The Annals of Statistics
1973, Vol. 1, No. 5, 964-972

AN ASYMPTOTIC EXPANSION OF THE DISTRIBUTION OF THE STUDENTIZED CLASSIFICATION STATISTIC W[1]

By T. W. Anderson

Stanford University and Center for Advanced Study in the Behavioral Sciences

The classification statistic W is used to classify an observation as coming from one of two multivariate normal populations with common covariance matrix and different means when these parameters are estimated from two samples, one from each population. The distribution of W depends on the Mahalanobis distance between the populations, α. When the sample sizes approach infinity, the limiting distribution of $(W - \frac{1}{2}\alpha)/\alpha^{\frac{1}{2}}$ is the standard normal distribution if the observation is from the first population; the same is true of $(W - \frac{1}{2}a)/a^{\frac{1}{2}}$, where a is an estimate of α. This paper gives an asymptotic expansion of the distribution of $(W - \frac{1}{2}a)/a^{\frac{1}{2}}$ with an error of the order of the square of the number of observations. The correction to the standard normal distribution function is the standard normal density times a third-degree polynomial in the argument divided by the sum of the observations (less 2).

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

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

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

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

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

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

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

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

Received September 1972; revised January 1973.

[1] This research was supported by Office of Naval Research Contract N00014-67-A-0112-0030, Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government. This paper was completed while the author was a Visiting Scholar at the Center for Advanced Study in the Behavioral Sciences.

Key words and phrases. Classification, classification statistic, asymptotic expansion, discriminant analysis.

which can be estimated by

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

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 $\mathbf{x}$ is from $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ and mean $-\frac{1}{2}\alpha$ if $\mathbf{x}$ is from $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$. Bowker and Sitgreaves (1961) for $N_1 = N_2$ and Okamoto (1963) (with correction, Okamoto (1968)) gave asymptotic expansions of the distribution of $(W - \frac{1}{2}\alpha)/\alpha^{\frac{1}{2}}$ for $\mathbf{x}$ coming from $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ and $(W + \frac{1}{2}\alpha)/\alpha^{\frac{1}{2}}$ for $\mathbf{x}$ coming from $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$ to terms of 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, $\Pr\{W \leqq 0\}$ was evaluated. (In Bowker and Sitgreaves (1961) the coefficients a_{31} and a_{32} should be replaced by $-a_{31}$ and $-a_{32}$, respectively.)

The statistician, who wants to classify $\mathbf{x}$, 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 $(W \pm \frac{1}{2}\alpha)/\alpha^{\frac{1}{2}}$ 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 $(W - \frac{1}{2}\alpha)/\alpha^{\frac{1}{2}}$ and $(W + \frac{1}{2}\alpha)/\alpha^{\frac{1}{2}}$ are $N(0, 1)$ when $\mathscr{E}\mathbf{x} = \boldsymbol{\mu}^{(1)}$ and $\mathscr{E}\mathbf{x} = \boldsymbol{\mu}^{(2)}$, respectively, a first approximation to the pair of misclassification probabilities is $\Phi(\frac{1}{2}\alpha + c\alpha^{\frac{1}{2}})$ and $\Phi(-\frac{1}{2}\alpha + c\alpha^{\frac{1}{2}})$, where $\Phi(\ )$ 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 distributions of $(W - \frac{1}{2}a)/a^{\frac{1}{2}}$ and $(W + \frac{1}{2}a)/a^{\frac{1}{2}}$ are $N(0, 1)$ when $\mathscr{E}\mathbf{x} = \boldsymbol{\mu}^{(1)}$ and $\mathscr{E}\mathbf{x} = \boldsymbol{\mu}^{(2)}$, respectively. In this paper we make asymptotic expansions of the distributions of $(W - \frac{1}{2}a)/a^{\frac{1}{2}}$ and $(W + \frac{1}{2}a)/a^{\frac{1}{2}}$ in these two cases, respectively.

