probabilities of misclassification than does $c(\sigma)$.

Let

$$(4.3.10) \qquad g_1(K, \alpha) = P(m) - P(\sigma).$$

Then when g_1 is positive, $c(\sigma)$ is the better c. For computational convenience the identity

$$(4.3.11) \qquad \int_{-\infty}^{a} N(0,1)dx \equiv \int_{a}^{\infty} N(0,1)dx \equiv \tfrac{1}{2} - \int_{0}^{a} N(0,1)dx$$

was used to put $g_1(K, \alpha)$ into the form

$$(4.3.12) \qquad g_1(K,\alpha) = 2\int_{0}^{K} N(0,1)dx - \int_{0}^{\frac{K(\alpha+1)}{2}} N(0,1)dx - \int_{0}^{\frac{K(\alpha+1)}{2\alpha}} N(0,1)dx.$$

The value of $g_1(K,\alpha)$ was tabulated for several values of K and α. This table appears as Table 1. A figure, Fig. 2, shows $g_1(K, \alpha)$ plotted against K for various α. Note that whenever a curve lies above the zero point on the ordinate, $c(\sigma)$ is the better c, since it then gives a smaller sum of conditional probabilities of misclassification than does c(m) and by definition a smaller maximum conditional probability of misclassification than does c(m). Fig. 3 exhibits regions of (K,α) in which the sum of conditional probabilities of misclassification using $AX' = c(\sigma)$ is greater, less than those using $AX' = c(m)$.

TABLE 1

DIFFERENCE IN SUM OF CONDITIONAL PROBABILITIES OF MISCLASSIFICATION FROM USING $c(m)$ OVER $c(\sigma')$ BASED ON PARAMETERS K AND α.

K \ α	1	2	4	10	∞
0	0	0	0	0	0
0.1	0	- .010	- .044	- .151	- .440
0.2	0	- .019	- .083	- .250	- .381
0.5	0	- .037	- .133	- .221	- .216
1.0	0	- .024	- .045	- .026	- .009
1.1	0	- .017	- .023	+ .002	+ .020
1.2	0	- .010	- .002	.025	.044
1.4	0	+ .003	+ .031	.059	.080
1.6	0	.014	. 049	.080	.102
1.8	0	.020	.059	.089	.112
2.0	0	.023	.060	.090	.113
3.0	0	.010	.028	.047	.064
4.0	0	.001	.006	.014	.023
∞	0	0	0	0	0

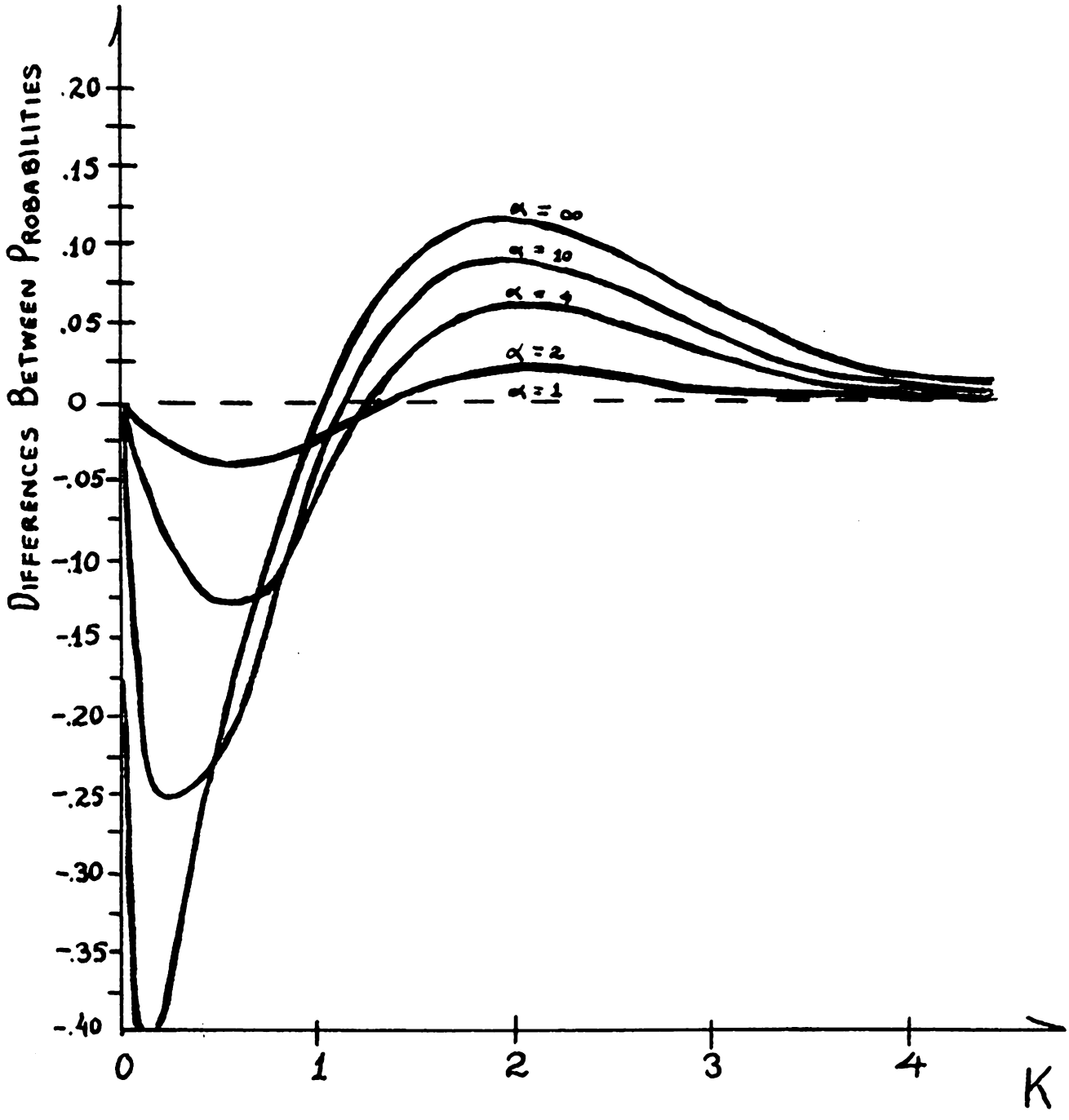


FIG. 2

Differences between conditional probabilities of misclassification from using c(m) over c(σ) expressed as a function of K and α.

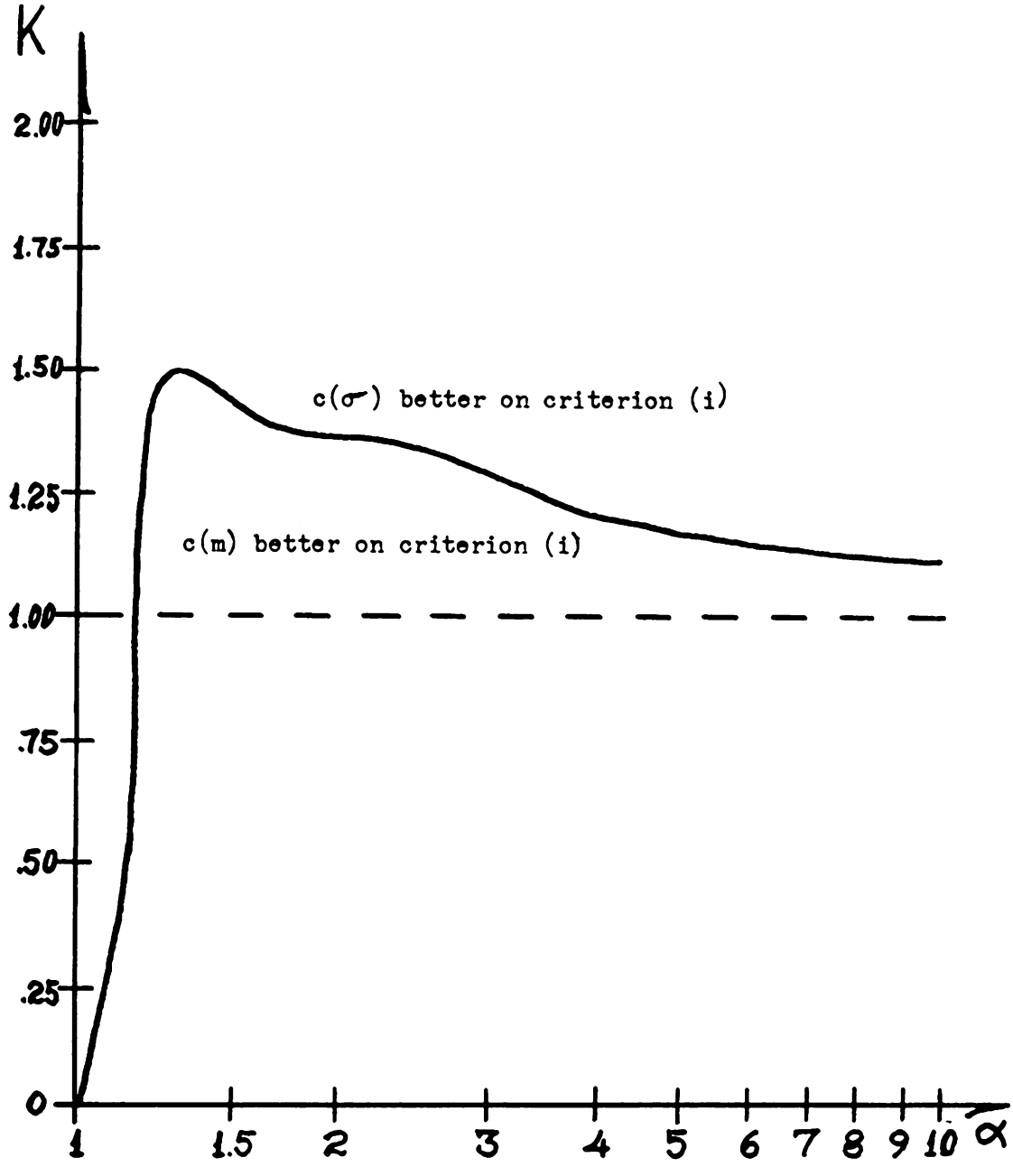


FIG. 3

Regions of (K, α) in which the sum of conditional probabilities of misclassification using $AX' = c(\sigma)$ is greater, less than those using $AX' = c(m)$.

4.4. Maximum conditional probability of misclassification using c(m), c(min)

Denote the larger conditional probability of misclassification using c(m), c(min) by $P_L(m)$, $P_L(\min)$, respectively.

Now $c(\sigma)$ is the point on either side of which the probabilities of misclassification are equal, so that a $c < c(\sigma)$ indicates $P_I = P_L$ and a $c > c(\sigma)$ indicates $P_{II} = P_L$.

We assumed in Section (4.1) with no loss of generality that $m_2 > m_1, \sigma_2 > \sigma_1$. Then

$$m_2 > m_1$$

$$\equiv \sigma_2 m_2 - \sigma_1 m_2 > \sigma_2 m_1 - \sigma_1 m_1$$

$$\equiv \frac{m_1 + m_2}{2} > \frac{\sigma_1 m_2 + \sigma_2 m_1}{\sigma_1 + \sigma_2}$$

$$\equiv c(m) > c(\sigma). \tag{4.4.1}$$

Therefore

$$P_L(m) = P_{II}(m) = \int_{-\infty}^{\frac{c(m)-m_2}{\sigma_2}} N(0,1)\,dx = \int_{-\infty}^{\frac{-K(\alpha+1)}{2\alpha}} N(0,1)\,dx \text{ (from 4.3.4)} \tag{4.4.2}$$

In a similar manner,

$$\sqrt{(m_2-m_1)^2 + 2(\sigma_2^2-\sigma_1^2)\ln \sigma_2/\sigma_1} > (m_2-m_1)$$

$$\equiv -\sqrt{(m_2-m_1)^2 + 2(\sigma_2^2-\sigma_1^2)\ln \sigma_2/\sigma_1} < (m_1-m_2)$$

$$\equiv \sigma_1^2 m_2 - \sigma_2^2 m_1 - \sigma_1\sigma_2\sqrt{(m_2-m_1)^2 + 2(\sigma_2^2-\sigma_1^2)\ln \sigma_2/\sigma_1} <$$

$$\frac{(m_1\sigma_2+m_2\sigma_1)(\sigma_1-\sigma_2)(\sigma_1+\sigma_2)}{\sigma_1+\sigma_2}$$

$$\equiv \frac{1}{\sigma_1^2-\sigma_2^2}\left[\sigma_1^2 m_2 - \sigma_2^2 m_1 - \sigma_1\sigma_2\sqrt{(m_2-m_1)+2(\sigma_2^2-\sigma_1^2)\ln\sigma_2/\sigma_1}\right] > \frac{m_1\sigma_2+m_2\sigma_1}{\sigma_1+\sigma_2}$$

$$\equiv c(\min) > c(\sigma). \tag{4.4.3}$$

Therefore

$$(4.4.4) \qquad P_L(\min) = P_{II}(\min) = \int_{-\infty}^{\frac{c(\min)-m_2}{\sigma_2}} N(0;1)\,dx \ ,$$

Now

$$\frac{c(\min) - m_2}{\sigma_2} = \frac{\sigma_1^2 m_2 - \sigma_2^2 m_1 - m_2\sigma_1^2 + m_2\sigma_2^2 - \sigma_1\sigma_2\sqrt{(m_2-m_1)^2 + 2(\sigma_2^2-\sigma_1^2)\ln \sigma_2/\sigma_1}}{\sigma_2(\sigma_1^2 - \sigma_2^2)}$$

$$= \frac{(m_2-m_1)\,\sigma_2^2}{(\sigma_1+\sigma_2)(\sigma_1-\sigma_2)\sigma_2} - \frac{\sigma_1\sqrt{\left(\frac{m_2-m_1}{\sigma_1+\sigma_2}\right)^2(\sigma_1+\sigma_2)^2 + 2(\sigma_1+\sigma_2)^2\left(\frac{\sigma_2-\sigma_1}{\sigma_2+\sigma_1}\right)\ln\frac{\sigma_2}{\sigma_1}}}{\sigma_1^2 - \sigma_2^2}$$

$$= K\frac{\sigma_2}{\sigma_1+\sigma_2} - \frac{\sigma_1}{\sigma_1-\sigma_2}\sqrt{K^2 + 2\,\frac{\sigma_2+\sigma_1}{\sigma_2+\sigma_1}\,\ln \sigma_2/\sigma_1} \qquad \text{(from 4.2.1)}$$

$$(4.4.5) \qquad = \frac{K\alpha - \sqrt{K^2 + 2\,\frac{\alpha-1}{\alpha+1}\ln\alpha}}{1-\alpha} \qquad \text{(from 4.2.2)}$$

Therefore

$$(4.4.6) \qquad R_L(\min) = \int_{-\infty}^{-\frac{K\alpha - \sqrt{K^2 + 2\,\frac{\alpha-1}{\alpha+1}\,\ln\alpha}}{\alpha-1}} N(0,1)\,dx \ .$$

Now let us define

$$g_2(K,\alpha) = P_L(m) - P_L(\min)$$

$$(4.4.7) \qquad = \int_{-\infty}^{-\frac{K(\alpha+1)}{2\alpha}} N(0,1)dx - \int_{-\infty}^{-\frac{K\alpha - \sqrt{K^2 + 2\frac{\alpha-1}{\alpha+1}\ln\alpha}}{\alpha-1}} N(0,1)dx.$$

When g_2 is positive, $P_L(m) > P_L(\min)$, and c(min) is better.

$g_2(K,\alpha)$ has the same sign as the inequality

$$(4.4.8) \qquad -\frac{K(\alpha+1)}{2\alpha} + \frac{K\alpha - \sqrt{K^2 + 2\frac{\alpha-1}{\alpha+1}\ln\alpha}}{\alpha - 1} > 0$$

$$\equiv \frac{K^2(1+\alpha^2)^2}{4\alpha^2} > K^2 + 2\frac{\alpha-1}{\alpha+1}\ln\alpha$$

$$\equiv K^2(\alpha^4 + 2\alpha^2 + 1) - 4\alpha^2K^2 - 8\alpha^2\frac{\alpha-1}{\alpha+1}\ln\alpha > 0$$

$$(4.4.9) \qquad \equiv K^2(\alpha^2 - 1) - \frac{8\alpha^2\ln\alpha}{(\alpha+1)^2} > 0.$$

It follows that whenever (4.4.9) is satisfied, c(min) is better than c(m).

Fig. 4 shows regions of (K,α) in which the larger probability of misclassification using AX' = c(min) is greater, less than those using AX' = c(m).

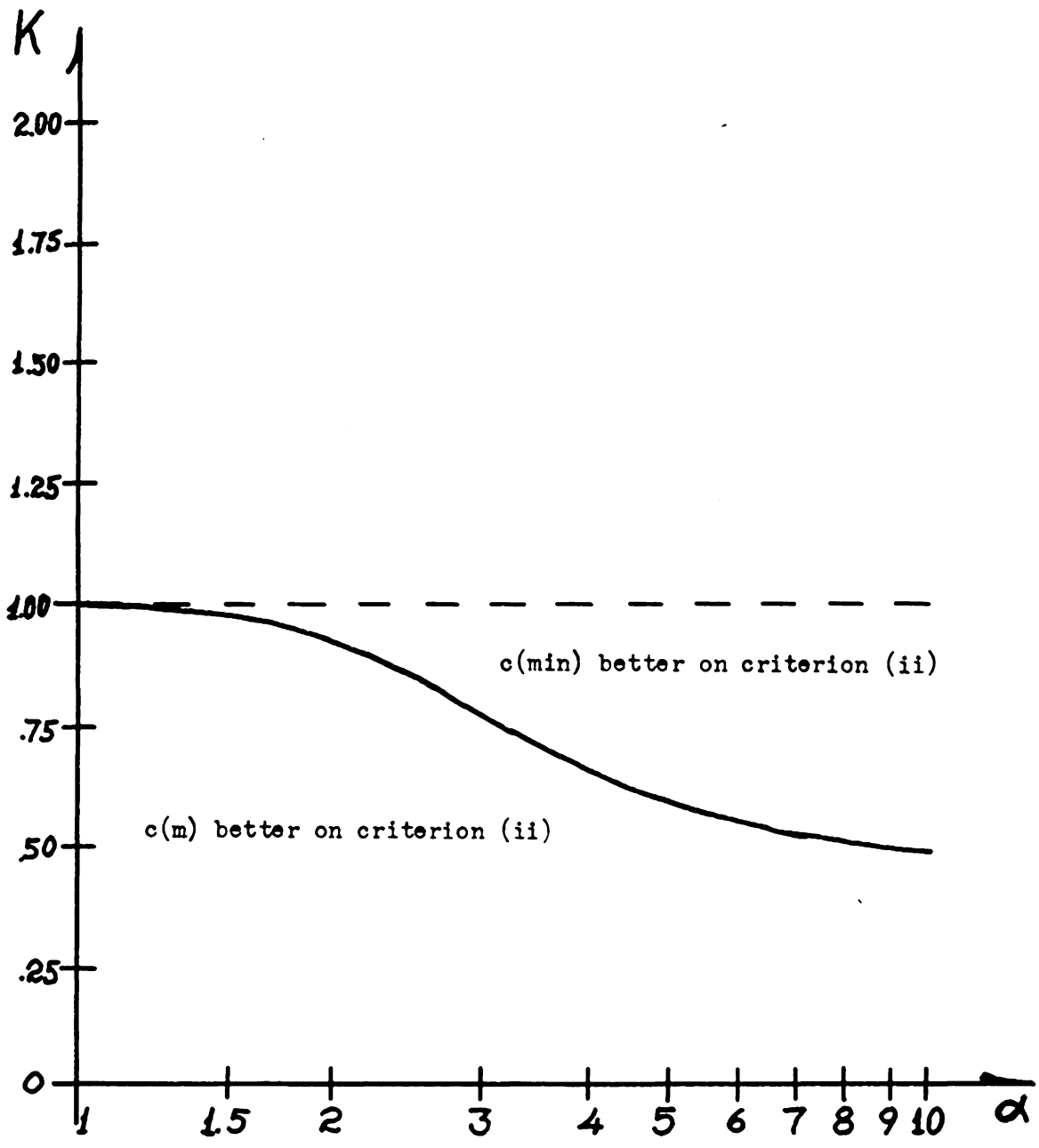


FIG. 4

Regions of (K, α) in which the larger conditional probability of misclassification using $AX' = c(\min)$ is greater, less than those using $AX' = c(m)$.

4.5. Combination of Figs. 2, 3, and 4

Fig. 5 combines Figs. 2, 3, and 4 on the same set of axes. By inspection of the curves in Fig. 5 it can be seen that c(m) is better than c(min) for a small region of (K, α). Within this region, represented by the shaded area in Fig. 5, c(m) may be considered as a compromise between c(min) and $c(\sigma)$, but elsewhere at least one is better on both criteria than is c(m) when linear discrimination is appropriate.

