▩ 연구논문

On an Equal Mean Quadratic Classification Rule With Unknown Prior Probabilities

Hea-Jung Kim
Dept. of Statistics, Dongguk University
Koichi Inada
Dept. of Mathematic, Kagoshima University

Abstract

We describe a formal approach to the construction of optimal classification rule for the two-group normal classification with equal population mean problem. Based on the utility function of Bernardo, we suggest a balanced design for the classification and construct the optimal rule under the balanced design condition. The rule is characterized by a constrained minimization of total risk of misclassification, the constraint of which is constructed by the process of equation between expected utilities of the two group conditional densities. The efficacy of the suggested rule is examined through numerical studies. This indicates that, in case little is known about the relative population sizes, dramatic gains in accuracy of classification result can be achieved.

1. Introduction

In the two-group classification problem, one wishes to assign an individual to one of the two populations Π_i, $i=1, 2$ on the basis of a vector $\boldsymbol{z}$ of observations. The most common assumption in the classification analysis is that $\boldsymbol{z}$ is a p-dimensional vector of observations, and that if it comes from Π_i then it follows a multivariate normal distribution with mean vector μ_i and covariance matrix Σ_i. Our concern in this paper is the two-group classification analysis with equal normal mean vectors, $\mu_1 = \mu_2 = \mu$. A practical example to which this model is applicable is in the discrimination between monozygotic and like-sexed dizygotic pairs of twins. The vector of differences between the measurements of various anatomical features on each pair will tend to be much smaller in the former case than in the latter one.

For this analysis, if we suppose a classification rule ϕ based on the observation z is defined by specifying $\phi(\Pi_i \mid z)$, the simplest and most usual Bayes (or optimal) rule for the equal mean two-group normal classification (cf. Fatti et. al., 1982; Anderson, 1984) with respect to fixed prior probabilities $\pi=(\pi_1, \pi_2)'$ of z coming from Π_i, $i=1, 2$, is

$$\begin{aligned}\phi(\Pi_i \mid z) &= 1 \text{ if and only if } \quad z \in R_i,\ R_i=\{z: \pi_j c_{ij} f(z \mid \Pi_j) \le \pi_i c_{ij} f(z \mid \Pi_i)\} \\ &= 0 \quad \text{otherwise,}\end{aligned} \tag{1}$$

where c_{ij} is the cost of misclassification of a sampling unit z from Π_j as being from Π_i, $c_{ii}=0$ for every i, $\pi_1+\pi_2=1$, $\pi_i>0$ and $f(z \mid \Pi_i)$ is pdf for z from $\Pi_i=N_p(\mu, \Sigma_i)$. As seen in (1), R_1 and $R_2(R_2=\overline{R}_1)$ denote classification regions so that if z falls in R_i, the rule classifies z into Π_i, $i=1, 2$. The risk $\gamma(\phi, \pi)$ of misclassification of a sampling unit z under the rule (1) is

$$\gamma(\phi, \pi) = \pi_1 c_{21} \int_{R_2} f(z \mid \Pi_1) dz + \pi_2 c_{12} \int_{R_1} f(z \mid \Pi_2) dz. \tag{2}$$

When the costs of misclassification are all equal, the Bayes rule (1) simplifies to classify z into Π_1 if and only if it falls into a region R_1:

$$\log(|\Sigma_2|/|\Sigma_1|) - (z-\mu)'(\Sigma_1^{-1}-\Sigma_2^{-1})(z-\mu) \ge 2\log\frac{\pi_2}{\pi_1}; \tag{3}$$

otherwise classify z into Π_2. Under the rule (3), the risk is proportional to the total probability of misclassification (TPM):

$$TPM = \pi_1 \int_{R_2} f(z \mid \Pi_1) dz + \pi_2 \int_{R_1} f(z \mid \Pi_2) dz. \tag{4}$$

It is also shown that (3) is equivalent to the optimal probabilistic classification rule which minimizes TPM among all possible classification rules derived from a purely probabilistic point of view(cf. Press, 1982).

