Covariance Adjustments in Discrimination of Mixed Discrete and Continuous Variables

Chi-Ying Leung

The Chinese University of Hong Kong, Hong Kong, China

Received December 12, 1997

Sufficient conditions are given to ensure a better performance of the plug-in version of the covariates adjusted location linear discriminant function in an asymptotic comparison of the overall expected error rate. Our findings generalize several earlier results on discriminant function with covariance. © 1999 Academic Press

AMS 1991 subject classifications: 62H10, 62H30.

Key words and phrases: plug-in location linear discriminant function; covariance adjustment; asymptotic expansion; overall expected error rate.

1. INTRODUCTION

Treatment of covariates or concomitant variables arises in many statistical investigations. In classification, covariates are often handled by using the residual vector of regression of the discriminators on the covariates in the discriminant function. See Cochran and Bliss [3]. Without much analytic investigation, Cochran [2] claims that the discriminators subject to the suggested treatment generally improve the performance of the discriminant function since internal correlations have been incorporated in the procedure. Memon and Okamoto [8] re-examines the same problem and reaches similar conclusions. In Memon and Okamoto [8], only continuous discriminators are considered in a two-population problem under homogeneous dispersion matrices for both populations with a zero threshold. Their conclusion draws criticism since the argument is based on an asymptotic expansion in Okamoto [9] which is invalid for a nonzero threshold with unknown Mahalanobis distance.

In this article, we consider a similar problem with mixed discrete and continuous variables. Classification of mixed discrete and continuous variables is prevalent in many situations. See Daudin [4], Krzanowski [5], and Vlachonikolis and Marriott [13] for instance. Specifically, we consider discrimination between two populations say, Π_1 and Π_2 with mixed covariates amongst the discrete and continuous variables. Under conditional homogeneity of discrete values specific dispersion matrices for

0047-259X/99 $30.00

both populations, an asymptotic overall expected error rate for the plug-in covariates adjusted discriminant function is derived. The result is compared to the corresponding error rate for the discriminant function without the adjustments. Sufficient conditions are obtained in support of the covariates adjusted discriminant function. Explicit statements are given in two special cases. One of our results generalizes Memon and Okamoto [8] and provides further theoretical justification of Cochran [2].

2. THE PROCEDURE

Existing procedures for classification between two populations say Π_1 and Π_2 using mixed discrete and continuous variables are based on the location probability model of Olkin and Tate [10]. To simplify the discussion, we adopt the formulation in Krzanowski [5]. Suppose that a vector measurement $u' = (y', z')$ is observed on an individual where $z' = (z_1, ..., z_r)$ is a multinomial variable with r discrete states and $y' = (y_1, ..., y_{p+q})$ is a vector of $p+q$ continuous variables. The vector z has only one nonzero entry equal to one which is the incidence for the corresponding state.

The continuous variables consist of p discriminators and q covariates. To simplify the discussion, suppose that the first p variables are the discriminators while the remaining q variables are covariates. Partition $y' = (y^{(1)'}, y^{(2)'})$; where $y^{(1)'} = (y_1, ..., y_p)$; $y^{(2)'} = (y_{p+1}, ..., y_{p+q})$. Similarly, let $z' = (z^{(1)'}, z^{(2)'})$; where $z^{(1)'} = (z_1, ..., z_{r_1})$ consists of r_1 discrete discriminators and $z^{(2)'} = (z_{r_1+1}, ..., z_r)$ denotes the discrete covariates $(0 < r_1 \leqslant r)$. Furthermore, it is assumed that for $i = 1, 2$, $Z \mid \Pi_i \sim \text{Multinomial}(1, p_i^{(1)}, p^{(2)})$; $p_i^{(1)'} = (p_{1i}, ..., p_{r_1 i})$; $p^{(2)} = (p_{r_1+1}, ..., p_r)$; $\sum_{m=1}^{r_1} p_{mi} + \sum_{l=r_1+1}^{r} p_l = 1$ and for $i = 1, 2$; $m = 1, ..., r$; $Y \mid Z_m = 1, \Pi_i \sim N_{p+q}(\nu_{mi}, \Sigma^{(m)})$ where ν_{mi} is a $p+q$ vector with the first p entries forming a p vector equal to $\mu_{mi} = E(Y^{(1)} \mid Z_m = 1, \Pi_i)$ and the last q entries forming a q vector equal to $\lambda_m = E(Y^{(2)} \mid Z_m = 1, \Pi_i)$.

Let $\Sigma^{(m)}$ be partitioned as

$$\Sigma^{(m)} = \begin{bmatrix} \Sigma_{11}^{(m)} & \Sigma_{12}^{(m)} \\ \Sigma_{21}^{(m)} & \Sigma_{21}^{(m)} \end{bmatrix}$$

and $\beta_m = \Sigma_{12}^{(m)} \Sigma_{22}^{(m)^{-1}}$; where $\Sigma_{ij}^{(m)} = \text{Cov}(Y^{(i)}, Y^{(j)})$, $i, j = 1, 2$; $m = 1, ..., r$. Notice that the role of $Z^{(2)}$ assumes that for $i = 1, 2$; $E(Z^{(2)} \mid \Pi_i) = p^{(2)}$ is known. Similarly, the role of $Y^{(2)}$ assumes that the state specific mean λ_m common to both Π_1 and Π_2 in state m, $m = r_1 + 1, ..., r$ is known.