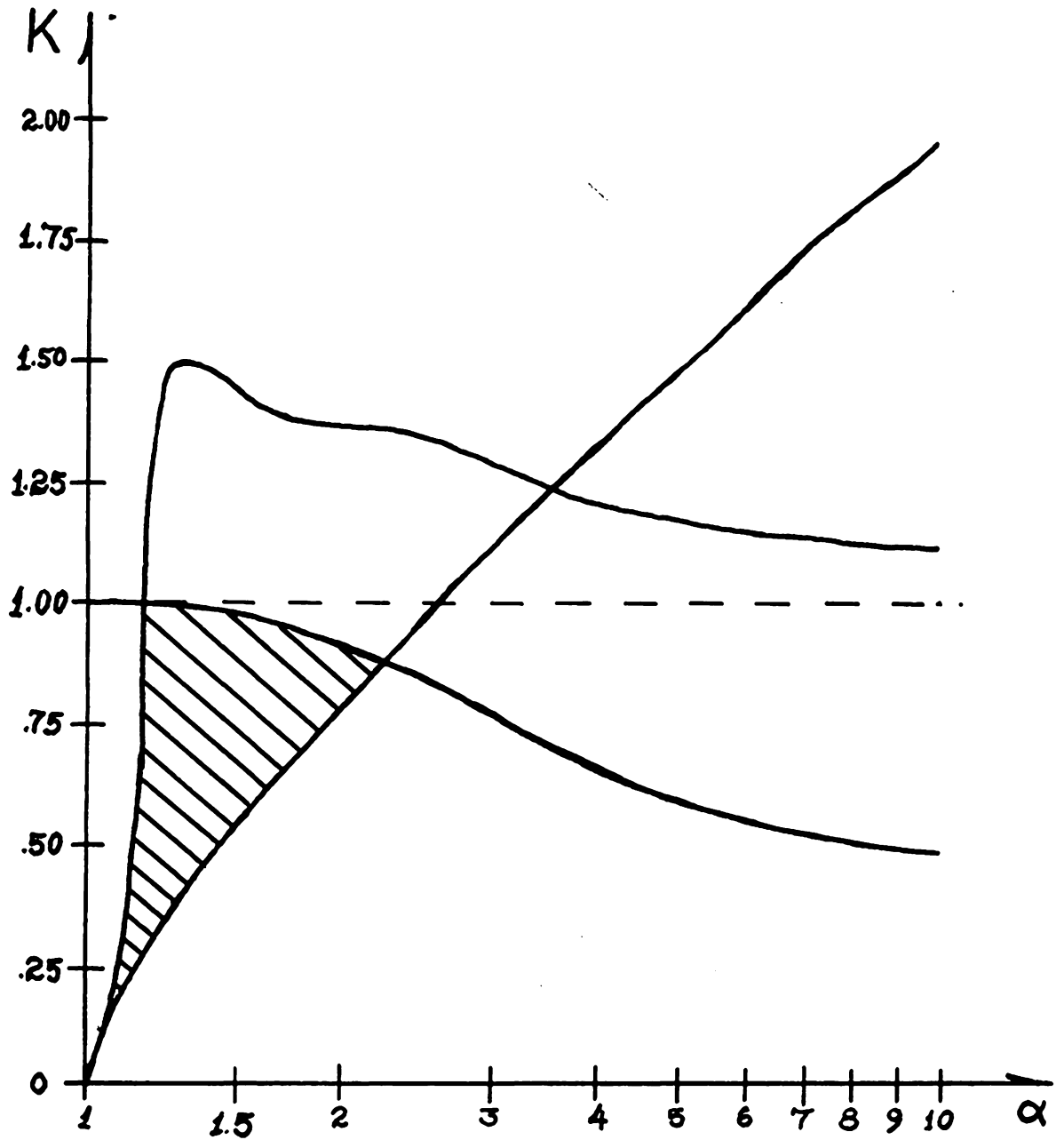


FIG. 5

Combination of Figures 1, 3, and 4.

4.6 A more general comparison of discriminators

To compare the differences in performance of any two discriminators, the difference of their classificatory errors is a reasonable and obvious criterion. Note that any two discriminators, of whatever type and for whatever distributions (the discriminators may be of different types), may be compared in this manner.

Let r and r' represent two classificatory rules, r_i being the region where an individual from population i is misclassified, i = I,II; then the conditional probability of misclassification over population i using r is given by

$$\int_{r_i} f_i(x)dX \tag{4.6.1}$$

and using r' is given by

$$\int_{r'_i} f_i(x)dX, \tag{4.6.2}$$

where $f_i(x)$ is the density of population i. Define the quantity H(r',r) to be the difference in sums of conditional probabilities of misclassification when the classification is based on r',r.

$$H(r',r) = \int_{r_I} f_I dX - \int_{r'_I} f_I dX + \int_{r_{II}} f_{II} dX - \int_{r'_{II}} f_{II} dX. \tag{4.6.3}$$

Note that $r_I + r_{II} = r'_I + r'_{II}$ = the entire k-space. If H is negative, it represents an increase in the sum of conditional probabilities of misclassification by using r' rather than r. If H is positive, it represents a decrease in the conditional probabilities of misclassification by using r' rather than r. When H is positive, we may say that r' is the better classificatory rule under criterion (i).

If f_i should have a bivariate normal distribution, and if we should devise the classificatory rule using likelihood ratios, then r is a conic section such that, if r' is any other region, r is better than r'. Comparison of the best linear discriminator with the best quadratic discriminator would involve integrating a bivariate normal over an area defined by the conic section. A simplification of this integral to a single integral with a polynomial as the exponent of the integrand and simple limits of integration is given in the Appendix (10.1).

To compare r',r under criterion (ii), we need to determine the largest of:

$$\int_{r_I} f_I dX, \quad \int_{r_{II}} f_{II} dX, \quad \int_{r'_I} f_I dX, \quad \int_{r'_{II}} f_{II} dX,$$

If one of the first two is the largest, then r' is better than r. If one of the second two is the largest, then r is better than r'. A comparison of the best linear discriminator with the best quadratic discriminator would again involve integrating a bivariate normal over an area defined by conic section. The simplification given in (10.1) will again be useful.

V. LARGE SAMPLE THEORY

5.1. Introduction

When the parameters of the populations are known, the conditional probabilities of misclassification of an individual from either population may be given exactly. In Chapters II and III, the coefficients of the linear discriminators were derived as functions of these parameters. If, however, the parameters are not known, they must be estimated, and sampling fluctuations affect the probabilities of misclassification. This chapter considers (a) large sample fluctuations of estimates of population parameters, (b) the effect these fluctuations have on the functions of these estimates involved in the discriminators, and, hence, (c) the resulting effects on the conditional probabilities of misclassification.

There is one sample of individuals from each population; the individuals composing each sample are well identified, i.e. there is no misclassification within these samples. Suppose we use maximum likelihood estimates of the parameters based on these samples; then as each sample size tends to infinity, the estimates converge stochastically to the parameters. If we use the coefficients of the linear discriminators which were derived for known parameters with the estimates replacing the parameters, then the conditional

probabilities of misclassification are the same as in the population case, by Slutsky's Theorem (5).

When the parameters have to be estimated for fixed finite samples, it is necessary to consider linear discriminators as functions of the samples. If we consider the discriminator to be a hyperplane, then we may say that it would be desirable to look for $\hat{c}$ and $\hat{a}_i$, i=1,...,k, which have, in some sense, optimum sampling distributions; this sense in which the sampling distributions are optimum may be one of several and need not be specified at the moment. Consider the sampling distributions of the statistics $\hat{c}$ and $\hat{a}_i$, i=1,...,k. Since the $\hat{c}$ and $\hat{a}_i$ are functions of the sample values, they have a joint sampling distribution. Technically speaking, it is possible to obtain this joint sampling distribution from knowledge of the parent distributions from which these samples were drawn.

Now

(5.1.1) $$\hat{a}_i = \sum_j \hat{d}_j \hat{\sigma}^{ij},$$

where (1)

(5.1.2) $$\hat{d}_j = \hat{\mu}_{jII} - \hat{\mu}_{jI},$$

and (2)

(5.1.3) $\hat{\sigma}^{ij}$ is the i,jth element of the inverse of the matrix with elements $\hat{\sigma}_{ij} = \hat{\sigma}_{ijI} + \hat{\sigma}_{ijII}$,

where $\hat{\mu}_{jI}$ and $\hat{\mu}_{jII}$ are maximum likelihood estimates of the population means for the jth variable, μ_{jI} and μ_{jII} respectively, $\hat{\sigma}_{ijI}$ and $\hat{\sigma}_{ijII}$ are maximum liklihood estimates of the i, jth covariances of the populations, σ_{ijI} and σ_{ijII} respectively, and the form

$\hat{\sigma}_{ij} = \hat{\sigma}_{ijI} + \hat{\sigma}_{ijII}$ is used rather than either $\hat{\sigma}_{ijI}$ or $\hat{\sigma}_{ijII}$ in order to make use of information from both samples. The $\hat{\mu}_{jI}$ and $\hat{\mu}_{jII}$ have multivariate normal distributions and the $\hat{\sigma}_{iiI}$ and $\hat{\sigma}_{iiII}$ have chi-square distributions.

Further

$$\hat{c}(m) = \frac{\hat{m}_1 + \hat{m}_2}{2}, \tag{5.1.4}$$

$$\hat{c}(\sigma) = \frac{\hat{\sigma}_1\hat{m}_2 + \hat{\sigma}_2\hat{m}_1}{\hat{\sigma}_1 + \hat{\sigma}_2}, \tag{5.1.5}$$

$$\hat{c}(\min) = \frac{1}{\hat{\sigma}_1^2 - \hat{\sigma}_2^2}\left[\hat{\sigma}_1^2\hat{m}_2 - \hat{\sigma}_2^2\hat{m}_1 - \hat{\sigma}_1\hat{\sigma}_2\sqrt{(\hat{m}_2-\hat{m}_1)^2 + 2(\hat{\sigma}_2^2-\hat{\sigma}_1^2)\ln\hat{\sigma}_2/\hat{\sigma}_1}\right], \tag{5.1.6}$$

$$\hat{m}_1 = \sum_i \hat{a}_i\hat{\mu}_{iI}, \tag{5.1.7}$$

$$\hat{m}_2 = \sum_i \hat{a}_i\hat{\mu}_{iII}, \tag{5.1.8}$$

$$\hat{\sigma}_1^2 = \sum_{i,j} \hat{a}_i\hat{\sigma}_{ijI}\hat{a}_j, \tag{5.1.9}$$

and

$$\hat{\sigma}_2^2 = \sum_{i,j} \hat{a}_i\hat{\sigma}_{ijII}\hat{a}_j. \tag{5.1.10}$$

If Σ_1 is known to be equal to Σ_2, some simplifications occur in that certain statistics are proportional to Mahalanobis' D^2, which has an F distribution. For general Σ_1 and Σ_2, (a) the exact parametric forms for the linear discriminator have not been derived, and (b) the joint sampling distributions of the suggested $\hat{a}_i$ will become complicated.

For general $\Sigma_1 = \alpha^2\Sigma_2, \alpha \neq 1$, the joint sampling distribution still becomes complicated. Consequently, we shall consider the sampling distributions (represented by the first two moments when necessary) of the statistics involved in the limits of integration

of the standard normal integrals which give the conditional probabilities of misclassification, P_I and P_{II}. We will restrict ourselves subsequently to proportional dispersion matrices.

Note that the integrals contain two types of functions of estimates, one containing only estimates, which we shall designate with a caret, and the other containing a mixture of estimates and parameters, which we shall designate with an inverted caret. For example, $\hat{m}_1 = \sum_i \hat{a}_i \hat{\mu}_{iI}$, but $\check{m}_1 = \sum_i \hat{a}_i \mu_{iI}$.

$$(5.1.11) \qquad \hat{P}_I = \int_{-\infty}^{\frac{\check{m}_1 - \hat{c}}{\check{\sigma}_I}} N(0,1)dx \quad ,$$

where

$$(5.1.12) \qquad \check{m}_1 = \sum_{i=1}^{k} \hat{a}_i \mu_{iI} \, ,$$

$$(5.1.13) \qquad \check{\sigma}_1^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} \hat{a}_i \hat{a}_j \sigma_{ijI} \, ,$$

and

$$(5.1.14) \qquad \hat{a}_i = \sum_{j=1}^{k} \hat{d}_j \hat{\sigma}^{ij} \, , \qquad i,j=1,\ldots,k \, ;$$

and $\hat{c}$ may be any one of $c(m)$, $c(\sigma)$, or $c(\min)$ with maximum likelihood estimates substituted for all parameter values, which will be referred to as $\hat{c}(m)$, $\hat{c}(\sigma)$, or $\hat{c}(\min)$ respectively. Also

$$(5.1.15) \qquad \hat{P}_{II} = \int_{-\infty}^{\frac{\hat{c} - \check{m}_2}{\check{\sigma}_2}} N(0,1)dx \quad ,$$

where definitions are similar, differing only in subscripts.

Recall that for criterion (i), $R = R_I + R_{II}$, $P^* = P_I^* + P_{II}^*$, $P = P_I + P_{II}$, and for criterion (ii), $R_L = \max(R_I, R_{II})$, $P_L^* = \max(P_I^*, P_{II}^*)$, $P_L = \max(P_I, P_{II})$, all defined in (1.2). Define $P = P_I + P_{II}$, where P_I and P_{II} are defined by (5.1.11) and (5.1.14). Similarly, define $\hat{P}_L = \max(\hat{P}_I, \hat{P}_{II})$. $\hat{P}_I(m)$ is $\hat{P}_I$ (5.1.11) with $\hat{c}$ taken to be $\hat{c}(m)$. $\hat{P}_{II}(m)$, $\hat{P}(m)$, and $\hat{P}_L(m)$ are defined correspondingly. $\hat{P}_I(\sigma)$ is $\hat{P}_I$ (5.1.11) with $\hat{c}$ taken to be $\hat{c}(\sigma)$. $\hat{P}_{II}(\sigma)$, $\hat{P}(\sigma)$, and $\hat{P}_L(\sigma)$ are defined correspondingly. $\hat{P}_I(\min)$ is $\hat{P}_I$ (5.1.11) with $\hat{c}$ taken to be $\hat{c}(\min)$. $\hat{P}_{II}(\min)$, $\hat{P}(\min)$, and $\hat{P}_L(\min)$ are defined correspondingly.

As in the population case, no solution can be given for general loss functions. The special case of loss functions proportionate to the multivariate densities of the populations from which the samples are being drawn ((b) of (1.2)), may be given by

$$R_I = p\beta_1 \int\cdots\int_{AX' > c} \left[f_I(X)\right]^{t_I+1} dX, \tag{5.1.16}$$

$$R_{II} = q\beta_2 \int\cdots\int_{AX' \leq c} \left[f_{II}(X)\right]^{t_{II}+1} dX, \tag{5.1.17}$$

where β_1 and β_2 are constants of proportionality. If the loss functions are uniform ((c) of (1.2)), then

$$R_I = \gamma_1 P_I^* , \quad R_{II} = \gamma_2 P_{II}^* , \tag{5.1.18}$$

where γ_1 and γ_2 are the uniform loss functions.

Recall that

$$P_I^* = pP_I , \quad P_{II}^* = qP_{II} , \tag{5.1.19}$$

where p and q are the a priori probabilities as defined in (1.2).

Results for certain special loss functions and for expected errors will be similar to results for conditional probabilities of misclassification, differing only in multiplicative constants. We will look at sampling distributions only for the conditional probabilities of misclassification, from which we can obtain results for the cases just mentioned.

To obtain the first two moments of the sampling distributions of the limits of integration in $\hat{P}_I$ and $\hat{P}_{II}$, we need first the means and covariances of $\hat{\sigma}_{ijI}$ and $\hat{\sigma}_{ghI}$, of $\hat{\sigma}_{ijII}$ and $\hat{\sigma}_{ghII}$, of $\hat{\mu}_{iI}$ and $\hat{\mu}_{jI}$, and of $\hat{\mu}_{iII}$ and $\hat{\mu}_{jII}$.

5.2. Description of samples from which estimates are to be made

The observations in the sample from population I will be represented by $x_i(r)$, $i=1,\ldots,k$, $r=1,\ldots,n_1$. Thus i denotes the particular measurements on the individual and r denotes the particular individual of the sample. Similarly, the observations in the sample from population II will be represented by $y_{i(s)}$, $i=1,\ldots,k$, $s=1,\ldots,n_2$. $x_{i(r)}$ is independent of $x_{j(s)}$, $r\neq s$; $y_{i(r)}$ is independent of $y_{j(s)}$, $r\neq s$; and $x_{i(r)}$ is independent of $y_{j(s)}$, all i,j,r,s. The maximum likelihood estimate of μ_{iI} is $\hat{\mu}_{iI} = \bar{x}_{iI} = \frac{1}{n_1}\sum_{r=1}^{n_1} x_{i(r)}$ and of μ_{iII} is $\hat{\mu}_{iII} = \bar{y}_{iII} = \frac{1}{n_2}\sum_{s=1}^{n_2} y_{i(s)}$. The maximum likelihood estimate of σ_{ijI} is $\hat{\sigma}_{ijI} = \frac{1}{n_1}\sum_{r=1}^{n_1}(x_{i(r)} - \bar{x}_i)(x_{j(r)} - \bar{x}_j)$ and of σ_{ijII} is $\hat{\sigma}_{ijII} = \frac{1}{n_2}\sum_{s=1}^{n_2}(y_{i(s)} - y_i)(y_{j(s)} - \bar{y}_j)$.

Write the observations from the samples from populations I and II as X and Y respectively, where

$$(5.2.1) \qquad X = \begin{bmatrix} x_{1(1)} & x_{2(1)} & \cdots & x_{k(1)} \\ x_{1(2)} & x_{2(2)} & \cdots & x_{k(2)} \\ \vdots & \vdots & & \vdots \\ x_{1(n_1)} & x_{2(n_1)} & & x_{k(n_1)} \end{bmatrix}$$

and

$$(5.2.2) \qquad Y = \begin{bmatrix} y_{1(1)} & y_{2(1)} & \cdots & y_{k(1)} \\ y_{1(2)} & y_{2(2)} & \cdots & y_{k(2)} \\ \vdots & \vdots & & \vdots \\ y_{1(n_2)} & y_{2(n_2)} & & y_{k(n_2)} \end{bmatrix} .$$

5.3. A transformation of the samples

Consider an orthogonal transformation of the samples, specifically a multivariate application of the Helmert transformation. Define

$$(5.3.1) \qquad H_u = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & \cdots & & & & & \\ \vdots & \vdots & \vdots & & & & & & \\ \frac{1}{\sqrt{r(r-1)}} & \cdots & & \frac{1}{\sqrt{r(r-1)}} & \frac{-r}{\sqrt{r(r-1)}} & 0 & & & \\ \vdots & \vdots & \vdots & & & & & & \\ \frac{1}{\sqrt{n_u(n_u-1)}} & \cdots & & & & & \frac{1}{\sqrt{n_u(n_u-1)}} & \frac{-(n_u-1)}{\sqrt{n_u(n_u-1)}} \\ \frac{1}{\sqrt{n_u}} & \cdots & & & & & \frac{1}{\sqrt{n_u}} & \frac{1}{\sqrt{n_u}} \end{bmatrix},$$

$u = 1,2$.

Make the transformation

(5.3.2) $$\underset{(n_1 x n_1)(n_1 x k)}{H_1} \quad \underset{(n_1 x k)}{X} = \underset{(n_1 x k)}{V} ,$$

where the last row of V has elements $v_{i(n_1)} = \frac{1}{\sqrt{n_1}} \sum_{r=1}^{n_1} x_{i(r)}$, $i=1,\ldots,k$,

and the other rows have elements $v_{i(r)} = \frac{1}{\sqrt{r(r-1)}}\left[x_{i(1)}+\ldots+x_{i(r-1)}-rx_{i(r)}\right]$,

$i=1,\ldots,k$, $r=1,\ldots,n_1-1$. Make the transformation

(5.3.3) $$\underset{(n_2 x n_2)(n_2 x k)}{H_2} \quad \underset{(n_2 x k)}{X} = \underset{(n_2 x k)}{V} ,$$

where the last row of W has elements $w_{i(n_2)} = \frac{1}{\sqrt{n_2}} \sum_{s=1}^{n_2} y_{i(s)}$, $i=1,\ldots,k$,

and the other rows have elements $w_{i(s)} = \frac{1}{\sqrt{s(s-1)}}\left[y_{i(1)}+\ldots+y_{i(s-1)}-ry_{i(s)}\right]$,

$i=1,\ldots,k$, $s=1,\ldots,n_2-1$. The v's and w's have multivariate normal distributions.

It can be shown that V and W have the following properties:

(5.3.4) (i) $E(v_{i(n_1)}) = \sqrt{n_1}\,\mu_{iI}$,

$E(w_{i(n_2)}) = \sqrt{n_2}\,\mu_{iII}$,

(5.3.5) (ii) $E(v_{i(r)}) = 0$, $r=1,2,\ldots,n_1-1$,

$E(w_{i(s)}) = 0$, $s=1,2,\ldots,n_2-1$,

(5.3.6) (iii) $\operatorname{cov}(v_{i(r)},v_{j(r)}) = \operatorname{cov}(x_{i(r)},x_{j(r)}) = \sigma_{ijI}$,

$\operatorname{cov}(w_{i(s)},w_{j(s)}) = \operatorname{cov}(y_{i(s)},y_{j(s)}) = \sigma_{ijII}$,

(5.3.7) (iv) $\operatorname{cov}(v_{i(r)}, v_{j(s)}) = 0,\ r \neq s,$

$\operatorname{cov}(w_{i(r)}, w_{j(s)}) = 0,\ r \neq s,$ and

(5.3.8) (v) $\operatorname{cov}(v_{i(r)}, w_{j(s)}) = 0$, all $i,j,r,s.$

The set of the first two moments specify the multivariate normal distribution. Thus the joint distribution of the $k(n_1+n_2)$ variates after the above transformation becomes a set of n_1+n_2 independent k-variate normal distributions.

