AN ASYMPTOTIC EXPANSION OF THE DISTRIBUTION OF THE "STUDENTIZED" CLASSIFICATION STATISTIC W

BY

T. W. ANDERSON

TECHNICAL REPORT NO. 9

SEPTEMBER 10, 1972

PREPARED UNDER CONTRACT

N00014-67-A-0112-0030 (NR-042-034)

FOR THE OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

AN ASYMPTOTIC EXPANSION OF THE DISTRIBUTION OF THE "STUDENTIZED" CLASSIFICATION STATISTIC W

BY

T. W. ANDERSON
Stanford University

TECHNICAL REPORT NO. 9

SEPTEMBER 10, 1972

PREPARED UNDER THE AUSPICES

OF

OFFICE OF NAVAL RESEARCH CONTRACT #N00014-67-A-0112-0030

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. Introduction

A sample $\underset{\sim}{x}_1^{(1)}, \ldots, \underset{\sim}{x}_{N_1}^{(1)}$ is drawn from the normal distribution $N(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$, and a sample $\underset{\sim}{x}_1^{(2)}, \ldots, \underset{\sim}{x}_{N_2}^{(2)}$ is drawn from $N(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$. The p-component mean vectors $\underset{\sim}{\mu}^{(1)}$ and $\underset{\sim}{\mu}^{(2)}$ and the common covariance matrix $\underset{\sim}{\Sigma}$ are unknown; it is assumed that $\underset{\sim}{\mu}^{(1)} \neq \underset{\sim}{\mu}^{(2)}$ and $\underset{\sim}{\Sigma}$ is non-singular. Another observation $\underset{\sim}{x}$ is drawn. It is desired to classify this observation as coming from $N(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$ or $N(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$. [See T. W. Anderson (1951) or T. W. Anderson (1958), Chapter 6.]

The observation $\underset{\sim}{x}$ may be classified by means of the classification statistic

$$W = (\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)})' \underset{\sim}{S}^{-1} [\underset{\sim}{x} - \tfrac{1}{2} (\bar{\underset{\sim}{x}}^{(1)} + \bar{\underset{\sim}{x}}^{(2)})] , \tag{1}$$

$$\bar{\underset{\sim}{x}}^{(1)} = \frac{1}{N_1} \sum_{j=1}^{N_1} \underset{\sim}{x}_j^{(1)} , \quad \bar{\underset{\sim}{x}}^{(2)} = \frac{1}{N_2} \sum_{j=1}^{N_2} \underset{\sim}{x}_j^{(2)} , \tag{2}$$

$$n\underset{\sim}{S} = \sum_{j=1}^{N_1} (\underset{\sim}{x}_j^{(1)} - \bar{\underset{\sim}{x}}^{(1)}) (\underset{\sim}{x}_j^{(1)} - \bar{\underset{\sim}{x}}^{(1)})' + \sum_{j=1}^{N_2} (\underset{\sim}{x}_j^{(2)} - \bar{\underset{\sim}{x}}^{(2)}) (\underset{\sim}{x}_j^{(2)} - \bar{\underset{\sim}{x}}^{(2)})' , \tag{3}$$

and $n = N_1 + N_2 - 2$. The rule is to classify $\underset{\sim}{x}$ as coming from $N(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$ if $W > c$ and from $N(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$ if $N \leq c$, where c may be a constant, particularly 0, or a function of $\bar{\underset{\sim}{x}}^{(1)}$, $\bar{\underset{\sim}{x}}^{(2)}$, and $\underset{\sim}{S}$.

The distribution of W depends on the parameters $\underset{\sim}{\mu}^{(1)}$, $\underset{\sim}{\mu}^{(2)}$, and $\underset{\sim}{\Sigma}$ through the squared Mahalanobis distance

$$\alpha = (\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\mu}^{(1)} - \underset{\sim}{\mu}^{(2)}) , \tag{4}$$

which can be estimated by

$$a = (\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)})' \underset{\sim}{S}^{-1} (\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)}) . \tag{5}$$

The limiting distribution of W as $N_1 \to \infty$ and $N_2 \to \infty$ is normal with variance α and mean $\frac{1}{2}\alpha$ if $\underset{\sim}{x}$ is from $N(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$ and mean $-\frac{1}{2}\alpha$ if $\underset{\sim}{x}$ is from $N(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$. Bowker and Sitgreaves* (1961) for $N_1 = N_2$ and Okamoto (1963) [with correction, Okamoto (1968)] gave asymptotic expansions of the distribution of $(W - \frac{1}{2}\alpha)/\sqrt{\alpha}$ for $\underset{\sim}{x}$ coming from $N(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$ and $(W + \frac{1}{2}\alpha)/\sqrt{\alpha}$ for $\underset{\sim}{x}$ coming from $N(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$ to terms or order $1/N_1^2$, $1/N_2^2$, and $1/n^2$ when $N_1 \to \infty$, $N_2 \to \infty$, and $N_2/N_1 \to k$, a finite positive constant. In particular, $\Pr\{W \leq 0\}$ was evaluated.