To briefly state the problem, let $\lambda_m = 0$ for $m = 1, ..., r$ in the sequel. With complete knowledge of the parameters, the Bayes rule is given by the location linear discriminant function. Specifically, for an object with measurement $(Y', Z')'$ with $Z_m = 1, m = 1, ..., r$, the rule with threshold $t \in (-\infty, \infty)$ assigns the object to Π_1 if and only if $U_m > t$, where

$$U_m = \begin{cases} D_m - \log(p_{m2}/p_{m1}), & \text{for} \quad m = 1, ..., r_1 \\ D_m, & \text{for} \quad m = r_1 + 1, ..., r \end{cases}$$

with $D_m = [Y^{(1)} - \beta_m Y^{(2)} - \frac{1}{2}(\mu_{m1} + \mu_{m2})]' \Sigma_{1.2}^{(m)^{-1}} (\mu_{m1} - \mu_{m2})$ and $\Sigma_{1.2}^{(m)} = \Sigma_{11}^{(m)} - \beta_m \Sigma_{22}^{(m)} \beta_m'$. Notice that D_m is the Fisher linear discriminant function adjusted for the continuous covariates $Y^{(2)}$ for state m, $m = 1, ..., r$ when all the parameters are known. See Memon and Okamoto [8].

The threshold $t = 0$ is a common choice and $\Sigma^{(1)} = \cdots = \Sigma^{(r)}$ is usually assumed. See Krzanowski [5]. In practice, an approximate sample based rule rather than the Bayes rule is used due to lack of knowledge of the parameters. Suppose that random training samples of sizes n_1 and n_2 respectively from Π_1 and Π_2 are available. Let n_{mi} observations from Π_i fall in state m, with $Y'_{mji} = (Y_{mji}^{(1)'}, Y_{mji}^{(2)'})$ denoting continuous measurements on the jth sample in state m from Π_i, $j = 1, ..., n_{mi}$; $m = 1, ..., r$; $i = 1, 2$. Let $n(m) = n_{m1} + n_{m2} - 2$, $m = 1, ..., r$. Unbiased continuous covariates adjusted estimates specific to state m are

$$\hat{\beta}_m = S_{12}^{(m)} S_{22}^{(m)^{-1}};$$

$$\hat{\Sigma}_{1.2}^{(m)} = (n(m) - q)^{-1} (S_{11}^{(m)} - S_{12}^{(m)} S_{22}^{(m)^{-1}} S_{21}^{(m)}); \qquad \hat{\mu}_{mi} = \bar{Y}_{mi}^{(1)} - \hat{\beta}_m \bar{Y}_{mi}^{(2)},$$

where

$$\bar{Y}'_{mi} = (\bar{Y}_{mi}^{(1)'}; \bar{Y}_{mi}^{(2)'}); \qquad \bar{Y}_{mi}^{(v)} = n_{mi}^{-1} \sum_{j=1}^{n_{mi}} Y_{mi}^{(v)}; \quad v = 1, 2; \quad m = 1, ..., r,$$

and

$$\sum_{i=1}^{2} \sum_{j=1}^{n_{mi}} (Y_{mji} - \bar{Y}_{mi})(Y_{mji} - \bar{Y}_{mi})' = \mathrm{S}^{(m)} = \begin{bmatrix} S_{11}^{(m)} & S_{12}^{(m)} \\ S_{21}^{(m)} & S_{22}^{(m)} \end{bmatrix}$$

is similarly partitioned as $\Sigma^{(m)}$ for $m = 1, ..., r$.

From Kshirsagar [6, p. 20, Eq. (4.12)],

$$E(n_{mi} \mid \Pi_i, n_{r_1+1i}, ..., n_{ri}) = (n_i - n_{r_1+1} - \cdots - n_{ri})\, p_{mi}(1 - p_{r_1+1} - \cdots - p_r)^{-1}.$$

Unbiased estimates of the state probabilities are obtained by adjusting the known state probabilities in the last $r - r_1$ multinomial cells for each of the two discrete samples and are as follows: $\hat{p}_{mi} = \tilde{p}_{mi}(1 - p_{r_1+1} - \cdots - p_r) \times (1 - \tilde{p}_{r_1+1}^{(i)} - \cdots - \tilde{p}_r^{(i)})^{-1}$; where $\tilde{p}_{mi}$ and $\tilde{p}_l^{(i)}$ are the sample proportions of the mth and lth multinomial cells from the ith sample for $m=1, ..., r_1$; $l=r_1+1, ..., r$ and $i=1, 2$. A popular sample based approximation to U_m is the plug-in version of U_m using above estimates due to its simplicity. In the next section, an asymptotic expansion of the overall expected error rate of the plug-in rule is given. The expansion provides an index of the long term performance of the procedure.

3. THE EXPECTED ERROR RATE

For $-\infty < t < \infty$, given $Z_m = 1$, $m=1, ..., r$; $i=1, 2$, the probability of misclassification is $e_{im}(t) = \Pr\{(-1)^i \hat{U}_m > (-1)^i t \mid \Pi_i\}$. With equal prior for Π_1 and Π_2, the overall expected error rate is given by $\bar{e}(t) = \frac{1}{2}\sum_{i=1}^{2}\sum_{m=1}^{r} p_{mi}\, e_{im}(t)$ which admits an asymptotic expansion given below. Details of the derivation are given in the appendix. To facilitate the derivation, the following results are needed.