Specifically, the joint distribution of $v_{i(n_1)}$, $i=1,\ldots,k$, is

(5.3.9) $\mathcal{N}([\sqrt{n_1}\,\mu_{iI}], [\sigma_{ijI}])$;

the joint distribution for $w_{i(n_2)}$, $i=1,\ldots k$, is

(5.3.10) $\mathcal{N}([\sqrt{n_2}\,\mu_{iII}], [\sigma_{ijII}])$;

the joint distributions for $v_{i(r)}$, $i=1,\ldots,k$, $r=1,\ldots,n_1-1$, are (n_1-1) independent, identical, k-variate distributions, each being

(5.3.11) $\mathcal{N}(0, [\sigma_{ijI}])$;

and the joint distributions for $w_{i(s)}$, $i=1,\ldots,k$, $s=1,\ldots,n_2-1$, are (n_2-1) independent, identical, k-variate distributions, each being

$\mathcal{N}(0, [\sigma_{ijII}])$.

5.4. Joint sampling distributions of $\hat{\mu}_{iI}$ and $\hat{\mu}_{jII}$

The maximum likelihood estimates of the means are

(5.4.1) $$\hat{\mu}_{iI} = \frac{1}{n_1}\sum_{r=1}^{n_1} x_{i(r)} = \frac{1}{\sqrt{n_1}}\, v_{i(n_1)}$$

and

$$(5.4.2) \qquad \hat{\mu}_{iII} = \frac{1}{n_2}\sum_{s=1}^{n_2} y_{i(s)} = \frac{1}{\sqrt{n_2}} w_{i(n_2)} .$$

Therefore it follows from Section (5.3) that the distributions of $[\hat{\mu}_{iI}], [\hat{\mu}_{jII}]$ are a pair of independent k-variate normal distributions

$$(5.4.3) \qquad \mathcal{N}([\mu_{iI}], [\tfrac{1}{n_1}\sigma_{ijI}]), \quad \mathcal{N}([\mu_{jII}], [\tfrac{1}{n_2}\sigma_{ijII}])$$

respectively, or

$$(5.4.4) \qquad \mathcal{N}([\mu_{iI}], \tfrac{1}{n_1}\Sigma_I), \quad \mathcal{N}([\mu_{jII}], \tfrac{1}{n_2}\Sigma_{II}) .$$

5.5. Joint sampling distribution of $\hat{d}_i$ and $\hat{d}_j$

$\hat{d}_i = \hat{\mu}_{iII} - \hat{\mu}_{iI} = \frac{1}{\sqrt{n_2}} w_{i(n_2)} - \frac{1}{\sqrt{n_1}} v_{i(n_1)}$. Therefore, the joint sampling distribution of the d_i is a k-variate normal distribution,

$$(5.5.1) \qquad \mathcal{N}([d_i], \tfrac{1}{n_1}\Sigma_I + \tfrac{1}{n_2}\Sigma_{II}) ,$$

since $d_i = \mu_{iII} - \mu_{iI}$.

5.6. Sampling distributions of $\hat{\sigma}_{ijI}$ and $\hat{\sigma}_{ijII}$

$\hat{\sigma}_{ijI}$ is the maximum likelihood estimate of σ_{ijI}, such that

$$\hat{\sigma}_{ijI} = \frac{1}{n_1}\sum_{r=1}^{n_1}(x_{i(r)} - \bar{x}_i)(x_{j(r)} - \bar{x}_j)$$

$$= \frac{1}{n_1}\sum_{r=1}^{n_1}(x_{i(r)}x_{j(r)}) - \bar{x}_i\bar{x}_j$$

$$= \frac{1}{n_1}\sum_{r=1}^{n_1}(v_{i(r)}v_{j(r)}) - \frac{1}{n_1}v_{i(1)}v_{j(1)}$$

$$= \frac{1}{n_1}\sum_{r=1}^{n_1-1}(v_{i(r)}v_{j(r)}). \tag{5.6.1}$$

Similarly,

$$\hat{\sigma}_{ghII} = \frac{1}{n_2}\sum_{s=1}^{n_2-1}(w_{g(s)}w_{h(s)}). \tag{5.6.2}$$

Hence, $n_1\hat{\sigma}_{ijI}$ has a $\sigma_{ijI}\chi^2_{n_1-1}$ distribution and $n_2\hat{\sigma}_{ggII}$ has a $\sigma_{ggII}\chi^2_{n_2-1}$ distribution. Note that $\hat{\sigma}_{ijI}$ and $\hat{\sigma}_{ghII}$ are independent; $\hat{\sigma}_{ijI}$ and $\hat{\sigma}_{ghI}$ are not necessarily independent and $\hat{\sigma}_{ijII}$ and $\hat{\sigma}_{ghII}$ are not necessarily independent.

Consider the first moments of $\hat{\sigma}_{ijI}$ and $\hat{\sigma}_{ijII}$.

$$E(\hat{\sigma}_{ijI}) = \frac{n_1-1}{n_1}\sigma_{ijI} \tag{5.6.3}$$

and

$$E(\hat{\sigma}_{ijII}) = \frac{n_2-1}{n_2}\sigma_{ijII}\ . \tag{5.6.4}$$

The second moments will be found by the use of moment generating function in Section (5.9).

5.7. Summation convention

In this chapter we shall use summation convention defined as follows:

In any algebraic form, when two elements contain the same subscript and/or superscript, the form is summed over the repeated subscript and/or superscript from 1 to k. For example, $\sigma_{ii}\sigma^{ij} = \sum_{i=1}^{k}\sigma_{ii}\sigma^{ij}$; $a_i\sigma_{ijI}a_j = \sum_{i,j=1}^{k} a_i\sigma_{ijI}a_j = A\Sigma A'$. If the same subscript appears in the top and bottom of a partial derivative, it is not to be taken as summed; for example, $\frac{\partial\sigma^{ij}}{\partial\sigma_{ig}} \neq \sum_{i=1}^{k}\frac{\partial\sigma^{ij}}{\partial\sigma_{ig}}$.

The repeated subscripts I and II do not imply summation.

5.8. Moment generating function

Consider the moment generating function for each $[v_{i(r)}]$, $r=1,\ldots,n_1-1$. All $[v_{i(r)}]$, for different r, have the same k-variate normal distribution, $\mathcal{N}(0, \Sigma_I)$. Thus, the moment generating function is given by

$$\text{(5.8.1)}\qquad \text{m.g.f.} = \exp \tfrac{1}{2}t_{i(r)}t_{j(r)}\sigma_{ijI}$$

$$= 1 + \tfrac{1}{2}(t_{i(r)}^{2}\sigma_{iiI} + t_{j(r)}^{2}\sigma_{jjI} + \ldots + 2t_{i(r)}t_{j(r)}\sigma_{ijI} + \ldots + 2t_{g(r)}t_{h(r)}\sigma_{ghI})$$

$$+ \frac{1}{8}\Big[t_{i(r)}^{4}\sigma_{iiI}^{2} + 2t_{i(r)}^{2}t_{j(r)}^{2}(\sigma_{iiI}\sigma_{jjI} + 2\sigma_{ijI}^{2})$$

$$+ 4t_{i(r)}^{3}t_{j(r)}\sigma_{iiI}\sigma_{ijI} + 4t_{i(r)}^{2}t_{j(r)}t_{g(r)}(\sigma_{iiI}\sigma_{jgI} + \sigma_{ijI}\sigma_{igI})$$

$$+ 8t_{i(r)}t_{j(r)}t_{g(r)}t_{h(r)}(\sigma_{ijI}\sigma_{ghI} + \sigma_{igI}\sigma_{jhI} + \sigma_{ihI}\sigma_{jgI}) + \ldots\Big] + \ldots.$$

Consider the moment generating function for each $\left[w_{i(s)}\right]$, $s=1,\ldots,n_2-1$. All $\left[w_{i(s)}\right]$, for different s, have the same k-variate normal distribution, $\mathcal{N}(0, \Sigma_{II})$, and the moment generating function is given by

$$(5.8.2) \qquad \text{m.g.f.} = \exp\left\{\tfrac{1}{2}t_{i(s)}t_{j(s)}\sigma_{ijII}\right\},$$

with an expansion similar to (5.8.1).

From (5.8.1),

$$(5.8.3) \qquad E(v_{i(r)}v_{j(r)}v_{g(r)}v_{h(r)}) = \sigma_{ijI}\sigma_{ghI}+\sigma_{igI}\sigma_{jhI}+\sigma_{ihI}\sigma_{jgI}$$

and

$$(5.8.4) \qquad E(v_{i(r)}v_{j(r)}) = \sigma_{ijI}\,.$$

Similarly

$$(5.8.5) \qquad E(w_{i(s)}w_{j(s)}w_{g(s)}w_{h(s)}) = \sigma_{ijII}\sigma_{ghII}+\sigma_{igII}\sigma_{jhII}+\sigma_{ihII}\sigma_{jgII}$$

and

$$(5.8.6) \qquad E(w_{i(s)}w_{j(s)}) = \sigma_{ijII}\,.$$

5.9. Covariances of $\hat{\sigma}_{ijI}$ and $\hat{\sigma}_{ghI}$ and of $\hat{\sigma}_{ijII}$ and $\hat{\sigma}_{ghII}$

$$(5.9.1) \qquad n_1^{\,2}\operatorname{cov}(\hat{\sigma}_{ijI},\hat{\sigma}_{ghI}) = E(\hat{\sigma}_{ijI}\hat{\sigma}_{ghI}) - E(\hat{\sigma}_{ijI})E(\hat{\sigma}_{ghI})$$

$$= E\left(\sum_{r=1}^{n_1-1} v_{i(r)}v_{j(r)} \sum_{s=1}^{n_1-1} v_{g(s)}v_{h(s)}\right) - E\left(\sum_{r=1}^{n_1-1} v_{i(r)}v_{j(r)}\right)\left(E\sum_{s=1}^{n_1-1} v_{g(s)}v_{h(s)}\right)$$

$$= E\left(\sum_{r=1}^{n_1-1} v_{i(r)}v_{j(r)}v_{g(r)}v_{h(r)} + 2\sum_{\substack{r,s=1\\ r<s}}^{n_1-1} v_{i(r)}v_{j(r)}v_{g(s)}v_{h(s)}\right)$$

$$-E\left(\sum_{r=1}^{n_1-1} v_{i(r)}v_{j(r)}\right)\, E\left(\sum_{s=1}^{n_1-1} v_{g(s)}v_{h(s)}\right)$$

$$= \sum_{r=1}^{n_1-1} E(v_{i(r)}v_{j(r)}v_{g(r)}v_{h(r)}) - 2\sum_{\substack{r,s=1 \\ r<s}}^{n_1-1} E(v_{i(r)}v_{j(r)})E(v_{g(s)}v_{h(s)})$$

$$-\sum_{r=1}^{n_1-1} E(v_{i(r)}v_{j(r)}) \sum_{s=1}^{n_1-1} E(v_{g(s)}v_{h(s)})$$

$$= (n_1-1)(\sigma_{ijI}\sigma_{ghI}+\sigma_{igI}\sigma_{jhI}+\sigma_{ihI}\sigma_{gjI}) + (n_1-1)(n_1-2)(\sigma_{ijI}\sigma_{ghI})$$

$$-(n_1-1)^2(\sigma_{ijI}\sigma_{ghI}) \quad \text{(from (5.8.3) and (5.8.4))}$$

$$= (n_1-1)(\sigma_{ijI}\sigma_{ghI}+\sigma_{igI}\sigma_{jhI}+\sigma_{ihI}\sigma_{gjI}-\sigma_{ijI}\sigma_{ghI}).$$

$$(5.9.2) \qquad \mathrm{Cov}(\hat{\sigma}_{ijI},\hat{\sigma}_{ghI}) = \frac{n_1-1}{n_1^2}(\sigma_{igI}\sigma_{jhI}+\sigma_{ihI}\sigma_{gjI}).$$

Similarly

$$(5.9.3) \qquad \mathrm{cov}(\hat{\sigma}_{ijII},\hat{\sigma}_{ghII}) = \frac{n_2-1}{n_2^2}(\sigma_{igII}\sigma_{jhII}+\sigma_{ihII}\sigma_{gjII}).$$

Although each covariance must be determined with the use of the moment generating functions independently of other results, the other results have been found to be of the forms (5.9.2) and (5.9.3) for the respective populations when the corresponding subscripts are employed. (5.9.2) and (5.9.3), then may be looked upon as if they were general results for all i, j, g, h.

5.10. Covariance of $\hat{\sigma}_{ij}$ and $\hat{\sigma}_{gh}$

As before, let $\sigma_{ij} = \sigma_{ijI} + \sigma_{ijII}$. $\hat{\sigma}_{ij} = \hat{\sigma}_{ijI} + \hat{\sigma}_{ijII}$.

$$(5.10.1)\qquad \operatorname{cov}(\hat{\sigma}_{ij},\hat{\sigma}_{gh}) = \operatorname{cov}(\hat{\sigma}_{ijI},\hat{\sigma}_{ghI}) + \operatorname{cov}(\hat{\sigma}_{ijII},\hat{\sigma}_{ghII}) + \operatorname{cov}(\hat{\sigma}_{ijI},\hat{\sigma}_{ghII}) + \operatorname{cov}(\hat{\sigma}_{ijII},\hat{\sigma}_{ghI}).$$

But since samples from population I are independent of samples from population II, the last two terms vanish. Then in general

$$(5.10.2)\qquad \operatorname{cov}(\hat{\sigma}_{ij},\hat{\sigma}_{gh}) = \frac{n_1-1}{n_1^2}(\sigma_{igI}\sigma_{jhI}+\sigma_{ihI}\sigma_{gjI}) + \frac{n_2-1}{n_2^2}(\sigma_{igII}\sigma_{jhII}+\sigma_{ihII}\sigma_{gjII}).$$

In particular, if $\sigma_{ijII} = \alpha^2\sigma_{ijI}$,

$$(5.10.3)\qquad \operatorname{cov}(\hat{\sigma}_{ij},\hat{\sigma}_{gh}) = \left[\frac{n_1-1}{n_1^2} + \frac{\alpha^4(n_2-1)}{n_2^2}\right](\sigma_{igI}\sigma_{jhI}+\sigma_{ihI}\sigma_{gjI})$$

$$= \frac{1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2} + \frac{\alpha^4(n_2-1)}{n_2^2}\right](\sigma_{ig}\sigma_{jh}+\sigma_{ih}\sigma_{gj}).$$

5.11. A method for finding approximate covariances

In order to obtain covariances for the other statistics involved in $\hat{P}_I$ and $\hat{P}_{II}$, we must resort to approximations. Suppose we wish to find $\operatorname{cov}\left[g(\{x_i\}), h(\{x_i\})\right]$, $g(\{x_i\})$, $h(\{x_i\})$ functions of k random variables x_i, i=1,...,k, where the joint distribution of g and h is not susceptible to evaluation by exact methods.

Suppose there are multivariate Taylor expansions for each of g and h about $\{E(x_i)\}$.

$$(5.11.1)\qquad g(\{x_i\}) = g_o + g_{1i}[x_i-E(x_i)] + \tfrac{1}{2}g_{2ij}[x_i-E(x_i)][x_j-E(x_j)] + \frac{1}{3!}g_{3ijg}[x_i-E(x_i)][x_j-E(x_j)][x_g-E(x_g)] + \ldots,$$

where

$$g_o = g\left[\left\{E(x_i)\right\}\right], \tag{5.11.2}$$

$$g_{1i} = \frac{\partial g(\{x_i\})}{\partial x_i}\,\Big|\,\{x_i\} = \{E(x_i)\}, \tag{5.11.3}$$

and

$$g_{2ij} = \frac{\partial^2 g(\{x_i\})}{\partial x_i \partial x_j}\,\Big|\,\{x_i\} = \{E(x_i)\}. \tag{5.11.4}$$

$$h(\{x_i\}) = h_o + h_{1i}[x_i - E(x_i)] + \tfrac{1}{2}h_{2ij}[x_i - E(x_i)][x_j - E(x_j)] + \frac{1}{3!}h_{3ijg}[x_i - E(x_i)][x_j - E(x_j)][x_g - E(x_g)] + \cdots, \tag{5.11.5}$$

where the definitions of h_o, h_{1i}, h_{2ij}, ..., correspond to the definitions of $g_o, g_{1i}, g_{2ij}, \cdots$.

Suppose further that

$$\frac{1}{3!}g_{3ijg}E[x_i - E(x_i)][x_j - E(x_j)][x_g - E(x_g)] + \cdots \tag{5.11.6}$$

and the corresponding terms in the expansion of h are negligible.

The product gh also has a Taylor series given by the product of the Taylor series for g and h, viz.

$$\begin{aligned} g(\{x_i\})h(\{x_i\}) = {} & g_o h_o + h_o g_{1i}[x_i - E(x_i)] + g_o h_{1i}[x_i - E(x_i)] \\ & + g_{1i}[x_i - E(x_i)]\, h_{1i}[x_i - E(x_i)] + \tfrac{1}{2}h_o g_{2ij}[x_i - E(x_i)][x_j - E(x_j)] \\ & + \tfrac{1}{2}g_o h_{2ij}[x_i - E(x_i)][x_j - E(x_j)] + \tfrac{1}{2}g_{1i}[x_i - E(x_i)]\, h_{2ij}[x_i - E(x_i)][x_j - E(x_j)] \\ & + \tfrac{1}{2}h_{1i}[x_i - E(x_i)]\, g_{2ij}[x_i - E(x_i)][x_j - E(x_j)] \\ & + \tfrac{1}{4}g_{2ij}[x_i - E(x_i)][x_j - E(x_j)]\, h_{2ij}[x_i - E(x_i)][x_j - E(x_j)] + \cdots \end{aligned} \tag{5.11.7}$$

$E[g(\{x_i\})]$ and $E[h(\{x_i\})]$ and $E[g(\{x_i\})][h(\{x_i\})]$ may be found by integrating term by term and from (5.11.6) can be approximated by the first three terms. Thus,

$$(5.11.8)\qquad E\left[g(\{x_i\})\right] = g_o + \tfrac{1}{2}g_{2ij}\,\mathrm{cov}(x_i,x_j) + \cdots,$$

$$(5.11.9)\qquad E\left[h(\{x_i\})\right] = h_o + \tfrac{1}{2}h_{2ij}\,\mathrm{cov}(x_i,x_j) + \cdots,$$

and

$$(5.11.10)\qquad E\left[g(\{x_i\})\right]\left[h(\{x_i\})\right] = g_o h_o + g_{1i}h_{1j}\,\mathrm{cov}(x_i,x_j) + \tfrac{1}{2}h_o g_{2ij}\,\mathrm{cov}(x_i,x_j) + \tfrac{1}{2}g_o h_{2ij}\,\mathrm{cov}(x_i,x_j) + \cdots.$$

Suppose that central moments, product-moments, and products of moments of order greater than two of $\{x_i\}$ are negligible. Then

$$(5.11.11)\qquad \mathrm{cov}\left[g(\{x_i\}),\, h(\{x_i\})\right] \doteq E\left[g(\{x_i\})\right]\left[h(\{x_i\})\right] - E\left[g(\{x_i\})\right]E\left[h(\{x_i\})\right]$$

$$= g_o h_o + g_{1i}h_{1i}\mathrm{cov}(x_i,x_j) + \tfrac{1}{2}h_o g_{2ij}\mathrm{cov}(x_i,x_j) + \tfrac{1}{2}g_o h_{2ij}\mathrm{cov}(x_i,x_j) - g_o h_o - \tfrac{1}{2}h_o g_{2ij}\mathrm{cov}(x_i,x_j) - \tfrac{1}{2}g_o h_{2ij}\mathrm{cov}(x_i,x_j)$$

$$= g_{1i}h_{1i}\mathrm{cov}(x_i,x_j)\ .$$

Writing this quadratic form in matrix notation, we obtain

$$(5.11.12)\qquad \mathrm{cov}\left[g(\{x_i\}),h(\{x_i\})\right] \doteq \left[\frac{\partial g}{\partial x_1} \cdots \frac{\partial g}{\partial x_k}\right]\begin{bmatrix}\mathrm{var}(x_1) & \cdots & \mathrm{cov}(x_1,x_k)\\ \vdots & & \vdots \\ \mathrm{cov}(x_1,x_k) & \cdots & \mathrm{var}(x_k)\end{bmatrix}\begin{bmatrix}\frac{\partial h}{\partial x_1}\\ \vdots \\ \frac{\partial h}{\partial x_k}\end{bmatrix}$$

where $x_i = E(x_i)$. To make (5.11.12) more general, suppose we have h functions, $g_1(\{x_i\})$, $g_2(\{x_i\})$, ..., $g_h(\{x_i\})$. The properties of the g's are the same as those of the previous g, h. Then under the same conditions and to the same order of approximation, the variance-covariance matrix of $g_r(\{x_i\})$, $g_s(\{x_i\})$, r,s=1,...,h, is given approximately by

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_i} \\ \cdots\cdots \\ \frac{\partial g_h}{\partial x_i} \end{bmatrix} \begin{bmatrix} \mathrm{var}(x_1) & \cdots & \mathrm{cov}(x_1, x_k) \\ \vdots & & \vdots \\ \mathrm{cov}(x_1, x_k) & \cdots & \mathrm{var}(x_k) \end{bmatrix} \begin{bmatrix} \frac{\partial g_1}{\partial x_i} \\ \cdots\cdots \\ \frac{\partial g_h}{\partial x_i} \end{bmatrix}' \Bigg|_{x_i = E(x_i)},$$

h x k k x k k x h

When this approximation is used henceforth, it is implied that all variables are evaluated at their expectations.