In most practical situations, the population quantities, π_1, μ, Σ_1, and Σ_2 are unknown, so the rule (1) (or (3)) must be modified to approximate the optimal classification rule. Since the risk and TPM are dependent on π_i, in addition to knowing the population parameters, it is also necessary to know the values of π_i. The unknown prior probabilities can be estimated in various ways. Sometimes

the prior probabilities might be well approximated from knowledge of the relative sizes of the two populations. When little is known about the relative population sizes, it is usual to set $\pi_1 = \pi_2 = 1/2$ (cf. Johnson and Wichern, 1992). We will call this as the little knowledge estimates (LKE). Besides LKE, one can illustrate a couple of other estimates; (i) intuitive estimates that use training sample proportions under the assumption of mixed sampling scheme(cf. Goldstein and Dillon, 1978); (ii) a trial-and-error method using minimax condition by Anderson (1984). However, all the estimates referred above are either intuitive or unclosed form estimates.

The purpose of this paper is to propose a formal and closed form estimates of π_i for the classification rule when little is known about the relative population sizes. This is achieved by introducing a balanced condition for the classification experiment which controls the population distributions to have equal expected utilities by Bernardo(1979). This leads to find an optimal(in a sense of minimizing the risk of misclassification) classification rule under the constraint that the expected log-likelihood ratio of two class conditional distributions with respect to a true unconditional population distribution is equal to zero. The resulting optimal classification rule, involving a formal and closed form estimates of the unknown prior probabilities, is named as balanced classification rule.

2. Balanced Classification Rule

Suppose there are two continuous multivariate populations Π_i, $i=1, 2$, with corresponding absolutely continuous probability densities $f_i(\boldsymbol{z})$, where $\boldsymbol{z}$ is an p vector observation from a particular population Π_i, and that there are unknown prior probabilities π_i that $\boldsymbol{z} \in \Pi_i$. Further suppose that $U\{f_i(\boldsymbol{z}), \boldsymbol{z}\}$ be a real valued function describing the utility associated with the choice of a density function $f_i(\boldsymbol{z})$ as that of the observation $\boldsymbol{z}$. Then its expected utility is defined as

$$EU\{f_i(\boldsymbol{z}), \boldsymbol{z}\} = \int U\{f_i(\boldsymbol{z}), \boldsymbol{z}\} f(\boldsymbol{z}) d\boldsymbol{z}, \tag{5}$$

where $f(\boldsymbol{z}) = \Sigma_{i=1}^{2} \pi_i f_i(\boldsymbol{z})$ denotes the unconditional true density of $\boldsymbol{Z}$ at $\boldsymbol{z}$. It is appropriate to assume that the utility function (5) satisfies the proper condition that the supremum of the expected utilities is attained if and only if $f(\boldsymbol{z})$ is chosen as the density of $\boldsymbol{z}$. Buehler(1971) mentioned a number of examples of proper utility functions. However following logarithmic proper utility function by

Bernardo(1979) is easier to handle to describe the preferences of the probability density functions involved in our problem of concern.

Lemma 1 (Bernardo, 1979). If the utility function U is smooth and proper, then, for some constant A and function B,

$$U\{f_i(\boldsymbol{z}), \boldsymbol{z}\} = A\log f_i(\boldsymbol{z}) + B(\boldsymbol{z}), \; i=1, 2. \tag{6}$$

Using the utility function (6), we can judge that the expected utility of $f_k(\boldsymbol{z})$ is larger in classifying the observation $\boldsymbol{z}$ than that of $f_i(\boldsymbol{z})$ if

$$EU\{f_k(\boldsymbol{z}), \boldsymbol{z}\} - EU\{f_i(\boldsymbol{z}), \boldsymbol{z}\} = A\int \log \frac{f_k(\boldsymbol{z})}{f_i(\boldsymbol{z})} f(\boldsymbol{z})d\boldsymbol{z} > 0. \tag{7}$$

In our classification experiment with little knowledge about the relative population sizes, we want the population distributions to have equal expected utilities in classifying an individual. So that the control of the utilities would enable an experimenter to classify an individual with profile $\boldsymbol{z}$ mainly based on the resemblance in its characteristics with a particular Π_i.