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

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

Then

$$\begin{aligned} \Pr\left\{\frac{W - \frac{1}{2}a}{a^{\frac{1}{2}}} \leqq u\right\} &= \Pr\{(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})'\mathbf{S}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leqq u[(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})'\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})]^{\frac{1}{2}} \\ &\qquad + (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})'\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(1)} - \boldsymbol{\mu})\} . \end{aligned} \tag{7}$$

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

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

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

$$\Pr\left\{\frac{W - \frac{1}{2}a}{a^{\frac{1}{2}}} \leqq u\right\} \tag{9}$$

$$= \Pr\left\{r \leqq \frac{u[(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})'\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})]^{\frac{1}{2}} + (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})'\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(1)} - \boldsymbol{\mu})}{[(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})'\mathbf{S}^{-1}\mathbf{\Sigma}\mathbf{S}^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})]^{\frac{1}{2}}}\right\}$$

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

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

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

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

$$\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)} = \boldsymbol{\delta} + \frac{1}{n^{\frac{1}{2}}}\mathbf{Y}, \qquad \bar{\mathbf{x}}^{(1)} = \frac{1}{n^{\frac{1}{2}}}\mathbf{Z}, \tag{10}$$

$$\mathbf{S} = \mathbf{I} + \frac{1}{n^{\frac{1}{2}}}\mathbf{V}. \tag{11}$$

The joint distribution of $(\mathbf{Y}'\ \mathbf{Z}')'$ is

$$N\left[\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix} n\left(\frac{1}{N_1} + \frac{1}{N_2}\right)\mathbf{I} & \frac{n}{N_1}\mathbf{I} \\ \frac{n}{N_1}\mathbf{I} & \frac{n}{N_1}\mathbf{I}\end{pmatrix}\right]. \tag{12}$$

Then (9) is

$$\begin{aligned} \Pr\left\{\frac{W - \frac{1}{2}a}{a^{\frac{1}{2}}} \leqq u\right\} = \mathscr{E}\Phi\Bigg[\Bigg\{u\left[\left(\boldsymbol{\delta} + \frac{1}{n^{\frac{1}{2}}}\mathbf{Y}\right)'\left(\mathbf{I} + \frac{1}{n^{\frac{1}{2}}}\mathbf{V}\right)^{-1}\left(\boldsymbol{\delta} + \frac{1}{n^{\frac{1}{2}}}\mathbf{Y}\right)\right]^{\frac{1}{2}} \\ + \frac{1}{n^{\frac{1}{2}}}\left(\boldsymbol{\delta} + \frac{1}{n^{\frac{1}{2}}}\mathbf{Y}\right)'\left(\mathbf{I} + \frac{1}{n^{\frac{1}{2}}}\mathbf{V}\right)^{-1}\mathbf{Z}\Bigg\} \\ \div \left\{\left[\left(\boldsymbol{\delta} + \frac{1}{n^{\frac{1}{2}}}\mathbf{Y}\right)'\left(\mathbf{I} + \frac{1}{n^{\frac{1}{2}}}\mathbf{V}\right)^{-2}\left(\boldsymbol{\delta} + \frac{1}{n^{\frac{1}{2}}}\mathbf{Y}\right)\right]^{\frac{1}{2}}\right\}\Bigg]. \end{aligned} \tag{13}$$

We can write

$$\begin{aligned}\left(\mathbf{I} + \frac{1}{n^{\frac{1}{2}}}\mathbf{V}\right)^{-1} = \mathbf{I} - \frac{1}{n^{\frac{1}{2}}}\mathbf{V} + \frac{1}{n}\mathbf{V}^2 - \frac{1}{n^{\frac{3}{2}}}\mathbf{V}^3 + \frac{1}{n^2}\mathbf{V}^4 \\ - \frac{1}{n^{\frac{5}{2}}}\mathbf{V}^5\left(\mathbf{I} + \frac{1}{n^{\frac{1}{2}}}\mathbf{V}\right)^{-1},\end{aligned} \tag{14}$$

$$\begin{aligned}\left(\mathbf{I} + \frac{1}{n^{\frac{1}{2}}}\mathbf{V}\right)^{-2} = \mathbf{I} - \frac{2}{n^{\frac{1}{2}}}\mathbf{V} + \frac{3}{n}\mathbf{V}^2 - \frac{4}{n^{\frac{3}{2}}}\mathbf{V}^3 + \frac{5}{n^2}\mathbf{V}^4 \\ - \frac{1}{n^{\frac{5}{2}}}\left(6\mathbf{V}^5 + \frac{5}{n^{\frac{1}{2}}}\mathbf{V}^6\right)\left(\mathbf{I} + \frac{1}{n^{\frac{1}{2}}}\mathbf{V}\right)^{-2}.\end{aligned} \tag{15}$$