5.12. Covariance of $\hat{\sigma}^{ij}$ and $\hat{\sigma}^{gh}$

In this section, Σ will denote the estimated dispersion matrix $[\hat{\sigma}_{ij}]$ and Σ_{ij} will denote the cofactor of σ_{ij} in Σ. Then

$$\hat{\sigma}^{ij} = f(\hat{\sigma}_{11}, \hat{\sigma}_{22}, \ldots, \hat{\sigma}_{kk}, \hat{\sigma}_{12}, \ldots, \hat{\sigma}_{k-1,k}) = \frac{|\Sigma_{ij}|}{|\Sigma|} . \tag{5.12.1}$$

$$\mathrm{cov}(\hat{\sigma}^{ij}, \hat{\sigma}^{gh}) = \begin{bmatrix} \frac{\partial \hat{\sigma}^{ij}}{\partial \hat{\sigma}_{rr}} & \vdots & \frac{\partial \hat{\sigma}^{ij}}{\partial \hat{\sigma}_{st}} \end{bmatrix} \begin{bmatrix} \mathrm{cov}(\hat{\sigma}_{rr}, \hat{\sigma}_{r'r'}) & \vdots & \mathrm{cov}(\hat{\sigma}_{st}, \hat{\sigma}_{r'r'}) \\ \cdots\cdots & \vdots & \cdots\cdots \\ \mathrm{cov}(\hat{\sigma}_{rr}, \hat{\sigma}_{s't'}) & \vdots & \mathrm{cov}(\hat{\sigma}_{st}, \hat{\sigma}_{s't'}) \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{\sigma}^{gh}}{\partial \hat{\sigma}_{r'r'}} \\ \cdots\cdots \\ \frac{\partial \hat{\sigma}^{gh}}{\partial \hat{\sigma}_{s't'}} \end{bmatrix}, \tag{5.12.2}$$

$s<t$, $s'<t'$. It should be noted that wherever there are double subscripts being used as single subscripts (for the matrix or vector) this pattern of partitioning will be used.

When $\Sigma_1 \neq \Sigma_2$ in general, explicit results are not obtainable. However, some results can be obtained for the case of proportionate dispersion matrices; the author feels intuitively that in some fields of science dispersion matrices very far from being proportionate would not be met as frequently as nearly proportionate matrices. For the rest of this chapter we will restrict ourselves to $\Sigma_2 = \alpha^2 \Sigma_1$, where α^2 is a constant of proportionality. When $\Sigma_2 = \alpha^2 \Sigma_1$, then $\sigma_1^2 = D\Sigma_1^{-1}D'$, and $\sigma_2^2 = D\alpha^2\Sigma_1^{-1}D' = \alpha^2\sigma_1^2$.

Consider (5.12.2). For a more convenient form, combine the $\frac{1}{2}k(k+1)$ vectors of partial derivatives of the $\hat{\sigma}_{ij}$ into a $\frac{1}{2}k(k+1)$ x $\frac{1}{2}k(k+1)$ matrix, as in (5.11.13); denote this matrix as B. Then

(5.12.3) (dispersion matrix for $\hat{\sigma}^{ij}$) = BVB'

where

(5.12.4) $$V = \frac{1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2} + \frac{\alpha^4(n_2-1)}{n_2^2}\right]$$

$$\begin{bmatrix} 2\sigma_{11}^2 & \cdots & 2\sigma_{1k}^2 & \vdots & 2\sigma_{11}\sigma_{12} & \cdots & 2\sigma_{1,k-1}\sigma_{1k} \\ \vdots & & \vdots & & \vdots & & \vdots \\ 2\sigma_{1k}^2 & \cdots & 2\sigma_{kk}^2 & \vdots & 2\sigma_{kk}\sigma_{12} & \cdots & 2\sigma_{k,k-1}\sigma_{kk} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 2\sigma_{11}\sigma_{12} & \cdots & 2\sigma_{kk}\sigma_{12} & \vdots & \sigma_{11}\sigma_{22}+\sigma_{12}^2 & \cdots & \sigma_{1,k-1}\sigma_{2k}+\sigma_{12}\sigma_{2,k-1} \\ \vdots & & \vdots & & \vdots & & \vdots \\ 2\sigma_{1,k-1}\sigma_{1k} & \cdots & 2\sigma_{k,k-1}\sigma_{kk} & \vdots & \sigma_{1,k-1}\sigma_{2k}+\sigma_{12}\sigma_{2,k-1} & \cdots & \sigma_{k-1,k-1}\sigma_{kk}+\sigma_{k-1,k}^2 \end{bmatrix}$$

Let us obtain an explicit expression for B. In general, by a proof in the appendix (10.2),

$$\frac{\partial \sigma^{ij}}{\partial \sigma_{gh}} = \frac{|\Sigma|\frac{\partial}{\partial \sigma_{gh}}|\Sigma_{ij}| - |\Sigma_{ij}|\frac{\partial}{\partial \sigma_{gh}}|\Sigma|}{|\Sigma|^2}$$

$$(5.12.5) \qquad = -\frac{|\Sigma_{ih}||\Sigma_{gj}| + |\Sigma_{ig}||\Sigma_{hj}|}{|\Sigma|^2} .$$

Substituting (5.12.5) in B, we obtain

$$(5.12.6) \qquad B = -\frac{1}{|\Sigma|^2}\begin{bmatrix} |\Sigma_{rr'}|^2 & \vdots & 2|\Sigma_{rs'}||\Sigma_{rt'}| \\ \cdots & \cdots & \cdots \\ |\Sigma_{sr'}||\Sigma_{tr'}| & \vdots & |\Sigma_{ss'}||\Sigma_{tt'}| + |\Sigma_{st'}||\Sigma_{ts'}| \end{bmatrix},$$

where $r,s,t,r',s',t' = 1,\ldots,k$.

Consider the element $(b\sigma)_{ij,gh}$ of BV.

$$(5.12.7) \qquad (b\sigma)_{ij,gh} = \frac{-1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2} + \frac{\alpha^4(n_2-1)}{n_2^2}\right]\left[|\Sigma_{ir}||\Sigma_{js}|(\sigma_{rg}\sigma_{sh}+\sigma_{rh}\sigma_{sg})\right] .$$

But

$$(5.12.8) \qquad |\Sigma_{ir}| \cdot \sigma_{rg} = |\Sigma|\delta_{ig},$$

where δ is Kronecker's delta. Therefore,

$$(5.12.9) \qquad (b\sigma)_{ij,gh} = \frac{-1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2} + \frac{\alpha^4(n_2-1)}{n_2^2}\right](\delta_{ig}\delta_{jh}+\delta_{ih}\delta_{jg}),$$

$i \leq j$, $g \leq h$. Then

$$(5.12.10) \qquad BV = \frac{-1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2} + \frac{\alpha^4(n_2-1)}{n_2^2}\right]\begin{bmatrix} 2I & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & I \end{bmatrix} .$$

$$BVB' = \frac{-1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2}+\frac{\alpha^4(n_2-1)}{n_2^2}\right]\left[\begin{array}{c:c} 2I & 0 \\ \hdashline 0 & I \end{array}\right] B'$$

$$(5.12.11) \qquad = \frac{1}{|\Sigma|^2(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2}+\frac{\alpha^4(n_2-1)}{n_2^2}\right]\left[\begin{array}{c:c} 2|\Sigma_{ir}|^2 & 2|\Sigma_{ir}||\Sigma_{jr}| \\ \hdashline 2|\Sigma_{ir}||\Sigma_{is}| & |\Sigma_{ir}||\Sigma_{js}| + |\Sigma_{is}||\Sigma_{jr}| \end{array}\right].$$

In general,

$$\mathrm{cov}(\hat{\sigma}^{ij},\hat{\sigma}^{gh}) = \frac{1}{|\Sigma|^2(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2}+\frac{\alpha^4(n_2-1)}{n_2^2}\right](|\Sigma_{ig}||\Sigma_{jh}|+|\Sigma_{ih}||\Sigma_{jg}|)$$

$$(5.12.12) \qquad = \frac{1}{(1+\alpha^2)}\left[\frac{n_1-1}{n_1^2}+\frac{\alpha^4(n_2-1)}{n_2^2}\right](\sigma^{ig}\sigma^{jh}+\sigma^{ih}\sigma^{jg})$$

where the choice of superscripts gives the desired variance or covariance.

5.13. Covariance of $\hat{a}_i$ and $\hat{a}_j$

Recall that $A = D\Sigma^{-1}$; $\Sigma = \Sigma_1 + \Sigma_2$.

$$\mathrm{Var}(a_i) \doteq \left[\frac{\partial \hat{a}_i}{\partial \hat{d}_r} \vdots \frac{\partial \hat{a}_i}{\partial \sigma^{is}}\right] \left[\begin{array}{c:c} \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2 & 0 \\ \hdashline 0 & \mathrm{BVB}' \end{array}\right] \left[\begin{array}{c} \frac{\partial \hat{a}_i}{\partial \hat{d}_r} \\ \hdashline \frac{\partial \hat{a}_i}{\partial \hat{\sigma}^{is}} \end{array}\right]$$

$$= \left[\sigma^{ir} \vdots d_s\right] \left[\begin{array}{c:c} \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2 & 0 \\ \hdashline 0 & \mathrm{BVB}' \end{array}\right] \left[\begin{array}{c} \sigma^{ir} \\ \hdashline d_s \end{array}\right] \qquad (5.13.1)$$

Now the r,sth element of $\frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2$ is

$$\frac{1}{n_1}\sigma_{rsI} + \frac{1}{n_2}\sigma_{rsII} = \left[\frac{1}{n_1} + \frac{\alpha^2}{n_2}\right]\sigma_{rsI} = \frac{n_1\alpha^2+n_2}{(1+\alpha^2)n_1n_2}\sigma_{rs} \quad ; \qquad (5.13.2)$$

the r,sth element of BVB' is

$$\frac{1}{(1+\alpha^2)}\left[\frac{n_1-1}{n_1^2} + \frac{\alpha^4(n_2-1)}{n_2^2}\right](\sigma^{ij}\sigma^{rs}+\sigma^{ir}\sigma^{js}). \qquad (5.13.3)$$

Then expansion of (5.13.1) gives

$$\mathrm{var}(\hat{a}_i) \doteq \left[\frac{n_1\alpha^2-n_2}{(1+\alpha^2)n_1n_2}\right]\sigma^{ir}\sigma^{is}\sigma_{rs} + \frac{1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2} + \frac{\alpha^4(n_2-1)}{n_2^2}\right] d_r d_s \,(\sigma^{ii}\sigma^{rs}+\sigma^{ir}\sigma^{is}). \qquad (5.13.4)$$

The first term of (5.13.4) may be simplified:

$$\sigma^{ir}\sigma^{is}\sigma_{rs} = \sigma^{ir}\delta_{ir} \qquad \text{(Kronecker's delta)}$$

$$= \sigma^{ii}. \qquad (5.13.5)$$

Then

$$(5.13.6)\qquad \operatorname{var}(\hat{a}_i) \doteq \left[\frac{n_1\alpha^2+n_2}{(1+\alpha^2)n_1n_2}\right]\sigma^{ii} + \frac{1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2}+\frac{\alpha^4(n_2-1)}{n_2^2}\right] d_r d_s(\sigma^{ii}\sigma^{rs}+\sigma^{ir}\sigma^{is}).$$

Covariance of $\hat{a}_i$ and $\hat{a}_j$ is found by a similar procedure, differing only in that the post-multiplied column vector consists of partial derivatives of a_j rather than of a_i. Going through corresponding operations, we obtain

$$(5.13.7)\qquad \operatorname{cov}(\hat{a}_i,\hat{a}_j) \doteq \left[\frac{n_1\alpha^2+n_2}{(1+\alpha^2)n_1n_2}\right]\sigma^{ij} + \frac{1}{(1+\alpha^2)^2}\left[\frac{n_1-1}{n_1^2}+\frac{\alpha^4(n_2-1)}{n_2^2}\right] d_r d_s(\sigma^{ij}\sigma^{rs}+\sigma^{ir}\sigma^{js}).$$

5.14. Covariances of $\hat{\mu}_{iI}$ and $\hat{a}_g$, and of $\hat{\mu}_{iII}$ and $\hat{a}_g$

$$\operatorname{cov}(\hat{\mu}_{iI},\hat{a}_g) \doteq \left[\frac{\partial\hat{\mu}_{iI}}{\partial\hat{\mu}_{jI}} \cdot \frac{\partial\hat{\mu}_{iI}}{\partial\hat{\mu}_{jII}} \cdot \frac{\partial\hat{\mu}_{iI}}{\partial\hat{\sigma}^{ej}}\right] \begin{bmatrix} \frac{1}{n_1}\Sigma_1 & 0 & 0 \\ 0 & \frac{1}{n_2}\Sigma_2 & 0 \\ 0 & 0 & BVB' \end{bmatrix} \begin{bmatrix} \frac{\partial\hat{a}_g}{\partial\hat{\mu}_{iI}} \\ \frac{\partial\hat{a}_g}{\partial\hat{\mu}_{iII}} \\ \frac{\partial\hat{a}_g}{\partial\hat{\sigma}^{ej}} \end{bmatrix}$$

$$= \left[0\ldots0\,0\ldots0\,0\ldots0\right] \begin{bmatrix} \frac{1}{n_1}\Sigma_1 & 0 & 0 \\ 0 & \frac{1}{n_2}\Sigma^2 & 0 \\ 0 & 0 & BVB' \end{bmatrix} \begin{bmatrix} \sigma^{jg} \\ \sigma^{jg} \\ d_j \end{bmatrix}$$

(where the "one" occurs in the ith position)

$$= \left[\operatorname{cov}(\hat{\mu}_{iI},\hat{\mu}_{iI})\ldots\operatorname{cov}(\hat{\mu}_{kI},\hat{\mu}_{iI})\ 0\ldots0\right] \begin{bmatrix} \sigma^{ig} \\ \cdots\cdots \\ -\sigma^{jg} \\ \cdots\cdots \\ d_j \\ \cdots \end{bmatrix}$$

$$= \frac{1}{n_1(1+\alpha^2)}\ \sigma_{ij}\sigma^{ij}$$

$$= \frac{1}{n_1(1+\alpha^2)} \quad . \tag{5.14.1}$$

Similarly,

$$\operatorname{cov}(\hat{\mu}_{iII},\hat{a}_g) = \frac{\alpha^2}{n_2(1+\alpha^2)} \quad . \tag{5.14.2}$$

5.15. Covariance of $\hat{m}_1$ and $\hat{m}_2$

For convenience let us define some dispersion matrices which will appear as sub-matrices of partitioned dispersion matrices in subsequent sections. Define

$$\operatorname{cov}(\hat{\sigma}_{ijI},\hat{\sigma}_{ghI}) = V_1,\ \operatorname{cov}(\hat{\sigma}_{ijII},\hat{\sigma}_{ghII}) = V_2 \tag{5.15.1}$$

$$\operatorname{cov}(\hat{a}_i,\hat{a}_j) = V_A \tag{5.15.2}$$

$$\operatorname{cov}(\hat{\mu}_{iI},\hat{a}_j) = V_{AM_1},\ \operatorname{cov}(\hat{\mu}_{iII},\hat{a}_j) = V_{AM_2} \tag{5.15.3}$$

$$\operatorname{cov}(\hat{\sigma}_{ijI},\hat{a}_g) = V_{A\sigma_1},\ \operatorname{cov}(\hat{\sigma}_{ijII},\hat{a}_g) = V_{A\sigma_2} \quad . \tag{5.15.4}$$

$$(5.15.5)\qquad \mathrm{Cov}(m_1, m_2) \doteq \begin{bmatrix} \frac{\partial \hat{m}_1}{\partial \hat{a}_i} & \frac{\partial \hat{m}_1}{\partial \hat{\mu}_{iI}} & \frac{\partial \hat{m}_1}{\partial \hat{\mu}_{iII}} \end{bmatrix} \begin{bmatrix} V_A & V_{AM_1} & V_{AM_2} \\ V_{AM_1} & \frac{1}{n_1}\Sigma_1 & 0 \\ V_{AM_2} & 0 & \frac{1}{n_2}\Sigma_2 \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{m}_2}{\partial \hat{a}_i} \\ \frac{\partial \hat{m}_2}{\partial \hat{\mu}_{iI}} \\ \frac{\partial \hat{m}_2}{\partial \hat{\mu}_{iII}} \end{bmatrix}$$

$$= \begin{bmatrix} \mu_{iI} & a_i & 0 \end{bmatrix} \begin{bmatrix} V_A & V_{AM_1} & V_{AM_2} \\ V_{AM_1} & \frac{1}{n_1}\Sigma_1 & 0 \\ V_{AM_2} & 0 & \frac{1}{n_2}\Sigma_2 \end{bmatrix} \begin{bmatrix} \mu_{jII} \\ 0 \\ a_j \end{bmatrix}$$

Now substituting (5.14.1) and (5.14.2), we obtain

$$(5.15.6)\qquad \mathrm{cov}(\hat{m}_1, \hat{m}_2) \doteq \sum_i \sum_j \mu_{iI}\mu_{jII}\ \mathrm{cov}(\hat{a}_i, \hat{a}_j) + \frac{1}{n_1(1+\alpha^2)} \sum_i \sum_j \mu_{iII} a_j + \frac{\alpha^2}{n_2(1+\alpha^2)} \sum_i \sum_j \mu_{iI} a_j ,$$

where the summation symbols are necessary because the form does not fit summation convention. Similarly,

$$(5.15.7)\qquad \mathrm{var}(\hat{m}_1) \doteq \sum_i \sum_j \mu_{iI}\mu_{jI}\ \mathrm{cov}(\hat{a}_i, \hat{a}_j) + \frac{2\sum_i \sum_j \mu_{iI} a_j}{n_1(1+\alpha^2)} + \frac{1}{n_1(1+\alpha^2)} \sum_i \sum_j a_i a_j \sigma_{ij}$$

and

$$(5.15.8)\qquad \mathrm{var}(\hat{m}_2) \doteq \sum_i \sum_j \mu_{iII}\mu_{jII}\ \mathrm{cov}(\hat{a}_i, \hat{a}_j) + \frac{2\alpha^2 \sum_i \sum_j \mu_{iII} a_j}{n_2(1+\alpha^2)} + \frac{\alpha^2}{n_2(1+\alpha^2)} \sum_i \sum_j a_i a_j \sigma_{ij} .$$

5.16. Covariances of $\hat{\sigma}_{ijI}$ and $\hat{a}_g$, and of $\hat{\sigma}_{ijII}$ and $\hat{a}_g$

$$\hat{a}_g = \hat{d}_h \hat{\sigma}^{gh}, \quad \hat{\sigma}_{ij} = \hat{\sigma}_{ijI} + \hat{\sigma}_{ijII}.$$

(5.16.1)
$$\operatorname{cov}(\hat{\sigma}_{ijI}, \hat{a}_g) \doteq \left[\frac{\partial \hat{\sigma}_{ijI}}{\partial \hat{\sigma}_{ghI}} \vdots \frac{\partial \hat{\sigma}_{ijI}}{\partial \hat{\sigma}_{ghII}} \vdots \frac{\partial \hat{\sigma}_{ijI}}{\partial \hat{d}_h}\right] \begin{bmatrix} V_1 & \vdots & 0 & \vdots & 0 \\ \cdots & & \cdots & & \cdots \\ 0 & \vdots & V_2 & \vdots & 0 \\ \cdots & & \cdots & & \cdots \\ 0 & \vdots & 0 & \vdots & \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2 \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{a}_g}{\partial \hat{\sigma}_{hkI}} \\ \cdots \\ \frac{\partial \hat{a}_g}{\partial \hat{\sigma}_{hkII}} \\ \cdots \\ \frac{\partial \hat{a}_g}{\partial \hat{d}_h} \end{bmatrix}$$

$$= \left[0 \ldots 0 1 0 \ldots 0 \vdots 0 \vdots 0\right] \begin{bmatrix} V_1 & \vdots & 0 & \vdots & 0 \\ \cdots & & \cdots & & \cdots \\ 0 & \vdots & V_2 & \vdots & 0 \\ \cdots & & \cdots & & \cdots \\ 0 & \vdots & 0 & \vdots & \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2 \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{a}_g}{\partial \hat{\sigma}_{hkI}} \\ \cdots \\ \frac{\partial \hat{a}_g}{\partial \hat{\sigma}_{hkII}} \\ \cdots \\ \frac{\partial \hat{a}_g}{\partial \hat{d}_h} \end{bmatrix}.$$

In the first sub-vector of the partitioned pre-multiplying vector, the elements are arranged with subscripts in the sequence: 11,...,kk,12,...,(k-1)k; the "one" appears in the i,jth position.

