ERRORS IN DISCRIMINATION

By S. John

Indian Statistical Institute, Calcutta

Summary. The probabilities of misclassification involved in the use of estimated discriminant functions are subject to chance variations. The author's purpose in this paper is to derive the distribution laws that the probabilities of misclassification follow and to obtain their expected values. The parent populations are assumed to be normal. The first part of the paper considers the univariate case and the second part the multivariate case. The discussion of the multivariate case proceeds in three stages of increasing complexity. When the exact results are complicated, asymptotic results or approximations are given. Finally, the problem of estimating the expected probabilities of misclassification is considered. Interval estimates as well as point estimates are given.

1. Introduction. Multivariate statistical methods have been found extremely useful in devising efficient procecures for the solution of taxonomic problems. About twenty-five years ago Sir Ronald A. Fisher was consulted by M. M. Barnard as to the best method of classifying skeletal remains unearthed by archaeological excavations. Fisher suggested the use of the now well-known discriminant function [4], [7]. A general mathematical theory of statistical taxonomy was built by Welch [23] on foundations laid by Neyman and Pearson's theory of tests of hypotheses. Subsequent authors introduced many refinements. For a fairly complete account of the theory as it has developed during these years see chapter six of [3] or chapter eight of [18] and literature cited therein.

The situation we are considering is the following: We have an individual who has come from one of the two populations $P^{(1)}$, $P^{(2)}$, but from which one is not known. It is required to devise a procedure that ensures a high probability of a correct classification of the individual. To come to a decision various characteristics of the individual are measured. Suppose we have measurements on p characteristics. Let the vector of measurements be $\mathbf{x} = (x_1, x_2, \cdots, x_p)$. Let the distribution of these measurements in $P^{(k)}$ have $\boldsymbol{\mu}^{(k)}$ as its mean vector. Assume that the dispersion matrix is the same in both the populations. Denote this common dispersion matrix by $\boldsymbol{\Sigma}$. The discriminant function is then the linear function $(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}\mathbf{x}'$. We shall set

$$D(\mathbf{x}; \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}; \boldsymbol{\Sigma}) = (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}\mathbf{x}'. \tag{1}$$

The procedure usually adopted[1] is to classify the individual as belonging to $P^{(1)}$ or $P^{(2)}$ according as

$$D(\mathbf{x}; \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}; \boldsymbol{\Sigma}) \lesseqgtr D(\tfrac{1}{2}[\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)}]; \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}; \boldsymbol{\Sigma}). \tag{2}$$

The above procedure is possible only if $\boldsymbol{\mu}^{(k)} (k = 1, 2)$ and $\boldsymbol{\Sigma}$ are known. But

Received April 29, 1959; revised November 26, 1960.

[1] Certain situations require a slightly modified procedure. See Section 12.

usually such is not the case. We may then try to estimate the unknown parameters $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$ and $\boldsymbol{\Sigma}$ from random samples from $P^{(1)}$ and $P^{(2)}$, substitute these estimates in the appropriate places and use the resulting function to classify individuals in exactly the same way as $D(\mathbf{x};\, \boldsymbol{\mu}^{(1)};\, \boldsymbol{\mu}^{(2)};\, \boldsymbol{\Sigma})$ is used.

Let $x_i^{(k)}(i = 1, 2, \cdots, p;\, r = 1, 2, \cdots, N_k)$ be a random sample of size N_k from population $P^{(k)}$. Put

$$(3)\quad \bar{x}_i^{(k)} = N_k^{-1}\sum_{r=1}^{N_k} x_{ir}^{(k)} \qquad i = 1, 2, \cdots, p;\, k = 1, 2,$$

$$(4)\quad s_{ij} = (N_1 + N_2 - 2)^{-1}\sum_{k=1}^{2}\sum_{r=1}^{N_k}[x_{ir}^{(k)} - \bar{x}_i^{(k)}][x_{jr}^{(k)} - \bar{x}_j^{(k)}], \quad i, j = 1, 2, \cdots p,$$

$$(5)\quad \bar{\mathbf{x}}^{(k)} = (\bar{x}_1^{(k)}, \bar{x}_2^{(k)}, \cdots, \bar{x}_p^{(k)}) \qquad k = 1, 2,$$

$$(6)\quad \mathbf{S} = (s_{ij}).$$

The vector $\bar{\mathbf{x}}^{(k)}$ is an estimate of $\boldsymbol{\mu}^{(k)}$ and the $p \times p$ matrix $\mathbf{S}$ is an estimate of the dispersion matrix $\boldsymbol{\Sigma}$. Substituting these estimates in $D(\mathbf{x};\, \boldsymbol{\mu}^{(1)},\, \boldsymbol{\mu}^{(2)};\, \boldsymbol{\Sigma})$ we get $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'$. Using this function we may assign individuals to $P^{(1)}$ or $P^{(2)}$ according as

$$(7)\qquad (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\bar{\mathbf{x}}' \lesseqgtr \tfrac{1}{2}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} + \bar{\mathbf{x}}^{(1)})'.$$

In any classification procedure there are chances for two kinds of errors: (1) we may classify an individual from $P^{(1)}$ as belonging to $P^{(2)}$; (2) we may classify an individual from $P^{(2)}$ as belonging to $P^{(1)}$. It is clear that if an individual is assigned to $P^{(1)}$ or $P^{(2)}$ depending on the value of a linear function $\sum c_i x_i$, these two chances will depend on the particular coefficients c_i used. Now, in $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'$, the coefficients of $x_1, x_2, \cdots, x_p$ are respectively the components of the vector $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}$. These components are random variables. Random fluctuations in the coefficients induce random fluctuations in the chances of committing either kind of error and it is of interest to study these random fluctuations. This is what we do in the present paper. We assume $P^{(k)}(k = 1, 2)$ to be normal.

Wald's paper [22] appears to be the earliest one to discuss problems connected with the classification of an individual to $P^{(1)}$ or $P^{(2)}$, when the distributions of the characteristics in $P^{(1)}$ and $P^{(2)}$ are not completely known. He considers the use of the statistic $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'$. Wald had visualized a way of using this statistic slightly different from the one which we described. He required the distribution of $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'$ to set up the classification procedure. Papers [2], [8], [9], [10], [19] and [22] are partly or wholly concerned with the derivation of this distribution.[2] In [2], [10], [17] and [19], other statistics which can be used similarly are considered.

[2] A referee informs the author that Elfving has given an expansion for the unconditional probability in the univariate case and that Bowker and Sitgreaves have given an asymptotic expansion for the distribution function of the classification statistic when all parameters are estimated, in papers written for a forthcoming publication, *Mathematical Studies in Item Selection and Classification*, to be published by the Stanford University Press.

2. Notation. Besides the symbols already introduced in the introduction, we use other symbols also. We shall here explain the manner in which these symbols are to be construed.

To distinguish vectors and matrices from scalars we shall employ small bold face type to denote row vectors and capital bold face letters to denote matrices. The same letters, when primed, stand for the transposes of the vectors or matrices.

The letter **I** will denote the identity matrix of order p. If $\mathbf{u} = (u_1, u_2, \cdots, u_p)$, we shall set

$$d\mathbf{u} = du_1\, du_2 \cdots du_p. \tag{8}$$

The symbol $g(x)$ will denote the standard normal density. The integral of $g(x)$ from $-\infty$ to x will be denoted by $G(x)$. The function inverse to $G(x)$ will get the symbol $G^{-1}(x)$. We define

$$\begin{aligned} &G(x_1, x_2; \rho) \\ &= \frac{(1-\rho^2)^{-\frac{1}{2}}}{2\pi} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \exp\left[-\tfrac{1}{2}(1-\rho^2)^{-1}(u_1^2 - 2\rho u_1 u_2 + u_2^2)\right] du_1\, du_2. \end{aligned} \tag{9}$$

The symbol $I_x(p, q)$ will stand for the incomplete beta function,

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x u^{p-1}(1-u)^{q-1}\, du. \tag{10}$$

Finally, we set

$$\delta^2 = (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)}).' \tag{11}$$

Besides the symbols introduced in this section, we use others locally. They will be explained at the appropriate places.

DISCRIMINATION USING A SINGLE CHARACTERISTIC

3. Introduction to univariate case. In the univariate case we shall for convenience write $\mu^{(k)}$ for $\mu_1^{(k)}$, $\bar{x}^{(k)}$ for $\bar{x}_1^{(k)}$, and x for x_1.