LEMMA 3.1. *Under the formulation in section* 2, *with* $\Sigma^{(1)} = \cdots = \Sigma^{(r)} = I_{p+q}$, *a* $(p+q)\times(p+q)$ *identity matrix, if* $Z_m = 1$, $\mu_{m1} = 0$, $\mu_{m2} = -\delta_m$, $\delta'_m = (\Delta_m, ..., 0)$, $0 < \Delta_m = [(\mu_{m1}-\mu_{m2})'\, \Sigma_{1.2}^{(m)^{-1}}(\mu_{m1}-\mu_{m2})]^{1/2}$, *then for given* n_{m1} *and* n_{m2}, $m=1, ..., r$, *the following hold.*

(i) $E_{2m}(\hat{\mu}_{m1}) = 0$;

(ii) $E_{2m}(\hat{\mu}_{m1} - \hat{\mu}_{m2}) = \delta_m$;

(iii) $E_{2m}(\hat{\beta}_m) = 0$;

(iv) $E_{2m}(\hat{\beta}_m \hat{\beta}'_m) = b_m I_p$, $b_m = q(n(m)-q-1)^{-1}$;

(v) $E_{2m}(\Sigma_{1.2}^{(m)}) = I_p$;

(vi) $E_{2m}(\hat{\mu}_{m1}\hat{\mu}'_{m1}) = n_{m1}^{-1}(1+b_m)\, I_p$;

(vii) $E_{2m}((\hat{\mu}_{m1}-\hat{\mu}_{m2}-\delta_m)(\hat{\mu}_{m1}-\hat{\mu}_{m2}-\delta_m)') = (n_{m1}^{-1}+n_{m2}^{-1})(1+b_m)\, I_p$;

(viii) $E_{2m}((\hat{\mu}_{m1}-\hat{\mu}_{m2}-\delta_m)\, \hat{\mu}'_{m1})) = n_{m1}^{-1}(1+b_m)\, I_p$;

(ix) $E_{2m}(\delta'_m(\Sigma_{1.2}^{(m)} - I_p)\, \delta_m) = (n(m))^{-1}\, (p+1)\, \Delta_m^2$;

(x) $E_{2m}((\delta'_m(\Sigma_{1.2}^{(m)} - I_p)\, \delta_m)^2) = 2(n(m))^{-1}\, \Delta_m^4$;

(xi) $E_{2m}((\delta'_m(\hat{\beta}_m\hat{\beta}'_m - b_m I_p)\, \delta_m)^2) = 2d_m\, \Delta_m^4$, $d_m = q[(n(m)-1)(n(m)-q)^{-1}\, (n(m)-q-1)^{-1}\, (n(m)-q-3)^{-1} + (n(m)-2)^{-2}\, (n(m)-4)^{-1}]$.

Proof. Part (i) follows from the fact that given n_{m1} and n_{m2} and $Z_m = 1$, $\bar{Y}_{mi}$ and $\hat{\Sigma}^{(m)}$ are independently distributed with $S_{11}^{(m)} - S_{12}^{(m)} S_{22}^{(m)^{-1}} S_{21}^{(m)} \sim W_p(I_p, n(m)-q)$, $\hat{\beta}_m \mid S_{22}^{(m)} \sim N_{p,q}(0, I_p, S_{22}^{(m)})$ and $S_{22}^{(m)} \sim W_q(I_p, n(m))$.

Part (ii) follows similarly as Part (i).

Part (iii) follows from $E_{2m}(\hat{\beta}_m)=E_{2m}(E_{2m,\, S_{22}^{(m)}}(\hat{\beta}_m \mid S_{22}^{(m)}))$, where $E_{2m,\, S_{22}^{(m)}}(\,.\,)$ denotes the conditional expectation with respect to $S_{22}^{(m)}$ for given n_{m1}, n_{m2}.

Part (iv) follows since given $S_{22}^{(m)}$, the rows of $\hat{\beta}_m$ are independent and identically distributed as $N_q(0, S_{22}^{(m)-1})$ with $E_{2m}(S_{22}^{(m)-1})=(n(m)-q-1)^{-1} I_p$.

Part (v) follows from the Wishart distribution of $S_{11}^{(m)}-S_{12}^{(m)} S_{22}^{(m)-1} S_{21}^{(m)}$.

Part (vi) follows from $E_{2m}(\hat{\mu}_{m1}\hat{\mu}'_{m1})=n_{m1}^{-1} E_{2m}([I_p+\hat{\beta}_m\hat{\beta}'_m])$.

Parts (vii) and (viii) follow similarly.

Parts (ix) and (x) follow from Anderson [1], Eq. (26) and Eq. (27), p. 969 respectively.

Part (xi) follows from $E_{2m}((\delta'_m(\hat{\beta}_m\hat{\beta}'_m-b_m I_p)\,\delta_m)^2)=E_1 V_2+V_1 E_2$, where E_2 and V_2 denote respectively the conditional expectation and variance given $S_{22}^{(m)}$ and E_1 and V_1 stand for the expectation and variance with respect to the distribution of $S_{22}^{(m)}$. Using $\hat{\beta}_m\delta_m \mid S_{22}^{(m)} \sim N_q(0, \Delta_m^2 S_{22}^{(m)-1})$ and Searle [11, Theorem 1, p. 55], $V_1E_2=2\Delta_m^4 q(n(m)-2)^{-2}(n(m)-4)^{-1}$. By Searle [11, Corollary 1.2, p. 57] and Srivastava and Khatri [12, problem 3.2(iv) p. 97], $E_1 V_2=2\,\Delta_m^4 q(n(m)-1)(n(m)-q)^{-1}(n(m)-q-1)^{-1}(n(m)-q-3)^{-1}$.

Lemma 3.2. *Under the assumptions in Section 2, given $Z_m=1$,*

(i) $E_{1m}(\tilde{p}_{mi})=p_{mi}$; $m=1, ..., r_1$; $i=1, 2$;

(ii) $E_{1m}((\tilde{p}_{mi}-\tilde{p}_{mi})^2)=n_i^{-1} p_{mi}(1-p_{mi})$; $m=1, ..., r_1$; $i=1, 2$;

(iii) $E_{1m}((\tilde{p}_l^{(i)}-p_l)(\tilde{p}_{l'}^{(j)}-p_{l'}))=0$; $l, l'=r_1+1, ..., r$; $i\neq j=1, 2$;

(iv) $E_{1m}((\tilde{p}_l^{(i)}-p_l)^2)=n_i^{-1} p_l(1-p_l)$; $l=r_1+1, ..., r$; $i=1, 2$;

(v) $E_{1m}((\tilde{p}_l^{(i)}-p_l)(\tilde{p}_{l'}^{(i)}-p_{l'}))=-n_i^{-1} p_l p_{l'}$; $l\neq l'=r_1+1, ..., r$; $i=1, 2$;

(vi) $E_{1m}((\tilde{p}_{mi}-p_{mi})(\tilde{p}_l^{(i)}-p_l))=-n_i^{-1} p_{mi} p_l$; $m=1, ..., r_1$; $l=r_1+1, ..., r$; $i=1, 2$.