(5.16.2)
$$\operatorname{cov}(\hat{\sigma}_{ijI}, \hat{a}_g) \doteq \operatorname{cov}(\hat{\sigma}_{ijI}, \hat{\sigma}_{rsI}) \frac{\partial \hat{a}_g}{\partial \hat{\sigma}_{rsI}}.$$

Now from Section (10.2),

$$\frac{\partial \hat{\sigma}^{gh}}{\partial \hat{\sigma}_{rs}} = -(\sigma^{hr}\sigma^{gs} + \sigma^{hs}\sigma^{gr}).$$

Also,

(5.16.3)
$$\frac{\partial \hat{\sigma}^{gh}}{\partial \hat{\sigma}_{rsI}} = \frac{\partial \hat{\sigma}^{gh}}{\partial \hat{\sigma}_{rs}}.$$

Then

$$(5.16.4)\qquad \frac{\partial \hat{a}_g}{\partial \hat{\sigma}_{rsI}} = -d_h\left[\sigma^{hr}\sigma^{gs}+\sigma^{hs}\sigma^{gr}\right].$$

Then

$$(5.16.5)\qquad \operatorname{cov}(\hat{\sigma}_{ijI},\hat{a}_g) \doteq \frac{(n_1-1)}{n_1^2(1+\alpha^2)^2}\, d_h\left[\sigma_{ir}\sigma_{js}+\sigma_{is}\sigma_{jr}\right]\left[\sigma^{hr}\sigma^{gs}+\sigma^{hs}\sigma^{gr}\right]$$

$$= \frac{(n_1-1)}{n_1^2(1+\alpha^2)^2}\, d_h\left[\delta_{ih}\delta_{jg}+\delta_{ig}\delta_{jh}+\delta_{ig}\delta_{jh}+\delta_{ih}\delta_{jg}\right]$$

$$= \frac{2(n_1-1)}{n_1^2(1+\alpha^2)^2}\left[d_i\delta_{jg}+d_j\delta_{ig}\right].$$

Specifically,

$$(5.16.6)\qquad \operatorname{cov}(\hat{\sigma}_{ijI},\hat{a}_j) \doteq \frac{2(n_1-1)d_i}{n_1^2(1+\alpha^2)^2},$$

$$(5.16.7)\qquad \operatorname{cov}(\hat{\sigma}_{iiI},\hat{a}_i) \doteq \frac{4(n_1-1)d_i}{n_1^2(1+\alpha^2)^2},$$

and other covariances are zero. Similarly,

$$(5.16.8)\qquad \operatorname{cov}(\hat{\sigma}_{ijII},\hat{a}_j) \doteq \frac{2(n_2-1)\alpha^4 d_i}{n_2^2(1+\alpha^2)^2}$$

and

$$(5.16.9)\qquad \operatorname{cov}(\hat{\sigma}_{iiII},\hat{a}_j) \doteq \frac{4(n_2-1)\alpha^4 d_i}{n_2^2(1+\alpha^2)^2}$$

5.17. Covariance of $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$

$$\hat{\sigma}_1^2 = \hat{a}_i \hat{\sigma}_{ijI} \hat{a}_j, \quad \hat{\sigma}_2^2 = \hat{a}_i \hat{\sigma}_{ijII} \hat{a}_j .$$

$$(5.17.1) \qquad \mathrm{Var}(\hat{\sigma}_1^2) \doteq \left[\frac{\partial \hat{\sigma}_1^2}{\partial \hat{a}_i} \vdots \frac{\partial \hat{\sigma}_1^2}{\partial \hat{\sigma}_{iiI}} \vdots \frac{\partial \hat{\sigma}_1^2}{\partial \hat{\sigma}_{ijI}} \right] \begin{bmatrix} V_A & \vdots & (V_{A\Sigma_1}) \\ \cdots & \vdots & \cdots \\ (V_{A\Sigma_1}) & \vdots & (V_1) \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{\sigma}_1^2}{\partial \hat{a}_i} \\ \cdots \\ \frac{\partial \hat{\sigma}_1^2}{\partial \hat{\sigma}_{iiI}} \\ \cdots \\ \frac{\partial \hat{\sigma}_1^2}{\partial \hat{\sigma}_{ijI}} \end{bmatrix}$$

$$= \left[2\sigma_{ij} a_j \vdots a_i^2 \vdots 2a_i a_j \right] \begin{bmatrix} V_A & \vdots & (V_{A\Sigma_1}) \\ \cdots & \vdots & \cdots \\ (V_{A\Sigma_1}) & \vdots & (V_1) \end{bmatrix} \begin{bmatrix} 2\sigma_{gh} a_h \\ \cdots \\ a_g^2 \\ \cdots \\ 2a_g a_h \end{bmatrix}$$

$$= 4\sigma_{ijI} a_j \sigma_{ghI} a_h \, \mathrm{cov}(\hat{a}_i, \hat{a}_g) + a_i^2 a_g^2 \, \mathrm{cov}(\hat{\sigma}_{iiI}, \hat{\sigma}_{ggI})$$
$$+ 4a_i^2 \sigma_{ghI} a_h \, \mathrm{cov}(\hat{a}_i, \hat{\sigma}_{ghI}) + 8a_i a_j \sigma_{gh} a_h \, \mathrm{cov}(\hat{a}_i, \hat{\sigma}_{ghI})$$
$$+ 4a_i a_j a_g a_h \, \mathrm{cov}(\hat{\sigma}_{ijI}, \hat{\sigma}_{ghI}) + 4a_i a_j a_g^2 \, \mathrm{cov}(\hat{\sigma}_{ggI}, \hat{\sigma}_{ijI}) .$$

These results may be simplified slightly, but not appreciably.

In a similar fashion, if we let

$$V^* = \begin{bmatrix} V_A & (V_A\Sigma_1) & (V_A\Sigma_2) \\ (V_A\Sigma_1) & (V_1) & 0 \\ (V_A\Sigma_2) & 0 & (V_2) \end{bmatrix},$$

then

$$\operatorname{cov}(\hat{\sigma}_1^2, \hat{\sigma}_2^2) = \begin{bmatrix} \frac{\partial \hat{\sigma}_1^2}{\partial a_i} & \frac{\partial \hat{\sigma}_1^2}{\partial \sigma_{iiI}} & \frac{\partial \hat{\sigma}_1^2}{\partial \sigma_{ijI}} & 0 & 0 \end{bmatrix} V^* \begin{bmatrix} \frac{\partial \hat{\sigma}_2^2}{\partial a_i} \\ 0 \\ 0 \\ \frac{\partial \hat{\sigma}_2^2}{\partial \sigma_{iiII}} \\ \frac{\partial \hat{\sigma}_2^2}{\partial \sigma_{ijII}} \end{bmatrix}$$

$$= \begin{bmatrix} 2\sigma_{ijI}a_j & a_i^2 & 2a_ia_j & 0 & 0 \end{bmatrix} V^* \begin{bmatrix} 2\sigma_{ghII}a_h \\ 0 \\ 0 \\ a_g^2 \\ 2a_ga_h \end{bmatrix}$$

$$= 4a_j\sigma_{ijI}a_h\sigma_{ghII}\operatorname{cov}(\hat{a}_i, \hat{a}_g) + 2a_i^2\sigma_{ghII}a_g\operatorname{cov}(\hat{\sigma}_{iiI}, \hat{a}_g) + 4a_ia_j\sigma_{ghII}a_g\operatorname{cov}(\hat{\sigma}_{ijI}, \hat{a}_g) + 2a_i^2\sigma_{ghI}a_g\operatorname{cov}(\hat{\sigma}_{iiII}, \hat{a}_g) + 4a_ia_j\sigma_{ghI}a_g\operatorname{cov}(\hat{\sigma}_{ijII}, \hat{a}_g).$$

This result again might be slightly simplified, but since it appears later only in another result not subject to simplification, it will not be reduced further here.

5.18. Covariance of $\hat{m}_1$ and $\hat{\sigma}_1^2$, and of $\hat{m}_2$ and $\hat{\sigma}_2^2$

$$(5.18.1)\qquad \operatorname{cov}(\hat{m}_1,\hat{\sigma}_1^2) \doteq \left[\frac{\partial\hat{m}_1}{\partial\hat{a}_i} \vdots \frac{\partial\hat{m}_1}{\partial\hat{\mu}_{iI}} \vdots \frac{\partial\hat{m}_1}{\partial\hat{\sigma}_{ijI}^2}\right] \left[\begin{array}{c:c:c} V_A & V_{AM_1} & V_{A\Sigma_1} \\ \hdashline V_{AM_1} & \frac{1}{n_1}\Sigma_1 & 0 \\ \hdashline V_{A\Sigma_1} & 0 & V_1 \end{array}\right] \left[\begin{array}{c} \frac{\partial\hat{\sigma}_1^2}{\partial\hat{a}_i} \\ \hdashline \frac{\partial\hat{\sigma}_1^2}{\partial\hat{\mu}_{iI}} \\ \hdashline \frac{\partial\hat{\sigma}_1^2}{\partial\hat{\sigma}_{ijI}^2} \end{array}\right]$$

$$= \left[\hat{\mu}_{iI} \vdots a_i \vdots 0\right] \left[\begin{array}{c:c:c} V_A & V_{AM_1} & V_{A\Sigma_1} \\ \hdashline V_{AM_1} & \frac{1}{n_1}\Sigma_1 & 0 \\ \hdashline V_{A\Sigma_1} & 0 & V_1 \end{array}\right] \left[\begin{array}{c} a_j\sigma_{jiI} \\ \hdashline 0 \\ \hdashline a_ia_j \end{array}\right]$$

$$= \hat{\mu}_{gI}a_j\sigma_{ijI}\operatorname{cov}(\hat{a}_i,\hat{a}_j) + \sigma_1^2\operatorname{cov}(\hat{a}_i,\hat{\mu}_{jI}) + a_ia_j\hat{\mu}_{gI}\operatorname{cov}(\hat{a}_i,\hat{\sigma}_{gjI}).$$

Similarly,

$$(5.18.2)\qquad \operatorname{cov}(\hat{m}_2,\hat{\sigma}_2^2) \doteq \hat{\mu}_{gII}a_i\sigma_{ijII}\operatorname{cov}(\hat{a}_i,\hat{a}_j) + \sigma_2^2\operatorname{cov}(\hat{a}_i,\hat{\mu}_{jII}) + a_ia_j\hat{\mu}_{gII}\operatorname{cov}(\hat{a}_i,\hat{\sigma}_{gjII}).$$

5.19. Variance of L_{1_m} and L_{2_m}

$$(5.19.1)\qquad L_{1_m} = \frac{\check{m}_1-\hat{c}(m)}{\check{\sigma}_1} = \frac{1}{2\check{\sigma}_1}\,(2\check{m}_1-\hat{m}_1-\hat{m}_2).$$

Note that

$$\check{m}_1 = \hat{a}_i\hat{\mu}_{iI},$$

$$\hat{m}_1 = \hat{a}_i\hat{\mu}_{iI},$$

and

$$\check{\sigma}_1^2 = \hat{a}_i \sigma_{ij} \hat{a}_j .$$

Then

$$L_{1_m} = \frac{\hat{a}_i(2\mu_{iI} - \hat{\mu}_{iI} - \hat{\mu}_{iII})}{2(\hat{a}_i \sigma_{ijI} \hat{a}_j)^{\frac{1}{2}}} . \tag{5.19.2}$$

$$\operatorname{var}(L_{1_m}) \doteq \frac{1}{4} \begin{bmatrix} \frac{\partial L_{1_m}}{\partial \hat{\mu}_{iI}} & \frac{\partial L_{1_m}}{\partial \hat{\mu}_{iII}} & \frac{\partial L_{1_m}}{\partial \hat{a}_i} \end{bmatrix} \begin{bmatrix} \frac{1}{n_1}\Sigma_1 & 0 & V_{AM_1} \\ 0 & \frac{1}{n_2}\Sigma_2 & V_{AM_2} \\ V_{AM_1} & V_{AM_2} & V_A \end{bmatrix} \begin{bmatrix} \frac{\partial L_{1_m}}{\partial \hat{\mu}_{iI}} \\ \frac{\partial L_{1_m}}{\partial \hat{\mu}_{iII}} \\ \frac{\partial L_{1_m}}{\partial \hat{a}_i} \end{bmatrix}$$

$$= \frac{1}{4} \begin{bmatrix} \frac{-a_i}{\sigma_1} & \frac{-a_i}{\sigma_1} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{n_1}\Sigma_1 & 0 & V_{AM_1} \\ 0 & \frac{1}{n_2}\Sigma_2 & V_{AM_2} \\ V_{AM_1} & V_{AM_2} & V_A \end{bmatrix} \begin{bmatrix} \frac{-a_j}{\sigma_1} \\ \frac{-a_j}{\sigma_1} \\ 0 \end{bmatrix},$$

since

$$\frac{\partial L_{1_m}}{\partial \hat{a}_i} = \frac{\sigma_1 d_i - a_i d_i a_j \sigma_{ijI}/\sigma_1}{\sigma_1^2} = \frac{\sigma_1^2 d_i - \sigma_1^2 d_i}{\sigma_1^3} = 0.$$

Then

$$\operatorname{Var}(L_{1_m}) \doteq \frac{1}{4} \frac{1}{\sigma_1^2} \left[\frac{1}{n_1}(a_i \sigma_{ijI} a_j) + \frac{1}{n_2}(a_i \sigma_{ijII} a_j) \right]$$

$$= \frac{1}{4}\left(\frac{1}{n_1} + \frac{\alpha^2}{n_2}\right) . \tag{5.19.3}$$

In a similar fashion,

$$(5.19.4) \qquad L_{2_m} = \frac{\hat{m}_1 - \hat{m}_2 - 2\check{m}_2}{2\alpha\check{\sigma}_1} ,$$

and

$$\text{var}\,(L_{2_m}) \doteq \frac{1}{4}\left[\frac{\partial L_{2_m}}{\partial \hat{\mu}_{iI}} \cdot \frac{\partial L_{2_m}}{\partial \hat{\mu}_{iII}}\right] \begin{bmatrix} \frac{1}{n_1}\Sigma_1 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \frac{1}{n_2}\Sigma_2 \end{bmatrix} \begin{bmatrix} \frac{\partial L_{2_m}}{\partial \hat{\mu}_{jI}} \\ \cdots \\ \frac{\partial L_{2_m}}{\partial \hat{\mu}_{jII}} \end{bmatrix}$$

$$= \frac{1}{4\alpha^2}\left[\frac{a_i}{\sigma_1} \cdot \frac{a_i}{\sigma_1}\right] \begin{bmatrix} \frac{1}{n_1}\Sigma_1 & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \frac{1}{n_2}\Sigma_2 \end{bmatrix} \begin{bmatrix} \frac{a_j}{\sigma_1} \\ \cdots \\ \frac{a_i}{\sigma_1} \end{bmatrix}$$

$$(5.19.5) \qquad = \frac{1}{4\alpha^2}\left[\frac{1}{n_1} + \frac{\alpha^2}{n_2}\right] .$$

5.20. Variance of L_{1_σ} and L_{2_σ} .

$$(5.20.1) \qquad L_{1_\sigma} = \frac{\check{m}_1 - \hat{c}(\sigma)}{\check{\sigma}_1} = \frac{\hat{\sigma}_1(\check{m}_1 - \hat{m}_2) + \hat{\sigma}_2(\check{m}_1 - \hat{m}_1)}{\check{\sigma}_1(\hat{\sigma}_1 + \hat{\sigma}_2)} ,$$

where

$$\check{m}_1 = \hat{a}_i \check{\mu}_{iI},$$
$$\hat{m}_1 = \hat{a}_i \hat{\mu}_{iI},$$
$$\check{\sigma}_1 = (\hat{a}_i \sigma_{ijI} \hat{a}_j)^{\frac{1}{2}},$$
$$\hat{\sigma}_2 = (\hat{a}_i \hat{\sigma}_{ijII} \hat{a}_j)^{\frac{1}{2}},$$

and

$$\hat{\sigma}_1 = (\hat{a}_i \hat{\sigma}_{ijI} \hat{a}_j)^{\frac{1}{2}} .$$

Let

$$V^{**} = \begin{bmatrix} \frac{1}{n_1}\Sigma_1 & 0 & 0 & 0 & V_{AM_1} \\ 0 & \frac{1}{n_2}\Sigma_2 & 0 & 0 & V_{AM_2} \\ 0 & 0 & V_1 & 0 & V_{A_1} \\ 0 & 0 & 0 & V_2 & V_{A_2} \\ V_{AM_1} & V_{AM_2} & V_{A_1} & V_{A_2} & V_A \end{bmatrix} ;$$

then

$$\text{(5.20.2)} \qquad \operatorname{var}(L_{1\sigma}) \doteq \begin{bmatrix} \frac{\partial L_{1\sigma}}{\partial \hat{\mu}_{iI}} & \frac{\partial L_{1\sigma}}{\partial \hat{\mu}_{iII}} & \frac{\partial L_{1\sigma}}{\partial \hat{\sigma}_{ijI}} & \frac{\partial L_{1\sigma}}{\partial \hat{\sigma}_{ijII}} & \frac{\partial L_{1\sigma}}{\partial \hat{a}_i} \end{bmatrix} V^{**} \begin{bmatrix} \frac{\partial L_{1\sigma}}{\partial \hat{\mu}_{iI}} \\ \frac{\partial L_{1\sigma}}{\partial \hat{\mu}_{iII}} \\ \frac{\partial L_{1\sigma}}{\partial \hat{\sigma}_{ijI}} \\ \frac{\partial L_{1\sigma}}{\partial \hat{\sigma}_{ijII}} \\ \frac{\partial L_{1\sigma}}{\partial \hat{a}_i} \end{bmatrix}$$

Now

$$\text{(5.20.3)} \qquad \frac{\partial L_{1\sigma}}{\partial \hat{\mu}_{iI}} = \frac{-\sigma_2 a_i}{\sigma_1(\sigma_1+\sigma_2)} = \frac{-\sigma_1 \alpha a_i}{\sigma_1\sigma_1(1+\alpha)} = \frac{-a_i \alpha}{\sigma_1(1+\alpha)} ,$$

$$\text{(5.20.4)} \qquad \frac{\partial L_{1\sigma}}{\partial \hat{\mu}_{iII}} = \frac{-\sigma_1 a_i}{\sigma_1(\sigma_1+\sigma_2)} = \frac{-a_i}{\sigma_1(1+\alpha)} ,$$

$$\text{(5.20.5)} \qquad \frac{\partial L_{1\sigma}}{\partial \hat{\sigma}_{ijI}} = \frac{(\sigma_1+\sigma_2)a_i a_j(m_1+m_2) - a_i a_j \sigma_1(m_1-m_2)}{\sigma_1^2(\sigma_1+\sigma_2)^2}$$

$$= \frac{(m_1-m_2)a_i a_j \alpha}{\sigma_1^3(1+\alpha)^2}$$

$$= \frac{-a_i a_j}{\sigma_1} \frac{\alpha}{(1+\alpha)^2} ,$$

since $m_2 - m_1 = D\Sigma_1^{-1}D' = \sigma_1^2$,

$$(5.20.6) \qquad \frac{\partial L_{1\sigma}}{\partial \hat{\sigma}_{ijII}} = \frac{-\sigma_1(m_1-m_2)\sigma_1 \frac{a_i a_j}{\sigma_1}}{\sigma_1^2\sigma_1^2(1+\alpha)^2} = \frac{a_i a_j}{\sigma_1} \frac{1}{(1+\alpha)^2} ,$$

and

$$(5.20.7) \qquad \frac{\partial L_{1\sigma}}{\partial \hat{a}_i} = \sigma_1^2(1+\alpha)\left[\sigma_1(\mu_{iI}-\mu_{iII})+(m_1-m_2)\frac{\sigma_{ijI}a_j}{\sigma_1}\right]$$

$$-\sigma_1(m_1-m_2)\left[\sigma_1\frac{\sigma_{ijI}a_j}{\sigma_1} + \frac{\sigma_{ijII}a_j}{\sigma_2}\right.$$

$$\left. + (\sigma_1+\sigma_2)\frac{\sigma_{ijI}a_j}{\sigma_1}\right] \Big/ \sigma_1^2\sigma_1^2(1+\alpha)^2$$

$$= \frac{\sigma_1(1+\alpha)\left[\mu_{iI}-\mu_{iII}-\sigma_{ijI}a_j\right] + \sigma_1\left[\sigma_{ijI}a_j + \frac{1}{\alpha}\sigma_{ijII}a_j + (1+\alpha)\sigma_{ijI}a_j\right]}{\sigma_1^2(1+\alpha)^2}$$

$$= \frac{\mu_{iI}-\mu_{iII}-\sigma_{ijI}a_j}{\sigma_1} + \frac{2(1+\alpha)\sigma_{ijI}a_j}{\sigma_1(1+\alpha)}$$

$$= \frac{\mu_{iI}-\mu_{iII}+\sigma_{ijI}a_j}{\sigma_1}$$

$$= \frac{-d_i + \sigma_{ijI}\sigma^{ijI}d_i}{\sigma_1}$$

$$= 0.$$

If V_1^* denotes V^{**} with the last row and column omitted, or

$$V_1^{**} = \begin{bmatrix} \frac{1}{n_1}\Sigma_1 & 0 & 0 & 0 \\ 0 & \frac{1}{n_2}\Sigma_2 & 0 & 0 \\ 0 & 0 & V_1 & 0 \\ 0 & 0 & 0 & V_2 \end{bmatrix},$$

then

$$\text{var}(L_{1\sigma}) \doteq \frac{1}{(1+\alpha)^2}\left[-\frac{\alpha a_i}{\sigma_1} \cdots -\frac{a_i}{\sigma_1} \cdots -\frac{a_i a_j \alpha}{\sigma_1(1+\alpha)} \cdots \frac{a_i a_j}{\sigma_1(1+\alpha)}\right] V_1^{**} \begin{bmatrix} -\frac{\alpha a_i}{\sigma_1} \\ \cdots \\ -\frac{a_i}{\sigma_1} \\ \cdots \\ -\frac{a_i a_j \alpha}{\sigma_1(1+\alpha)} \\ \cdots \\ \frac{a_i a_j}{\sigma_1(1+\alpha)} \end{bmatrix}$$

$$= \frac{1}{(1+\alpha)^2}\left\{\alpha^2 \frac{a_i \sigma_{ijI} a_j}{n_1 \sigma_1^2} + \frac{a_i \sigma_{ijII} a_j}{n_2 \sigma_1^2}\right.$$