It is easy to see that the general classification procedure described in the introduction reduces in the univariate case to the following: If $\bar{x}^{(2)} > \bar{x}^{(1)}$, assign the individual to $P^{(1)}$ or $P^{(2)}$ according as $x \lesseqgtr [\bar{x}^{(1)} + \bar{x}^{(2)}]/2$. If $\bar{x}^{(2)} \leqq \bar{x}^{(1)}$, assign the individual to $P^{(1)}$ or $P^{(2)}$ according as $x \gtreqless [\bar{x}^{(1)} + \bar{x}^{(2)}]/2$.

Suppose $\bar{x}^{(1)}$ and $\bar{x}^{(2)}$ are given. We shall denote by $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ the conditional probability of assigning an individual from $P^{(1)}$ to $P^{(2)}$ and by $e_{21}(\bar{x}^{(1)}, \bar{x}^{(2)})$ the conditional probability of assigning an individual from $P^{(2)}$ to $P^{(1)}$.

$$e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) = \begin{cases} 1 - G([\sigma^{(1)}]^{-1}[\tfrac{1}{2}\{\bar{x}^{(1)} + \bar{x}^{(2)}\} - \mu^{(1)}]) & \text{if } \bar{x}^{(1)} < \bar{x}^{(2)}, \\ G([\sigma^{(1)}]^{-1}[\tfrac{1}{2}\{\bar{x}^{(1)} + \bar{x}^{(2)}\} - \mu^{(1)}]) & \text{if } \bar{x}^{(1)} \geqq \bar{x}^{(2)}. \end{cases} \tag{12}$$

Here $\sigma^{(k)}$ denotes the standard deviation of the characteristic under consideration, in population $P^{(k)}$. A similar equation for $e_{21}(\bar{x}^{(1)}, \bar{x}^{(2)})$ can be written down at once. We shall obtain the distribution and expected value of $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$. A discussion of $e_{21}(\bar{x}^{(1)}, \bar{x}^{(2)})$ would be completely analogous.

The classification procedure we have described is usually adopted only if $\sigma^{(1)} = \sigma^{(2)}$. However, in obtaining the distribution and expected value of $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ we shall not assume that $\sigma^{(1)} = \sigma^{(2)}$; there is some interest in studying the chances of errors under the more general set-up, since, although the classification procedure was designed on the assumption that $\sigma^{(1)} = \sigma^{(2)}$, there is a possibility that the assumption was false.

4. The distribution of $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$. The quantity $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ can be less than z if and only if either of the following two events happen:

$$\bar{x}^{(1)} < \bar{x}^{(2)} \quad \text{and} \quad \tfrac{1}{2}[\bar{x}^{(1)} + \bar{x}^{(2)}] - \mu^{(1)} > -\sigma^{(1)}G^{-1}(z)$$

or

$$\bar{x}^{(1)} \geqq \bar{x}^{(2)} \quad \text{and} \quad \tfrac{1}{2}[\bar{x}^{(1)} + \bar{x}^{(2)}] - \mu^{(1)} < \sigma^{(1)}G^{-1}(z).$$

The distribution function of $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ is therefore given by the equation

$$\Pr\,(e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) < z) = G(h_{11}, h_{21};\rho) + G(h_{12}, h_{22};\rho), \tag{13}$$

where

$$h_{11} = (N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-\frac{1}{2}}(\mu^{(2)} - \mu^{(1)}) = -h_{12}, \tag{14}$$

$$h_{21} = (N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-\frac{1}{2}}[2\sigma^{(1)}G^{-1}(z) + \mu^{(2)} - \mu^{(1)}], \tag{15}$$

$$h_{22} = (N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-\frac{1}{2}}[2\sigma^{(1)}G^{-1}(z) - \mu^{(2)} + \mu^{(1)}], \tag{16}$$

and

$$\rho = (N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-1}(N_2^{-1}[\sigma^{(2)}]^2 - N_1^{-1}[\sigma^{(1)}]^2). \tag{17}$$

The expression on the right hand side in equation (13) can be evaluated using the tables of $G(x_1, x_2;\rho)$ given in [15]. If $N_1^{-1}[\sigma^{(1)}]^2 = N_2^{-1}[\sigma^{(2)}]^2$,

$$\Pr\,(e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) < z) = G(h_{11})G(h_{21}) + G(h_{12})G(h_{22}), \tag{18}$$

and hence can be evaluated with the help of the tables of $G(x)$ given in [14].

5. Expected value of $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$. The expected value of $e_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ can be calculated from the equation

$$Ee_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) = G(a_{11}, a_{21};\rho) + G(a_{12}, a_{22};\rho), \tag{19}$$

where

$$\begin{aligned} a_{11} &= -(N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-\frac{1}{2}}(\mu^{(2)} - \mu^{(1)}) = -a_{12}, \\ a_{21} &= \tfrac{1}{2}\{[\sigma^{(1)}]^2 + \tfrac{1}{4}(N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)\}^{-\frac{1}{2}}(\mu^{(2)} - \mu^{(1)}) = -a_{22}, \end{aligned} \tag{20}$$

and

$$\rho = \tfrac{1}{2}(N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)^{-\frac{1}{2}}$$

$$\{[\sigma^{(1)}]^2 + \tfrac{1}{4}(N_1^{-1}[\sigma^{(1)}]^2 + N_2^{-1}[\sigma^{(2)}]^2)\}^{-\frac{1}{2}} (N_1^{-1}[\sigma^{(1)}]^2 - N_2^{-1}[\sigma^{(2)}]^2). \tag{21}$$

If $N_1^{-1}[\sigma^{(1)}]^2 = N_2^{-1}[\sigma^{(2)}]^2$, $Ee_{12}(\bar{x}^{(1)}, \bar{x}^{(2)})$ can be evaluated using only tables of $G(x)$; for, in this case,

$$Ee_{12}(\bar{x}^{(1)}, \bar{x}^{(2)}) = G(a_{11})G(a_{21}) + G(a_{12})G(a_{22}). \tag{22}$$

Equation (19) is easily established if we observe that a wrong assignment of an individual from $P^{(1)}$ corresponds to the occurrence of either of the following two events:

$$\bar{x}^{(1)} < \bar{x}^{(2)} \quad \text{and} \quad x \geqq \tfrac{1}{2}[\bar{x}^{(1)} + \bar{x}^{(2)}]$$

or,

$$\bar{x}^{(1)} \geqq \bar{x}^{(2)} \quad \text{and} \quad x \leqq \tfrac{1}{2}[\bar{x}^{(1)} + \bar{x}^{(2)}].$$

The reader may wish to compare our treatment of the univariate case with that of [11].

DISCRIMINATION USING MORE THAN ONE CHARACTERISTIC

6. Introduction to the multivariate case. We now take up for consideration the multivariate case. The procedure discussed is the one described in the introduction. It is an adaptation of the standard discriminant function analysis to situations where the parameters required for the construction of the discriminant function are unknown.

Classification procedures based on the correct discriminant function are known to be the best possible when the distributions in the two populations are multivariate normal with identical dispersion matrices. We shall, throughout our discussion, assume that the distributions in the two populations, do, in fact, satisfy these conditions.

The discussion will proceed in several stages. We shall, at stage number one, assume that only $\boldsymbol{\mu}^{(2)}$ is unknown. The case where only $\boldsymbol{\mu}^{(1)}$ is unknown is comletely analogous and does not require separate consideration. At stage two we shall only assume that the dispersion matrix $\boldsymbol{\Sigma}$ is known. In the third stage we shall not assume that $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$, or $\boldsymbol{\Sigma}$ are known.

7. Case one: only $\mu^{(2)}$ is unknown. Before starting discussion of this case let us note that we shall not err seriously if we take $\bar{\mathbf{x}}^{(1)}$ to be the true value of $\boldsymbol{\mu}^{(1)}$ and S to be the true value of $\boldsymbol{\Sigma}$, provided N_1 is sufficiently large.

