ASYMPTOTIC DISTRIBUTION OF DISCRIMINANT FUNCTION WHEN COVARIANCE MATRICES ARE PROPORTIONAL AND UNKNOWN

CHIEN-PAI HAN

(Received July 21, 1971; revised Oct. 9, 1973)

1. Introduction

Let $\boldsymbol{X}_{p\times 1}$ be an observation to be classified into one of two p-variate normal populations denoted by $\pi_1: N(\boldsymbol{\mu}_1, \underline{\boldsymbol{\Sigma}})$, and $\pi_2: N(\boldsymbol{\mu}_2, \sigma^2\underline{\boldsymbol{\Sigma}})$. When $\underline{\boldsymbol{\Sigma}}$ arises in the intraclass correlation models, the discriminant function has been derived in Bartlett and Please [1] and Han [2]. The distribution of the discriminant function was obtained in Han [3] when the covariance matrices are arbitrary and known. When σ^2 and $\underline{\boldsymbol{\Sigma}}$ are both unknown, the problem is very difficult. However if we have partial information about the covariance matrices, the asymptotic distribution of the discriminant function can be found. We shall assume in this paper that $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ are unknown and the covariance matrices are partially known, i.e. either σ^2 is known or $\underline{\boldsymbol{\Sigma}}$ is known. The asymptotic expansion of the distributions for the two cases can be obtained by the "studentization" method of Hartley [4] and of Welch [8]. Section 2 will derive the asymptotic distribution when σ^2 is known and $\underline{\boldsymbol{\Sigma}}$ is unknown; Section 3 considers the case when $\underline{\boldsymbol{\Sigma}}$ is known but σ^2 is unknown, in the latter case the covariance matrix is completely specified under π_1. The asymptotic distributions are obtained up to the first order with respect to the sample sizes.

2. $\underline{\boldsymbol{\Sigma}}$ unknown and σ^2 known

Since $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\underline{\boldsymbol{\Sigma}}$ are unknown, we shall find estimators for them. Suppose a sample of size n_i is taken from π_i, $i=1, 2$, then the estimators of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\underline{\boldsymbol{\Sigma}}$ are $\bar{\boldsymbol{X}}_1$, $\bar{\boldsymbol{X}}_2$ and $\underline{\boldsymbol{S}}$ respectively where

$$\begin{aligned} &\bar{\boldsymbol{X}}_1=\frac{1}{n_1}\sum_{j=1}^{n_1}\boldsymbol{X}_{1j}\ , \\ (2.1)\qquad &\bar{\boldsymbol{X}}_2=\frac{1}{n_2}\sum_{j=1}^{n_2}\boldsymbol{X}_{2j}\ , \\ &\underline{\boldsymbol{S}}=\frac{1}{m}\left\{\sum_j(\boldsymbol{X}_{1j}-\bar{\boldsymbol{X}}_1)(\boldsymbol{X}_{1j}-\bar{\boldsymbol{X}}_1)'+\frac{1}{\sigma^2}\sum_j(\boldsymbol{X}_{2j}-\bar{\boldsymbol{X}}_2)(\boldsymbol{X}_{2j}-\bar{\boldsymbol{X}}_2)'\right\} \end{aligned}$$

where $m=n_1+n_2-2$. The distribution of $\bar{X}_1$, $\bar{X}_2$ and $m\underline{S}$ are independent $N(\mu_1, (1/n_1)\underline{\Sigma})$, $N(\mu_2, (\sigma^2/n_2)\underline{\Sigma})$ and $W(\underline{\Sigma}, m)$ respectively.