Then (as Taylor series expansions) we have

$$\begin{aligned}
&\left[\left(\delta + \frac{1}{n^{\frac12}}\mathbf{Y}\right)'\left(\mathbf{I} + \frac{1}{n^{\frac12}}\mathbf{V}\right)^{-1}\left(\delta + \frac{1}{n^{\frac12}}\mathbf{Y}\right)\right]^{\frac12} \\
&\quad = \left[\delta'\delta + \frac{1}{n^{\frac12}}(2\delta'\mathbf{Y} - \delta'\mathbf{V}\delta) + \frac{1}{n}(\delta'\mathbf{V}^2\delta + \mathbf{Y}'\mathbf{Y} - 2\delta'\mathbf{V}\mathbf{Y}) \right.\\
&\qquad \left. + r_{1n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})\right]^{\frac12} \qquad (16)\\
&\quad = \Delta + \frac{1}{2\Delta n^{\frac12}}(2\delta'\mathbf{Y} - \delta'\mathbf{V}\delta) + \frac{1}{n}\left[\frac{1}{2\Delta}(\delta'\mathbf{V}^2\delta + \mathbf{Y}'\mathbf{Y} - 2\delta'\mathbf{V}\mathbf{Y}) \right.\\
&\qquad \left. - \frac{1}{8\Delta^3}(2\delta'\mathbf{Y} - \delta'\mathbf{V}\delta)^2\right] + r_{2n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V}),
\end{aligned}$$

$$\begin{aligned}
(17) \quad &\frac{1}{n^{\frac12}}\left(\delta + \frac{1}{n^{\frac12}}\mathbf{Y}\right)'\left(\mathbf{I} + \frac{1}{n^{\frac12}}\mathbf{V}\right)^{-1}\mathbf{Z} \\
&\quad = \frac{1}{n^{\frac12}}\delta'\mathbf{Z} + \frac{1}{n}(\mathbf{Y}'\mathbf{Z} - \delta'\mathbf{V}\mathbf{Z}) + r_{3n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V}),
\end{aligned}$$

$$\begin{aligned}
&\left[\left(\delta + \frac{1}{n^{\frac12}}\mathbf{Y}\right)'\left(\mathbf{I} + \frac{1}{n^{\frac12}}\mathbf{V}\right)^{-2}\left(\delta + \frac{1}{n^{\frac12}}\mathbf{Y}\right)\right]^{-\frac12} \\
&\quad = \left[\delta'\delta + \frac{1}{n^{\frac12}}(2\delta'\mathbf{Y} - 2\delta'\mathbf{V}\delta) + \frac{1}{n}(3\delta'\mathbf{V}^2\delta + \mathbf{Y}'\mathbf{Y} - 4\delta'\mathbf{V}\mathbf{Y}) \right.\\
&\qquad \left. + r_{4n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})\right]^{-\frac12} \qquad (18)\\
&\quad = \frac{1}{\Delta} - \frac{1}{\Delta^3 n^{\frac12}}(\delta'\mathbf{Y} - \delta'\mathbf{V}\delta) - \frac{1}{n}\left[\frac{1}{2\Delta^3}(3\delta'\mathbf{V}^2\delta + \mathbf{Y}'\mathbf{Y} - 4\delta'\mathbf{V}\mathbf{Y}) \right.\\
&\qquad \left. - \frac{3}{2\Delta^5}(\delta'\mathbf{Y} - \delta'\mathbf{V}\delta)^2\right] + r_{5n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V}).
\end{aligned}$$

Here $r_{jn}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})$, $j = 1, \cdots, 5$, is a remainder term consisting of $1/n^{\frac32}$ times a homogeneous polynomial (not depending on n) of degree 3 in the elements of $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{V}$ plus $1/n^2$ times a homogeneous polynomial of degree 4 plus a remainder term which is $O(n^{-\frac52})$ for fixed $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{V}$.