Definition 1. A design for the two-group classification experiment is balanced, if the expected utility of probability densities $f_i(\boldsymbol{z})$ characterized by the populations Π_i, $i=1, 2$ are equal :

$$EU\{f_1(\boldsymbol{z}), \boldsymbol{z}\} - EU\{f_2(\boldsymbol{z}), \boldsymbol{z}\} = 0 \tag{8}$$

Under the balanced design for the classification experiment, we may have following optimal classification rule.

Theorem 2. Suppose $\boldsymbol{z}:p\times 1$ is an observation from one of populations Π_i with density $f_i(\boldsymbol{z})$ with prior probabilities for Π_i of π_i, $\Sigma_{i=1}^{2}\pi_i = 1$, and costs of misclassification c_{ij}, $i, j=1, 2$; $i\neq j$. Under the balanced design, optimal(minimum risk) classification rule is to classify $\boldsymbol{z}$ into Π_1 if

$$\pi_2 c_{12} f_2(\boldsymbol{z}) \leq \pi_1 c_{21} f_1(\boldsymbol{z}) \tag{9}$$

$$s.t. \quad \int \log \frac{f_1(\boldsymbol{z})}{f_2(\boldsymbol{z})} \sum_{j=1}^{2} \pi_j f_j(\boldsymbol{z}) d\boldsymbol{z} = 0. \tag{10}$$

Proof. Define classification regions R_1 and R_2 in the sample space generated by the random vector $\boldsymbol{Z}$ so that if $\boldsymbol{z} \in R_i$, classify $\boldsymbol{z}$ into Π_i, $i=1, 2$. Then, if we denotes the risk ρ, the problem is to determine the regions R_1 and R_2 that minimize ρ

$$\rho = c_{12}\pi_2 \int_{R_1} f_2(\boldsymbol{z}) d\boldsymbol{z} + c_{21}\pi_1 \int_{R_2} f_1(\boldsymbol{z}) d\boldsymbol{z},$$

subject to the balanced design condition (8). Through the logarithmic utility function (6), (8) can be expressed as

$$\int \log \frac{f_1(\boldsymbol{z})}{f_2(\boldsymbol{z})} \sum_{j=1}^{2} \pi_j f_j(\boldsymbol{z}) d\boldsymbol{z} = 0. \tag{11}$$

Since $\int_{R_1} f_i(\boldsymbol{z}) d\boldsymbol{z} + \int_{R_2} f_i(\boldsymbol{z}) d\boldsymbol{z} = 1$,

$$\rho = \int_{R_1} c_{12}\pi_2 f_2(\boldsymbol{z}) - c_{21}\pi_1 f_1(\boldsymbol{z}) d\boldsymbol{z} + c_{21}\pi_1,$$

and the balanced condition (11) is independent of the space of $\boldsymbol{z}$. By the Neyman Pearson lemma(see, for instance, Kendall and Stuart, 1966), ρ is minimized if R_1 is selected to include all those $\boldsymbol{z}$'s for which $c_{12}\pi_2 f_2(\boldsymbol{z}) - c_{21}\pi_1 f_1(\boldsymbol{z}) \leq 0$, and R_1 excludes those $\boldsymbol{z}$'s for which the reverse in equality holds subject to the values of π_i satisfying the condition (11).

When the costs of misclassification c_{ij} are all equal, the balanced classification rule assigns $\boldsymbol{z}$ to Π_1 if

$$\begin{aligned} &\pi_1 f_1(\boldsymbol{z}) \geq \pi_2 f_2(\boldsymbol{z}) \\ &s.t. \quad \int \log \frac{f_1(\boldsymbol{z})}{f_2(\boldsymbol{z})} \sum_{j=1}^{2} \pi_j f_j(\boldsymbol{z}) d\boldsymbol{z} = 0. \end{aligned} \tag{12}$$

It is noted that the condition (11) for the balanced design defines two linearly independent equations in π_i; one is $E[\log f_1(\boldsymbol{z}) - \log f_2(\boldsymbol{z})] = 0$ and the other is

$\Sigma_{i=1}^{2}\pi_i=1$. As they are linear in π_i, the condition gives unique solution of π_i.