For constructing the discriminant function, $\boldsymbol{\mu}^{(2)}$ has to be estimated. Substituting $\bar{\mathbf{x}}^{(2)}$ for $\boldsymbol{\mu}^{(2)}$ we have the discriminant function,

$$D(\mathbf{x}; \boldsymbol{\mu}^{(1)}, \bar{\mathbf{x}}^{(2)}; \boldsymbol{\Sigma}) = (\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}\mathbf{x}'. \tag{23}$$

An individual with measurements $\mathbf{x}$ is assigned to $P^{(1)}$ or $P^{(2)}$ according as

$$D(\mathbf{x}; \boldsymbol{\mu}^{(1)}, \bar{\mathbf{x}}^{(2)}; \boldsymbol{\Sigma}) \lesseqqgtr D(\tfrac{1}{2}[\boldsymbol{\mu}^{(1)} + \bar{\mathbf{x}}^{(2)}]; \boldsymbol{\mu}^{(1)}, \bar{\mathbf{x}}^{(2)}; \boldsymbol{\Sigma}). \tag{24}$$

7.1. *Distribution of* $e_{12}(\bar{\mathbf{x}}^{(2)})$. Given $\bar{\mathbf{x}}^{(2)}$, the probability of misclassifying an individual from $P^{(1)}$ is $1 - G(y)$ where

$$(25) \qquad y = \tfrac{1}{2}[(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})']^{\frac{1}{2}}.$$

We shall denote this probability by $e_{12}(\bar{\mathbf{x}}^{(2)})$. Clearly, $e_{12}(\bar{\mathbf{x}}^{(2)})$ is a random variable since it depends on $\bar{\mathbf{x}}^{(2)}$. The distribution function of $e_{12}(\bar{\mathbf{x}}^{(2)})$ is given by the equation

$$(26) \qquad \Pr\,(e_{12}(\bar{\mathbf{x}}^{(2)}) < z) = \Pr\,(4N_2y^2 > 4N_2[G^{-1}(z)]^2) \quad (0 \leqq z \leqq \tfrac{1}{2}).$$

Now, $4N_2y^2$ is a noncentral chisquare variable with p degrees of freedom and noncentrality[3] equal to $(N_2\delta^2)/2$. $\Pr\,(e_{12}(\bar{\mathbf{x}}^{(2)}) < z)$ can therefore be determined from tables of the noncentral chisquare distribution[4].

It is interesting to note that $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$ and $\boldsymbol{\Sigma}$ enter into the distribution of $e_{12}(\bar{\mathbf{x}}^{(2)})$ only in the form of δ. For any given z, $\Pr\,(e_{12}(\bar{\mathbf{x}}^{(2)}) < z)$ is a monotonic function of δ and therefore can be asserted to lie between certain bounds provided we know upper and lower bounds for δ.

7.2. *Expected value of* $e_{12}(\bar{\mathbf{x}}^{(2)})$. From the preceding section we see that

$$(27) \qquad e_{12}(\bar{\mathbf{x}}^{(2)}) = 1 - G(\tfrac{1}{2}[v/N_2]^{\frac{1}{2}})$$

where

$$(28) \qquad v = N_2(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})'.$$

The random variable v has the density function

$$(29) \qquad 2^{-\frac{1}{2}p}e^{-\frac{1}{2}(2\lambda+v)}v^{\frac{1}{2}p-1}\sum_{r=0}^{\infty}[\Gamma(\tfrac{1}{2}p + r)]^{-1}\frac{(\frac{1}{2}\lambda v)^r}{r!},$$

where

$$(30) \qquad \lambda = \tfrac{1}{2}N_2\delta^2.$$

Therefore,

$$(31) \qquad \begin{aligned} Ee_{12}(\bar{\mathbf{x}}^{(2)}) &= 2^{-\frac{1}{2}p}e^{-\lambda}\int_0^{\infty} v^{\frac{1}{2}p-1}e^{-\frac{1}{2}v}\left\{\sum_{r=0}^{\infty}[\Gamma(\tfrac{1}{2}p + r)]^{-1}\frac{(\frac{1}{2}\lambda v)^r}{r!}\right\} \\ &\qquad \left\{\int_{\frac{1}{2}(v/N_2)^{\frac{1}{2}}}^{\infty} g(x)\,dx\right\} dv, \\ &= \tfrac{1}{2}\,e^{-\lambda}\sum_{r=0}^{\infty}\frac{\lambda^r}{r!}I_a(\tfrac{1}{2}p + r, \tfrac{1}{2}) \end{aligned}$$

where $a = 4N_2/(1 + 4N_2)$.

[3] Some authors use the term "noncentrality" for twice this number.

[4] Tables now available are not exactly in the form we require. Editors of [14] have announced that tables of the probability integral of the noncentral chisquare distribution are among the tables considered for inclusion in Vol. II. For the present, recourse must be had to approximate methods developed in [1] and [13].

The justification for the last step is the fact that

$$[\Gamma(\tfrac{1}{2}p + r)]^{-1}\, 2^{-\frac{1}{2}(p+2r)} \int_0^\infty v^{\frac{1}{2}p+r-1} e^{-\frac{1}{2}v}\, dv \int_{\frac{1}{2}(v/N_2)^{\frac{1}{2}}}^\infty g(x)\, dx \tag{32}$$

is equal to half the probability that a random variable having the F-distribution with degrees of freedom one and $p + 2r$ takes a value greater than $(4N_2)^{-1}(p + 2r)$.

It is possible to give several other expressions for $Ee_{12}(\bar{\mathbf{x}}^{(2)})$; the one we have given above appeared to be the most convenient.

7.3. *Distribution of* $e_{21}(\bar{\mathbf{x}}^{(2)})$. Thus far we have been discussing the chances of wrongly assigning an individual from $P^{(1)}$ to $P^{(2)}$. We now take up consideration of the probability of wrongly assigning an individual from $P^{(2)}$ to $P^{(1)}$.

Given $\bar{\mathbf{x}}^{(2)}$, the probability of misclassifying an individual from $P^{(2)}$ is $G(w)$ where

$$\begin{aligned} w = {} & \tfrac{1}{2}[(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})']^{\frac{1}{2}} \\ & - [(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})']^{-\frac{1}{2}}[(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})']. \end{aligned} \tag{33}$$

This probability we denote by $e_{21}(\bar{\mathbf{x}}^{(2)})$. Obviously $e_{21}(\bar{\mathbf{x}}^{(2)})$ is a random variable. We shall derive its distribution.

The distribution function of $e_{21}(\bar{\mathbf{x}}^{(2)})$ is given by the equation

$$\Pr\,(e_{21}(\bar{\mathbf{x}}^{(2)}) < z) = \Pr\,(w < G^{-1}(z)). \tag{34}$$

This equation shows that it suffices to derive the distribution of w.

Observe that w is a function of

$$(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})', \quad \text{and} \quad (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})'.$$

Set

$$t_1 = (\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})' \tag{35}$$

and

$$t_2 = (\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})'. \tag{36}$$

Without loss of generality we may assume that $\boldsymbol{\mu}^{(1)} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$. The density function of $\bar{\mathbf{x}}^{(2)}$ is then

$$\begin{aligned} & (N_2/2\pi)^{\frac{1}{2}p} \exp\,[-\tfrac{1}{2}N_2(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(2)})(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(2)})'] \\ & \qquad = (N_2/2\pi)^{\frac{1}{2}p} \exp\,[-\tfrac{1}{2}N_2(t_1 - 2t_2 + \delta^2)]. \end{aligned} \tag{37}$$

Therefore, if we denote by $f(t_1, t_2)$ the joint density function of t_1 and t_2,

$$f(t_1, t_2)\, dt_1\, dt_2 = \int_{\substack{t_1 < \bar{\mathbf{x}}^{(2)}\bar{\mathbf{x}}^{(2)\prime} < t_1 + dt_1 \\ t_2 < \mathbf{u}^{(2)}\bar{\mathbf{x}}^{(2)\prime} < t_2 + dt_2}} \cdots \int (N_2/2\pi)^{\frac{1}{2}p} \exp\left[-\tfrac{1}{2}N_2(t_1 - 2t_2 + \delta^2)\right] d\bar{\mathbf{x}}^{(2)},$$

(38)
$$= (N_2/2\pi)^{\frac{1}{2}p} \exp\left[-\tfrac{1}{2}N_2(t_1 - 2t_2 + \delta^2)\right] \int_{\substack{t_1 < \bar{\mathbf{x}}^{(2)}\bar{\mathbf{x}}^{(2)\prime} < t_1 + dt_1 \\ t_2 < \mathbf{u}^{(2)}\bar{\mathbf{x}}^{(2)\prime} < t_2 + dt_2}} \cdots \int d\bar{\mathbf{x}}^{(2)},$$

$$= (N_2/2)^{\frac{1}{2}p}\pi^{-\frac{1}{2}}[\delta\Gamma(\tfrac{1}{2}(p-1))]^{-1} \cdot [t_1 - (t_2/\delta)^2]^{\frac{1}{2}(p-3)} \exp\left[-\tfrac{1}{2}N_2(t_1 - 2t_2 + \delta^2)\right] dt_1\, dt_2,$$

since [21]

(39)
$$\int_{\substack{t_1 < \bar{\mathbf{x}}^{(2)}\bar{\mathbf{x}}^{(2)\prime} < t_1 + dt_1 \\ t_2 < \mathbf{u}^{(2)}\bar{\mathbf{x}}^{(2)\prime} < t_2 + dt_2}} \cdots \int d\bar{\mathbf{x}}^{(2)} = \pi^{\frac{1}{2}(p-1)}[\Gamma(\tfrac{1}{2}(p-1))]^{-1}\delta^{-1}[t_1 - (t_2/\delta)^2]^{\frac{1}{2}(p-3)}\, dt_1\, dt_2.$$