The argument of $\Phi(\ )$ in (13) is the product of

$$\begin{aligned}
&u\Delta + \frac{1}{n^{\frac12}}\left[\frac{u}{2\Delta}(2\delta'\mathbf{Y} - \delta'\mathbf{V}\delta) + \delta'\mathbf{Z}\right] \\
(19) \qquad &\quad + \frac{1}{n}\left[\frac{u}{2\Delta}(\delta'\mathbf{V}^2\delta + \mathbf{Y}'\mathbf{Y} - 2\delta'\mathbf{V}\mathbf{Y}) \right.\\
&\qquad \left. - \frac{u}{8\Delta^3}(2\delta'\mathbf{Y} - \delta'\mathbf{V}\delta)^2 + \mathbf{Y}'\mathbf{Z} - \delta'\mathbf{V}\mathbf{Z}\right] \\
&\qquad + r_{6n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})
\end{aligned}$$

and (18), which is

$$
\begin{aligned}
& u + \frac{1}{n^{\frac{1}{2}}}\left(\frac{u}{2\Delta^2}\,\delta'\mathbf{V}\delta + \frac{1}{\Delta}\,\delta'\mathbf{Z}\right) + \frac{1}{n}\left[\frac{u}{\Delta^2}\,(\delta'\mathbf{V}\mathbf{Y} - \delta'\mathbf{V}^2\delta)\right. \\
& \qquad + \frac{u}{\Delta^4}\,(-\delta'\mathbf{Y}\delta'\mathbf{V}\delta + \tfrac{7}{8}(\delta'\mathbf{V}\delta)^2) + \frac{1}{\Delta}\,\mathbf{Y}'\mathbf{Z} \qquad (20) \\
& \qquad \left. - \frac{1}{\Delta}\,\delta'\mathbf{V}\mathbf{Z} - \frac{1}{\Delta^3}\,\delta'\mathbf{Y}\mathbf{Z}'\delta + \frac{1}{\Delta^3}\delta'\mathbf{Z}\delta'\mathbf{V}\delta\right] + r_{7n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) \\
& \qquad = u + \frac{1}{n^{\frac{1}{2}}}\,C(\mathbf{Z}, \mathbf{V}) + \frac{1}{n}\,D(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) + r_{7n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})\,,
\end{aligned}
$$

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

A Taylor series expansion of $\Phi(\ )$ in (13) gives

$$
\begin{aligned}
& \Phi\left[u + \frac{1}{n^{\frac{1}{2}}}\,C(\mathbf{Z}, \mathbf{V}) + \frac{1}{n}\,D(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) + r_{7n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})\right] \\
& \qquad = \Phi(u) + \phi(u)\left\{\frac{1}{n^{\frac{1}{2}}}\,C(\mathbf{Z}, \mathbf{Y}) + \frac{1}{n}\left[D(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) - \tfrac{1}{2}uC^2(\mathbf{Z}, \mathbf{V})\right]\right\} \qquad (21) \\
& \qquad + \frac{1}{n^{\frac{3}{2}}}\,r_8(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) + \frac{1}{n^2}\,r_9(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) + r_{10n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})\,,
\end{aligned}
$$

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

Let J_n be the set of $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{V}$ such that $|y_i| < g(\log n)^{\frac{1}{2}}$, $|z_i| < g(\log n)^{\frac{1}{2}}$, $i = 1, \cdots, p$, and $|v_{ij}| < 2 \log n$, $i, j = 1, \cdots, p$, where $g > 2(1 + k)/k^{\frac{1}{2}}$. As shown in the Appendix,