Corollary 1. The condition for the balanced two-group classification experiment always yields the unique solutions for π_i, $0<\pi_i<1$, $i=1, 2$.

Proof. For the two population case, (11) reduces to two equations;

$$\pi_1\int \log\frac{f_1(z)}{f_2(z)}f_1(z)dz-\pi_2\int \log\frac{f_2(z)}{f_1(z)}f_2(z)dz=0$$

and $\pi_1+\pi_2=1$. Using the inequality(Lindley, 1965, Theorem 1);

$$\int \log\frac{f_i(z)}{f_k(z)}f_i(z)dz>0 \quad \text{foll all} \quad f_i(z)\neq f_k(z);\ i\neq k, \tag{13}$$

we can see that the two equations lead to unique solution

$$\pi_1=\frac{\int f_2(z)\log(f_2(z)/f_1(z))dz}{\int f_1(z)\log(f_1(z)/f_2(z))dz+\int f_2(z)\log(f_2(z)/f_1(z))dz}$$

and $\pi_2=1-\pi_1$, where $0<\pi_i<1$, $i=1, 2$. Hence, the result follows.

3. Equal Mean Normal Classification Rule

In practice, the probability density functions $f_i(z)$, $i=1, 2$ are seldom known. As usual, we assume that they have multivariate normal distributions. Further we will assume that the costs of misclassification are all equal, so that the resulting decision-theoretic and probabilistic classification rules are the same.

Lemma 3. Let Z follow p-dimensional multivariate normal distribution $N_p(\mu, \Sigma)$. Then, for any nonsingular $\Omega: p\times p$,

$$E(Z-\mu)'\Omega^{-1}(Z-\mu)=tr(\Sigma\Omega^{-1}). \tag{14}$$

Proof. Under the distribution, the left-hand side of (14) can be expanded as $EZ'\Omega^{-1}Z-\mu'\Omega^{-1}\mu$. Thus the statement follows from the fact that

$$EZ'\Omega^{-1}Z = trE(ZZ')\Omega^{-1} = \mu'\Omega^{-1}\mu + tr(\Sigma\Omega^{-1}).$$

Theorem 4. Suppose $\Pi_i = N_p(\mu, \Sigma_i)$, $\Sigma_i > 0$ with density $f(z \mid \Pi_i)$, where μ and Σ_i are known, $i = 1, 2$ and suppose the costs of misclassification are all equal. Then the balanced optimal classification rule in Theorem 2 classifies z into Π_1 if and only if it falls in the region R_1:

$$\log(|\Sigma_2|/|\Sigma_1|) - (z-\mu)'(\Sigma_1^{-1} - \Sigma_2^{-1})(z-\mu) \geq 2\log\frac{\pi_2}{\pi_1}, \tag{15}$$

where

$$\pi_1 = \frac{\log(|\Sigma_2|/|\Sigma_1|) + p - tr(\Sigma_2\Sigma_1^{-1})}{2p - tr(\Sigma_1\Sigma_2^{-1}) - tr(\Sigma_2\Sigma_1^{-1})} \tag{16}$$

Proof. Under the hypothesis, the optimal classification rule in Theorem 2 is equal to

$$\pi_1 f(z \mid \Pi_1) \geq \pi_2 f(z \mid \Pi_2),$$

$$s.t. \quad \int \log\frac{f(z \mid \Pi_1)}{f(z \mid \Pi_2)} \sum_{j=1}^{2} \pi_j f(z \mid \Pi_j) dz = 0.$$

Direct substitution of exact functional form of the multivariate normal density and evaluation of the integral by applying Lemma 3 give the result.