The statistician, who wants to classify $\underset{\sim}{x}$, may take c to be a constant, perhaps 0, and accept the pair of misclassification probabilities that result. The asymptotic expansion of the distribution of $(W \pm \frac{1}{2}\alpha)/\sqrt{\alpha}$ gives approximate evaluations of these probabilities, which are functions of the unknown parameter α as well as of c.

On the other hand the statistician may want to determine the cut-off point c to adjust the probabilities of misclassification. Since the limiting distribution of $(W-\alpha)/\sqrt{\alpha}$ and $(W+\alpha)/\sqrt{\alpha}$ are $N(0, 1)$ when $\mathcal{E}\underset{\sim}{x} = \underset{\sim}{\mu}^{(1)}$ and $\mathcal{E}\underset{\sim}{x} = \underset{\sim}{\mu}^{(2)}$, respectively, a first approximation to the pair of misclassification probabilities is $\Phi(\frac{1}{2}\alpha + c\sqrt{\alpha})$ and $\Phi(-\frac{1}{2}\alpha + c\sqrt{\alpha})$, where $\Phi(a)$ is the cumulative distribution function of the standard normal variate. Since a is an estimate of α, one might base his choice of c on the fact that the limiting distribution of $(W - \frac{1}{2}a)/\sqrt{a}$ and $(W + \frac{1}{2}a)/\sqrt{a}$ are $N(0, 1)$ when $\mathcal{E}\underset{\sim}{x} = \underset{\sim}{\mu}^{(1)}$ and $\mathcal{E}\underset{\sim}{x} = \underset{\sim}{\mu}^{(2)}$, respectively. In this paper we make asymptotic expansions of the distribution of $(W - \frac{1}{2}a)/\sqrt{a}$ and $(W + \frac{1}{2}a)/\sqrt{a}$ in these two cases, respectively.

---

*The coefficients a_{31} and a_{32} should be replaced by $-a_{31}$ and $-a_{32}$, respectively.

2. The Asymptotic Expansion

The statistics x, $\bar{x}^{(1)}$, $\bar{x}^{(2)}$, and S are independently distributed according to $N(\mu, \Sigma)$, $N[\mu^{(2)}, (1/N_1)\Sigma]$, $N[\mu^{(2)}, (1/N_2)\Sigma]$, and $W(\Sigma, n)$, respectively; here $\mu = \mathcal{E}x$ and $W(\Sigma, n)$ denotes the Wishart distribution with n degrees of freedom. We write

$$W - \frac{1}{2}a = (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (x - \bar{x}^{(1)}) . \tag{6}$$

Then

$$\Pr\left\{\frac{W - \frac{1}{2}a}{\sqrt{a}} \leq u\right\} = \Pr\left\{(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (x - \mu) \leq u\sqrt{(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})} + (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \mu)\right\} . \tag{7}$$

Since x has the distribution $N(\mu, \Sigma)$ independently of $\bar{x}^{(1)}$, $\bar{x}^{(2)}$, and S, the conditional distribution of $(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (x - \mu)$ is $N[0, (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} \Sigma S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})]$, and

$$r = \frac{(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (x - \mu)}{\sqrt{(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} \Sigma S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})}} \tag{8}$$

has the distribution $N(0, 1)$. Then (7) is

$$\Pr\left\{\frac{W - \frac{1}{2}a}{\sqrt{a}} \leq u\right\} = \Pr\left\{r \leq \frac{u\sqrt{(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})} + (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \mu)}{\sqrt{(\bar{x}^{(1)} - \bar{x}^{(2)}) S^{-1} \Sigma S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})}}\right\} \tag{9}$$

$$= \mathcal{E}\Phi\left[\frac{u\sqrt{(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})} + (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \mu)}{\sqrt{(\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} \Sigma S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)}}}\right]$$

where the expectation is with respect to $\bar{x}^{(1)}$, $\bar{x}^{(2)}$, and S.