$$
\Pr\{J_n\} = 1 - o(n^{-2})\,. \tag{22}
$$

The difference between $\mathscr{E}\Phi(\ )$ and the integral of $\Phi(\ )$ times the density of $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{V}$ over J_n is $o(n^{-2})$, because $0 \leqq \Phi(\ ) \leqq 1$. In J_n each element of $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{V}$ divided by $n^{\frac{1}{2}}$ is less than a constant times $\log n/n^{\frac{1}{2}}$. The part of the remainder $r_{jn}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})$, $j = 1, \cdots, 7$, that is $O(n^{-\frac{5}{2}})$ for fixed $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{V}$ can be written as a homogeneous polynomial of degree 5 in the elements of $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{V}$ with coefficients possibly depending on $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{V}$ (by use of Taylor series with remainder); each coefficient is bounded in J_n (for sufficiently large n). The same holds for $r_{10n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})$. Hence, in J_n

$$
|r_{10n}(\mathbf{Y}, \mathbf{Z}, \mathbf{V})| < \text{constant} \times \left(\frac{\log n}{n^{\frac{1}{2}}}\right)^5\,, \tag{23}
$$

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

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

$$\begin{aligned} \Pr\left\{\frac{W - \frac{1}{2}a}{a^{\frac{1}{2}}} \leqq u\right\} &= \Phi(u) + \phi(u)\left\{\frac{1}{n^{\frac{1}{2}}}\mathscr{E}C(\mathbf{Z}, \mathbf{V}) + \frac{1}{n}\left[\mathscr{E}D(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) - \frac{u}{2}\mathscr{E}C^2(\mathbf{Z}, \mathbf{V})\right]\right\} \\ &\quad + \frac{1}{n^{\frac{3}{2}}}\mathscr{E}r_8(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) + O(n^{-2}) \\ &= \Phi(u) + \phi(u)\left\{\frac{1}{n^{\frac{1}{2}}}\mathscr{E}C(\mathbf{Z}, \mathbf{V}) + \frac{1}{n}\left[\mathscr{E}D(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) - \frac{u}{2}\mathscr{E}C^2(\mathbf{Z}, \mathbf{V})\right]\right\} \\ &\quad + O(n^{-2}) \end{aligned} \tag{24}$$

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

Since $C(\mathbf{Z}, \mathbf{V})$ is linear and homogeneous, $\mathscr{E}C(\mathbf{Z}, \mathbf{V}) = 0$. Since $(\mathbf{Y}, \mathbf{Z})$ and **V** are independent,

$$\begin{aligned} \mathscr{E}D(\mathbf{Y}, \mathbf{Z}, \mathbf{V}) &= -\frac{u}{\Delta^2}\mathscr{E}\boldsymbol{\delta}'\mathbf{V}^2\boldsymbol{\delta} + \frac{7}{8}\frac{u}{\Delta^4}\mathscr{E}(\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta})^2 \\ &\quad + \frac{1}{\Delta}\mathscr{E}\mathbf{Y}'\mathbf{Z} - \frac{1}{\Delta^3}\mathscr{E}\boldsymbol{\delta}'\mathbf{Y}\mathbf{Z}'\boldsymbol{\delta} \\ &= -(p - \tfrac{3}{4})u + (p - 1)\frac{n}{N_1}\frac{1}{\Delta} \end{aligned} \tag{25}$$

since

$$\begin{aligned} \mathscr{E}\boldsymbol{\delta}'\mathbf{V}^2\boldsymbol{\delta} = \mathscr{E}\boldsymbol{\delta}'\mathbf{V}\mathbf{V}'\boldsymbol{\delta} &= \Delta^2\mathscr{E}\textstyle\sum_{i=1}^p v_{1i}^2 \\ &= \Delta^2(\mathscr{E}v_{11}^2 + \textstyle\sum_{i=2}^p \mathscr{E}v_{1i}^2) = \Delta^2(p + 1)\,, \end{aligned} \tag{26}$$

$$\mathscr{E}(\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta})^2 = \Delta^4\mathscr{E}v_{11}^2 = 2\Delta^4\,. \tag{27}$$

We have

$$\begin{aligned} \mathscr{E}C^2(\mathbf{Z}, \mathbf{V}) &= \frac{u^2}{4\Delta^4}\mathscr{E}(\boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta})^2 + \frac{1}{\Delta^2}\mathscr{E}\boldsymbol{\delta}'\mathbf{Z}\mathbf{Z}'\boldsymbol{\delta} \\ &= \tfrac{1}{2}u^2 + \frac{n}{N_1}\,. \end{aligned} \tag{28}$$