Substituting

(40)
$$t_1 = u^2, \qquad w = \tfrac{1}{2}t_1^{\frac{1}{2}} - t_1^{-\frac{1}{2}}t_2,$$

we find that the joint density function of u and w is

(41)
$$Cu^{p-1}[1 - \delta^{-2}(w - \tfrac{1}{2}u)^2]^{\frac{1}{2}(p-3)} \exp\left[-\tfrac{1}{2}N_2(2uw + \delta^2)\right]$$

where

(42)
$$C = 2^{-\frac{1}{2}(p-2)}\pi^{-\frac{1}{2}}[\delta\Gamma(\tfrac{1}{2}(p-1))]^{-1}N_2^{\frac{1}{2}p}.$$

Integrating out u we obtain as the density function of w the function

(43)
$$h(w) = \begin{cases} Ce^{-\frac{1}{2}(N_2\delta^2)} \displaystyle\int_{2(w-\delta)}^{2(w+\delta)} u^{p-1}[1 - \delta^{-2}(w - \tfrac{1}{2}u)^2]^{\frac{1}{2}(p-3)} e^{-N_2uw}\, du & \text{if } w \geqq \delta, \\ Ce^{-\frac{1}{2}(N_2\delta^2)} \displaystyle\int_{0}^{2(w+\delta)} u^{p-1}[1 - \delta^{-2}(w - \tfrac{1}{2}u)^2]^{\frac{1}{2}(p-3)} e^{-N_2uw}\, du & \\ & \text{if } -\delta \leqq w \leqq \delta, \\ 0 & \text{if } w < -\delta. \end{cases}$$

If p is odd, the expression within square brackets in the integrand may be expanded and each term integrated by parts. For example, if $p = 3$,

(44)
$$h(w) = \begin{cases} (2N_2)^{\frac{1}{2}}\pi^{-\frac{1}{2}}(\delta w)^{-1}e^{-\frac{1}{2}(N_2\delta^2)}[e^{-2N_2w(w-\delta)}\{2(w-\delta)^2 \\ \quad + 2(N_2w)^{-1}(w-\delta) + (N_2w)^{-2}\} \\ \quad - e^{-2N_2w(w+\delta)}\{2(w+\delta)^2 \\ \quad + 2(N_2w)^{-1}(w+\delta) + (N_2w)^{-2}\}] & \text{if } w \geqq \delta \\ (2N_2)^{\frac{1}{2}}\pi^{-\frac{1}{2}}(\delta w)^{-1}e^{-\frac{1}{2}(N_2\delta^2)}[(N_2w)^{-2} - e^{-2N_2w(w+\delta)}\{2(w+\delta)^2 \\ \quad + 2(N_2w)^{-1}(w+\delta) + (N_2w)^{-2}\}] & \text{if } -\delta \leqq w < \delta \text{ but } w \neq 0 \\ 2^{\frac{5}{2}}3^{-1}\pi^{-\frac{1}{2}}N_2^{\frac{3}{2}}e^{-\frac{1}{2}(N_2\delta^2)}\delta^2 & \text{if } w = 0, \\ 0 & \text{if } w < -\delta. \end{cases}$$

For even values of p either recourse must be had to numerical integration or percentage points must be obtained by interpolation from corresponding percentage points for distributions with p an odd integer.

Here we observe that $w > -\delta$ with probability one. Therefore, with probability one

$$e_{21}(\bar{\mathbf{x}}^{(2)}) > G(-\delta). \tag{45}$$

From equations (27) and (28) we see that $e_{12}(\bar{\mathbf{x}}^{(2)})$, on the other hand, can go down even to zero.

Observe that besides N_2 and p, δ is the only parameter entering into $h(w)$.

7.4. *Asymptotic distribution of* $e_{21}(\bar{\mathbf{x}}^{(2)})$. Since the exact distribution of $e_{21}(\bar{\mathbf{x}}^{(2)})$ is somewhat complicated, it may be useful to note that, as $N_2 \to \infty$, the distribution of $2N_2^{\frac{1}{2}}[g(\delta/2)]^{-1}[e_{21}(\bar{\mathbf{x}}^{(2)}) - G(-\delta/2)]$ tends (weakly) to the normal distribution with mean zero and variance unity.

Perhaps it is better to use the asymptotic distribution of w together with equation (34). The limiting distribution of $2N_2^{\frac{1}{2}}(w + \frac{1}{2}\delta)$ is normal with mean zero and unit variance. Hence we have, using equation (34),

$$\Pr\,(e_{21}(\bar{\mathbf{x}}^{(2)}) < z) \approx G(2N_2^{\frac{1}{2}}[G^{-1}(z) + \tfrac{1}{2}\delta]). \tag{46}$$

8. Case two: only $\mathbf{\Sigma}$ is known.

8.1. *Introduction to Case two.* Since $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ are unknown we shall construct the discriminant function using $\bar{\mathbf{x}}^{(1)}$ and $\bar{\mathbf{x}}^{(2)}$. The resulting discriminant function is

$$D(\mathbf{x}; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{\Sigma}) = (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}\mathbf{x}'. \tag{47}$$

The classification procedure consists in assigning individuals to $P^{(1)}$ or $P^{(2)}$ according as

$$D(\mathbf{x}; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{\Sigma}) \lesseqgtr D(\tfrac{1}{2}[\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}]; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{\Sigma}). \tag{48}$$

Given $\bar{\mathbf{x}}^{(1)}$ and $\bar{\mathbf{x}}^{(2)}$, the probability of misclassifying an individual from $P^{(1)}$ is $1 - G(u_1)$ where

$$\begin{aligned} u_1 = {} & \tfrac{1}{2}[(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})']^{\frac{1}{2}} \\ & + [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})']^{-\frac{1}{2}}[(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}^{(1)} - \boldsymbol{\mu}^{(1)})']. \end{aligned} \tag{49}$$

We shall denote this by $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$. Similarly let $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ denote the conditional probability of misclassifying an individual from $P^{(2)}$. Being functions of random variables, $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ and $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ are themselves random variables. We shall obtain the distribution of $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$. Since we are free to regard either of the two populations as $P^{(1)}$ it is not necessary to consider $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ separately.

8.2. *The distribution of* $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$. Since

$$\Pr\,(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) < z) = \Pr\,(u_1 > -G^{-1}(z)), \tag{50}$$

the distribution function of $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ will be determined when the distribution of u_1 is obtained. In deriving the distribution of u_1 we shall assume that $\mathbf{\Sigma} = \mathbf{I}$ and $\boldsymbol{\mu}^{(1)} = \mathbf{0}$. Clearly, there is no loss of generality in doing so.

From equation (49) we observe that u_1 is a function of $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}$ and $\bar{\mathbf{x}}^{(1)}$. The joint distribution of $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)}$ and $\bar{\mathbf{x}}^{(1)}$ is multivariate normal with means $(\boldsymbol{\mu}^{(2)}, \mathbf{0})$ and variance-covariance matrix

$$\begin{pmatrix} [N_1^{-1} + N_2^{-1}]\mathbf{I} & -N_1^{-1}\mathbf{I} \\ -N_1^{-1}\mathbf{I} & N_1^{-1}\mathbf{I} \end{pmatrix}. \tag{51}$$

Therefore, the distribution of $\bar{\mathbf{x}}^{(1)}$, given $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)} = \mathbf{y}$, is multivariate normal with mean vector

$$-(N_1 + N_2)^{-1}N_2(\mathbf{y} - \boldsymbol{\mu}^{(2)})$$

and dispersion matrix $(N_1 + N_2)^{-1}\mathbf{I}$. It follows, therefore, that, given $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)} = \mathbf{y}$, u_1 has the normal distribution with mean

$$\begin{aligned} &\tfrac{1}{2}(N_1 + N_2)^{-1}(N_1 - N_2)(\mathbf{yy}')^{\frac{1}{2}} + (N_1 + N_2)^{-1}N_2(\boldsymbol{\mu}^{(2)}\mathbf{y}')(\mathbf{yy}')^{-\frac{1}{2}} \\ &\qquad = \tfrac{1}{2}(N_1 + N_2)^{-1}(N_1 - N_2)t_3^{\frac{1}{2}} + (N_1 + N_2)^{-1}N_2t_3^{-\frac{1}{2}}t_4 \quad \text{(say)} \end{aligned} \tag{52}$$

and variance $(N_1 + N_2)^{-1}$. Hence, if $h(u_1)$ denotes the density function of u_1 and $f(t_3, t_4)$ the joint density function of t_3 and t_4, we have the equation

$$\begin{aligned} h(u_1) = \iint \left(\frac{N_1 + N_2}{2\pi}\right)^{\frac{1}{2}} \exp\Bigg[-\frac{N_1 + N_2}{2} \\ \cdot \left(u_1 - \tfrac{1}{2}\frac{N_1 - N_2}{N_1 + N_2}t_3^{\frac{1}{2}} - \frac{N_2}{N_1 + N_2}t_3^{-\frac{1}{2}}t_4\right)^2\Bigg] f(t_3, t_4)\, dt_3\, dt_4 \end{aligned} \tag{53}$$

where the region of integration is the entire domain of variation of t_3 and t_4. It is thus necessary to obtain the joint density function $f(t_3, t_4)$ of t_3 and t_4. This is done in the next section.