The distribution of W and a is invariant with respect to the transformations $\underset{\sim}{x}^* = \underset{\sim}{A}\,\underset{\sim}{x} + \underset{\sim}{b}$, $\underset{\sim}{x}_j^{*(1)} = \underset{\sim}{A}\,\underset{\sim}{x}_j^{(1)} + \underset{\sim}{b}$, $j = 1, \ldots, N_1$, and $\underset{\sim}{x}_j^{*(2)} = \underset{\sim}{A}\,\underset{\sim}{x}_j^{(2)} + \underset{\sim}{b}$, where $\underset{\sim}{A}$ is nonsingular. The maximal invariant of these transformations is the distance α, given by (4). We can choose $\underset{\sim}{A}$ and $\underset{\sim}{b}$ to transform $\underset{\sim}{\Sigma}$ to $\underset{\sim}{I}$, $\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2$ to $\delta = (\Delta, 0, \ldots, 0)$, where $\Delta = \sqrt{\alpha}$, and $\underset{\sim}{\mu}_1$ to $\underset{\sim}{0}$. We shall first treat the case where $\underset{\sim}{\mu} = \underset{\sim}{\mu}_1$.

The vectors $\bar{\underset{\sim}{x}}^{(1)}$ and $\bar{\underset{\sim}{x}}^{(2)}$ and the matrix $\underset{\sim}{S}$ are distributed independently. The distribution of $[(\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)})', \bar{\underset{\sim}{x}}^{(1)\prime}]'$ is

$$N\left[\begin{pmatrix} \underset{\sim}{\delta} \\ \underset{\sim}{0} \end{pmatrix}, \begin{pmatrix} \left(\frac{1}{N_1} + \frac{1}{N_2}\right)\underset{\sim}{I} & \frac{1}{N_1}\underset{\sim}{I} \\ \frac{1}{N_1}\underset{\sim}{I} & \frac{1}{N_1}\underset{\sim}{I} \end{pmatrix}\right]. \tag{10}$$

Let $\underset{\sim}{Y}$, $\underset{\sim}{Z}$ and $\underset{\sim}{V}$ be defined by

$$\bar{\underset{\sim}{x}}^{(1)} - \bar{\underset{\sim}{x}}^{(2)} = \underset{\sim}{\delta} + \frac{1}{\sqrt{n}}\underset{\sim}{Y}, \quad \bar{\underset{\sim}{x}}^{(1)} = \frac{1}{\sqrt{n}}\underset{\sim}{Z}, \tag{11}$$

$$\underset{\sim}{S} = \underset{\sim}{I} + \frac{1}{\sqrt{n}}\underset{\sim}{V}. \tag{12}$$

Then the joint distribution of $(\underset{\sim}{Y}', \underset{\sim}{Z}')'$ is

$$N\left[\begin{pmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{pmatrix}, \begin{pmatrix} n\left(\frac{1}{N_1} + \frac{1}{N_2}\right)\underset{\sim}{I} & \frac{n}{N_1}\underset{\sim}{I} \\ \frac{n}{N_1}\underset{\sim}{I} & \frac{n}{N_1}\underset{\sim}{I} \end{pmatrix}\right]. \tag{13}$$

Then (9) is

$$(14)\qquad \Pr\left\{\frac{W - \frac{1}{2}a}{\sqrt{a}} \leq u\right\}$$

$$= \mathcal{E}\Phi\left[\frac{u\sqrt{(\delta + \frac{1}{\sqrt{n}}Y)'(I + \frac{1}{\sqrt{n}}V)^{-1}(\delta + \frac{1}{\sqrt{n}}Y)} + \frac{1}{\sqrt{n}}(\delta + \frac{1}{\sqrt{n}}Y)'(I + \frac{1}{\sqrt{n}}V)^{-1}Z}{\sqrt{(\delta + \frac{1}{\sqrt{n}}Y)'(I + \frac{1}{\sqrt{n}}V)^{-2}(\delta + \frac{1}{\sqrt{n}}Y)}}\right].$$

We can write

$$(15)\qquad (I + \frac{1}{\sqrt{n}}V)^{-1} = I - \frac{1}{\sqrt{n}}V + \frac{1}{n}V^2 - \frac{1}{n^{3/2}}V^3 + \frac{1}{n^2}V^4 - \frac{1}{n^{5/2}}V^5\,(I + \frac{1}{\sqrt{n}}V)^{-1},$$

$$(16)\qquad (I + \frac{1}{\sqrt{n}}V)^{-2} = I - \frac{2}{\sqrt{n}}V + \frac{3}{n}V^2 - \frac{4}{n^{3/2}}V^3 + \frac{5}{n^{3/2}}V^4$$

$$- \frac{1}{n^{5/2}}(6V^5 + \frac{5}{\sqrt{n}}V^6)(I + \frac{1}{\sqrt{n}}V)^{-1}.$$