Replacing n/N_1 by its limit $1 + k$ and substituting in (24), we have

$$\begin{aligned} \Pr\left\{\frac{W - \frac{1}{2}a}{a^{\frac{1}{2}}} \leqq u\right\} &= \Phi(u) + \frac{1}{n}\phi(u)\left[\frac{(p - 1)}{\alpha^{\frac{1}{2}}}(1 + k) - (p - \tfrac{1}{4} + \tfrac{1}{2}k)u - \tfrac{1}{4}u^3\right] \\ &\quad + O(n^{-2}) \end{aligned} \tag{29}$$

when $\mathscr{E}\mathbf{x} = \boldsymbol{\mu}^{(1)}$. Interchanging N_1 and N_2 gives

$$\Pr\left\{\frac{W + \frac{1}{2}a}{a^{\frac{1}{2}}} \leqq u\right\}$$

$$= \Phi(v) - \frac{1}{n}\phi(v)\left[\frac{p-1}{\alpha^{\frac{1}{2}}}\left(1 + \frac{1}{k}\right) + \left(p - \tfrac{1}{4} + \frac{1}{2k}\right)v + \tfrac{1}{4}v^3\right] + O(n^{-2}), \tag{30}$$

when $\mathscr{E}\mathbf{x} = \boldsymbol{\mu}^{(2)}$.

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 $W \leqq 0$ is equivalent to $(W - \frac{1}{2}a)/a^{\frac{1}{2}} \leqq -\frac{1}{2}a^{\frac{1}{2}}$, and $-\frac{1}{2}a^{\frac{1}{2}}$ estimates $-\frac{1}{2}\alpha^{\frac{1}{2}} = -\frac{1}{2}\Delta$. For most purposes, then, one is interested in $u \leqq 0$. Then the correction term to $\Phi(u)$ is nonnegative; 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 $\Pr\{(W - \frac{1}{2}\alpha^{\frac{1}{2}})/\alpha^{\frac{1}{2}} \leqq u \mid \mu = \mu^{(1)}\}$ and $\Pr\{(W + \frac{1}{2}\alpha^{\frac{1}{2}})/\alpha^{\frac{1}{2}} \leqq u \mid \mu = \mu^{(2)}\}$ given by Okamoto (1963) can be obtained by the method of this paper. It is interesting that the expansions for $(W \pm \frac{1}{2}a^{\frac{1}{2}})/a^{\frac{1}{2}}$ here are much simpler than the expansions for $(W \pm \frac{1}{2}\alpha^{\frac{1}{2}})/\alpha^{\frac{1}{2}}$ as given by Okamoto. At $\mu = -\frac{1}{2}\Delta = -\frac{1}{2}\alpha^{\frac{1}{2}}$ (corresponding to the cut-off point 0) the correction term of order $1/n$ to the probability for $(W \pm \frac{1}{2}\alpha^{\frac{1}{2}})/\alpha^{\frac{1}{2}}$ is about $\frac{1}{2}$ as much as for $(W \pm \frac{1}{2}a^{\frac{1}{2}})/a^{\frac{1}{2}}$.

As indicated in the introduction, the statistician may want to use the evaluation of $\Pr\{(W - \frac{1}{2}a^{\frac{1}{2}})/a^{\frac{1}{2}} \leqq u\}$ in order to set the cut-off point $c = ua^{\frac{1}{2}} + \frac{1}{2}a$ 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 $1/n^{\frac{3}{2}}$. 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

$$\delta u = -\frac{1}{n}\left[\frac{(p-1)(1+k)}{a^{\frac{1}{2}}} - (p - \tfrac{1}{4} + \tfrac{1}{2}k)u - \tfrac{1}{4}u^3\right]. \tag{31}$$

Then the actual probability is the specified one with an error of order n^{-2}.

For further discussion, see Anderson (1972).

Acknowledgments. 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 $|y_j| < 2(\log n)^{\frac{1}{2}}$, $|z_j| < 2(\log n)^{\frac{1}{2}}$, $j = 1, \cdots, p$, and $|v_{ij}| < 2 \log n$, $i, j = 1, \cdots, p$. We want to show $\Pr\{J_n\} = 1 - o(n^{-2})$.