8.3. *The joint density function of t_3 and t_4.*

$$t_3 = (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})' = \mathbf{yy}'; \tag{54}$$

$$t_4 = \boldsymbol{\mu}^{(2)}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})' = \boldsymbol{\mu}^{(2)}\mathbf{y}'. \tag{55}$$

The distribution of $\mathbf{y}$ is multivariate normal with mean $\boldsymbol{\mu}^{(2)}$ and variance-covariance matrix $(N_1^{-1} + N_2^{-1})\mathbf{I}$. Therefore,

$$\begin{aligned} f(t_3, t_4)\, dt_3\, dt_4 = (2\pi)^{-p/2}\left(\frac{N_1 N_2}{N_1 + N_2}\right)^{\frac{1}{2}p} \int_{\substack{t_3 < \mathbf{yy}' < t_3 + dt_3 \\ t_4 < \boldsymbol{\mu}^{(2)}\mathbf{y}' < t_4 + dt_4}} \cdots \int \\ \cdot \exp\left[-\frac{N_1 N_2}{2(N_1 + N_2)}(\mathbf{y} - \mathbf{u}^{(2)})(\mathbf{y} - \mathbf{u}^{(2)})'\right] d\mathbf{y} \end{aligned}$$

$$= (2\pi)^{-p/2}\left(\frac{N_1 N_2}{N_1+N_2}\right)^{\frac{1}{2}p}$$

(56) $$\cdot\exp\left[-\tfrac{1}{2}\frac{N_1 N_2}{N_1+N_2}(t_3 - 2t_4 + \delta^2)\right] \times \int_{\substack{t_3<\mathbf{y}\mathbf{y}'<t_3+dt_3 \\ t_4<\mathbf{u}^{(2)}\mathbf{y}'<t_4+dt_4}} \cdots \int d\mathbf{y}$$

$$= (2\pi)^{-\frac{1}{2}p}\left(\frac{N_1 N_2}{N_1+N_2}\right)^{\frac{1}{2}p} \cdot\exp\left[-\tfrac{1}{2}\cdot\frac{N_1 N_2}{N_1+N_2}\cdot(t_3 - 2t_4 + \delta^2)\right] \times \frac{\pi^{\frac{1}{2}(p-1)}}{\Gamma^{\frac{1}{2}}(\frac{1}{2}(p-1))}\cdot\delta^{-1}[t_3 - (t_4/\delta)^2]^{\frac{1}{2}(p-3)}\,dt_3\,dt_4.$$

8.4. *The distribution of u_1 when $N_1 = N_2$*. If the two sample sizes N_1 and N_2 are equal to N (say), the distribution of u_1 takes a simpler form. In this case, given $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)} = \mathbf{y}$, the conditional distribution of u_1 is normal with mean

(57) $$\tfrac{1}{2}t_3^{-\frac{1}{2}}t_4 = \tfrac{1}{2}\delta t \quad \text{(say)}$$

and variance $(2N)^{-1}$. Therefore, if $f(t)$ is the density function of t, we may write

(58) $$h(u_1) = \int_{-1}^{1}(N/\pi)^{\frac{1}{2}}\exp[-N(u_1 - \tfrac{1}{2}\delta t)^2]f(t)\,dt.$$

It is now necessary to obtain the density function of t. For this we have only to use the joint density function $f(t_3, t_4)$ of t_3 and t_4 to find the joint density of t_3 and t and integrate out t_3 from this joint density. The resulting expression for $f(t)$ is given by the equation

(59) $$f(t) = \pi^{-\frac{1}{2}}[\Gamma(\tfrac{1}{2}(p-1))]^{-1}e^{-\frac{1}{2}N\delta^2}(1-t^2)^{\frac{1}{2}(p-3)}\sum_{r=0}^{\infty}\frac{\Gamma(\frac{1}{2}(p+r))}{r!}N^{\frac{1}{2}r}(\delta t)^r.$$

8.5. *An alternative derivation of the distribution of u_1 in the case $N_1 = N_2$*. In the general case we gave the distribution of u_1 as a double integral. When $N_1 = N_2$, we are able to give the density function of u_1 as a single integral. In this case it is also possible to give the density function of u_1 in a more explicit form using a different method.

Let $\varphi_{u_1}(\theta)$ be the characteristic function of u_1. We have seen that, given $\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)} = \mathbf{y}$, u_1 has the normal distribution with mean $(\delta t/2)$ and variance $(2N)^{-1}$. Therefore, the conditional characteristic function is

(60) $$\exp[\tfrac{1}{2}i\delta t\theta - (4N)^{-1}\theta^2].$$

Hence we have

(61) $$\varphi_{u_1}(\theta) = \int_{-1}^{+1}\exp[\tfrac{1}{2}i\delta t\theta - (4N)^{-1}\theta^2]f(t)\,dt.$$

For the sake of convenience we now change over from u_1 to the variable v_1 defined by the equation

$$v_1 = (2N)^{\frac{1}{2}} u_1 . \tag{62}$$

Let $\varphi_{v_1}(\theta)$ denote the characteristic function of v_1. Then,

$$\begin{aligned} \varphi_{v_1}(\theta) &= \varphi_{u_1}([2N]^{\frac{1}{2}}\theta), \\ &= \int_{-1}^{1} \exp\,[i(N/2)^{\frac{1}{2}}\delta t\theta - \theta^2/2] f(t)\, dt, \\ &= \pi^{-\frac{1}{2}}\{\Gamma(\tfrac{1}{2}[p-1])\}^{-1} \exp\,(-\tfrac{1}{4}N\delta^2 - \tfrac{1}{2}\theta^2) \\ &\quad \cdot \int_{-1}^{1} \left[\sum_{r=0}^{\infty} \frac{\Gamma(\frac{1}{2}[p+r])}{r!} N^{r/2}(\delta t)^r\right](1-t^2)^{(p-3)/2} e^{i(N/2)^{\frac{1}{2}}\delta t\theta}\, dt. \end{aligned} \tag{63}$$

Expanding the exponential factor of the integrand and integrating term by term we obtain the following equation for $\varphi_{v_1}(\theta)$:

$$\begin{aligned} \varphi_{v_1}(\theta) = \pi^{-\frac{1}{2}} \exp\,(-\tfrac{1}{4}N\delta^2 - \tfrac{1}{2}\theta^2) \sum{}' \frac{\Gamma(\frac{1}{2}[p+r])\Gamma(\frac{1}{2}[r+m+1])}{\Gamma(\frac{1}{2}[p+r+m])} \\ \cdot \frac{(N\delta^2)^{(r+m)/2}(-\frac{1}{2}\theta^2)^{m/2}}{r!\, m!} \end{aligned} \tag{64}$$

where $\sum'$ denotes summation over all non-negative integral values of r and m such that $r + m$ is an even integer.

The inversion formula for characteristic functions now readily yields an expression for the density function of v_1. This expression is

$$\begin{aligned} \pi^{-1} e^{-N\delta^2/4 - v_1^2/2} \sum{}' \frac{\Gamma(\frac{1}{2}[p+r])\Gamma(\frac{1}{2}[r+m+1])}{\Gamma(\frac{1}{2}[p+r+m])} \\ \cdot \frac{(N\delta^2)^{(r+m)/2} 2^{-(m-1)/2}}{r!\, m!} \cdot H_m(v_1). \end{aligned} \tag{65}$$

Here $H_r(x)$ denotes the Hermite polynomial of degree r defined by the equation

$$\left(-\frac{d}{dx}\right)^r e^{-\frac{1}{2}x^2} = H_r(x) e^{-\frac{1}{2}x^2}. \tag{66}$$

8.6. *The asymptotic distribution of* $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$. As N_1 and N_2 tend to infinity, the distribution of $2(N_1^{-1} + N_2^{-2})^{-\frac{1}{2}}[g(\frac{1}{2}\delta)]^{-1}[e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) - G(-\frac{1}{2}\delta)]$ tends weakly to the normal distribution with zero mean and unit variance.