Then (as Taylor series expansions) we have

$$(17)\qquad \left[(\delta + \frac{1}{\sqrt{n}}Y)'(I + \frac{1}{\sqrt{n}}V)^{-1}(\delta + \frac{1}{\sqrt{n}}Y)\right]^{1/2}$$

$$= \left[\delta'\delta + \frac{1}{\sqrt{n}}(2\delta'Y - \delta'V\delta) + \frac{1}{n}(\delta'V^2\delta + Y'Y - 2\delta'VY) + r_{1n}(Y,Z,V)\right]^{1/2}$$

$$= \Delta + \frac{1}{2\Delta\sqrt{n}}(2\delta'Y - \delta'V\delta) + \frac{1}{n}\left[\frac{1}{2\Delta}(\delta'V^2\delta + Y'Y - 2\delta'VY)\right.$$

$$\left. - \frac{1}{8\Delta^3}(2\delta'Y - \delta'V\delta)^2\right] + r_{2n}(Y,Z,V),$$

(18) $$\frac{1}{\sqrt{n}}\left(\delta + \frac{1}{\sqrt{n}} Y\right)'\left(I + \frac{1}{\sqrt{n}} V\right)^{-1} Z = \frac{1}{\sqrt{n}} \delta' Z + \frac{1}{n} (Y'Z - \delta' VZ) + r_{3n}(Y,Z,V) \;,$$

(19) $$\left[\left(\delta + \frac{1}{\sqrt{n}} Y\right)'\left(I + \frac{1}{\sqrt{n}} V\right)^{-2}\left(\delta + \frac{1}{\sqrt{n}} Y\right)\right]^{-1/2}$$

$$= \left[\delta'\delta + \frac{1}{\sqrt{n}} (2\delta' Y - 2\delta' V\delta) + \frac{1}{n} (3\delta' V^2 \delta + Y'Y - 4\delta' VY) + r_{4n}(Y,Z,V)\right]^{1/2}$$

$$= \frac{1}{\Delta} - \frac{1}{\Delta^3 \sqrt{n}} (\delta' Y - \delta' V\delta) - \frac{1}{n}\left[\frac{1}{2\Delta^3} (3\delta' V^2 \delta + Y'Y - 4\delta' VY)\right.$$

$$\left. - \frac{3}{2\Delta^5} (\delta' Y - \delta' V\delta)^2\right] + r_{5n}(Y,Z,V) \;.$$

Here $r_{jn}(Y,Z,V)$, $j = 1, \ldots, 5$ is a remainder term consisting of $1/n^{3/2}$ times a homogeneous polynomial (not depending on n) of degree 3 in the elements of Y, Z, and V plus $1/n^2$ times a homogeneous polynomial of degree 4 plus a remainder term which is $O(n^{-5/2})$ for fixed Y, Z, and V.

The argument of $\Phi(\;)$ in (14) is the product of

(20) $$u\Delta + \frac{1}{\sqrt{n}} \frac{u}{2\Delta}[(2\delta' Y - \delta' V\delta) + \delta' Z] + \frac{1}{n}\left[\frac{u}{2\Delta} (\delta' V^2 \delta + Y'Y - 2\delta' VY)\right.$$

$$\left. - \frac{u}{8\Delta^3} (2\delta' Y - \delta' V\delta)^2 + Y'Z - \delta' VZ\right] + r_{6n}(Y,Z,V)$$

and (19), which is

(21) $$u + \frac{1}{\sqrt{n}} \left(\frac{u}{2\Delta^2} \delta' V\delta + \frac{1}{\Delta} \delta' Z\right) + \frac{1}{n}\left[\frac{u}{\Delta^2} (\delta' Vy - \delta' V^2 \delta) + \frac{u}{\Delta^4} \left(-\delta' Y\delta' V\delta + \frac{7}{8} (\delta' V\delta)^2\right)\right.$$

$$\left. + \frac{1}{\Delta} Y'Z - \frac{1}{\Delta} \delta' VZ - \frac{1}{\Delta^3} \delta' YZ'\delta + \frac{1}{\Delta^3} \delta' Z\delta' V\delta\right] + r_{7n}(Y,Z,V)$$

$$= u + \frac{1}{\sqrt{n}} C(Z,V) + \frac{1}{n} D(Y,Z,V) + r_{7n}(Y,Z,V) \;,$$

say [as the definition of $C(\underset{\sim}{Z},\underset{\sim}{V})$ and $D(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$] and $r_{6n}(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$ and $r_{7n}(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$ have the same properties as $r_{jn}(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$, $j = 1, \ldots, 5$.

A Taylor series expansion of $\Phi(\ )$ in (14) gives

$$\begin{aligned} (22)\qquad \Phi[u &+ \frac{1}{\sqrt{n}} C(\underset{\sim}{Z},\underset{\sim}{V}) + \frac{1}{n} D(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) + r_{7n}(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})] \\ &= \Phi(u) + \phi(u) \left\{\frac{1}{\sqrt{n}} C(\underset{\sim}{Z},\underset{\sim}{V}) + \frac{1}{n} [D(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) - \frac{1}{2} u\, C^2(\underset{\sim}{Z},\underset{\sim}{V})]\right\} \\ &\quad + \frac{1}{n^{3/2}} r_8(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) + \frac{1}{n^2} r_9(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) + r_{10n}(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}), \end{aligned}$$

where $r_8(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$ is a homogeneous polynomial (not depending on n but depending on u) of degree 3 in the elements of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$, $r_9(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$ is a polynomial of degree 4, and $r_{10n}(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$ is a remainder term, which is $O(n^{-5/2})$ for fixed $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$ (and u).