(vii) $E_{1m}((\tilde{p}_{mi}-p_{mi})(\tilde{p}_l^{(j)}-p_l))=0$, $m=1, ..., r_1$; $i\neq j=1, 2$.

Proof. This is obvious.

Lemma 3.3. *For the two random training samples, suppose that the following conditions are satisfied.*

(C1) *$n_{m2} n_{m1}^{-1}$ converges in probability to $k_m>0$, $m=1, ..., r$ as n_1 and n_2 tend to infinity.*

(C2) *$n_{s1} n_{m1}^{-1}$ converges in probability to $k_{s,m}>0$, $s, m=1, ..., r$ as n_1 and n_2 tend to infinity. Then given $Z_m=1$, $E_{1m}(\Phi(\Delta_m^{-1}[t+\log(\hat{p}_{m2}/\hat{p}_{m1})-\Delta_m^2/2]))=\Phi(\eta_{1mt})+n^{-1}\,\Delta_m^{-1}\phi(\eta_{1mt})\,\zeta(p_{m1}, p_{m2})+O(n^{-2})$;* where

$$\zeta(p_{m1}, p_{m2}) = \{nn_1^{-1} p_{m1}^{-1}(1-p_{m1}) + nn_2^{-1} \bar{p}(1-\bar{p})^{-1}\}$$
$$\times \{\tfrac{3}{4} - \Delta_m^{-2}/2[t + \log(p_{m2}/p_{m1}]\}$$
$$- \{nn_2^{-1} p_{m2}^{-1}(1-p_{m2}) + nn_1^{-1} \bar{p}(1-\bar{p})^{-1}\}$$
$$\times \{\tfrac{1}{4} + \Delta_m^{-2}/2[t + \log(p_{m2}/p_{m1}]\}$$
$$- n(n_1^{-1} + n_2^{-1})\, \bar{p}(1-\bar{p})^{-1}$$
$$\times \{\tfrac{1}{4} - \Delta_m^{-2}/2[t + \log(p_{m2}/p_{m1})]\};$$

with $\bar{p} = \sum_{l=r_1+1}^{r} p_l$.

Proof. Given $Z_m = 1$, $m = 1, ..., r_1$, the result follows from a Taylor series expansion of $\Phi(\Delta_m^{-1}[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \Delta_m^{-2}/2])$ about $\tilde{p}_{mi} = p_{mi}$, $\tilde{p}_l^{(i)} = p_l$, $l = r_1 + 1, ..., r$; $i = 1, 2$. Under (C1) and (C2), the remainder term has order $O(n^{-2})$. The result follows from Lemma 3.2.

Remark 3.1. It should be pointed out that Lemma 3.1 and Lemma 3.2 ensure that the expansions in the following theorem have the indicated order of approximation.

THEOREM 3.1 (Main Result). *Suppose that* (C1) *and* (C2) *in Lemma* 3.3 *are satisfied. Let* $n = n_1 + n_2 - 2r$. *Then*

(a) $n(n(m))^{-1}$ *converges in probability to* $1 + k_m^* \geq 0$ *as both* n_1 *and* n_2 *tend to infinity and* $\lim_{n_1, n_2 \to \infty} n_2 n_1^{-1} = k > 0$.

(b) *for* $t \in (-\infty, \infty)$ *and given* $Z_m = 1$, $m = 1, ..., r_1$,

$$e_{1m}(t) = \Phi(\eta_{1mt}) + n^{-1}\phi(\eta_{1mt})(\alpha_{1mt} + \tau_{1mt} + \gamma_{1mt}) + O(n^{-2}); \qquad (3.1)$$

and for $m' = r_1 + 1, ..., r$,

$$e_{1m'}(t) = \Phi(\eta_{1m't}^*) + n^{-1}\phi(\eta_{1m't}^*)(\tau_{1m't}^* + \gamma_{1m't}^*) + O(n^{-2}); \qquad (3.2)$$

where

$$\eta_{1mt} = \Delta_m^{-1}[t + \log(p_{m2}/p_{m1}) - \Delta_m^2/2];$$
$$\alpha_{1mt} = \Delta_m^{-1}\{(1+k)\, p_{m1}^{-1}(1-p_{m1}) + (1+k^{-1})\, \bar{p}(1-\bar{p})^{-1}\}$$
$$\times \{\tfrac{3}{4} - \Delta_m^{-2}/2[t + \log(p_{m2}/p_{m1}]\} - \Delta_m^{-1}\{(1+k^{-1})\, p_{m2}^{-1}(1-p_{m2})$$
$$+ (1+k)\, \bar{p}(1-\bar{p}))^{-1}\}\{\tfrac{1}{4} + \Delta_m^{-2}/2[t + \log(p_{m2}/p_{m1})]\}$$
$$- \Delta_m^{-1}(2 + k + k^{-1})\, \bar{p}(1-\bar{p})^{-1}\, \{\tfrac{1}{4} - \Delta_m^{-2}/2[t + \log(p_{m2}/p_{m1})]\};$$

$$\bar{p} = \sum_{l=r_1+1}^{r} p_l;$$

$$\tau_{1mt} = -\tfrac{1}{2} q(1 + k_m^*)\, \Delta_m^{-1}[t + \log(p_{m2}/p_{m1}) - \Delta_m^2/2];$$