The discriminant function, by using the likelihood ratio procedure, is given in equation (3.1) of Han [3]. With $\underline{\Sigma}$ estimated by $\underline{S}$, we obtain the discriminant function

$$U=[X-\bar{X}_1+\alpha(\bar{X}_2-\bar{X}_1)]'\underline{S}^{-1}[X-\bar{X}_1+\alpha(\bar{X}_2-\bar{X}_1)] -\alpha(\alpha+1)(\bar{X}_1-\bar{X}_2)'\underline{S}^{-1}(\bar{X}_1-\bar{X}_2) \tag{2.2}$$

where $\alpha=1/(\sigma^2-1)$. It is easily seen that U is invariant under any linear transformation. Hence without loss of generality, we shall derive the distribution of U by letting $\mu_1=0$, $\mu_2=\mu_0=(D, 0, \cdots, 0)'$ and $\underline{\Sigma}=\underline{I}$, where D^2 denotes the Mahalanobis squared distance.

The cumulative distribution function (c.d.f.) of U given that X comes from π_i is denoted by $F_i(u)$ for $i=1, 2$. Let us first derive $F_1(u)$. The characteristic function (c.f.) of U is $\varphi(t)=\mathrm{E}\,(e^{itU}\,|\,\pi_1)$ which can be written as

$$\varphi(t)=E^{\bar{X}_1, \bar{X}_2, \underline{S}}\{\mathrm{E}\,[e^{itU}\,|\,\bar{X}_1, \bar{X}_2, \underline{S}; \pi_1]\}\ , \tag{2.3}$$

where the expectation in the curled bracket is the conditional c.f. given $\bar{X}_1$, $\bar{X}_2$, and $\underline{S}$. Let us denote it by $\phi(\bar{X}_1, \bar{X}_2, \underline{S})$. We find that

$$\phi(\bar{X}_1, \bar{X}_2, \underline{S})=\exp\left\{-it\alpha(\alpha+1)Q-\frac{1}{2}\sum_{j=1}^{p} v_j^2+\frac{1}{2}\sum_{j=1}^{p}\frac{v_j^2}{1-2ith_j} -\frac{1}{2}\sum_{j=1}^{p}\log\,(1-2ith_j)\right\} \tag{2.4}$$

where

$$Q=(\bar{X}_1-\bar{X}_2)'\underline{S}^{-1}(\bar{X}_1-\bar{X}_2)\ ,$$

$$v=\underline{M}'[-(\alpha+1)\bar{X}_1+\alpha\bar{X}_2]\ ,$$

$\underline{M}$ is a $p\times p$ orthogonal matrix such that $\underline{M}'\underline{S}^{-1}\underline{M}=\underline{H}$ and $\underline{H}$ is the diagonal matrix of eigen-values h_j of $\underline{S}^{-1}$.

Since the function ϕ is analytic about the point $(\bar{X}_1, \bar{X}_2, \underline{S})=(0, \mu_0, \underline{I})$, expanding ϕ into a Taylor's series, we have

$$\phi(\bar{X}_1, \bar{X}_2, \underline{S})=\exp\left[\sum_{j=1}^{p}\bar{x}_{1j}\frac{\partial}{\partial\mu_{1j}}+\sum_{j=1}^{p}(\bar{x}_{2j}-\mu_{0j})\frac{\partial}{\partial\mu_{2j}} +\sum_{j\leq k=1}^{p}(s_{jk}-\delta_{jk})\frac{\partial}{\partial\sigma_{jk}}\right]\phi(\mu_1, \mu_2, \underline{\Sigma})|_0 \tag{2.5}$$

where $\bar{x}_{1j}$, $\bar{x}_{2j}$, s_{jk}, μ_{1j}, μ_{2j}, μ_{0j} are elements of $\bar{X}_1$, $\bar{X}_2$, $\underline{S}$, μ_1, μ_2, and μ_0 respectively; δ_{jk} is the Kronecker delta; and $|_0$ denotes that the expression is evaluated at the point $(\mathbf{0}, \boldsymbol{\mu}_0, \underline{\boldsymbol{I}})$. Hence the c.f. of U is

$$\varphi(t)=E^{\bar{X}_1,\bar{X}_2,\underline{S}}\{\phi(\bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \underline{\boldsymbol{S}})\}=\Theta\phi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \underline{\boldsymbol{\Sigma}})|_0 \tag{2.6}$$

where Θ is the differential operator

$$\Theta=E^{\bar{X}_1,\bar{X}_2,\underline{S}}\left\{\exp\left[\sum \bar{x}_{1j}\frac{\partial}{\partial\mu_{1j}}+\sum(\bar{x}_{2j}-\mu_{0j})\frac{\partial}{\partial\mu_{2j}}+\sum(s_{kj}-\delta_{kj})\frac{\partial}{\partial\sigma_{kj}}\right]\right\}. \tag{2.7}$$