Again it may be better to use the asymptotic distribution of u_1 together with equation (50). The limiting distribution of $2(N_1^{-1} + N_2^{-2})^{-\frac{1}{2}}(u_1 - \frac{1}{2}\delta)$ is normal with zero mean and unit variance. Hence we have

$$\Pr(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) < z) \approx G(2[N_1^{-1} + N_2^{-1}]^{-\frac{1}{2}}[G^{-1}(z) + \tfrac{1}{2}\delta]). \tag{67}$$

9. Case three: $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$ and $\boldsymbol{\Sigma}$ unknown. In this case the discriminant function as constructed from sample estimates of $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$ and $\boldsymbol{\Sigma}$ is

$$D(\mathbf{x};\ \bar{\mathbf{x}}^{(1)},\ \bar{\mathbf{x}}^{(2)};\ \mathbf{S}) = (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\mathbf{x}'. \tag{68}$$

The classification procedure consists in assigning individuals to $P^{(1)}$ or $P^{(2)}$ according as

$$D(\mathbf{x}; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) \lesseqqgtr D(\tfrac{1}{2}[\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}]; \bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}). \tag{69}$$

Given $\bar{\mathbf{x}}^{(1)}$, $\bar{\mathbf{x}}^{(2)}$ and $\mathbf{S}$, the probability of misclassifying an individual from $P^{(1)}$ is $1 - G(w_1)$ where

$$w_1 = [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})']^{-\frac{1}{2}} \qquad [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}(\tfrac{1}{2}[\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)}] - \boldsymbol{\mu}^{(1)})']. \tag{70}$$

We shall denote this probability by $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S})$. Clearly, $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S})$ is a random variable. The exact distribution of $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S})$ is complicated and we shall be content with giving its asymptotic distribution. To be slightly more general, we shall suppose that $\mathbf{S}$ is some estimate of $\boldsymbol{\Sigma}$ with n degrees of freedom and independent of $\bar{\mathbf{x}}^{(1)}$ and $\bar{\mathbf{x}}^{(2)}$, not necessarily obtained exclusively from the same samples as those from which $\bar{\mathbf{x}}^{(1)}$ and $\bar{\mathbf{x}}^{(2)}$ were obtained. It can then be shown that as N_1, N_2 and n tend to infinity, the distribution of $2[N_1^{-1} + N_2^{-1}]^{-\frac{1}{2}}[g(\tfrac{1}{2}\delta)]^{-1}$ $[e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) + G(\tfrac{1}{2}\delta) - 1]$ tends (weakly) to the normal distribution with mean zero and unit variance. We have also, corresponding to equations (46) and (67) of previous sections, for large values of N_1, N_2 and n,

$$\Pr(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) < z) \approx G(2[N_1^{-1} + N_2^{-1}]^{-\frac{1}{2}}[G^{-1}(z) + \tfrac{1}{2}\delta]). \tag{71}$$

10. Expected probability of misclassification.

10.1. *Introduction.* We have been discussing in previous sections sampling fluctuations in the chances of misclassification involved in using estimated discriminant functions. Distributions and expected values were the objects of investigation. The expected values were evaluated in certain special cases using *ad hoc* methods. We shall, in this section, give a unified treatment.[5] The method of this section is capable of yielding expected values in all cases where the dispersion matrix is known.

Besides exact expressions some simple approximations also will be given.

10.2. *Exact expressions.* Consider the case where only the variance-covariance matrix $\boldsymbol{\Sigma}$ is assumed to be known. We require expected values of $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ and $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$. We shall evaluate $Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ only, since $Ee_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ can be evaluated on similar lines.

Using results contained in [10], it is possible to prove that

$$Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = \Pr(v_2^{-1}v_1 > [1 + \rho]^{-1}[1 - \rho]) \tag{72}$$

[5] Though the method of this section has the merit of greater generality, the expressions obtained in the earlier sections are somewhat simpler.

where v_1 and v_2 are independent non-central chisquare variables each having p degrees of freedom and non-centralities given respectively by

$$\lambda_1 = [4(1+\rho)]^{-1}N_1N_2[(N_1+N_2)^{-\frac{1}{2}} - (N_1+N_2+4N_1N_2)^{-\frac{1}{2}}]^2\delta^2 \tag{73}$$

and

$$\lambda_2 = [4(1-\rho)]^{-1}N_1N_2[(N_1+N_2)^{-\frac{1}{2}} + (N_1+N_2+4N_1N_2)^{-\frac{1}{2}}]^2\delta^2. \tag{74}$$

Here

$$\rho = [(N_1+N_2)(N_1+N_2+4N_1N_2)]^{-\frac{1}{2}}(N_2-N_1). \tag{75}$$

Using the expression for the density function of $v_2^{-1}v_1$ we may write

$$\begin{aligned} Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) &= e^{-(\lambda_1+\lambda_2)} \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \frac{\Gamma(p+r+s)}{\Gamma(\frac{1}{2}p+r)\Gamma(\frac{1}{2}p+s)} \cdot \frac{\lambda_1^r\lambda_2^s}{r!\,s!} \\ &\qquad \cdot \int_{(1-\rho)/(1+\rho)}^{\infty} (1+u)^{-(p+r+s)} u^{\frac{1}{2}p+r-1}\,du. \end{aligned} \tag{76}$$

We shall put equation (76) in the slightly different form

$$\begin{aligned} Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) &= e^{-(\lambda_1+\lambda_2)} \Bigg[\sum_{r=0}^{\infty}\sum_{s=0}^{r} \frac{\lambda_1^r\lambda_2^s}{r!\,s!}\{1 - I_{\frac{1}{2}(1-\rho)}(\tfrac{1}{2}p+r, \tfrac{1}{2}p+s)\} \\ &\qquad + \sum_{r=0}^{\infty}\sum_{s=r+1}^{\infty} \frac{\lambda_1^r\lambda_2^s}{r!\,s!} I_{\frac{1}{2}(1+\rho)}(\tfrac{1}{2}p+s, \tfrac{1}{2}p+r)\Bigg]. \end{aligned} \tag{77}$$

The various terms on the right hand side in equation (77) can be evaluated using tables of $I_x(p, q)$ given in [16].

In case $\boldsymbol{\mu}^{(1)}$ also is known with $\boldsymbol{\Sigma}$, we have only to take, in equation (77),

$$\begin{aligned} \lambda_1 &= [4(1+\rho)]^{-1}N_2[1 - (1+4N_2)^{-\frac{1}{2}}]^2\delta^2, \\ \lambda_2 &= [4(1-\rho)]^{-1}N_2[1 + (1+4N_2)^{-\frac{1}{2}}]^2\delta^2, \end{aligned} \tag{78}$$

and

$$\rho = -[1+4N_2]^{-\frac{1}{2}} \tag{79}$$

to get the expected value of $e_{12}(\bar{\mathbf{x}}^{(2)})$. If the known parameters are $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}^{(2)}$, we take

$$\begin{aligned} \lambda_1 &= [4(1+\rho)]^{-1}N_1[1 - (1+4N_1)^{-\frac{1}{2}}]^2\delta^2, \\ \lambda_2 &= [4(1-\rho)]^{-1}N_1[1 + (1+4N_1)^{-\frac{1}{2}}]^2\delta^2, \end{aligned} \tag{80}$$

and

$$\rho = (1+4N_1)^{-\frac{1}{2}}. \tag{81}$$

10.3. *An approximation.* In the previous section we obtained an exact expression for the expected value of $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$. That expression is not quite convenient for numerical evaluation. For this reason we now give an approximation which permits evaluation of the expected value using only tables of $G(x)$.