Let J_n be the set of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$ such that $|y_i| < 2\sqrt{\log n}$, $|z_i| < 2\sqrt{\log n}$, $i = 1, \ldots, p$, and $|v_{ij}| < 2 \log n$, $i, j = 1, \ldots, p$. As shown in the Appendix,

$$(23)\qquad \Pr\{J_n\} = 1 - o(n^{-2}) .$$

The difference between $\mathcal{E}\Phi(\ )$ and the integral of $\Phi(\ )$ times the density of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$ over J_n is $o(n^{-2})$, because $0 \le \Phi(\ ) \le 1$. In J_n each element of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$ divided by $\sqrt{n}$ is less than a constant times $(\log n/\sqrt{n})^5$. Hence

$$(24)\qquad r_{10n}(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) < \text{constant} \times \left(\frac{\log n}{\sqrt{n}}\right)^5 ,$$

and the integral of this times the density of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$ over J_n is $o(n^{-2})$.

Since fourth-order absolute moments of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$ exist and are bounded, the integral of $r_9(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$ times the density of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$ over J_n is bounded; hence, the contribution of this term (with the factor n^{-2}) is $O(n^{-2})$.

The differences between $n^{-1/2}\,\mathcal{E}C(\underset{\sim}{Z},\underset{\sim}{V})$, $n^{-1}[\mathcal{E}\,D(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) - \frac{1}{2}uC^2(\underset{\sim}{Z},\underset{\sim}{V})]$ and $n^{-3/2}\,r_8(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$ and the integrals over J_n of $n^{-1/2}\,C(\underset{\sim}{Z},\underset{\sim}{V})$, $n^{-1}[D(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) - \frac{1}{2}uC^2(\underset{\sim}{Z},\underset{\sim}{V})]$ and $n^{-3/2}\,r_8(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$ times the density of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$, respectively, are $O(n^{-2})$. Thus

$$(25)\quad \Pr\left\{\frac{W-\frac{1}{2}a}{\sqrt{a}} \le u\right\} = \Phi(u) + \phi(u)\left\{\frac{1}{\sqrt{n}}\,\mathcal{E}C(\underset{\sim}{Z},\underset{\sim}{V}) + \frac{1}{n}[\mathcal{E}D(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) - \frac{u}{2}\,\mathcal{E}C^2(\underset{\sim}{Z},\underset{\sim}{V})]\right\}$$
$$+ \frac{1}{n^{3/2}}\,\mathcal{E}\,r_8(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) + O(n^{-2})$$
$$= \Phi(u) + \phi(u)\,\{\frac{1}{\sqrt{n}}\,\mathcal{E}\,C(\underset{\sim}{Z},\underset{\sim}{V}) + \frac{1}{n}\,[\,\mathcal{E}D(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V})$$
$$- \frac{u}{2}\,\mathcal{E}\,C^2(\underset{\sim}{Z},\underset{\sim}{V})]\} + O(n^{-2})\ .$$

because the third-order moments of the elements of $\underset{\sim}{Y}$, $\underset{\sim}{Z}$, and $\underset{\sim}{V}$ are either 0 or $O(n^{-2})$.

Since $C(\underset{\sim}{Z},\underset{\sim}{V})$ is linear and homogeneous, $\mathcal{E}C(\underset{\sim}{Z},\underset{\sim}{V}) = 0$. Since $(\underset{\sim}{Y},\underset{\sim}{Z})$ and $\underset{\sim}{V}$ are independent

$$(26)\quad \mathcal{E}D(\underset{\sim}{Y},\underset{\sim}{Z},\underset{\sim}{V}) = -\frac{u}{\Delta^2}\,\mathcal{E}\,\underset{\sim}{\delta}'\underset{\sim}{V}^2\underset{\sim}{\delta} + \frac{7}{8}\frac{u}{\Delta^4}\,\mathcal{E}(\underset{\sim}{\delta}'\underset{\sim}{V}\underset{\sim}{\delta})^2 + \frac{1}{\Delta}\,\mathcal{E}\,\underset{\sim}{Y}'\underset{\sim}{Z} - \frac{1}{\Delta^3}\,\mathcal{E}\,\underset{\sim}{\delta}'\underset{\sim}{Y}\underset{\sim}{Z}'\underset{\sim}{\delta}$$
$$= -\frac{u}{\Delta^2}\,\Delta^2(p+1) + \frac{7}{8}\,\frac{u}{\Delta^4}\,2\Delta^4 + \frac{1}{\Delta}\frac{n}{N_1}\,p - \frac{1}{\Delta^3}\,\Delta^2\,\frac{n}{N_1}$$
$$= -(p-\frac{3}{4})u + (p-1)\,\frac{n}{N_1}\,\frac{1}{\Delta}$$