We have when $\mathbf{\Sigma} = \mathbf{I}$ and $h_n = g(\log n)^{\frac{1}{2}}/[n(N_1^{-1} + N_2^{-1})]^{\frac{1}{2}}$

$$\begin{aligned} \Pr\{|y_j| > g(\log n)^{\frac{1}{2}}\} &= \frac{2}{(2\pi)^{\frac{1}{2}}} \int_{h_n}^{\infty} e^{-\frac{1}{2}v^2}\, dv \\ &< \frac{2e^{-\frac{1}{2}h_n^2}}{(2\pi)^{\frac{1}{2}} h_n} \\ &= \left[\frac{2n(n+2)}{\pi g^2 N_1 N_2 \log n}\right]^{\frac{1}{2}} n^{-\frac{1}{2}g^2 N_1 N_2/[n(n+2)]} \\ &= o(n^{-2}) \end{aligned} \tag{A.1}$$

by use of Mill's ratio. Then

$$\Pr\{|y_j| < g(\log n)^{\frac{1}{2}}, |z_j| < g(\log n)^{\frac{1}{2}}, j = 1, \cdots, p\} = 1 - o(n^{-2}) . \tag{A.2}$$

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

$$\begin{aligned} \mathscr{E} \exp[\operatorname{tr} \mathbf{\Theta} n\mathbf{S}] &= \mathscr{E} \exp[n \textstyle\sum_{i,j=1}^{p} \theta_{ij} s_{ij}] \\ &= |\mathbf{I} - 2\mathbf{\Theta}|^{-\frac{1}{2}n} \end{aligned} \tag{A.3}$$

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

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

Then

$$\begin{aligned} \Pr\{v_{ii} > 2 \log n\} &= \Pr\{n^{\frac{1}{2}} s_{ii} - n^{\frac{1}{2}} > 2 \log n\} \\ &= \Pr\{n s_{ii} > n + 2n^{\frac{1}{2}} \log n\} \\ &\leqq (1 - 2\theta)^{-\frac{1}{2}n} \exp[-\theta(n + 2n^{\frac{1}{2}} \log n)] \end{aligned} \tag{A.5}$$

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

$$\begin{aligned} \Pr\{v_{ii} > 2 \log n\} &\leqq (1 - 2\gamma/n^{\frac{1}{2}})^{-\frac{1}{2}n} \exp[-2\gamma \log n - kn^{\frac{1}{2}}] \\ &\leqq \text{constant} \times \exp[-2\gamma \log n] \\ &= O(n^{-2\gamma}) \\ &= o(n^{-2}) . \end{aligned} \tag{A.6}$$

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

$$\begin{aligned} \Pr\{v_{ij} > 2 \log n\} &= \Pr\{n^{\frac{1}{2}} s_{ij} > 2 \log n\} \\ &= \Pr\{n s_{ij} > 2n^{\frac{1}{2}} \log n\} \\ &\leqq \exp[-\theta 2n^{\frac{1}{2}} \log n](1 - \theta^2)^{-\frac{1}{2}n} \end{aligned} \tag{A.7}$$

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

$$\begin{aligned} \Pr\{v_{ij} > 2 \log n\} &\leqq \text{constant} \times \exp[-2\gamma \log n] \\ &= o(n^{-2}) . \end{aligned} \tag{A.8}$$

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

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

REFERENCES

ANDERSON, T. W. (1951). Classification by multivariate analysis. *Psychometrika* **16** 31–50.

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

ANDERSON, T. W. (1972). Asymptotic evaluation of the probabilities of misclassification by linear discriminant functions. Technical Report No. 10, Stanford Univ. To be published in the Proceedings of the NATO Advanced Study Institute on Discriminant Analysis and Applications.

BOWKER, ALBERT H. and SITGREAVES, ROSEDITH (1961). An asymptotic expansion for the distribution function of the W-classification statistic. *Studies in Item Analysis and Prediction* (H. Solomon, ed.). Stanford Univ. Press.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

OKAMOTO, MASASHI (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.* **34** 1286–1301, Correction **39** 1358–1359.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305