$$+ \frac{\alpha^2(n_1-1)}{(1+\alpha)^2} \frac{a_i a_j (\sigma_{ihI}\sigma_{jgI} + \sigma_{igI}\sigma_{jhI}) a_g a_h}{n_1^2 \sigma_1^2}$$

$$\left. + \frac{(n_2-1)}{\alpha^2(1+\alpha)^2} \frac{a_i a_j (\sigma_{ihII}\sigma_{igII} + \sigma_{jgII}\sigma_{jhII}) a_g a_h}{n_2^2 \sigma_1^2}\right\}$$

$$= \frac{1}{(1+\alpha)^2}\left\{\alpha^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) + \frac{\alpha^2(n_1-1)}{(1+\alpha)^2}\left(\frac{\sigma_1^2+\sigma_1^2}{n_1^2\sigma_1^2}\right)\right.$$

$$\left. + \frac{(n_2-1)}{\alpha^2(1+\alpha)^2}\left(\frac{\sigma_2^2+\sigma_2^2}{n_2^2\sigma_1^2}\right)\right\}$$

$$= \frac{1}{(1+\alpha)^2}\left[\alpha^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)+\frac{2(n_1-1)\alpha^2}{n_1^2(1+\alpha)^2}+\frac{2(n_2-1)}{n_2^2(1+\alpha)^2}\right]$$

$$= \frac{1}{(1+\alpha)^4}\left[\frac{\alpha^4+2\alpha^3+\alpha^2\left(\frac{3n_1-2}{n_1}\right)}{n_1}+\frac{\alpha^2(\alpha^2+2\alpha(+1)+2\left(\frac{n_2-1}{n_2}\right)}{n_2}\right] . \tag{5.20.8}$$

$$(L_{2\sigma}) = \frac{\hat{c}(\sigma)-\check{m}_2}{\check{\sigma}_2} = \frac{\hat{\sigma}_1(\hat{m}_2-\check{m}_2)+\hat{\sigma}_2(\hat{m}_1-\check{m}_2)}{\check{\sigma}_2(\hat{\sigma}_1+\hat{\sigma}_2)} . \tag{5.20.9}$$

$$\mathrm{Var}(L_{2\sigma}) \doteq \left[\frac{\partial L_{2\sigma}}{\partial \hat{\mu}_{iI}} \vdots \frac{\partial L_{2\sigma}}{\partial \hat{\mu}_{iII}} \vdots \frac{\partial L_{2\sigma}}{\partial \hat{\sigma}_{ijI}} \vdots \frac{\partial L_{2\sigma}}{\partial \hat{\sigma}_{ijII}} \vdots \frac{\partial L_{2\sigma}}{\partial \hat{a}_i}\right] V^{**} \begin{bmatrix} \frac{\partial L_{2\sigma}}{\partial \hat{\mu}_{iI}} \\ \cdots\cdots \\ \frac{\partial L_{2\sigma}}{\partial \hat{\mu}_{iII}} \\ \cdots\cdots \\ \frac{\partial L_{2\sigma}}{\partial \hat{\sigma}_{ijI}} \\ \cdots\cdots \\ \frac{\partial L_{2\sigma}}{\partial \hat{\sigma}_{ijII}} \\ \cdots\cdots \\ \frac{\partial L_{2\sigma}}{\partial \hat{a}_i} \end{bmatrix} , \tag{5.20.10}$$

$$\frac{\partial L_{2\sigma}}{\partial \hat{\mu}_{iI}} = \frac{\sigma_2 a_i}{\sigma_2(\sigma_1+\sigma_2)} = \frac{a_i}{\sigma_1(1+\alpha)} , \tag{5.20.11}$$

$$\frac{\partial L_{2\sigma}}{\partial \hat{\mu}_{iII}} = \frac{\sigma_1 a_i}{\sigma_2(\sigma_1+\sigma_2)} = \frac{a_i}{\sigma_1\alpha(1+\alpha)} , \tag{5.20.12}$$

$$\frac{\partial L_{2\sigma}}{\partial \hat{\sigma}_{ijI}} = \frac{-\sigma_2(m_1-m_2)\sigma_2\frac{a_i a_j}{\sigma_1}}{\sigma_2^2(\sigma_1+\sigma_2)^2} = \frac{d\, a_i a_j \alpha}{\sigma_1^3\alpha(1+\alpha)^2} = \frac{a_i a_j}{\sigma_1(1+\alpha)^2} , \tag{5.20.13}$$

$$(5.20.14)\qquad \frac{\partial L_{2\sigma}}{\partial \sigma_{ijII}} = \frac{\sigma_2(\sigma_2+\sigma_2)(m_1-m_2)\ \frac{a_i a_j}{\sigma_2} - \sigma_2(m_1-m_2)\sigma_2\frac{a_i a_j}{\sigma_2}}{\sigma_2^{\,2}(\sigma_1+\sigma_2)^2}\ ,$$

$$= \frac{(m_1-m_2)a_i a_j \sigma_1(1+\alpha-\alpha)}{\alpha\sigma_1^{\,4}(1+\alpha)^2} = \frac{da_i a_j}{\sigma_1^{\,3}(1+\alpha)^2\alpha}$$

$$= \frac{a_i a_j}{\sigma_1\alpha(1+\alpha)^2}\ ,$$

and

$$(5.20.15)\qquad \frac{\partial L_{2\sigma}}{\partial \hat{a}_i} = 0\ ,$$

by a derivation similar to that before. Then

$$\operatorname{var}(L_{2_\sigma}) \doteq \frac{1}{\sigma_1^{\,2}(1+\alpha)^2}\left[a_i \;\vdots\; \frac{a_i}{\alpha} \;\vdots\; \frac{a_i a_j}{(1+\alpha)} \;\vdots\; \frac{a_i a_j}{\alpha^2(1+\alpha)}\right] V_1^{**} \begin{bmatrix} a_j \\ \cdots\cdots \\ \frac{a_j}{\alpha} \\ \cdots\cdots \\ \frac{a_g a_h}{1+\alpha} \\ \cdots\cdots \\ \frac{a_g a_h}{\alpha^2(1+\alpha)} \end{bmatrix}$$

$$\text{Var}(L_{2_\sigma}) \doteq \frac{1}{\sigma_1^2(1+\alpha)^2}\left[\frac{a_i\sigma_{i,jI}a_j}{n_1} + \frac{a_i\sigma_{i,jII}a_j}{n_2} + \frac{2\sigma_1^2(n_1-1)}{(1+\alpha)^2 n_1^2} + \frac{2\sigma_2^2(n_2-1)}{\alpha^2(1+\alpha)^2 n_2^2}\right]$$

$$= \frac{1}{(1+\alpha)^2}\left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{n-1}{n_1^2}\,\frac{2}{(1+\alpha)^2} + \frac{n-1}{n_2^2}\,\frac{2}{(1+\alpha)^2}\right]$$

(5.20.16)
$$= \frac{1}{(1+\alpha)^4}\left[\frac{\alpha^2+2\alpha+\frac{(3n_1-2)}{n_1}}{n_1} + \frac{\alpha^2+2\alpha+\frac{(3n_2-2)}{n_2}}{n_2}\right].$$

5.21. Variance of $L_{1_{min}}$ and $L_{2_{min}}$

$$L_{1_{min}} = \frac{\check{m}_1 - \hat{c}(\min)}{\check{\sigma}_1} =$$

$$\frac{\check{m}_1\hat{\sigma}_1^2 - \check{m}_1\hat{\sigma}_2^2 - \hat{\sigma}_1^2\hat{m}_2 + \hat{\sigma}_2^2\hat{m}_1 + \hat{\sigma}_1\hat{\sigma}_2\sqrt{(\hat{m}_2-\hat{m}_1)^2 + 2(\hat{\sigma}_2^2-\hat{\sigma}_1^2)\ln\hat{\sigma}_2/\hat{\sigma}_1}}{\check{\sigma}_1(\hat{\sigma}_1^2-\hat{\sigma}_2^2)}.$$

Let

$$V^{***} = \begin{bmatrix} \text{var}(\hat{m}_1) & \text{cov}(\hat{m}_1,\hat{m}_2) & \text{cov}(\hat{m}_1,\hat{\sigma}_1) & \text{cov}(\hat{m}_1,\hat{\sigma}_2) \\ \text{cov}(\hat{m}_1,\hat{m}_2) & \text{var}(\hat{m}_2) & \text{cov}(\hat{m}_2,\hat{\sigma}_1) & \text{cov}(\hat{m}_2,\hat{\sigma}_2) \\ \text{cov}(\hat{m}_1,\hat{\sigma}_1) & \text{cov}(\hat{m}_2,\hat{\sigma}_1) & \text{var}(\hat{\sigma}_1) & \text{cov}(\hat{\sigma}_1,\hat{\sigma}_2) \\ \text{cov}(\hat{m}_1,\hat{\sigma}_2) & \text{cov}(\hat{m}_2,\hat{\sigma}_2) & \text{cov}(\hat{\sigma}_1,\hat{\sigma}_2) & \text{var}(\hat{\sigma}_2) \end{bmatrix}.$$

Then

$$\text{var}(L_{1_{min}}) \doteq \left[\frac{\partial L_{1_{min}}}{\partial\hat{m}_1}\;\frac{\partial L_{1_{min}}}{\partial\hat{m}_2}\;\frac{\partial L_{1_{min}}}{\partial\hat{\sigma}_1}\;\frac{\partial L_{1_{min}}}{\partial\hat{\sigma}_2}\right] V^{***} \begin{bmatrix} \frac{\partial L_{1_{min}}}{\partial\hat{m}_1} \\ \frac{\partial L_{1_{min}}}{\partial\hat{m}_2} \\ \frac{\partial L_{1_{min}}}{\partial\hat{\sigma}_1} \\ \frac{\partial L_{1_{min}}}{\partial\hat{\sigma}_2} \end{bmatrix}$$

$$\frac{\partial L_{1_{min}}}{\partial \hat{m}_1} = \frac{\sigma_2^2+\sigma_1\sigma_2\left[(m_1-m_2)^2+2(\sigma_2^2-\sigma_1^2)\ln \sigma_2/\sigma_1\right]^{-\frac{1}{2}}(m_1-m_2)}{\sigma_1(\sigma_1^2-\sigma_2^2)}$$

$$= \alpha^2\sigma_1 + \frac{\sigma_1(m_1-m_2)}{\left[(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha\right]^{\frac{1}{2}}}$$

$$\sigma_1^2(1-\alpha^2)$$

$$(5.21.3) \qquad = \frac{\alpha^2}{(1-\alpha^2)\sigma_1} + \frac{\alpha(m_1-m_2)}{\sigma_1(1-\alpha^2)\left[(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha\right]^{\frac{1}{2}}},$$

$$\frac{\partial L_{1_{min}}}{\partial \hat{m}_2} = \frac{-\sigma_1^2+(-1)\sigma_1\sigma_2\left[(m_1-m_2)^2+2(\sigma_2^2-\sigma_1^2)\ln\alpha\right]^{-\frac{1}{2}}(m_1-m_2)}{\sigma_1^3(1-\alpha^2)}$$

$$(5.21.4) \qquad = \frac{1}{(1-\alpha^2)\sigma_1} - \frac{\alpha(m_1-m_2)}{\sigma_1(1-\alpha^2)\left[(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha\right]^{\frac{1}{2}}},$$

$$\frac{\partial L_{1_{min}}}{\partial \hat{\sigma}_1} = \frac{1}{\sigma_1^2(\sigma_1^2-\sigma_2^2)^2}\Bigg\{\sigma_1(\sigma_1^2-\sigma_2^2)\Bigg[(m_1-m_2)2\sigma_1+\sigma_2\sqrt{(m_1-m_2)^2+2(\sigma_2^2-\sigma_1^2)\ln\frac{\sigma_2}{\sigma_1}}$$

$$+\sigma_1\sigma_2\left((m_1-m_2)^2+2(\sigma_2^2-\sigma_1^2)\ln\frac{\sigma_2}{\sigma_1}\right)^{-\frac{1}{2}}\left(-2\sigma_1\ln\frac{\sigma_2}{\sigma_1}-(\sigma_2^2-\sigma_1^2)\frac{1}{\sigma_1}\right)\Bigg]$$

$$-2\left[\sigma_1^2(m_2-m_1)+\sigma_1\sigma_2\sqrt{(m_1-m_2)^2+2(\sigma_2^2-\sigma_1^2)\ln\frac{\sigma_2}{\sigma_1}}\right]\sigma_1^2\Bigg\}$$

$$(5.21.5) \qquad = \frac{1}{\sigma_1^2(1-\alpha^2)^2}\left\{(\alpha^2-1)\left[2(m_2-m_1) - \alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}\right.\right.$$

$$\left.\left. - \frac{\sigma_1^2\alpha(2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right] - 2\left[m_2-m_1+\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}\right]\right\},$$

and

$$\frac{\partial L_{1_{min}}}{\partial \hat{\sigma}_2} = \frac{1}{\sigma_1^2(\sigma_1^2-\sigma_2^2)}\left\{\sigma_1(\sigma_2^2-\sigma_1^2)\left[\sigma_1\sigma_2\left((m_1-m_2)^2+2(\sigma_2^2-\sigma_1^2)\ln\frac{\sigma_2}{\sigma_1}\right)^{-\frac{1}{2}}\right.\right.$$

$$\left.\left(2\sigma_2\ln\,\sigma_2/\sigma_1+(\sigma_2^2-\sigma_1^2)\frac{1}{\sigma_2}\right)+\sigma_1\sqrt{(m_1-m_2)^2+2(\sigma_2^2-\sigma_1^2)\ln\frac{\sigma_2}{\sigma_1}}\right]$$

$$\left.+\left[\sigma_1^2(m_2-m_1)+\sigma_2\sigma_1\sqrt{(m_1-m_2)^2+2(\sigma_2^2-\sigma_1^2)\ln\frac{\sigma_2}{\sigma_1}}\right]2\sigma_1\sigma_2\right\}$$

$$(5.21.6) \qquad = \frac{1}{\sigma_1^2(1-\alpha^2)^2}\left\{(\alpha^2-1)\left[\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}\right.\right.$$

$$\left. + \frac{\sigma_1^2(2\alpha^2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right.$$

$$\left.\left. + 2\alpha\left[m_2-m_1+\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha-1)\ln\alpha}\right]\right\}.$$

Then substituting and expanding matrices,

$$(5.21.7)\qquad \operatorname{var}(L_{1_{\min}}) \doteq \frac{\alpha^2}{\sigma_1^2(1-\alpha^2)^2}\left(\alpha + \frac{m_1-m_2}{(1-\alpha^2)\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right)^2\left(\operatorname{var}(\hat{m}_1)\right)$$

$$+ \frac{1}{\sigma_1^2(1-\alpha^2)^2}\left(1 + \frac{\alpha(m_1-m_2)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right)^2\left(\operatorname{var}(\hat{m}_2)\right)$$

$$- \frac{1}{\sigma_1^4(1-\alpha^2)^2}\left[2(m_2-m_1) - \alpha\sqrt{(m_2-m_1)^2 + 2\sigma_1^2(\alpha^2-1)\ln\alpha}\right.$$

$$\left. + \frac{\sigma_2^2\,\alpha\,(2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2 + 2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right.$$

$$\left. - \frac{2(m_2-m_1)+2\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}{1-\alpha^2}\right]^2\left(\operatorname{var}(\hat{\sigma}_1)\right) - \frac{1}{\sigma_1^4(1-\alpha^2)^2}$$

$$\left[\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha} + \frac{\sigma_1^2(2\alpha^2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right.$$

$$\left. + \frac{2\alpha}{\alpha^2-1}(m_2-m_1+\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}\right]^2\left(\operatorname{var}(\hat{\sigma}_2)\right)$$

$$+ 2\frac{\alpha}{\sigma_1^2(1-\alpha^2)^2}\left(\alpha + \frac{m_1-m_2}{(1-\alpha^2)\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right)$$

$$\left(1 + \frac{\alpha(m_1-m_2)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right)\left(\operatorname{cov}(\hat{m}_1,\hat{m}_2)\right)$$

$$+2\frac{\alpha}{\sigma_1^3(1-\alpha^2)^3}\left(\alpha+\frac{m_1-m_2}{(1-\alpha^2)\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right)\left[2(m_2-m_1)\right.$$

$$-\alpha\sqrt{(m_2-m_1)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}+\frac{\sigma_2^2\alpha(2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}$$

$$\left.-\frac{2(m_2-m_1)+2\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}{1-\alpha^2}\right]\left(\operatorname{cov}(\hat{m}_1,\hat{\sigma}_1)\right)$$

$$+2\frac{\alpha}{\sigma_1^3(1-\alpha^2)^3}\left(\alpha+\frac{m_1-m_2}{(1-\alpha^2)\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right)^2\cdot$$

$$\cdot\left[\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}+\frac{\sigma_1^2(2\alpha^2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right.$$

$$\left.+\frac{2\alpha}{\alpha^2-1}(m_2-m_1)+\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}\right]\left(\operatorname{cov}(\hat{m}_1,\hat{\sigma}_2)\right)$$

$$+2\frac{1}{\sigma_1^3(1-\alpha^2)^3}\left(1+\frac{(m_1-m_2)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right)\left[2(m_2-m_1)\right.$$

$$-\alpha\sqrt{(m_2-m_1)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}+\frac{\sigma_2^2\alpha(2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}$$

$$\left.+\frac{2(m_2-m_1)+2\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}{1-\alpha^2}\right]\left(\operatorname{cov}(\hat{m}_2,\hat{\sigma}_1)\right)$$

$$+2\frac{1}{\sigma_1^3(1-\alpha^2)^3}\left(1+\frac{\alpha(m_1-m_2)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right.$$

$$\left[\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}+\frac{\sigma_1^2(2\alpha^2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right.$$

$$\left.+\frac{2\alpha}{\alpha^2-1}(m_2-m_1+\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}\right]\left(\operatorname{cov}(\hat{m}_2,\hat{\sigma}_2)\right)$$

$$+2\frac{1}{\sigma_1^2(1-\alpha^2)^3}\left[2(m_2-m_1)-\alpha\sqrt{(m_2-m_1)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}\right.$$

$$+\frac{\sigma_2^2\alpha(2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}$$

$$\left.+\frac{2(m_2-m_1)+2\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}{1-\alpha^2}\right]$$

$$\left[\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}+\frac{\sigma_1^2(2\alpha^2\ln\alpha+\alpha^2-1)}{\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}}\right.$$

$$\left.+\frac{2\alpha}{\alpha^2-1}(m_2-m_1+\alpha\sqrt{(m_1-m_2)^2+2\sigma_1^2(\alpha^2-1)\ln\alpha}\right]\left(\operatorname{cov}(\hat{\sigma}_1,\hat{\sigma}_2)\right)$$

where the variances and covariances of $\hat{m}_1, \hat{m}_2, \hat{\sigma}_1$, and $\hat{\sigma}_2$ were given in previous sections. They include variances and covariances of the $\hat{a}$'s, and are therefore not written out.

It becomes clear that $\text{var}(L_{1_{min}})$ is not a simple form and will not simplify appreciably.

$\text{Var}(L_{2_{min}})$ will be of a similar form with similar complications.

5.22. $E(\hat{P})$

We have found the first two moments of the limits of the integrals giving the conditional probabilities of misclassification, $\hat{P}_I$ and $\hat{P}_{II}$. Let us look at the expectations of $\hat{P}$'s, where $\hat{P} = \hat{P}_I + \hat{P}_{II}$. $\hat{P}(m)$, $\hat{P}(\sigma)$, and $\hat{P}(\text{min})$ were defined in 5.1.

Recall that the upper limit of integration of $\hat{P}_I$, $\hat{P}_{II}$ are designated $\hat{L}_1$, $\hat{L}_2$ respectively. Then

$$E(\hat{P}_I) \doteq P_I + \tfrac{1}{2} \frac{\partial^2 \hat{P}_I}{\partial \hat{L}_1^2} \Bigg|_{\hat{L}_1 = L_1} \left(\text{var}(\hat{L}_1)\right) \qquad \text{(from 5.11)}$$

$$= P_I + \tfrac{1}{2} \frac{\partial}{\partial L_1} \left(e^{-\frac{1}{2}L_1^2}\right) \left(\text{var}(\hat{L}_1)\right)$$

$$= P_I - L_1 e^{-\frac{1}{2}L_1^2} \left(\text{var}(\hat{L}_1)\right) . \qquad (5.22.1)$$

Similarly,

$$E(\hat{P}_{II}) \doteq P_{II} - L_2 e^{-\frac{1}{2}L_2^2} \left(\text{var}(\hat{L}_2)\right) . \qquad (5.22.2)$$

Then

$$(5.22.3) \qquad E(\hat{P}) \doteq P_I + P_{II} - L_1 e^{-\frac{1}{2}L_1^2}\left(\mathrm{var}(L_1)\right) - L_2 e^{-\frac{1}{2}L_2^2}\left(\mathrm{var}(L_2)\right) .$$

Note that L_1 and L_2 are negative, so that $E(\hat{P}) > \hat{P}_I + \hat{P}_{II}$.