since

$$(27) \qquad \mathcal{E}\,\delta' V^2 \delta = \mathcal{E}\,\delta' V V' \delta = \Delta^2 \; \mathcal{E} \sum_{i=1}^{p} v_{1i}^2 = \Delta^2 (\mathcal{E}\, v_{11}^2 + \sum_{i=2}^{p} \mathcal{E}\, v_{1i}^2) = \Delta^2 (p+1) \; ,$$

$$(28) \qquad \mathcal{E}(\delta' V \delta)^2 = \Delta^4 \; \mathcal{E}\, v_{11}^2 = 2\Delta^4 \; .$$

We have

$$(29) \qquad \mathcal{E}\, C^2(Z,V) = \frac{u^2}{4\Delta^4} \; \mathcal{E}(\delta' V \delta)^2 + \frac{1}{\Delta^2} \; \mathcal{E}\,\delta' Z Z' \delta$$

$$= \frac{u^2}{4\Delta^4} \cdot 2\Delta^4 + \frac{1}{\Delta^2} \frac{n}{N_1} \Delta^2 = \frac{1}{2} u^2 + \frac{n}{N_1} \; .$$

Thus

$$(30) \qquad \mathcal{E}\, D(Y,Z,V) - \frac{u}{2} \; \mathcal{E}\, C^2(Z,V) = (p-1) \frac{n}{N_1} \frac{1}{\Delta} - (p - \frac{3}{4} + \frac{1}{2} \frac{n}{N_1})u - \frac{1}{4} u^3 \; .$$

Replacing n/N_1 by its limit $1+k$, we have

$$(31) \qquad \Pr\left\{ \frac{W-a}{\sqrt{a}} \leq u \right\} = \Phi(u) + \frac{1}{n} \; \phi(u) \left[\frac{(p-1)}{\sqrt{\alpha}} (1+k) - (p - \frac{1}{4} + \frac{1}{2} k)u - \frac{1}{4} u^3 \right] + O(n^{-2})$$

when $\mathcal{E}\, x = \mu^{(1)}$. Interchanging N_1 and N_2 gives

$$(32) \qquad \Pr\left\{ \frac{W + \frac{1}{2} a}{\sqrt{a}} \leq v \right\} = \phi(v) - \frac{1}{n} \phi(v) \left[\frac{p-1}{\sqrt{\alpha}} (1 + \frac{1}{k}) + (p - \frac{1}{4} + \frac{1}{2k})v + \frac{1}{4} v^3 \right]$$

$$+ O(n^{-2}) \; ,$$

when $\mathcal{E}\, x = \mu^{(2)}$.

3. Discussion

If $N_1 = N_2$ and costs of misclassification are equal, the minimax classification procedure is defined by the cut-off point 0 for W; a cut-off point different from 0 will increase one probability of misclassification and decrease the other. The inequality $W \leq 0$ is equivalent to $(W - \frac{1}{2} a)/\sqrt{a} \leq -\frac{1}{2}\sqrt{a}$, and $-\frac{1}{2}\sqrt{a}$ estimates $-\frac{1}{2}\sqrt{\alpha} = -\frac{1}{2}\Delta$. For most purposes, then, one is interested in $u \leq 0$. Then the correction term to $\Phi(u)$ is nonpositive; use of the normal approximation alone tends to underestimate the probability of misclassification. The correction term decreases as the distance Δ between the two populations increases if $p > 1$ and for nonpositive u the correction term increases with the number of coordinates p (for fixed Δ).

The expansions of $\Pr\{(W - \frac{1}{2}\sqrt{\alpha})/\sqrt{\alpha} \leq u | \mu=\mu^{(1)}\}$ and $\Pr\{(W + \frac{1}{2}\sqrt{\alpha})/\sqrt{\alpha} \leq u | \mu=\mu^{(2)}\}$ given by Okamoto (1963) can be obtained by the method of this paper. It is interesting that the expansions for $(W \pm \frac{1}{2}\sqrt{a})/\sqrt{a}$ here are much simpler than the expansions for $(W \pm \frac{1}{2}\sqrt{\alpha})/\sqrt{\alpha}$ as given by Okamoto. At $\mu = -\frac{1}{2}\Delta = -\frac{1}{2}\sqrt{\alpha}$ (corresponding to the cut-off point 0) the correction term of order $1/n$ to the probability for $(W \pm \frac{1}{2}\sqrt{\alpha})/\sqrt{\alpha}$ is about $\frac{1}{2}$ as much as for $(W \pm \frac{1}{2}\sqrt{a}/\sqrt{a}$.