Okamoto [6] has shown that Θ can be expanded into

$$\Theta=1+\frac{1}{2n_1}\sum\frac{\partial^2}{\partial\mu_{1j}^2}+\frac{\sigma^2}{2n_2}\sum\frac{\partial^2}{\partial\mu_{2j}^2}+\frac{1}{m}\sum\frac{\partial^2}{\partial\sigma_{kj}^2}+O_2 \tag{2.8}$$

where O_2 stands for the terms of the second order with respect to $(n_1^{-1}, n_2^{-1}, m^{-1})$. Following Okamoto [6] and Memon [5], we can find term by term in (2.8), hence, with $\theta=-it$,

$$\varphi(t)=\{1+b_1(\theta, D)+b_2(\theta, D)+b_3(\theta, D)+O_2\}\phi(\mathbf{0}, \boldsymbol{\mu}_0, I), \tag{2.9}$$

where

$$b_1(\theta, D)=\frac{1}{n_1}\left[p\alpha(\alpha+1)\theta-\frac{p(\alpha+1)^2\theta}{1-2it}+2\left(-\alpha(\alpha+1)D\theta+\frac{\alpha(\alpha+1)D\theta}{1-2it}\right)^2\right],$$

$$b_2(\theta, D)=\frac{\sigma^2}{n_2}\left[p\alpha(\alpha+1)\theta-\frac{p\alpha^2\theta}{1-2it}+2\left(\alpha(\alpha+1)D\theta-\frac{\alpha^2D\theta}{1-2it}\right)^2\right], \tag{2.10}$$

$$b_3(\theta, D)=\frac{1}{m}\left[(p+1)\alpha(\alpha+1)D^2\theta-\frac{p(p+1)\theta}{1-2it}+\frac{p(p+1)\theta^2}{(1-2it)^2}-\frac{(p+1)\alpha^2D^2\theta}{(1-2it)^3}+\left(-\alpha(\alpha+1)D^2\theta+\frac{p\theta}{1-2it}+\frac{\alpha^2D^2\theta}{(1-2it)^2}\right)^2\right].$$

It is checked that the principal term given here is the same as that given in Han [3], i.e. the principal term is the c.f. of a non-central chi-square variate plus a constant.

To invert the characteristic function for the c.d.f. $F_1(u)$, we use the technique for inverting a c.f. of the form $(-it)^r\varphi(t)$ (see Wallace [7], p. 638). If $F(x)$ is the c.d.f. of a statistic and $\varphi(t)$ is the c.f., then the c.d.f. corresponding to $(-it)^r\varphi(t)$ is $F^{(r)}(x)$ where $F^{(r)}(x)$ is the rth deviate of $F(x)$. Now let $G_p(x)$ be the c.d.f. of a non-central chi-square variate with c.f. $\phi(\mathbf{0}, \boldsymbol{\mu}_0, \underline{\boldsymbol{I}})$, where p denotes the degrees of freedom, then the c.d.f. of U given that $\boldsymbol{X}$ comes from π_1 is

$$F_1(u)=w_1(d, D)G_p(u)+w_2(d, D)G_{p+2}(u)+w_3(d, D)G_{p+4}(u) + w_4(d, D)G_{p+6}(u)+w_5(d, D)G_{p+8}(u)+O_2 \tag{2.11}$$

where

$$w_1(d, D)=1+\left[\frac{p}{n_1}+\frac{p\sigma^2}{n_2}+\frac{(p+1)D^2}{m}\right]\alpha(\alpha+1)d+\left[\frac{2}{n_1}+\frac{2\sigma^2}{n_2}+\frac{D^2}{m}\right] \cdot \alpha^2(\alpha+1)^2D^2d^2 ,$$