We start with the result

$$Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = \Pr(v_2^{-1}v_1 > [1 + \rho]^{-1}[1 - \rho]). \tag{82}$$

Now,

$$\begin{aligned} \Pr(v_2^{-1}v_1 > [1 + \rho]^{-1}[1 - \rho]) &= \Pr(v_2^{-\frac{1}{3}} v_1^{\frac{1}{3}} > [1 + \rho]^{-\frac{1}{3}}[1 - \rho]^{\frac{1}{3}}), \\ &= \Pr([1 + \rho]^{\frac{1}{3}} v_1^{\frac{1}{3}} - [1 - \rho]^{\frac{1}{3}} v_2^{\frac{1}{3}} > 0). \end{aligned} \tag{83}$$

From [1] we know that if $\chi^{2\prime}$ has a noncentral chisquare distribution with f degrees of freedom and noncentrality parameter equal to λ, the variable $(\chi^{2\prime}/r)^{\frac{1}{3}}$, where

$$r = f + 2\lambda, \tag{84}$$

has approximately a normal distribution with expectation $1 - 2(1 + b)/9r$ and variance $2(1 + b)/9r$. (Here b stands for $2[f + 2\lambda]^{-1}\lambda$). Using this result and also equations (82) and (83) we may now write

$$Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) \approx G(a) \tag{85}$$

where

$$\begin{aligned} a = (18)^{-\frac{1}{2}}[r_1^{-\frac{1}{3}}(1 + \rho)^{\frac{2}{3}}(1 + b_1) + r_2^{-\frac{1}{3}}(1 - \rho)^{\frac{2}{3}}(1 + b_2)]^{-\frac{1}{2}} \\ [2\{r_2^{-\frac{2}{3}}(1 - \rho)^{\frac{1}{3}}(1 + b_2) - r_1^{-\frac{2}{3}}(1 + \rho)^{\frac{1}{3}}(1 + b_1)\} \\ + 9\{r_1(1 + \rho)\}^{\frac{1}{3}} - 9\{r_2(1 - \rho)\}^{\frac{1}{3}}], \end{aligned} \tag{86}$$

the quantities r_1, r_2, b_1 and b_2 being defined by the equations

$$\begin{aligned} r_i &= p + 2\lambda_i \qquad (i = 1, 2), \\ b_i &= 2(p + 2\lambda_i)^{-1}\lambda_i \qquad (i = 1, 2). \end{aligned} \tag{87}$$

The question of the closeness of the approximation (85) now arises. We should expect that the approximation involved is of the same order as that involved in assuming that the cube root of a noncentral chisquare variable has the normal distribution. Though it would be interesting to compare the approximate values with the corresponding exact values using numerical computations, we shall not embark on this venture at the present moment. Some numerical results given in [1] may be found enlightening.

11. Estimation of the expected probabilities of misclassification. The expressions for the expected probabilities of misclassification are found to be functions of δ. The problem of estimating these expected probabilities now arises. The empirical method of estimating the probability of an event by computing the proportion of outcomes favourable to the event in a number of repetitions of the experiment is available to us. This method is suggested in [20]. If the problem is one of estimating the conditional probabilities of misclassification, the empirical method is a simple way of solving it. What we have to do is to use the estimated discriminant function to classify the N_i individuals known to belong to $P^{(i)}$

and note down the proportion of these individuals assigned to $P^{(j)}(i = 1$ and $j = 2$ or $i = 2$ and $j = 1)$.

If the problem is one of estimating the unconditional probabilities of misclassification, the empirical method can prove exasperating. Fortunately, the maximum likelihood estimator is simple enough. To obtain the maximum likelihood estimate, all we have to do is to substitute $(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})'$, or $(\boldsymbol{\mu}^{(2)} - \bar{\mathbf{x}}^{(1)})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(2)} - \bar{\mathbf{x}}^{(1)})'$, or $(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})'$, or

$$(N_1 + N_2 - 2)^{-1}(N_1 + N_2)(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})'$$

for δ^2 in the expressions for the expected error (or approximations to them), depending on which parameters are known.

The exact value of the expected probability is certain to differ from its estimate. An idea of the magnitude of the difference that may be expected can be obtained from the variance of the estimator. But the expression for the variance happens to be quite cumbersome. Besides, it involves unknown parameters. Fortunately another approach is open to us. We indicate below how intervals which enclose the true value of the probability with any preassigned degree of certainty can be constructed.

The procedure to be adopted is the following: Suppose the desired confidence level is α. Set up for δ a confidence interval of confidence coefficient α. Suppose the upper and lower bounds are respectively δ_1 and δ_2. Evaluate the expression for the expected probability of misclassification substituting in turn δ_1 and δ_2 for δ. The two values thus obtained will enclose the true value of that quantity with probability α.

Confidence bounds for δ can be set up if we remember that

$$N_1(\bar{\mathbf{x}}^{(1)} - \boldsymbol{\mu}^{(2)})\,\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(1)} - \boldsymbol{\mu}^{(2)})',\ N_2(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \boldsymbol{\mu}^{(1)})',$$
$$(N_1 + N_2)^{-1}N_1N_2\,(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})'$$

have noncentral chisquare distributions with p degrees of freedom and noncentrality parameters equal respectively to $N_1\frac{1}{2}\delta^2$, $N_2\frac{1}{2}\delta^2$, and $(N_1 + N_2)^{-1}N_1N_2\frac{1}{2}\delta^2$ and that $[p(N_1 + N_2)(N_1 + N_2 - 2)]^{-1}N_1N_2(N_1 + N_2 - p - 1)(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})$ $\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})'$ has the noncentral F-distribution with degrees of freedom p and $N_1 + N_2 - p - 1$ and noncentrality equal to $(N_1 + N_2)^{-1}N_1N_2\frac{1}{2}\delta^2$.

For the sake of definiteness, let us suppose that we are dealing with a situation where $\boldsymbol{\Sigma}$ is known and $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ are unknown. Let v be a noncentral chisquare variable with p degrees of freedom and noncentrality equal to $(N_1 + N_2)^{-1}$ $N_1N_2\frac{1}{2}\delta^2$. Set

$$F_\delta(z) = \Pr(v < z). \tag{88}$$

As δ_1 we can take[6] the least upper bound of the set of all numbers δ satisfying the

[6] Marakathavalli [12] should also be consulted. She discusses how unbiased critical regions can be set up for testing hypotheses specifying the value of the noncentrality parameter. Inversion will give an unbiased confidence interval. Methods of approximately evaluating the probability integral of the noncentral chisquare density developed in [1] and [13] will be required. Tables given in [6] and [12] will be found useful.

condition

$$F_\delta([N_1 + N_2]^{-1}N_1N_2(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})') \geqq \tfrac{1}{2}(1 + \alpha) \tag{89}$$

and as δ_2 we can take the greatest lower bound of the set of all numbers δ satisfying the condition

$$F_\delta([N_1 + N_2]^{-1}N_1N_2(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})') \leqq \tfrac{1}{2}(1 - \alpha). \tag{90}$$

12. About a more general classification procedure.[7] There are situations where the two kinds of errors are not of equal importance. In some cases it may even be possible to determine the different losses consequent on each type of mistake. Suppose c_{12} is the loss incurred in assigning an individual from $P^{(1)}$ to $P^{(2)}$ and c_{21} is the loss incurred in assigning an individual from $P^{(2)}$ to $P^{(1)}$. Suppose further that the *a priori* probability of an individual coming from $P^{(k)}$ is $\pi^{(k)}$. Then the procedure with minimum expected loss is that of assigning an individual with measurements $\mathbf{x}$ to $P^{(1)}$ or $P^{(2)}$ according as

$$(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\mathbf{\Sigma}^{-1}\mathbf{x}' \lesseqqgtr \tfrac{1}{2}(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(1)})\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}^{(2)} + \boldsymbol{\mu}^{(1)})' + c \tag{91}$$

where

$$c = \log_e \frac{\pi^{(1)}c_{12}}{\pi^{(2)}c_{21}} \tag{92}$$

([3], p. 134). The procedure we considered earlier is a special case of this more general procedure. It corresponds to the case $c = 0$. A sufficient condition for c to be zero is that $c_{12} = c_{21}$ and $\pi^{(1)} = \pi^{(2)}$.

The procedure mentioned above can be carried out only if all the parameters $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$ and $\mathbf{\Sigma}$ are known. If such is not the case we may assign the individual to $P^{(1)}$ or $P^{(2)}$ according as

$$(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}\bar{\mathbf{x}}' \lesseqqgtr \tfrac{1}{2}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(2)} + \bar{\mathbf{x}}^{(1)})' + c. \tag{93}$$

The sampling fluctuations of $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})$ and $e_{21}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}, \mathbf{S})$, the conditional probabilities of the two kinds of errors, are again of interest. We shall briefly discuss the case of $\mathbf{\Sigma}$ known and give indications of the changes to be made in some of the earlier formulae, leaving a fuller discussion of the procedure to another occasion.

The distribution of $e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ is given by the equation

$$\Pr\,(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) < z) = \Pr\,(u_1 > -G^{-1}(z)) \tag{94}$$

where

$$\begin{aligned} u_1 = {} & \tfrac{1}{2}[\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})']^{\frac{1}{2}} \\ & + [(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})']^{-\frac{1}{2}}[(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}}^{(1)})\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}^{(1)} - \boldsymbol{\mu}^{(1)})' + c]. \end{aligned} \tag{95}$$

[7] The results of earlier sections can be generalized in other ways, which we hope to indicate in a later communication.

This equation shows that it suffices to obtain the distribution of u_1. The density function $h(u_1)$ of u_1 is

$$\int\int\left(\frac{N_1+N_2}{2\pi}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2}(N_1+N_2)\left\{u_1-\frac{1}{2}\cdot\frac{N_1-N_2}{N_1+N_2}t_3^{\frac{1}{2}} - \left(\frac{N_2\,t_4}{N_1+N_2}+c\right)t_3^{-\frac{1}{2}}\right\}^2\right] f(t_3,t_4)\,dt_3\,dt_4 \tag{96}$$

where $f(t_3, t_4)$ is to be taken from equation (56).