Specifically, for $\hat{P}(m)$,

$$(5.22.4) \qquad E\left[\hat{P}(m)\right] = P_I(m) + P_{II}(m) - L_{1_m} e^{-\frac{1}{2}L_{1_m}^2}\left(\mathrm{var}(L_{1_m})\right) - L_{2_m} e^{-\frac{1}{2}L_{2_m}^2}\left(\mathrm{var}(L_{2_m})\right)$$

$$= P_I(m) + P_{II}(m) - \alpha^3 L_{2_m} e^{-\frac{1}{2}\alpha^2 L_{2_m}^2}\left(\mathrm{var}(L_{2_m})\right) - L_{2_m} e^{-\frac{1}{2}L_{2_m}}\left(\mathrm{var}(L_{2_m})\right)$$

$$= P_I(m) + P_{II}(m) - L_{2_m}\left(\mathrm{var}(L_{2_m})\right)\left\{\alpha^3 e^{-\frac{1}{2}L_{2_m}^2\alpha^2} + e^{-\frac{1}{2}L_{2_m}^2}\right\}$$

$$= \int_{-\infty}^{\frac{m_1-m_2}{2\sigma_1}} N(0,1)\,dx + \int_{-\infty}^{\frac{m_1-m_2}{2\alpha\sigma_1}} N(0,1)\,dx$$

$$- \frac{m_1-m_2}{8\alpha^3\sigma_1}\left[\frac{1}{n_1} + \frac{\alpha^2}{n_2}\right]\left[\alpha^3 e^{-\frac{(m_1-m_2)^2}{8\sigma_1^2}} + e^{-\frac{(m_1-m_2)^2}{8\alpha^2\sigma_1^2}}\right] .$$

For $\hat{P}(\sigma)$ (recalling that $L_{1_\sigma} = L_{2_\sigma}$),

$$(5.22.5) \qquad E\,\hat{P}(\sigma) = P_I(\sigma) + P_{II}(\sigma) - L_{1_\sigma} e^{-\frac{1}{2}L_1^2\sigma}\left(\mathrm{var}(L_{1_\sigma})\right) - L_{2_\sigma} e^{-\frac{1}{2}L_2^2\sigma}\left(\mathrm{var}(L_{2_\sigma})\right)$$

$$= 2P_I(\sigma) - L_{1_\sigma} e^{-\frac{1}{2}L_{1_\sigma}^2}\left(\mathrm{var}(L_{1_\sigma}) + \mathrm{var}(L_{2_\sigma})\right)$$

$$= 2\int_{-\infty}^{\frac{m_1-m_2}{\sigma_1+\sigma_2}} N(0,1)\,dx - \left(\frac{m_1-m_2}{\sigma_1+\sigma_2}\right) e^{-\frac{1}{2}\left(\frac{m_1-m_2}{\sigma_1+\sigma_2}\right)^2}$$

$$\left[\frac{1+\alpha^2}{(1+\alpha)^4}\left(\frac{\alpha^2+2\alpha+\frac{3n_1-2}{n_1}}{n_1} + \frac{\alpha^2+2\alpha+\frac{3n_2-2}{n_2}}{n_2}\right)\right] .$$

For $\hat{P}(\min)$, we note that $\text{var}(L_{1_{\min}})$ and $\text{var}(L_{2_{\min}})$ are present, as are complicated forms in $L_{1_{\min}}$ and $L_{2_{\min}}$ themselves. Furthermore, while the other two forms can be reduced to functions of the parameters α, $\frac{m_2-m_1}{\sigma_1}$, n_1 and n_2, inspection of $\hat{P}(\min)$ reveals that these plus the additional parameter σ_1^2 are necessary. Further evaluation of $\hat{P}(\min)$ is seen to be unfeasible at this time.

Now $E(\hat{P})$ has a lower bound, say P_o.

(5.22.6) $$P_o = P(\min) = P_I(\min) + P_{II}(\min)$$
$$= \int_{-\infty}^{\frac{m_1-c(\min)}{\sigma_1}} N(0,1)dx + \int_{-\infty}^{\frac{c(\min)-m_2}{\sigma_2}} N(0,1)dx.$$

Inspection of $E\left[\hat{P}(\min)\right]$, $E\left[\hat{P}(\sigma)\right]$, and $E\left[\hat{P}(m)\right]$ reveals that $E\left[\hat{P}(\min)\right] = P_o + O(1/n_1, 1/n_2)$, whereas $E\left[\hat{P}(m)\right]$, $E\left[\hat{P}(\sigma)\right]$ are $P(m) + O(1/n_1, 1/n_2)$, $P(\sigma) + O(1/n_1, 1/n_2)$ respectively, where $P(m), P(\sigma) > P_o$. $\hat{P}$ has a range bounded below at P_o irrespective of which c is used.

Assume that samples are large enough that gross mistakes do not occur; by a gross mistake is meant that for a linear discriminator $(\hat{m}_1 - \hat{m}_2)(\check{m}_1 - \check{m}_2) < 0$, i.e. that the wrong side of the discriminator is used. Now

(5.22.7) $$\text{Prob}\left[\hat{P} \geq P_o + a\right] \leq \frac{h}{a},$$

where h is small. If, for $\hat{P}$ and any other estimate of P, say $\hat{P}_1$, $E(\hat{P}) < E(\hat{P}_1)$, then $\hat{P}$ has the lesser upper bound. Consider the

difference between the two upper bounds. If this difference is less than $E(\hat{P})$, a comparison of $\hat{P}$ and $\hat{P}_1$ is trivial. If this difference is large compared to $E(\hat{P}) - P_o$, then $\hat{P}$ will be said to have a sampling distribution better than that of $\hat{P}_1$.

Thus, we should like to compare $E\left[\hat{P}(m)\right]$ with $E\left[\hat{P}(\sigma)\right]$. If $E\left[\hat{P}(\sigma)\right] < E\left[\hat{P}(m)\right]$, say, then $\hat{c}(\sigma)$ is better than $\hat{c}(m)$. Since $E\left[\hat{P}(\min)\right]$ has such a complicated form and involves an extra parameter, comparisons involving $E\left[\hat{P}(\min)\right]$ will not be made.

5.23. Comparison of $E\left[\hat{P}(\min)\right]$ and $E\left[\hat{P}(\sigma)\right]$

As was noted previously, when $E\left[\hat{P}(m)\right] > E\left[\hat{P}(\sigma)\right]$, $\hat{c}(\sigma)$ is the estimate preferred. Similarly, when $E\left[\hat{P}(m)\right] < E\left[\hat{P}(\sigma)\right]$, $\hat{c}(m)$ is preferred. Let us consider the function

$$g_3\left[\hat{P}(m), \hat{P}(\sigma)\right] = E\left[\hat{P}(m)\right] - E\left[\hat{P}(\sigma)\right]. \tag{5.23.1}$$

Let us use again the parameter $K = \frac{m_2 - m_1}{\sigma_1 + \sigma_2}$. Then combining integrals and putting in a form similar to that of section 4.2.,

$$g_3\left[\hat{P}(m),\hat{P}(\sigma)\right] = g_3(K,\alpha) = 2\int_0^K N(0,1)\,dx - \int_0^{\frac{\alpha+1}{2\alpha}K} N(0,1)\,dx - \int_0^{\frac{\alpha+1}{2}K} N(0,1)\,dx$$

$$- \frac{(1+\alpha)K}{8\alpha^3}\left[\frac{1}{n_1} + \frac{\alpha^2}{n_2}\right]\left[\alpha^3 e^{-\frac{(1+\alpha)^2}{8}K^2} + e^{-\frac{(1+\alpha)^2}{8\alpha^2}K^2}\right]$$

$$-Ke^{-\frac{1}{2}K^2}\frac{1+\alpha^2}{(1+\alpha)^4}\left[\frac{\alpha^2+2\alpha+\frac{3n_1-2}{n_1}}{n_1} + \frac{\alpha^2+2\alpha+\frac{3n_2-2}{n_2}}{n_2}\right]$$

$$(5.23.2) \qquad = 2\int_0^K N(0,1)\,dx - \int_0^{\frac{\alpha+1}{2\alpha}K} N(0,1)\,dx - \int_0^{\frac{\alpha+1}{2}K} N(0,1)\,dx$$

$$- \frac{\sqrt{2\pi}\,K}{n_1 n_2}\left\{ \frac{(1+\alpha)(n_1\alpha^2+n_2)}{8\alpha^3}\left[\alpha^3\phi\left(\frac{1+\alpha}{2}K\right) + \phi\left(\frac{1+\alpha}{2\alpha}K\right)\right]\right.$$

$$\left. - \frac{1+\alpha^2}{(1+\alpha)^4}\left[n_1\left(\alpha^2+2\alpha+\frac{3n_2-2}{n_2}\right) + n_2\left(\alpha^2+2\alpha+\frac{3n_1-2}{n_1}\right)\right]\phi(K)\right\},$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ for evaluation from tables.

Two tables of values of $g_3(K,\alpha)$ have been constructed, one for equal sample sizes and the other for sample sizes proportionate to the variances.

Table 2 consists of values of $g_3(K,\alpha)$ for various K, especially those K-values near the change of sign of g_3, and for various $\alpha \geq 1$. In Table 2, $n_1 = n_2 = 50$. Since the Table 2 entries do not differ markedly from those of the asymptotically large sample case, Table 1, values for large sample sizes, where $n_1 \doteq n_2$, may be interpolated to obtain an approximate value. This interpolated value will be a very rough approximation, but should be adequate for practical usage.

Table 3 is similar to Table 2, differing only in that $n_2 \doteq \alpha^2 n_1$. Recalling that $\Sigma_2 = \alpha^2\Sigma_1$, we note that $n_2 \doteq \alpha^2 n_1$ yields sample sizes proportionate to the variances. Again the results are not markedly different from those of the asymptotically

large sample case, Table 1; again interpolation will yield approximations which are very rough but which should be adequate for practical usage.

It should be noted that in both tables, for increasing α the values tend to become exaggerated, but that the sign of g_3 changes at about the same K, i.e. $1.1 < K < 1.4$. Generally the values of K and α associated with the change in the sign of g_3 do not differ greatly from parameter values for the change of sign for the asymptotic case.

TABLE 2

VALUES OF $g_3(K, \alpha)$ COMPARING $E\left[\hat{P}(m)\right]$ AND $E\left[\hat{P}(\sigma)\right]$ FOR VARIOUS VALUES OF K AND α WHERE $n_1 = n_2 = 50$.

K \ α	1	2	4	10
0	0	0	0	0
0.5	.002	- .042	- .196	- .271
1.0	.004	- .028	- .051	- .027
1.2	.003	- .012	+ .003	+ .026
1.4	.002	+ .002	.032	.060
2.0	.001	.023	.060	.090
∞	0	0	0	0

TABLE 3

VALUES OF $g_3(K, \alpha)$ COMPARING $E\left[\hat{P}(m)\right]$ AND $E\left[\hat{P}(\sigma)\right]$ FOR VARIOUS VALUES OF K AND α WHERE $n_2 = \alpha^2 n_1$.

α	1	2	4	10
n_1	50	50	50	50
n_2	50	200	800	5000
K				
0	0	0	0	0
0.5	.002	- .133	- .138	- .222
1.0	.004	- .042	- .039	- .024
1.2	.003	- .006	+ .004	+ .027
1.4	.002	+ .027	.037	.061
2.0	.001	.021	.063	.091
∞	0	0	0	0

5.24. $E(\hat{P}_L)$

Using criterion (ii), we wish to compare $E(\hat{P}_L)$, where $\hat{P}_L = \max(\hat{P}_I, \hat{P}_{II})$. $\hat{P}_L(m) = \hat{P}_{II}(m)$ (4.4.2) and $\hat{P}_L(\min) = \hat{P}_{II}(\min)$ (4.4.4) when samples are large; $\hat{P}_{II}(\min)$ is not of a form suitable for comparison.

Now $P_I(\sigma) = P_{II}(\sigma)$, so that sampling fluctuation may yield either $\hat{P}_L(\sigma) = \hat{P}_I(\sigma)$ or $\hat{P}_L(\sigma) = \hat{P}_{II}(\sigma)$.

$$E\left[\hat{P}_L(\sigma)\right] = P_I(\sigma) + \text{some function of the variances, covariance of } \hat{P}_I(\sigma) \text{ and } \hat{P}_{II}(\sigma).$$

Thus, the identification of $\hat{P}_L(\sigma)$ depends on the samples and has not been made. Then no comparison is made on the basis of criterion (ii).

VI. SUMMARY

6.1. Summary

Linear discriminant analysis is the classification of an individual as having arisen from one or the other of two populations on the basis of a scalar linear function of measurements of the individual. This paper is a population and large sample study of linear discriminant analysis. The population study is carried out at three levels:

(6.1.1) (a) with loss functions and prior probabilities,

(6.1.2) (b) without loss functions but with prior probabilities,

(6.1.3) (c) with neither.

The first level leads to consideration of risks which may be split into two components, one for each type of misclassification, i.e. classification of an individual into population I given it arose from population II, and classification of it into II given it arose from I. Similarly, the second level leads to consideration of expected errors and the third level leads to consideration of conditional probabilities of misclassification, both again which may be divided into the same two components. At each level the "optimum" discriminator should jointly minimize the two probability components. These quantities are all positive for all hyperplanes. Either one of any pair may be made equal to zero by classifying all individuals of a sample into the appropriate population; but

this maximizes the other one. Consequently, joint minimization must be obtained by some compromise, e.g. by selecting a single criterion to be minimized. Two types of criteria for judging discriminators are considered at each level:

(6.1.4) (i) Total risk (a)

(6.1.5) Total expected errors (b)

(6.1.6) Sum of conditional probabilities of misclassification (c)

(6.1.7) (ii) Larger risk (a)

(6.1.8) Larger expected error (b)

(6.1.9) Larger conditional probability of misclassification (c).

These criteria are not particularly new, but have not been applied to linear discrimination and not been all used jointly.

If A is a k-dimensional row vector of direction numbers, X a k-dimensional row vector of variables, and c a constant, a linear discriminator is

(6.1.10) $AX' = c,$

which also represents a hyperplane in k-space. An individual is classified as being from one or the other population on the basis of its position relative to the hyperplane.

Let M_1, M_2 be the row vectors of k means from populations I, II respectively, Σ_1, Σ_2 be the dispersion matrices for populations I, II respectively, $D = M_2 - M_1$, and p, (1 - p) be the a priori probabilities that an individual has arisen from populations I, II respectively. Let $m = AM_1'$, $m_2 = AM_2'$, $\sigma_1^2 = A\Sigma_1A'$, and $\sigma_2^2 = A\Sigma_2A'$.

Using the optimum criteria (i) and (ii), the same set A is obtained for each independently, and is independent of c;

(6.1.11) $$A = D\Sigma^{-1},$$

where $\Sigma = \Sigma_1$ or $\Sigma = \Sigma_2$ if $\Sigma_1 = \alpha^{-2}\Sigma_2$, a constant. Explicit solutions are not available for general non-proportionate matrices.

The appropriate c is derived for each criterion. Criterion (i) yields a c with minimum total probability of misclassification,

(6.1.12) $$c(\min) = \frac{1}{\sigma_1^2-\sigma_2^2}\left[\sigma_1^2 m_2 - \sigma_2^2 m_1 - \sigma_1\sigma_2\sqrt{(m_2-m_1)^2+2(\sigma_2^2-\sigma_1^2)\ln\frac{p\sigma_2}{(1-p)\sigma_1}}\right].$$

Criterion (ii) yields a c for which the probabilities of misclassification are equal for the two populations,

(6.1.13) $$c(\sigma) = \frac{m_1\sigma_2 + m_2\sigma_1}{\sigma_1 + \sigma_2}.$$

If $\Sigma_1 = \Sigma_2$, $c(\sigma)$ and $c(\min)$ both reduce to

(6.1.14) $$c(m) = \frac{m_1 + m_2}{2},$$

the c used in practice heretofore. Loss functions are considered for both A and c and for both criteria, but explicit results may be obtained only for special cases of loss functions.

At level (c), the c's for criteria (i) and (ii), c(min) and $c(\sigma)$ respectively, were compared to c(m) to discover the conditions under which it was better(i.e. having lesser criteria) than both c(min) and $c(\sigma)$ on criteria (ii) and (i) respectively. The comparisons were made by an inequality reduced to a function of

two parameters, $K = (m_2 - m_1)/(\sigma_1 + \sigma_2)$ and $\alpha = \sigma_2/\sigma_1$, and values for this inequality were tabulated.

In the large sample study, variances and covariances were found (in many cases approximately) for all estimates of the parameters entering into the conditional probabilities of misclassification (level (c)). Extension of results to level (b) and to special cases of level (a) were given. From these variances and covariances were derived expectations of these probabilities for both criteria, at level (c), and comparisons were made where feasible. Results were tabulated.

VII. ACKNOWLEDGMENTS

The author wishes to express his deep and sincere appreciation to Professor Charles W. Clunies-Ross for his unending assistance, guidance, and encouragement and for the many things he taught the author. The author also wishes to thank Dr. Boyd Harshbarger and the other members of the Department of Statistics for their continual help and cooperation.

The author wishes to express appreciation to the National Cancer Institute of the United States Public Health Service whose two-year research grant provided much of the financial aid necessary for him to continue his studies.

The author is also grateful to his wife, Gerrye, who helped type this thesis.

REFERENCES

1. Aitken, A. C. (1948). Determinants and Matrices, 5th edit., Oliver and Boyd, London.

2. Barnard, M. M. (1935). "The Secular Variations of Skull Characters in Four Series of Egyptian Skulls", Annals of Eugenics, 6, 352-71.

3. Brown, G. W. (1950). "Basic Principles for Construction and Application of Discriminators", Journal of Clinical Psychology, 6, 58-61

4. Clunies-Ross, C. W. (1957). Unpublished paper, Va. Polytechnic Institute.

5. Cramer, H. (1946). Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey.

6. Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems", Annals of Eugenics, 7, 179-88.

7. Fisher, R. A. (1938). "The Statistical Utilization of Multiple Measurements", Annals of Eugenics, 8, 376-86.

8. Lindley, D. V. (1953). "Statistical Inference", Journal of the Royal Statistical Society, Series B, 15, 30-65.

9. Polya, G. (1945). "Remarks on Computing the Probability Integral in One and Two Dimensions", Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley.

10. Smith, C. A. B. (1947). "Some Examples of Discrimination", Annals of Eugenics, 13, 272-82.

11. Smith, C. A. B. (1954). Biomathematics, Hafner Pub. Co., New York.

12. Welch, B. L. (1939). "Note on Discriminant Functions", Biometrika, 31, 218-20.

The vita has been removed from the scanned document

X. APPENDIX

10.1. Simplification of a certain bivariate integral

Suppose it is required to evaluate

$$(10.1.1)\qquad I = \iint_{f(x,y,)<c} (2\pi\sigma_x\sigma_y\sqrt{1-\rho^2})^{-1} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dx\, dy$$

$$(10.1.2)\qquad = \int_{k_1(x)}^{k_o(x)} dx \int_{f(x,y)}^{h_o(x)} (2\pi\sigma_x\sigma_y\sqrt{1-\rho^2})^{-1} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dy$$

where $f(x,y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$. The following sequence of transformations will reduce the multiple integral to a single integral having simple limits of integration.