and

$$\begin{aligned}
\gamma_{1mt} = {} & \tfrac{1}{4}(p-1)(1+k_m^*)\, \Delta_m + \tfrac{1}{4}(p-1)\, \Delta_m^{-1} \\
& \times \{3(1+k)\, p_{m1}^{-1} - (1+k^{-1})\, p_{m2}^{-1}\} - [t + \log(p_{m2}/p_{m1})] \\
& \times \{\tfrac{3}{2}(p-1)(1+k_m^*)\, \Delta_m^{-1} + \tfrac{1}{2}(p-3)\, \Delta_m^{-3}[(1+k)\, p_{m1}^{-1} \\
& + (1+k^{-1})\, p_{m2}^{-1}]\} - \Delta_m^{-1}/2[t + \log(p_{m2}/p_{m1}) - \Delta_m^2/2] \\
& \times \{\tfrac{1}{4}[(1+k)\, p_{m1}^{-1} + (1+k^{-1})\, p_{m2}^{-1}] + \Delta_m^{-2}[t + \log(p_{m2}/p_{m1})] \\
& \times [(1+k^{-1})\, p_{m2}^{-1} - (1+k)\, p_{m1}^{-1}] + \Delta_m^{-4}[t + \log(p_{m2}/p_{m1})]^2 \\
& \times [(1+k)\, p_{m1}^{-1} + (1+k^{-1})\, p_{m2}^{-1} + 2(1+k_m^*)\, \Delta_m^2]\}.
\end{aligned}$$

$\eta_{1m't}^*$, $\tau_{1m't}^*$, and $\gamma_{1m't}^*$ are obtained by putting $p_{m1} = p_{m2} = p_{m'}$ in η_{1mt}, τ_{1mt}, and γ_{1mt} respectively in (3.1) for $m' = r_1 + 1, ..., r$.

In Eq. (3.1) and Eq. (3.2), $\Phi(.)$ and $\phi(.)$ denote respectively the standard normal distribution function and the density function.

Proof. The proof is given in the appendix.

COROLLARY 3.1. *For $m = 1, ..., r_1$, $e_{2m}(t)$ is obtained by interchanging m_1 and m_2, k and k^{-1} and substituting $-t$ for t throughout Eq. (3.1). For $m' = r_1 + 1, ..., r$, $e_{2m'}(t)$ is similarly obtained from Eq. (3.2).*

Proof. The result follows from the fact that interchanging m_1 and m_2 in $\hat{U}_m$ changes $\hat{U}_m$ to $-\hat{U}_m$.

4. ASYMPTOTIC COMPARISON

To investigate the effect of covariate adjustments due to both discrete and continuous variables, we need a similar expression for the overall expected error rate say $\bar{e}'(t)$, where $\bar{e}'(t)=\frac{1}{2}\sum_{i=1}^{2}(\sum_{m=1}^{r_1} p_{mi}e'_{im}(t)+\sum_{m'=1}^{r} p_{m'}e'_{im'}(t))$; and for $i=1,2$; $m=1,...,r$; $e'_{im}(t)$ is the probability of misclassification when all covariates are considered as discriminators. It follows from Leung [8] that $e'_{im}(t)$ can be obtained by dropping τ_{imt}, setting $\bar{p}=0$ in α_{1mt} and replacing p by $p+q$ throughout γ_{1mt} for $m=1,...,r_1$ in Eq. (3.1). For $m'=r_1+1,...,r$, $e_{1m'}(t)$ can be similarly obtained from $e_{1m}(t)$ using Eq. (3.2) and deleting τ^*_{1mt}, $m=1,...,r$. For $m=1,...,r$, $e'_{2m}(t)$ can be obtained similarly from Corollary 3.1.

Combining the above results, we have an asymptotic expression for the difference $\bar{e}'(t)-\bar{e}(t)$ which can be used to assess the effect of covariate adjustments in the plug-in location linear discriminant function. We highlight the assessment in two interesting cases where concrete conclusions can be drawn. The first case is the classical problem of Cochran and Bliss [3]. The second case examines the roles played by discrete covariates in mixed variables discrimination. The results are stated in the following corollaries.

Remark 4.1. It is of practical importance to retain all variables including covariates in classification. Covariates not only provide information on their own but also carry useful correlations to be used in classification. Omitting the covariates amounts to throwing away essential information.

COROLLARY 4.1. *Under the assumptions in Theorem* 3.1, *for* $t=0$, $r=1$, $p_{11}=p_{12}=1$ *and* $\Delta_1=\Delta>0$,

$$\bar{e}'(0)-\bar{e}(0)=\left[\frac{qn^{-1}\Delta^{-1}(3k^{-1}-k+2)}{4}\right]\phi\left(-\frac{\Delta}{2}\right)+O(n^{-2}).$$

COROLLARY 4.2. *For* $k=1$, $\bar{e}'(0)-\bar{e}(0)>0$ *up to the order of approximation in Corollary* 4.1.

Remark 4.2. Above result justifies the claim in Cochran [2]. The same conclusion is reached in Memon and Okamoto [8] via efficiency consideration.