$$w_2(d, D)=-\left[\frac{(\alpha+1)^2}{n_1}+\frac{\alpha^2\sigma^2}{n_2}+\frac{p+1}{m}\right]pd - \left[\frac{4\alpha(\alpha+1)}{n_1}+\frac{4\alpha^2\sigma^2}{n_2}+\frac{2p}{m}\right]\alpha(\alpha+1)D^2d^2 ,$$

$$w_3(d, D)=2\left[\frac{(\alpha+1)^2}{n_1}+\frac{\alpha^2\sigma^2}{n_2}-\frac{\alpha(\alpha+1)D^2}{m}\right]\alpha^2D^2d^2 + \left[\frac{p(p+1)}{m}+\frac{p^2}{m}\right]d^2 , \tag{2.12}$$

$$w_4(d, D)=\frac{2p}{m}\alpha^2D^2d^2-\frac{p+1}{m}\alpha^2D^2d ,$$

$$w_5(d, D)=\frac{\alpha^4D^4d^2}{m} ,$$

and d denotes the differential operator d/du.

Now we shall find the distribution of U when $\boldsymbol{X}$ comes from π_2. Using a similar procedure, we obtain the c.d.f. of U given that $\boldsymbol{X}$ comes from π_2,

$$F_2(u)=W_1(d, D)G_p(u)+W_2(d, D)G_{p+2}(u)+W_3(d, D)G_{p+4}(u) + W_4(d, D)G_{p+6}(u)+W_5(d, D)G_{p+8}(u)+O_2 , \tag{2.13}$$

where

$$W_1(d, D)=1+\left[\frac{p}{n_1}+\frac{p\sigma^2}{n_2}+\frac{(p+1)D^2}{m\sigma^2}\right]\alpha(\alpha+1)d + \left[\frac{2}{n_1}+\frac{2\sigma^2}{n_2}+\frac{D^2}{m\sigma^4}\right]\alpha^2(\alpha+1)^2D^2d^2 ,$$

$$W_2(d, D)=-\left[\frac{(\alpha+1)^2}{n_1}+\frac{\alpha^2\sigma^2}{n_2}+\frac{(p+1)\sigma^2}{m}\right]pd - \left[\frac{4(\alpha+1)^2}{n_1}+\frac{4\alpha(\alpha+1)\sigma^2}{n_2}+\frac{2p}{m}\right]\alpha(\alpha+1)D^2d^2 ,$$

$$W_3(d, D)=2\left[\frac{(\alpha+1)^2}{n_1}+\frac{\alpha^2\sigma^2}{n_2}-\frac{\alpha(\alpha+1)D^2}{m\sigma^2}\right](\alpha+1)^2D^2d^2 \tag{2.14}$$

$$+\left[\frac{p(p+1)}{m}+\frac{p^2}{m}\right]\sigma^4 d^2 ,$$

$$W_4(d, D)=\frac{2p}{m}(\alpha+1)^2\sigma^2 D^2 d^2-\frac{p+1}{m}(\alpha+1)^2 D^2 d ,$$

$$W_5(d, D)=\frac{(\alpha+1)^4 D^4 d^2}{m} .$$

3. σ^2 unknown and $\underline{\boldsymbol{\Sigma}}$ known

This section will derive the asymptotic distribution when σ^2 is unknown but $\underline{\boldsymbol{\Sigma}}$ is known. Hence the covariance matrix is completely specified under π_1. The estimators of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and σ^2 are $\bar{\boldsymbol{X}}_1$, $\bar{\boldsymbol{X}}_2$ and

$$\hat{\sigma}^2=\frac{1}{p(n_2-1)}\sum_{j=1}^{n_2}(\boldsymbol{X}_{2j}-\bar{\boldsymbol{X}}_2)'\underline{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X}_{2j}-\bar{\boldsymbol{X}}_2) \tag{3.1}$$