Transformation (1) (this transformation normalizes the density):

$$(10.1.3)\qquad u = \frac{x-\mu_x}{\sigma_x}\,, \quad v = \frac{y-\mu_y}{\sigma_y}\,, \quad dx = \sigma_x du\,, \quad dy = \sigma_y\, dv.$$

Then

$$(10.1.4)\qquad I = \iint_R \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right\} du\, dv,$$

where R is the conic section $A'u^2 + B'uv + C'v^2 + D'u + E'v + F' = 0$,

where $A' = A\sigma_x^2$,

$B' = B\sigma_x\sigma_y$,

$C' = C\sigma_y^2$,

$D' = 2A\sigma_x\mu_x + B\sigma_x\mu_y + D\sigma_x$,

$$E' = B\sigma_y\mu_x + 2C\sigma_y\mu_y + E\sigma_y ,$$

and

$$F' = A\mu_x^2 + B\mu_x\mu_y + C\mu_y^2 + D\mu_x + E\mu_y + F .$$

Transformation (2) (this transformation eliminates ρ from the integrand):

$$u = s\sqrt{\frac{1+\sqrt{1-\rho^2}}{2}} + t\sqrt{\frac{1-\sqrt{1-\rho^2}}{2}} \qquad (10.1.5)$$

and

$$v = s\sqrt{\frac{1-\sqrt{1-\rho^2}}{2}} + t\sqrt{\frac{1+\sqrt{1-\rho^2}}{2}} \qquad (10.1.6)$$

$$J = \sqrt{1-\rho^2} .$$

The region of integration is a conic of the form

$A'' s^2 + B'' st + C'' t^2 + D'' s + E'' t + F'' = 0$, where

$$A'' = A' \tfrac{1}{2}(1+\sqrt{1-\rho^2}) + B' \tfrac{1}{2}\rho + C' \tfrac{1}{2}(1-\sqrt{1-\rho^2}) ,$$

$$B'' = A' \tfrac{1}{2}\rho + B' + C' \tfrac{1}{2}\rho ,$$

$$C'' = A' \tfrac{1}{2}(1-\sqrt{1-\rho^2}) + B' \tfrac{1}{2}\rho + C' \tfrac{1}{2}(1+\sqrt{1-\rho^2}) ,$$

$$D'' = D' \tfrac{1}{2}(1+\sqrt{1-\rho^2}) + E' \tfrac{1}{2}(1-\sqrt{1-\rho^2}),$$

$$E'' = D' \tfrac{1}{2}(1-\sqrt{1-\rho^2}) + E' \tfrac{1}{2}(1+\sqrt{1-\rho^2}),$$

and

$$F'' = F' .$$

The probability element becomes

$$\text{p.e.} = \frac{1}{2\pi} \exp\left\{-\tfrac{1}{2}(s^2+t^2)\right\} ds\, dt . \qquad (10.1.7)$$

Transformation (3) (this transformation eliminates the cross-product term from the limit of integration):

(10.1.8) $s = g \cos \theta - w \sin \theta$

and

(10.1.9) $t = g \sin \theta + w \cos \theta,$

where θ is a constant such that

$$\theta = \frac{1}{2} \tan^{-1} \frac{B''}{A'' - C''}, \quad A'' \neq C''$$

$$= 45^{\circ}, \quad A'' = C''.$$

$$J = \left| \frac{\partial(s,t)}{\partial(g,w)} \right| = \begin{vmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{vmatrix} = \cos^2 \theta + \sin^2 \theta = 1.$$

Then

(10.1.11) $$e.p. = \frac{1}{2\pi} \exp\left\{ -\frac{1}{2} (g^2 + w^2) \right\} dg\, dw;$$

the region of integration becomes $A''' g^2 + C''' w^2 + D''' g + E''' w + F''' = 0,$

where $A''' = A'' \cos^2 \theta + C'' \sin^2 \theta + B'' \sin \theta \cos \theta,$

$C''' = A'' \sin^2 \theta + C'' \cos^2 \theta - B'' \sin \theta \cos \theta,$

$D''' = D'' \cos \theta + E'' \sin \theta,$

$E''' = E'' \cos \theta - D'' \sin \theta,$

and

$F''' = F'' = F'.$

Transformation (4) (this transformation converts the variables to polar coordinates):

(10.1.12) $g = r \cos \phi$

and

(10.1.13) $w = r \sin \phi.$

$$J = r.$$

$$\text{(10.1.14)} \qquad e.p. = \frac{r}{2\pi} \exp\left\{-\frac{1}{2} r^2\right\} dr\, d\phi;$$

the region of integration becomes

$$A''' r^2 \cos^2 \phi + C''' r^2 \sin^2 \phi + D''' r \cos \phi + E''' r \sin \phi + F''' = 0.$$

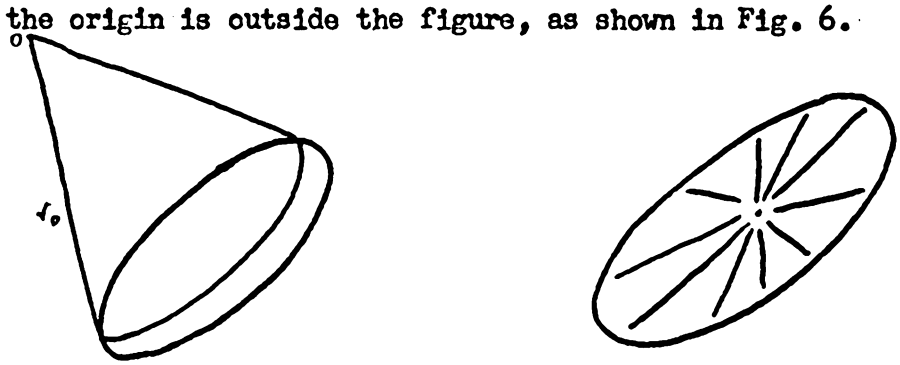
The region of integration may be found for r as a function of ϕ.

$$(A''' \cos^2 \phi + C''' \sin^2 \phi)\, r^2 + (D''' \cos \phi + E''' \sin \phi)\, r + F''' = 0.$$

Then

$$\text{(10.1.15)} \qquad r = \frac{-(D''' \sin\phi + E''' \cos\phi) \pm \sqrt{(D'''^2 - 4F'''A''')\sin^2\phi + (E'''^2 - 4C'''F''')\cos^2\phi - 2D'''E''' \sin\phi\cos\phi}}{2(A''' \sin^2 \phi + C''' \cos^2 \phi)}$$

Call the value r of (10.1.15) R_+ when the positive root is used and R_- when the negative root is used. There are two cases of integrals, one in which the polar origin is inside the figure and the other in which the origin is outside the figure, as shown in Fig. 6.



outside inside

Fig. 6

For the inside case,

$$\text{(10.1.16)} \qquad I = \int_0^{2\pi} \int_0^{R_+} \frac{r}{2\pi} \exp\left\{-\frac{1}{2} r^2\right\} dr\, d\phi$$

$$= 1 - \int_0^{2\pi} \frac{1}{2\pi} \exp\left\{-\frac{1}{2} R_+^{\,2}\right\} d\phi\,.$$

For the outside case,

$$(10.1.17)\qquad I = \int_{\phi_o}^{\phi_1}\int_{R_-}^{r_o} \frac{r}{2\pi}\exp\left\{-\tfrac{1}{2}r^2\right\} dr\, d\phi + \int_{\phi_o}^{\phi_1}\int_{r_o}^{R_+} \frac{r}{2\pi}\exp\left\{-\tfrac{1}{2}r^2\right\} dr\, d\phi$$

$$= \frac{1}{2\pi}\left[\int_{\phi_o}^{\phi_1} \exp\left\{-\tfrac{1}{2}R_+^{\,2}\right\} d\phi - \int_{\phi_o}^{\phi_1} \exp\left\{-\tfrac{1}{2}R_-^{\,2}\right\} d\phi\right].$$

The basic integral is now

$$(10.1.18)\qquad \int_{\phi_o}^{\phi_1} \exp\left\{-\tfrac{1}{2}\left(\frac{1}{2(A'''\sin^2\phi + C'''\cos^2\phi}\right)\left[-(D'''\sin\phi + E'''\cos\phi)\right.\right.$$

$$\left.\left. + \sqrt{(D'''^2 - 4F'''A''')\sin^2\phi + (E'''^2 - 4C'''F''')\cos^2\phi - 2D'''E'''\sin\phi\cos\phi}\right]\right\}^2 d\phi$$

$$= \int_{\phi_o}^{\phi_1} \exp\left\{-\tfrac{1}{8}\left(\frac{1}{k_9\sin^4\phi + k_{10}\sin^2\cos^2\phi + k_{11}\cos^4\phi}\right)\left[k_1\sin^2\phi + k_2\cos^2\phi + k_3\sin\phi\cos\phi\right.\right.$$

$$\left.\left. + \sqrt{k_4\sin^4\phi + k_5\cos^4\phi + k_6\sin^2\phi\cos^2\phi + k_7\sin^3\phi\cos\phi + k_8\sin\phi\cos^3\phi}\right]\right\} d\phi$$

where

$$K_1 = 2D'''^2 - 4F'''A'''$$

$$K_2 = 2E'''^2 - 4C'''F'''$$

$$K_3 = 0$$

$$K_4 = 4D'''^2\,(D'''^2 - 4A'''F''')$$

$$K_5 = 4E'''^2\,(E'''^2 - 4C'''F''')$$

$$K_6 = 4D'''^2\,(E'''^2 - 4C'''F''') + 4E'''^2\,(D'''^2 - 4A'''F''')$$

$$K_7 = -8D'''^3\,E'''$$

$$K_8 = -8D''' E'''^3$$

$$K_9 = 4A'''^2$$

$$K_{10} = 8A''' C'''$$

and

$$K_{11} = 4C'''^2$$

If we transform

$$\sin \phi = \frac{2z}{1+z^2}, \quad \cos \phi = \frac{1-z^2}{1+z^2}, \quad d\phi = \frac{2dz}{1+z^2},$$

We obtain

$$(10.1.19) \quad \int_{z_o}^{z_1} \exp\left\{ -\frac{(1+z^2)}{8}^2 \left(\frac{1}{h_9 z^8 + h_{10} z^6 + h_{11} z^4 + h_{10} z^2 + h_9} \right) \left[h_1 z^4 + h_3 z^2 - h_2 z + h_1 + \sqrt{h_4 z^8 - h_5 z^7 + h_6 z^6 + h_7 z^5 + h_8 z^4 - h_7 z^3 + h_6 z^2 + h_5 z + h_4} \right] \right\} dz$$

where

$$h_1 = K_2$$

$$h_2 = 2K_3 = 0$$

$$h_3 = (4K_1 - 2K_2)$$

$$h_4 = K_5$$

$$h_5 = 2K_8$$

$$h_6 = (4K_6 - 4K_5)$$

$$h_7 = (6K_8 - 8K_7)$$

$$h_8 = (16K_4 + 6K_5 - 8K_6)$$

$$h_9 = K_{11}$$

$$h_{10} = (4K_{10} - 4K_{11})$$

$$h_{11} = (16K_9 - 8K_{10} + 6K_{11})$$

From here a numerical approximation may be effected. It may be possible in some cases to expand the fraction by partial fractions, write as the integral of a product of exponential terms, and integrate by Laplace Transforms.

It may be of interest to note that after the third transformation bounds can be put on the value of the integral, I.

When the conic is an ellipse, a set of bounds may consist of the circumscribed rectangle and the inscribed rectangle of the largest area. The circumscribed rectangle will have length equal to the major axis of the ellipse and breadth equal to the minor axis. If the equation of the ellipse is put in the form $\frac{(g-h)^2}{a^2} + \frac{(w-k)^2}{b^2} = 1$, the inscribed rectangle has breadth $\sqrt{2}\,b$ and length $\sqrt{2}\,a$. Using Pearson's Tables of cumulative normalized bivariate normal densities, and $\rho = 0$, the probability over the area enclosed by the circumscribing rectangle is given by

$$(10.1.20)\qquad \int_{-\infty}^{k+\frac{\sqrt{2}}{2}b}\int_{-\infty}^{h+\frac{\sqrt{2}}{2}a} f(g,w)\,dg\,dw + \int_{-\infty}^{k-\frac{\sqrt{2}}{2}b}\int_{-\infty}^{h-\frac{\sqrt{2}}{2}a} f(g,w)\,dg\,dw$$

$$- \int_{-\infty}^{k+\frac{\sqrt{2}}{2}b}\int_{-\infty}^{h-\frac{\sqrt{2}}{2}a} f(g,w)\,dg\,dw - \int_{-\infty}^{k-\frac{\sqrt{2}}{2}b}\int_{-\infty}^{h+\frac{\sqrt{2}}{2}a} f(g,w)\,dg\,dw\,.$$

The probability enclosed by the inscribed rectangle is similar in form with slightly different constants added to k and h.

Other types of polygons may be used for bounds, when the conic is a parabola or an hyperbola, after integration limits are obtained. A paper by Pólya in the Berkeley Symposium [9] will allow us to integrate f(g,w) over areas inside or outside general polygons.

10.2. Derivation of $\frac{\partial \sigma^{ij}}{\partial \sigma_{gh}}$.

We need $\frac{\partial}{\partial \sigma_{gh}} |\Sigma|$ and $\frac{\partial}{\partial \sigma_{gh}} |\Sigma_{ij}|$. It will be convenient to use the concepts of bordered determinants.

Note that

$$(10.2.1) \qquad \begin{vmatrix} \sigma_{11} \cdots & \sigma_{1j} & \cdots \sigma_{1k} \\ \vdots & \vdots & \vdots \\ \sigma_{i1} \cdots & \sigma_{ij} & \cdots \sigma_{ik} \\ \vdots & \vdots & \vdots \\ \sigma_{k1} \cdots & \sigma_{kj} & \cdots \sigma_{kk} \end{vmatrix} = \sigma_{ij} |\Sigma_{ij}| + \begin{vmatrix} \sigma_{11} \cdots & \sigma_{1j} \cdots & \sigma_{1k} \\ \vdots & \vdots & \vdots \\ \sigma_{i1} \cdots & 0 & \cdots \sigma_{ik} \\ \vdots & \vdots & \vdots \\ \sigma_{k1} & \sigma_{kj} & \sigma_{kk} \end{vmatrix}$$

Further, $i \neq j$,

$$(10.2.2) \qquad \begin{vmatrix} \sigma_{11} \cdots & \sigma_{1i} \cdots & \sigma_{1j} & \cdots \sigma_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{i1} \cdots & \sigma_{ii} \cdots & 0 & \cdots \sigma_{ik} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{j1} \cdots & \sigma_{ji} \cdots & \sigma_{jj} \cdots & \sigma_{jk} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{k1} \cdots & \sigma_{ki} \cdots & \sigma_{kj} \cdots & \sigma_{kk} \end{vmatrix} = \sigma_{ji} \left| \Sigma^{*}_{ji} \right| + \begin{vmatrix} \sigma_{11} \cdots & \sigma_{1i} \cdots & \sigma_{1j} \cdots & \sigma_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{i1} \cdots & \sigma_{ii} \cdots & 0 & \cdots \sigma_{ik} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{j1} \cdots & 0 & \cdots \sigma_{jj} \cdots & \sigma_{jk} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{k1} \cdots & \sigma_{ki} \cdots & \sigma_{kj} \cdots & \sigma_{kk} \end{vmatrix}$$

where * indicates that the ijth element is zero. Then

$$(10.2.3)\quad |\Sigma| = \sigma_{ij}|\Sigma_{kj}| + \sigma_{ji}|\Sigma^{*}_{ji}| + \begin{vmatrix} \sigma_{11}\cdots & \sigma_{1i}\cdots & \sigma_{1j}\cdots & \sigma_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{i1}\cdots & \sigma_{ii}\cdots & 0 \cdots & \sigma_{ik} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{j1}\cdots & 0 \cdots & \sigma_{jj}\cdots & \sigma_{jk} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{k1}\cdots & \sigma_{ki}\cdots & \sigma_{kj}\cdots & \sigma_{kk} \end{vmatrix}.$$

$$(10.2.4)\quad \frac{\partial|\Sigma|}{\partial \sigma_{ij}} = \sigma_{ij}|\Sigma_{ij,ji}| + |\Sigma_{ij}| + |\Sigma_{ji}{}^{*}|,$$

since $\sigma_{ij} = \sigma_{ji}$, all i, j.

But

$$(10.2.5)\quad |\Sigma_{ji}| = \sigma_{ij}|\Sigma_{ji,ij}| + |\Sigma_{ji}{}^{*}| = \sigma_{ij}|\Sigma_{ij,ji}| + |\Sigma_{ji}{}^{*}|.$$

Then

$$(10.2.6)\quad \frac{\partial|\Sigma|}{\partial \sigma_{ij}} = |\Sigma_{ij}| + |\Sigma_{ji}|.$$

Consider now

$$\frac{\partial|\Sigma|}{\partial \sigma_{ii}}.$$

$$(10.2.7)\quad |\Sigma| = \sigma_{ii}|\Sigma_{ii}| + \begin{vmatrix} \sigma_{11}\cdots & \sigma_{1i}\cdots & \sigma_{1k} \\ \vdots & \vdots & \vdots \\ \sigma_{i1}\cdots & 0 \cdots & \sigma_{ik} \\ \vdots & \vdots & \vdots \\ \sigma_{k1}\cdots & \sigma_{ki}\cdots & \sigma_{kk} \end{vmatrix}$$

and

$$(10.2.8) \qquad \frac{\partial |\Sigma|}{\partial \sigma_{ii}} = |\Sigma_{ii}| .$$

$$(10.2.9) \qquad \frac{\partial |\Sigma_{ii}|}{\partial \sigma_{ij}} = 0 .$$

$$(10.2.10) \qquad \frac{\partial |\Sigma_{ii}|}{\partial \sigma_{jh}} = |\Sigma_{ii,jh}| + |\Sigma_{ii,hj}| = 2 |\Sigma_{ii,jh}| ,$$

by the same argument as above, since Σ_{ii} is symmetric.

$$(10.2.11) \qquad \frac{\partial |\Sigma_{ii}|}{\partial \sigma_{ij}} = |\Sigma_{ij,ji}| .$$

$$(10.2.12) \qquad \frac{\partial |\Sigma_{ii}|}{\partial \sigma_{ji}} = 0 .$$

$$(10.2.13) \qquad \frac{\partial |\Sigma_{ii}|}{\partial \sigma_{ih}} = |\Sigma_{ij,hi}| .$$

$$(10.2.14) \qquad \frac{\partial |\Sigma_{ii}|}{\partial \sigma_{gh}} = |\Sigma_{kj,gh}| + |\Sigma_{ij,hg}| ,$$

by a proof similar to that on the preceeding pages.

$$(10.2.15) \qquad \frac{\partial |\Sigma_{ii}|}{\partial \sigma_{hh}} = |\Sigma_{ij,hh}| .$$

Now by a lemma due to Jacobi [1],

$$(10.2.16) \qquad |\Sigma||\Sigma_{ij,gh}| = |\Sigma_{ij}| \, |\Sigma_{gh}| - |\Sigma_{ih}| \, |\Sigma_{gj}| .$$

Recall that

$$\sigma^{ii} = \frac{|\Sigma_{ii}|}{|\Sigma|}$$

and

$$\sigma^{ij} = \frac{|\Sigma_{ij}|}{|\Sigma|}.$$

Then

$$(10.2.17)\quad \frac{\partial \sigma^{ii}}{\partial \sigma_{ii}} = -\frac{|\Sigma_{ii}|^2}{|\Sigma|^2}.$$

$$(10.2.18)\quad \frac{\partial \sigma^{ii}}{\partial \sigma_{jj}} = -\frac{|\Sigma_{ii}|^2}{|\Sigma|^2}.$$

$$(10.2.19)\quad \frac{\partial \sigma^{ii}}{\partial \sigma_{ij}} = -\frac{|\Sigma_{ii}|\{|\Sigma_{ij}| + |\Sigma_{ji}|\}}{|\Sigma|^2} = -\frac{2|\Sigma_{ii}||\Sigma_{ij}|}{|\Sigma|^2}.$$

$$(10.2.20)\quad \frac{\partial \sigma^{ii}}{\partial \sigma_{jh}} = \frac{|\Sigma|\{|\Sigma_{ii,ih}| + |\Sigma_{ii,hi}|\} - |\Sigma_{ii}|\{|\Sigma_{ih}| + |\Sigma_{hi}|\}}{|\Sigma|^2}$$

$$= \frac{|\Sigma_{ii}||\Sigma_{ih}| - |\Sigma_{ih}||\Sigma_{ii}| + |\Sigma_{ii}||\Sigma_{hi}| - |\Sigma_{ii}||\Sigma_{hi}| - |\Sigma_{ii}||\Sigma_{ih}| - |\Sigma_{ii}||\Sigma_{hi}|}{|\Sigma|^2}$$

$$= -\frac{|\Sigma_{ih}||\Sigma_{ii}|}{|\Sigma|^2}.$$

$$(10.2.21)\quad \frac{\partial \sigma^{ij}}{\partial \sigma_{ii}} = -\frac{|\Sigma_{ij}||\Sigma_{ii}|}{|\Sigma|^2}.$$

$$(10.2.22)\quad \frac{\partial \sigma^{ij}}{\partial \sigma_{ij}} = \frac{|\Sigma||\Sigma_{ij,ij}| - |\Sigma_{ij}|\{|\Sigma_{ij}| + |\Sigma_{ji}|\}}{|\Sigma|^2}$$

$$= \frac{|\Sigma_{ii}||\Sigma_{jj}| - |\Sigma_{ij}||\Sigma_{ji}| - |\Sigma_{ij}||\Sigma_{ij}| - |\Sigma_{ij}||\Sigma_{ji}|}{|\Sigma|^2}$$

$$= -\frac{|\Sigma_{ii}||\Sigma_{jj}| + |\Sigma_{ij}|^2}{|\Sigma|^2}.$$

$$(10.2.23)\quad \frac{\partial \sigma^{ij}}{\partial \sigma_{ih}} = \frac{|\Sigma|\,|\Sigma_{ij,hi}| - |\Sigma_{ij}|\{|\Sigma_{ih}| + |\Sigma_{hi}|\}}{|\Sigma|^2}$$

$$= \frac{|\Sigma_{ij}|\,|\Sigma_{hi}| - |\Sigma_{ii}|\,|\Sigma_{hj}| - |\Sigma_{ij}|\,|\Sigma_{jh}| - |\Sigma_{ij}|\,|\Sigma_{hi}|}{|\Sigma|^2}$$

$$= -\frac{|\Sigma_{ii}|\,|\Sigma_{hj}| + |\Sigma_{ij}|\,|\Sigma_{jh}|}{|\Sigma|^2} .$$

$$(10.2.24)\quad \frac{\partial \sigma^{jh}}{\partial \sigma_{ii}} = \frac{|\Sigma|\,|\Sigma_{jh,ii}| - |\Sigma_{jh}|\,|\Sigma_{ii}|}{|\Sigma|^2}$$

$$= \frac{|\Sigma_{jh}|\,|\Sigma_{ii}| - |\Sigma_{ji}|\,|\Sigma_{ih}| - |\Sigma_{jh}|\,|\Sigma_{ii}|}{|\Sigma|^2}$$

$$= -\frac{|\Sigma_{ij}|\,|\Sigma_{ih}|}{|\Sigma|^2} .$$

$$(10.2.25)\quad \frac{\partial \sigma^{ij}}{\partial \sigma_{gh}} = \frac{|\Sigma|\{|\Sigma_{ij,gh}| + |\Sigma_{ij,hg}|\} - |\Sigma_{ij}|\{|\Sigma_{gh}| + |\Sigma_{hg}|\}}{|\Sigma|^2}$$

$$= \frac{|\Sigma_{ij}|\,|\Sigma_{gh}| - |\Sigma_{ih}|\,|\Sigma_{gj}| + |\Sigma_{ij}|\,|\Sigma_{hg}| - |\Sigma_{ig}|\,|\Sigma_{hj}| - |\Sigma_{ij}|\,|\Sigma_{gh}| - |\Sigma_{ij}|\,|\Sigma_{gh}|}{|\Sigma|^2}$$

$$= -\frac{|\Sigma_{ih}|\,|\Sigma_{gj}| + |\Sigma_{ig}|\,|\Sigma_{hj}|}{|\Sigma|^2} .$$

This last result may be used in general, adjusting the subscripts to obtain the desired partials.