COROLLARY 4.3. *Under the assumptions in Theorem* 3.1, *for* $q = 0$,

$$\bar{e}'(t)-\bar{e}(t)=\left[\frac{n^{-1}\bar{p}(1-\bar{p})^{-1}(k-k^{-1})}{4}\right]$$

$$\times\left[\sum_{m=1}^{r_1}\Delta_m^{-1}(p_{m1}\phi(\eta_{1mt})-p_{m2}\phi(\eta_{2mt}))\right]+O(n^{-2}).$$

Proof. Observe that

$$\bar{e}'(t)-\bar{e}(t)=\tfrac{1}{2}\sum_{i=1}^{2}\left(\sum_{m=1}^{r_1} p_{mi}[e'_{im}(t)-e_{im}(t)] + \sum_{m'=r_1+1}^{r} p_{m'}[e'_{im'}(t)-e_{im'}(t)]\right).$$

From Theorem 3.1 and the arguments before Corollary 4.1, for $q=0$ and $m=1, ..., r_1$,

$$e'_{1m}(t)-e_{1m}(t)=\tfrac{1}{2}[n^{-1}\bar{p}(1-\bar{p})^{-1}(k-k^{-1})]\,\Delta_m^{-1}\phi(\eta_{1mt})+O(n^{-2})$$

and

$$e'_{2m}(t)-e_{2m}(t)=\tfrac{1}{2}[n^{-1}\bar{p}(1-\bar{p})^{-1}(k^{-1}-k)]\,\Delta_m^{-1}\phi(\eta_{2mt})+O(n^{-2});$$

and for $m'=r_1+1, ..., r$,

$$e'_{1m'}(t)-e_{1m'}(t)=O(n^{-2}); \qquad \text{and} \qquad e'_{2m'}(t)-e_{2m'}(t)=O(n^{-2}).$$

Hence, the result.

COROLLARY 4.4. *Up to the order of approximation in Corollary* 4.3,

1. *If* $n_1=n_2$, *then* $\bar{e}'(t)-\bar{e}(t)=O(n^{-2})$ *for all* t, $-\infty<t<\infty$;
2. *If* $n_1>n_2$, *then* $\bar{e}(t)<\bar{e}'(t)$, *if and only if* $\sum_{m=1}^{r_1}\Delta_m^{-1}(p_{m1}\phi(\eta_{1mt})-p_{m2}\phi(\eta_{2mt}))<0$; *and*
3. *If* $n_1<n_2$, *then* $\bar{e}(t)<\bar{e}'(t)$, *if and only if* $\sum_{m=1}^{r_1}\Delta_m^{-1}(p_{m1}\phi(\eta_{1mt})-p_{m2}\phi(\eta_{2mt}))>0$.

Remark 4.3. From Corollary 4.4, adjustment of discrete covariates in discrimination of mixed variables without continuous covariates is essential only if the two training samples are of very different sizes.

5. NUMERICAL RESULTS

In this section, selected values of $\bar{e}'(t)-\bar{e}(t)$ are computed and reported to pinpoint the implication of Corollary 4.3 in practice. To achieve this and have the results conveniently presented, $\Delta_1=\cdots=\Delta_{r_1}=\Delta=0.5,\ 1.0,\ 1.5$; $t=-0.5, 0, 0.5$; $\bar{p}=0.2, 0.3$ and $n_2=kn_1$; $k=0.8, 1.5$; $n_1=50, 100, 200$ are used throughout the study. Only small values of r, $r-r_1$ and $\bar{p}$ are considered because discrete discriminators including covariates are rare in practice. The thresholds $t=-0.5, 0$, and $t=0.5$ are chosen so that the effect of departure from the zero threshold and symmetry of $\bar{e}'(t) - \bar{e}(t)$ about zero can be examined. To summarize the results of the computations in a readable form, only the cases where $(r_1, r) = (1, 2)$, and $(r_1, r) = (2, 4)$ are tabulated. Results for unequal values of Δ_m, $m = 1, ..., r_1$ are unlikely to cause much difference due to the rather large values of n. Only positive gains due to the adjustment are reported in Tables I and II. In our study, for each given set of values of (r_1, r); $\bar{p}$ and (p_{m1}, p_{m2}); for $m = 1, ..., r_1$; and $i = 1, 2$, there are 54 cases for all the combinations of t; Δ; k; and n_1. The number of cases showing positive gains, no gains and negative gains are equally divided amongst each set of the 54 cases considered in our study. An examination of Table I and Table II indicates that improvement occurs only at nonzero thresholds. For all the cases considered, positive gains occur at a negative threshold for $n_1 > n_2$. This feature is observed again at a positive threshold for $n_1 < n_2$. To summarize, up to the order of approximation given in Corollary 4.3, we have the following:

(i) A nonzero threshold and a substantial difference in the sizes of the two training samples are crucial to a positive gain due to the adjustment.

(ii) Adjustment is beneficial when either $n_1 > n_2$ and a negative threshold is adopted or $n_1 < n_2$ and a positive threshold is adopted.

(iii) At a zero threshold, there is practically no improvement by adjustment no matter the sizes of the two training samples.

(iv) The gain due to the adjustment is unlikely to be dramatic considering the large sample sizes of the two training samples and the other values of the relevant quantities in $\bar{e}'(t) - \bar{e}(t)$.

TABLE I

Values[a] of $\bar{e}'(t) - \bar{e}(t)$ for $\Delta_1 = \cdots = \Delta_{r_1} = \Delta = 0.5, 1.0, 1.5$; $t = -0.5, 0.5$; $\bar{p} = 0.2, 0.3$; $n_2 = kn_1$, $k = 0.8, 1.5$; $n_1 = 50, 100, 200$ and $(r_1, r) = (1, 2)$ for Cases Where Improvement Is Observed

			$k = 8$ $t = -0.5$			$k = 1.5$ $t = 0.5$	
	$n_1 =$	50	100	200	50	100	200
$\bar{p} = 0.3$ $p_{11} = 0.7$ $p_{12} = 0.7$	$\Delta = 0.5$	9.30	4.54	2.25	12.24	6.02	2.99
	$= 1.0$	6.16	3.01	1.49	8.11	3.99	1.98
	$= 1.5$	3.77	1.84	0.91	4.96	2.44	1.21
$\bar{p} = 0.2$ $p_{11} = 0.8$ $p_{12} = 0.8$	$\Delta = 0.5$	6.20	3.03	1.50	8.16	4.01	1.99
	$= 1.0$	4.11	2.01	0.99	5.41	2.66	1.32
	$= 1.5$	2.51	1.23	0.61	3.30	1.63	0.81

[a] Actual figures equal 10^{-5} times the tabulated values.