If N_1, N_2 and n are large, we have, corresponding to equation (71),

$$\Pr\,(e_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}; \mathbf{S}) < z) \approx G(2\delta^2[(\delta^2+2c)^2N_1^{-1} + (\delta^2-2c)^2N_2^{-1} + 2(2c\delta)^2n^{-1}]^{-\frac{1}{2}}[G^{-1}(z) + c\delta^{-1} + \tfrac{1}{2}\delta]). \tag{97}$$

If more information is available, the results become simpler. Thus if $\boldsymbol{\mu}^{(1)}$ is also known with $\boldsymbol{\Sigma}$, the distribution function of $e_{12}(\bar{\mathbf{x}}^{(2)})$ is given by the formula

$$\Pr\,(e_{12}(\bar{\mathbf{x}}^{(2)}) < z) = \begin{cases} \Pr\,(v < A_1(z)) + \Pr\,(v > A_2(z)) & \text{if } z < 1 - G([2c]^{\frac{1}{2}}), \\ 1 & \text{if } z \geqq 1 - G([2c]^{\frac{1}{2}}), \end{cases} \quad (c \geqq 0) \tag{98}$$

where v is a random variable having the noncentral chisquare distribution with p degrees of freedom and noncentrality $N_2\frac{1}{2}\delta^2$ and

$$A_1(z) = N_2[G^{-1}(z) + \{[G^{-1}(z)]^2 - 2c\}^{\frac{1}{2}}]^2 \tag{99}$$

and

$$A_2(z) = N_2[G^{-1}(z) - \{[G^{-1}(z)]^2 - 2c\}^{\frac{1}{2}}]^2. \tag{100}$$

Similarly we have, for the distribution function of $e_{21}(\bar{\mathbf{x}}^{(2)})$, the equation

$$\Pr\,(e_{21}(\bar{\mathbf{x}}^{(2)}) < z) = \Pr\,(w < G^{-1}(z)) \tag{101}$$

where the random variable w has the density function $h(w)$ defined below.

$$h(w) = \begin{cases} Ce^{-N_2\frac{1}{2}(-2c+\delta^2)}\left[\int_{m_1(w)}^{m_2(w)} + \int_{m_3(w)}^{m_4(w)}\right][1-\delta^{-2}(w-\tfrac{1}{2}u - cu^{-1})^2]^{\frac{1}{2}(p-3)}u^{p-1}e^{-N_2uw}\,du & \text{if } w \geqq \delta + (2c)^{\frac{1}{2}} \\ Ce^{-N_2\frac{1}{2}(-2c+\delta^2)}\int_{m_1(w)}^{m_4(w)}[1-\delta^{-2}(w-\tfrac{1}{2}u-cu^{-1})^2]^{\frac{1}{2}(p-3)} u^{p-1}e^{-N_2uw}\,du & \text{if } (2c)^{\frac{1}{2}} - \delta \leqq w < (2c)^{\frac{1}{2}} + \delta \\ 0 & \text{if } w < (2c)^{\frac{1}{2}} - \delta \end{cases} \quad (c \geqq 0) \tag{102}$$

Here

$$
\begin{aligned}
m_1(w) &= w + \delta - [(w+\delta)^2 - 2c]^{\frac{1}{2}}, \\
m_2(w) &= w - \delta - [(w-\delta)^2 - 2c]^{\frac{1}{2}}, \\
m_3(w) &= w - \delta + [(w-\delta)^2 - 2c]^{\frac{1}{2}}, \\
m_4(w) &= w + \delta + [(w+\delta)^2 - 2c]^{\frac{1}{2}},
\end{aligned} \tag{103}
$$

and C is the C of equation (42).

Observe that with probability one,

$$e_{12}(\bar{\mathbf{x}}^{(2)}) \leqq 1 - G([2c]^{\frac{1}{2}}) \qquad (c \geqq 0) \tag{104}$$

and

$$e_{21}(\mathbf{x}^{(2)}) \geqq G([2c]^{\frac{1}{2}} - \delta) \qquad (c \geqq 0). \tag{105}$$

If $c < 0$, we have, for the distribution function of $e_{12}(\bar{\mathbf{x}}^{(2)})$, the equation

$$\Pr\,(e_{12}(\bar{\mathbf{x}}^{(2)}) < z) = \Pr\,(v > A_2(z)) \tag{106}$$

instead of equation (98) and, for the density function of w, the equation

$$
\begin{aligned}
h(w) = Ce^{-N_2\frac{1}{2}(-2c+\delta^2)} \int_{m_3(w)}^{m_4(w)} [1 - \delta^{-2}(w - \tfrac{1}{2}u \\
- cu^{-1})^2]^{\frac{1}{2}(p-3)} u^{p-1} e^{-N_2uw}\, du
\end{aligned} \tag{107}
$$

instead of equation (102).

Note that statements corresponding to equations (104) and (105) cannot be made if $c < 0$.

For $Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)})$ we have the following result:

$$Ee_{12}(\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}) = \int_d^\infty f_1(z)\, dz \tag{108}$$

where $f_1(z)$ is the function defined by equation (4.9) of [10] and

$$d = 4N_1N_2(N_1 + N_2)^{-\frac{1}{2}}(N_1 + N_2 + 4N_1N_2)^{-\frac{1}{2}}c. \tag{109}$$

As the expression for $f_1(z)$ is complicated we shall not reproduce it here.

Acknowledgment. The author wishes to express his gratitude to Dr. C. R. Rao for his kind encouragement.

REFERENCES

[1] S. H. ABDEL-ATY, "Approximate formula for the percentage points and the probability integral of the noncentral χ^2 distribution," *Biometrika*, Vol. 41 (1954), pp. 538–540.

[2] T. W. ANDERSON, "Classification by multivariate analysis," *Psychometrika*, Vol. 16 (1951), pp. 31–50.

[3] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, John Wiley and and Sons, New York, 1958.

[4] M. M. Barnard, "The secular variations of skull characters in four series of Egyptian skulls," *Ann. Eugen.*, Vol. 6 (1935), pp. 352–371.

[5] R. C. Bose and S. N. Roy, "The exact distribution of the studentized D^2-statistic," *Sankhyā*, Vol. 4 (1938), pp. 19–38.

[6] R. A. Fisher, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc.*, Ser. A. Vol. 121, (1928), pp. 654–673.

[7] R. A. Fisher, "The use of multiple measurements in taxonomic problems', *Ann. Eugen.*, Vol. 7 (1936), pp. 179–188.

[8] Harman Leon Harter, "On the distribution of Wald's classification statistic," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 58–67.

[9] S. John, "The distribution of Wald's classification statistic when the dispersion matrix is known," *Sankhyā*, Vol. 21 (1959), pp. 371–376.

[10] S. John, "On some classification statistics," *Sankhyā*, Vol. 22 (1960), pp. 309–317.

[11] Carl F. Kossack, "Some techniques for simple classification," *Proceedings of the Berkeley Symposium in Statistics and Probability*, pp. 345–352, University of California Press, Berkeley, 1949.

[12] N. Marakathavalli, "Unbiased test for a specified value of a parameter in the noncentral F-distribution," *Sankhyā*, Vol. 15 (1955), pp. 321–330.

[13] P. B. Patnaik, "The noncentral χ^2 and F distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.

[14] E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, Vol. 1, Cambridge University Press, 1956.

[15] National Bureau of Standards, *Tables of the Bivariate Normal Distribution and Related Functions*," Washington, D. C., 1959.

[16] Karl Pearson, *Tables of the Incomplete Beta-Function*, Cambridge, Cambridge University Press, 1934.

[17] C. Radhakrishna Rao, "On a general theory of discrimination when the information on alternate hypotheses is based on samples," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 651–669.

[18] C. Radhakrishna Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952.

[19] Rosedith Sitgreaves, "On the distribution of two random matrices used in classification procedures," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 263–270.

[20] C. A. B. Smith, "Some examples of discrimination," *Ann. Eugen*, Vol. 13 (1947), pp. 272–282.

[21] Erling Sverdrup, "Derivation of the Wishart distribution of the second order sample moments by straightforward integration of a multiple integral," *Skandinavisk Aktuarietidskrift*, Vol. 30 (1947), pp. 151–166.

[22] A. Wald, "On a statistical problem arising in the classification of an individual into one of two groups," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 145–163.

[23] B. L. Welch, "Note on discriminant functions," *Biometrika*, Vol. 31 (1939), pp. 218–220.