respectively. The distribution of $\hat{\sigma}^2$ is $\sigma^2[p(n_2-1)]^{-1}\chi^2_{p(n_2-1)}$ and independent of $\bar{\boldsymbol{X}}_1$ and $\bar{\boldsymbol{X}}_2$. An analogy to the discriminant function (2.2) is

$$\begin{aligned} U=&[\boldsymbol{X}-\bar{\boldsymbol{X}}_1+\hat{\alpha}(\bar{\boldsymbol{X}}_2-\bar{\boldsymbol{X}}_1)]'\underline{\boldsymbol{\Sigma}}^{-1}[\boldsymbol{X}-\bar{\boldsymbol{X}}_1+\hat{\alpha}(\bar{\boldsymbol{X}}_2-\bar{\boldsymbol{X}}_1)] \\ &-\hat{\alpha}(\hat{\alpha}+1)(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)'\underline{\boldsymbol{\Sigma}}^{-1}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2) , \end{aligned} \tag{3.2}$$

where $\hat{\alpha}=1/(\hat{\sigma}^2-1)$.

The derivation of the asymptotic distribution of U is no more complicated than that of Section 2. Again U is invariant under any linear transformation, we may derive the distribution by letting $\boldsymbol{\mu}_1=\boldsymbol{0}$, $\boldsymbol{\mu}_2=\boldsymbol{\mu}_0$, and $\underline{\boldsymbol{\Sigma}}=\underline{\boldsymbol{I}}$. The c.f. of U when $\boldsymbol{X}$ comes from π_1 is $\varphi(t)=E^{\bar{X}_1,\bar{X}_2,\hat{\sigma}^2}\{\phi(\bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \hat{\sigma}^2)\}$ where

$$\phi(\bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \hat{\sigma}^2)=\exp\left\{-it\hat{\alpha}(\hat{\alpha}+1)Q_*+\frac{it}{1-2it}\boldsymbol{v}_*'\boldsymbol{v}_*-\frac{p}{2}\log(1-2it)\right\} \tag{3.3}$$

and

$$Q_*=(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)'(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2) ,$$

$$\boldsymbol{v}_*=-(\hat{\alpha}+1)\bar{\boldsymbol{X}}_1+\hat{\alpha}\bar{\boldsymbol{X}}_2 .$$

Using a similar expansion as (2.6), we have

$$\varphi(t)=\Theta\phi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \sigma^2)|_0 , \tag{3.4}$$

where

$$\Theta=1+\frac{1}{2n_1}\sum_j\frac{\partial^2}{\partial\mu_{1j}^2}+\frac{\sigma^2}{2n_2}\sum_j\frac{\partial^2}{\partial\mu_{2j}^2}+\frac{\sigma^4}{p(n_2-1)}\frac{\partial^2}{\partial(\sigma^2)^2}+O_2 \tag{3.5}$$

and $|_0$ denotes the expression is evaluated at $(\mu_1, \mu_2, \sigma^2)=(0, \mu_0, \sigma^2)$. $\varphi(t)$ are easily found to be

$$\varphi(t)=\{1+b_1(\theta, D)+b_2(\theta, D)+b_4(\theta, D)+O_2\}\psi(0, \mu_0, \sigma^2) , \tag{3.6}$$

where $b_1(\theta, D)$ and $b_2(\theta, D)$ are given in (2.10);

$$b_4(\theta, D)=\frac{\sigma^4}{p(n_2-1)}\left[2\alpha^3(\alpha+1)D^2\theta-\frac{6\alpha^4D^2\theta}{1-2it}+\left(-\alpha^2(2\alpha+1)D^2\theta+\frac{2\alpha^3D^2\theta}{1-2it}\right)^2\right] . \tag{3.7}$$