As indicated in the introduction, the statistician may want to use the evaluation of $\Pr\{(W - \frac{1}{2}\sqrt{a})/\sqrt{a} \leq u\}$ in order to set the cut-off point $c = u\sqrt{a} + \frac{1}{2}\sqrt{a}$ in order to obtain a specified probability of misclassification or at least approximate a specified probability. The crudest approximation is to take u so $\Phi(u)$ is the specified probability.

This approximation, however, is not very good; the error of the approximation is evaluated above to order $1/n^{3/2}$. The error depends on the unknown parameter if $p > 1$. To get a better approximation let $\Phi(u+\Delta u)$ be the specified probability, where

$$\Delta u = -\frac{1}{n} \frac{(p-1)(1+k)}{\sqrt{a}} - (p - \frac{1}{4} + \frac{1}{2} k)u - \frac{1}{4} u^3 \quad .$$

Then the actual probability is the specified one with an error of order $n^{-3/2}$ (because $\sqrt{a}$ is $\sqrt{\alpha}$ with an error of order $n^{-1/2}$).

For further discussion, see Anderson (1972).

Acknowledgements. The author is indebted to John Miller and Fred Nold for assistance in preparing this paper.

APPENDIX

To control the errors of approximation we define the set J_n by $|y_j| < 2\sqrt{\log n}$, $|z_j| < 2\sqrt{\log n}$, $j = 1, \dots, p$, and $|v_{ij}| < 2 \log n$, $i, j = 1, \dots, p$. We want to show $\Pr\{J_n\} = 1 - O(n^{-2})$.

We have when $\underset{\sim}{\Sigma} = \underset{\sim}{I}$

$$\text{(A.1)} \qquad \Pr\{|y_j| > 2\sqrt{\log n}\} = \frac{2}{\sqrt{2\pi}} \int_{2\sqrt{\log n}}^{\infty} e^{-\frac{1}{2}} dv$$

$$< \frac{e^{-\frac{1}{2}(2\sqrt{\log n})^2}}{\sqrt{2\pi \log n}}$$

$$= \frac{1}{n^2 \sqrt{2\pi \log n}}$$

$$= o(n^{-2})$$

by use of Mill's ratio. Then

$$\text{(A.2)} \qquad \Pr\{|y_j| < 2\sqrt{\log n},\ |z_j| < 2 \log n,\ j = 1, \dots, p\}$$

$$= 1 - o(n^{-2}) .$$

Now consider $\underset{\sim}{V} = (v_{ij})$. The moment generating function of $n\underset{\sim}{S}$ when $\underset{\sim}{\Sigma} = \underset{\sim}{I}$ is

$$\text{(A.3)} \qquad \mathcal{E} e^{\operatorname{tr}\underset{\sim}{\Theta} n \underset{\sim}{S}} = \mathcal{E} e^{n\sum_{i,j=1}^{p} \theta_{ij} s_{ij}}$$

$$= |\underset{\sim}{I} - 2\underset{\sim}{\Theta}|^{-1/2\, n}$$

where $\underset{\sim}{\Theta} = \underset{\sim}{\Theta}'$. We use the Tchebycheff-type inequality [Chernoff (1952), for example] for an arbitrary random variable X and $\theta > 0$

(A.4)
$$e^{-\theta a} \mathscr{E} e^{\theta X} = \mathscr{E} e^{\theta(X-a)} \geq \Pr\{X \geq a\} .$$

Then

(A.5)
$$\begin{aligned} \Pr\{v_{ii} > 2 \log n\} &= \Pr\{\sqrt{n}\, s_{ii} - \sqrt{n} > 2 \log n\} \\ &= \Pr\{n\, s_{ii} > n + 2\sqrt{n} \log n\} \\ &\leq (1-2\theta)^{-1/2\, n}\, e^{-\theta(n+2\sqrt{n} \log n)} \end{aligned}$$

for $0 < \theta < \frac{1}{2}$. Let $\theta = k/\sqrt{n}$, where $k > 1$. For $n > 4k^2$

(A.6)
$$\begin{aligned} \Pr\{v_{ii} > 2 \log n\} &\leq (1-2k/\sqrt{n})^{-1/2\, n}\, e^{-2k \log n - k\sqrt{n}} \\ &\leq \text{constant} \times e^{-2k \log n} \\ &= O(n^{-2k}) \\ &= o(n^{-2}). \end{aligned}$$