TABLE II

Values of $\bar{e}'(t)-\bar{e}(t)$ for $\Delta_1=\cdots=\Delta_{r_1}=\Delta=0.5$, 1.0, 1.5; $t=-0.5$, 0.5; $\bar{p}=0.2$, 0.3; $n_2=kn_1$, $k=0.8$, 1.5; $n_1=50$, 100, 200 and $(r_1, r)=(2{,}4)$ for Cases Where Improvement Is Observed.

			$k=0.8$ $t=-0.5$			$k=1.5$ $t=0.5$	
$\bar{p}=0.3$	$n_1=$	50	100	200	50	100	200
$p_{11}=0.4$ $p_{12}=0.2$	$\Delta=0.5$	9.27	4.42	2.16	9.32	4.51	2.22
$p_{21}=0.3$ $p_{22}=0.5$	$=1.0$	5.50	2.62	1.28	6.92	3.35	1.65
	$=1.5$	3.54	1.69	0.82	4.56	2.21	1.08
$p_{11}=0.1$ $p_{12}=0.3$	$\Delta=0.5$	4.13	1.97	0.96	1.44	0.70	0.34
$p_{21}=0.6$ $p_{22}=0.4$	$=1.0$	4.92	2.34	1.15	7.28	3.52	1.73
	$=1.5$	3.39	1.62	0.79	4.54	2.20	1.08
$p_{11}=0.2$ $p_{12}=0.4$	$\Delta=0.5$	7.18	3.43	1.67	12.03	5.82	2.86
$p_{21}=0.5$ $p_{22}=0.3$	$=1.0$	5.33	2.54	1.24	7.14	3.45	1.70
	$=1.5$	3.51	1.68	0.82	4.59	2.22	1.09
$p_{11}=0.3$ $p_{12}=0.1$	$\Delta=0.5$	11.08	5.28	2.58	5.36	2.59	1.27
$p_{21}=0.4$ $p_{22}=0.6$	$=1.0$	5.61	2.67	1.31	6.38	3.08	1.52
	$=1.5$	3.50	1.67	0.82	4.40	2.13	1.05
			$k=0.8$ $t=-0.5$			$k=1.5$ $t=0.5$	
$\bar{p}=0.2$	$n_1=$	50	100	200	50	100	200
$p_{11}=0.5$ $p_{12}=0.3$	$\Delta=0.5$	5.86	2.79	1.37	7.61	3.68	1.81
$p_{21}=0.3$ $p_{22}=0.5$	$=1.0$	3.78	1.80	0.88	4.91	2.37	1.17
	$=1.5$	2.42	1.15	0.56	3.14	1.52	0.75
$p_{11}=0.6$ $p_{12}=0.1$	$\Delta=0.5$	1.61	0.77	0.38	0.17	0.08	0.04
$p_{21}=0.2$ $p_{22}=0.7$	$=1.0$	1.83	0.87	0.43	1.48	0.72	0.35
	$=1.5$	1.41	0.67	0.33	1.62	0.78	0.39
$p_{11}=0.3$ $p_{12}=0.5$	$\Delta=0.5$	5.86	2.79	1.37	7.61	3.68	1.81
$p_{21}=0.5$ $p_{22}=0.3$	$=1.0$	3.78	1.80	0.88	4.91	2.37	1.17
	$=1.5$	2.42	1.15	0.56	3.14	1.52	0.75
$p_{11}=0.1$ $p_{12}=0.6$	$\Delta=0.5$	0.13	0.06	0.03	2.10	1.01	0.50
$p_{21}=0.7$ $p_{22}=0.2$	$=1.0$	1.14	0.54	0.27	2.37	1.15	0.56
	$=1.5$	1.25	0.59	0.29	1.83	0.89	0.44

Thus, discrete covariate adjustment is essential for a nonzero threshold and is recommended in situations which are considered appropriate.

APPENDIX

In this section, we prove Theorem 3.1.

Proof. A simple calculation gives part (a). It remains to derive Eq. (3.1). Given $Z_m=1$, $m=1, \ldots, r_1$, define T_m, W_m, H_m and V_m as follows

$$\hat{\mu}_{m1} - \hat{\mu}_{m2} = \delta_m + (n(m))^{-1/2}\, T_m;$$
$$\hat{\mu}_{m1} = (n(m))^{-1/2}\, W_m;$$
$$\hat{\beta}_m \hat{\beta}'_m = b_m I_p + (n(m))^{-1/2}\, V_m; \qquad b_m = q(n(m) - q - 1)^{-1};$$
$$\hat{\Sigma}^{(m)}_{1.2} = I_p + (n(m))^{-1/2}\, V_m.$$

A conditional argument shows that $e_{1m}(t) = E_{1m}(E_{2m}(\Phi(G_m)))$ where $E_{2m}(.)$ and $E_{1m}(.)$ denote respectively the conditional expectation given n_{m1} and n_{m2} and the expectation with respect to n_{m1} and n_{m2} with

$$G_m = a_m\, \Delta_m^{-1} \left[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \frac{\Delta_m^2}{2} \right] + (n(m))^{-1/2}\, L_m + (n(m))^{-1}\, Q_m + r_{1m};$$

where

$$a_m = [(n(m) - q - 1)(n(m) - 1)^{-1}]^{1/2};$$

$$L_m = a_m\, \Delta_m^{-1} \left[\delta'_m W_m - \delta'_m T_m + \frac{\delta'_m V_m \delta_m}{2} \right] - a_m^3 \Delta_m^3 \left[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \frac{\Delta_m^2}{2} \right] \times \left[a_m^{-2}(\delta'_m T_m - \delta'_m V_m \delta_m) - \frac{\delta'_m H_m \delta_m}{2} \right];$$