Inverting $\varphi(t)$, we obtain the c.d.f., $F_1(u)$, of U when X comes from π_1 to be

$$F_1(u)=v_1(d, D)G_p(u)+v_2(d, D)G_{p+2}(u)+v_3(d, D)G_{p+4}(u)+O_2 , \tag{3.8}$$

where

$$\begin{aligned}
v_1(d, D)&=1+\left(\frac{1}{n_1}+\frac{\sigma^2}{n_2}\right)p\alpha(\alpha+1)d+\frac{\sigma^4}{p(n_2-1)}2\alpha^3(3\alpha+1)D^2d \\
&\quad+\left(\frac{1}{n_1}+\frac{\sigma^2}{n_2}\right)2\alpha^2(\alpha+1)^2D^2d^2+\frac{\sigma^4}{p(n_2-1)}\alpha^4(2\alpha+1)^2D^4d^2 ,
\end{aligned}$$

$$\begin{aligned}
v_2(d, D)&=-\left[\frac{1}{n_1}p(\alpha+1)^2+\frac{\sigma^2}{n_2}p\alpha^2+\frac{\sigma^4}{p(n_2-1)}6\alpha^4D^2\right]d \\
&\quad-\left[\frac{1}{n_1}(\alpha+1)+\frac{\sigma^2}{n_2}\alpha\right]4\alpha^2(\alpha+1)D^2d^2 \\
&\quad-\frac{\sigma^4}{p(n_2-1)}4\alpha^5(2\alpha+1)D^4d^2 ,
\end{aligned} \tag{3.9}$$

$$v_3(d, D)=\left[\frac{1}{n_1}(\alpha+1)^2+\frac{\sigma^2}{n_2}\alpha^2+\frac{\sigma^4}{p(n_2-1)}2\alpha^4D^2\right]2\alpha^2D^2d^2 ,$$

and $G_p(u)$ is the c.d.f. corresponding to the c.f. $\psi(0, \mu_0, \sigma^2)$ which is a non-central chi-square plus a constant.

The c.d.f. of U when X comes from π_2, $F_2(u)$, is obtained in a similar way.

$$F_2(u)=V_1(d, D)G_p(u)+V_2(d, D)G_{p+2}(u)+V_3(d, D)G_{p+4}(u)+O_2 , \tag{3.10}$$

where

$$V_1(d, D)=v_1(d, D) ,$$

$$V_2(d, D)=-\left[\frac{1}{n_1}p(\alpha+1)^2+\frac{\sigma^2}{n_2}p\alpha^2+\frac{\sigma^4}{p(n_2-1)}2\alpha^3(3\alpha+1)D^2\right]d \tag{3.11}$$

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

$$-\frac{\sigma^4}{p(n_2-1)}4\alpha^4(\alpha+1)(2\alpha+1)D^4d\ ,$$

$$V_3(d, D)=\left[\frac{1}{n_1}(\alpha+1)^2+\frac{\sigma^2}{n_2}\alpha^2+\frac{\sigma^4}{p(n_2-1)}2\alpha^4D^2\right]2(\alpha+1)^2D^2d^2\ .$$

This research was partially supported by National Institutes of Health Grant No. GM0034-11. The author wishes to thank the referee for his comments.

IOWA STATE UNIVERSITY

REFERENCES

[1] Bartlett, M. S. and Please, N. W. (1963). Discrimination in the case of zero mean differences, *Biometrika*, **50**, 17-21.

[2] Han, C. P. (1968). A note on discrimination in the case of unequal covariance matrices, *Biometrika*, **55**, 586-587.

[3] Han, C. P. (1969). Distribution of discriminant function when covariance matrices are proportional, *Ann. Math. Statist.*, **40**, 979-985.

[4] Hartley, H. O. (1938). Studentization and large-sample theory, *J. Roy. Statist. Soc.*, Suppl., **5**, 80-88.

[5] Memon, A. Z. (1968). *Z Statistic in Discriminant Analysis*, Unpublished Ph.D. Thesis, Iowa State University.

[6] Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function, *Ann. Math. Statist.*, **34**, 1286-1301.

[7] Wallace, D. L. (1958). Asymptotic approximations to distributions, *Ann. Math. Statist.*, **29**, 635-654.

[8] Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved, *Biometrika*, **34**, 28-35.