Similarly $\Pr\{-v_{ii} > 2 \log n\} = o(n^{-2})$. We have for $i \neq j$

(A.7)
$$\begin{aligned} \Pr\{v_{ij} > 2 \log n\} &= \Pr\{\sqrt{n}\, s_{ij} > 2 \log n\} \\ &= \Pr\{n\, s_{ij} > 2\sqrt{n} \log n\} \\ &\leq e^{-\theta 2\sqrt{n} \log n}\, (1-\theta^2)^{-1/2\, n} \end{aligned}$$

for $0 < \theta < \frac{1}{2}$. Let $\theta = k/\sqrt{n}$, where $k > 1$. For $n > 4k^2$

(A.8)
$$\begin{aligned} \Pr\{v_{ij} > 2 \log n\} &\leq \text{constant} \times e^{-2k \log n} \\ &= o(n^{-2}). \end{aligned}$$

Similarly $\Pr\{-v_{ij} > 2 \log n\} = o(n^{-2})$. Then

(A.9)
$$\Pr\{|v_{ij}| < 2 \log n,\ i, j = 1, \ldots, p\} = 1-o(n^{-2}).$$

REFERENCES

Anderson, T. W. (1951), Classification by multivariate analysis, Psychometrika, 16, 31-50.

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

Anderson, T. W. (1972), Asymptotic evaluation of the probability of misclassification by the linear discriminant function, Technical Report No. 10, September, 1972, Stanford University.

Bowker, Albert H., and Rosedith Sitgreaves (1961), An asymptotic expansion for the distribution function of the W-classification statistic, Chapter 19 in Solomon, H. (ed.), Studies in Item Analysis and Prediction, Stanford University Press.

Chernoff, H. (1952), A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Annals of Mathematical Statistics, 23, 493-507.

Okamoto, Masashi (1963), An asymptotic expansion for the distribution of the linear discriminant function, Annals of Mathematical Statistics, 34, 1286-1301; correction (1968), 39, 1358-1359.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) DEPARTMENT OF STATISTICS STANFORD UNIVERSITY, CALIF.	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP
3. REPORT TITLE AN ASYMPTOTIC EXPANSION OF THE DISTRIBUTION OF THE "STUDENTIZED" CLASSIFICATION STATISTIC W	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) TECHNICAL REPORT	
5. AUTHOR(S) (Last name, first name, initial) ANDERSON, T. W.	
6. REPORT DATE September 10, 1972	7a. TOTAL NO. OF PAGES 14 7b. NO. OF REFS 6
8a. CONTRACT OR GRANT NO. N000-4167-A-0112-0030 b. PROJECT NO. NR-042-034 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
10. AVAILABILITY/LIMITATION NOTICES Reproduction in whole or in part is permitted for any purpose of the UNITED STATES GOVERNMENT	
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY OFFICE OF NAVAL RESEARCH ARLINGTON, VIRGINIA

13. ABSTRACT.

An observation $\underset{\sim}{x}$ is to be classified as coming from $N(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma})$ or $N(\underset{\sim}{\mu}^{(2)}, \underset{\sim}{\Sigma})$. The parameters, which are unknown, may be estimated by $\bar{\underset{\sim}{x}}^{(1)}$, the mean of a sample of N_1 from the first population, $\bar{\underset{\sim}{x}}^{(2)}$, the mean of a sample of N_2 from the second population, and $\underset{\sim}{S}$, the pooled sample covariance. The classification statistic $W = (\bar{\underset{\sim}{x}}^{(1)}-\bar{\underset{\sim}{x}}^{(2)})' \underset{\sim}{S}^{-1} [\underset{\sim}{x} - \frac{1}{2}(\bar{\underset{\sim}{x}}^{(1)} + \bar{\underset{\sim}{x}}^{(2)})]$ has a distribution which depends on $\alpha = (\underset{\sim}{\mu}^{(1)}-\underset{\sim}{\mu}^{(2)})' \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\mu}^{(1)}-\underset{\sim}{\mu}^{(2)})$ and its limiting distribution as $N_1 \to \infty$ and $N_2 \to \infty$ is $N(\frac{1}{2}\alpha, \alpha)$ or $N(-\frac{1}{2}\alpha, \alpha)$. An asymptotic expansion of $(W - \frac{1}{2}a)/\sqrt{a}$ and $(W + \frac{1}{2}a)/\sqrt{a}$ is made to order $n^{-3/2}$, where $a = (\bar{\underset{\sim}{x}}^{(1)}-\bar{\underset{\sim}{x}}^{(2)})'\underset{\sim}{S}^{-1}(\bar{\underset{\sim}{x}}^{(1)}-\bar{\underset{\sim}{x}}^{(2)})$ and $n = N_1 + N_2 - 2$.

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
classification						
classification statistic						
asymptotic expansion						
discriminant analysis						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, *or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.