$$\begin{aligned} Q_m = {} & a_m\, \Delta_m^{-1} \left[T'_m W_m - \delta'_m V_m W_m + \delta'_m V_m T_m - \frac{T'_m T_m}{2} - \frac{\delta'_m V_m^2 \delta_m}{2} \right] \\ & - \left[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \frac{\Delta_m^2}{2} \right] \\ & \times \left[\frac{\Delta_m^{-3} a_m^3}{2} \{ a_m^{-2}(T'_m T_m - 4\delta'_m V_m T_m + 3\delta'_m V_m^2 \delta_m) \right. \\ & + 2\delta'_m H_m T_m - 3\delta'_m (H_m V_m + V_m H_m)\, \delta_m \} \\ & \left. - \frac{3\Delta_m^{-5} a_m^5}{2} \left\{ a_m^{-2}(\delta'_m T_m - \delta'_m V_m \delta_m) + \frac{\delta'_m H_m \delta_m}{2} \right\}^2 \right] \\ & - a_m^3\, \Delta_m^{-3} \left[a_m^{-2}(\delta'_m T_m - \delta'_m V_m \delta_m) + \frac{\delta'_m H_m \delta_m}{2} \right] \\ & \times \left[\delta'_m W_m - \delta'_m T_m + \frac{\delta'_m V_m \delta_m}{2} \right]; \end{aligned}$$

and r_{1m} is a remainder term such that $E_{1m}(E_{2m}(r_{1m})) = O(n^{-2})$ under (C1) and (C2). It follows from Anderson [1, p. 968, Eq. (21)] that

$$e_{1m}(t) = E_{1m}\left(\Phi\left(a_m \Delta_m^{-1}\left[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \frac{\Delta_m^2}{2}\right]\right)\right)$$
$$+ n^{-1} E_{1m}(A_m) + O(n^{-2}); \tag{1}$$

where

$$A_m = n\phi\left(a_m\, \Delta_m^{-1}\left[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \frac{\Delta_m^2}{2}\right]\right)$$
$$\times\Bigg[(n(m))^{-1/2}\, E_{2m}(L_m) + (n(m))^{-1}$$
$$\times\left\{E_{2m}(Q_m) - \frac{a_m \Delta_m^{-1}}{2}\left[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \frac{\Delta_m^2}{2}\right] E_{2m}(L_m^2)\right\}\Bigg].$$

By Lemma 3.1, $E_{2m}(L_m) = 0$. Using the probability limits of a_m, b_m and d_m, we have

$$n^{-1} E_{1m}(A_m) = n^{-1}\phi(\eta_{1mt})\, \gamma_{1mt} + O(n^{-2}). \tag{2}$$

An application of Lemma 3.3 with a similar calculation in the expansions of $\Phi(a_m \Delta_m^{-1}[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \Delta_m^2/2])$ and $\phi(a_m \Delta_m^{-1}[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \Delta_m^2/2])$ leads to

$$E_{1m}\left(\Phi\left(a_m \Delta_m^{-1}\left[t + \log(\hat{p}_{m2}/\hat{p}_{m1}) - \frac{\Delta_m^2}{2}\right]\right)\right)$$
$$= \Phi(\eta_{1mt}) + n^{-1}\phi(\eta_{1mt})(\alpha_{1mt} + \tau_{1mt}) + O(^{-2}). \tag{3}$$

Combining Eq. (2) and Eq. (3) gives Eq. (3.1). This proves part (b).

ACKNOWLEDGMENTS

The author is grateful to the editor in charge and two referees for many constructive comments that helped improve the revision.

REFERENCES

1. T. W. Anderson, An asymptotic expansion of the distribution of the studentized classification statistics W, *Ann. Statist.* **1** (1973), 964–972.
2. W. G. Cochran, Comparison of two methods of handling covariates in discriminant function, *Ann. Instit. Statist. Math.* **16** (1964), 43–53.

3. W. G. Cochran and C. T. Bliss, Discriminant functions with covariance, *Ann. Math. Statist.* **19** (1948), 151–176.
4. J. J. Daudin, Selection of variables in mixed-variable discriminant analysis, *Biometrics* **42** (1986), 473–481.
5. W. J. Krzanowski, Discrimination and classification using both binary and continuous variables, *J. Amer. Statist. Assoc.* **70** (1975), 782–790.
6. A. M. Kshirsagar, "Multivariate Analysis," Marcel Dekker, New York, 1972.
7. C. Y. Leung, The location linear discriminant for classifying observations with unequal variances, *Statist. Probab. Lett.* **31** (1996), 23–29.
8. A. Z. Memon and M. Okmoto, The classification statistic W^* in covariate discriminant analysis, *Ann. Math. Statist.* **41** (1970), 1491–1499.
9. M. Okamoto, An asymptotic expansion for the distribution of the linear discriminant function, *Ann. Math. Statist.* **34** (1963), 1281–1301; Correction, *Ann. Math. Statist.* **39** (1968), 1358–1359.
10. I. Olkin and R. F. Tate, Multivariate correlation models with mixed discrete and continuous variables, *Ann. Math. Statist.* **32** (1961), 442–465; Correction, *Ann. Math. Statist.* **36** (1965), 343–344.
11. S. R. Searle, "Linear Models," Wiley, New York, 1971.
12. M. S. Srivastava and C. G. Khatri, "An Introduction to Multivariate Statistics," North Holland, Amsterdam, 1979.
13. I. G. Vlachonikolis and F. H. C. Marriott, Discrimination with mixed binary and continuous data, *Appl. Statist.* **31** (1982), 23–31.