We now evaluate the probabilities of misclassification for the balanced classification rule in Theorem 4. If Z is a random observation, we consider the random variables

$$U_1 = \log(|\Sigma_2|/|\Sigma_1|) - (Z-\mu)'(\Sigma_1^{-1} - \Sigma_2^{-1})(Z-\mu). \tag{17}$$

Since a linear transformation leaves (17) invariant, there is no loss of generality in considering the case $\Pi_1 \sim N_p(0, I)$ and $\Pi_2 \sim N_p(0, D)$. This canonical form are obtained via the transformation suggested by Dunn and Holloway(1967):

$$Y = A'\Sigma_1^{-1/2}(Z-\mu), \tag{18}$$

where A is an orthogonal matrix such that $A'\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}A = D$, a diagonal matrix

If D is a $p\times p$ matrix with diagonal elements $d_1, d_2, \cdots, d_p$, then the canonical form for (17) is

$$U_{12} = \sum_{k=1}^{p} \left(\frac{1-d_k}{d_k}\right) Z_k^2 + \sum_{k=1}^{p} \log d_k. \tag{19}$$

Theorem 4 shows that following total probability of misclassification(TPM) will be minimized by the balanced classification rule which assigns an individual to Π_1 whenever (15) satisfies, and to Π_2 otherwise. TPM for the rule is

$$\begin{aligned} TPM &= \pi_1 \int_{R_2} f(\boldsymbol{z} \mid \Pi_1) d\boldsymbol{z} + \pi_2 \int_{R_1} f(\boldsymbol{z} \mid \Pi_2) d\boldsymbol{z} \\ &= \pi_1 Pr(U_{12} < C_{\mathrm{BAL}} \mid \Pi_1) + \pi_2 Pr(U_{12} \geq C_{\mathrm{BAL}} \mid \Pi_2), \end{aligned} \tag{20}$$

where

$$C_{\mathrm{BAL}} = 2\log\left\{\frac{p + \Sigma_{k=1}^{p} (\log 1/d_k - 1/d_k)}{p + \Sigma_{k=1}^{p} (\log d_k - d_k)}\right\} \tag{21}$$

and

$$\pi_1 = \frac{p + \Sigma_{k=1}^{p} (\log d_k - d_k)}{2p - \Sigma_{k=1}^{p} (d_k + 1/d_k)}.$$

Suppose without loss of generality that the first q of the d_i are greater than or equal to one and the remainder are less than one. Then the probabilities in (20) become

$$Pr(U_{12} < C_{\mathrm{BAL}} \mid \Pi_1) = Pr\left(\sum_{k=1}^{p} T_k^2 - \sum_{k=q+1}^{p} T_k^2 > K_{\mathrm{BAL}} \mid \Pi_1\right). \tag{22}$$

$$Pr(U_{12} \geq C_{\mathrm{BAL}} \mid \Pi_2) = Pr\left(\sum_{k=1}^{q} T_k^2 - \sum_{k=q+1}^{p} T_k^2 \leq K_{\mathrm{BAL}} \mid \Pi_2\right). \tag{23}$$

where

$$T_k^2 = \frac{|d_k - 1|}{d_k} Z_k^2 \quad \text{and} \quad K_{\mathrm{BAL}} = -C_{\mathrm{BAL}} + \sum_{k=1}^{p} \log d_k.$$

In the particular case when all $d_k > 1$ or all $d_k < 1$ so that p is either q or 0, (22) and (23) can be obtained only from the distribution of $\Sigma_{k=1}^{p} T_k^2$. Using a result of Patnaik(1949), sum of squared normal random variables, $T = \Sigma_{k=1}^{p} T_k^2$ in Π_i, $i = 1, 2$ can be approximated by a multiple, α_i, of central χ^2 distribution with f_i degree of freedom, where α_i and f_i are chosen to satisfy

$$\mu_{T_i} = E(T \mid \Pi_i) = E(\alpha_i \chi^2_{f_i}) = \alpha_i f_i$$

and

$$\sigma^2_{T_i} = Var(T \mid \Pi_i) = Var(\alpha_i \chi^2_{f_i}) = 2\alpha_i^2 f_i, \quad i = 1, 2,$$

where

$$\mu_{T_1} = \sum_{k=1}^{p} \frac{|d_k - 1|}{d_k}, \quad \sigma^2_{T_1} = 2 \sum_{k=1}^{p} \frac{(d_k - 1)^2}{d_k},$$

$$\mu_{T_2} = \sum_{k=1}^{p} |d_k - 1|, \quad \sigma^2_{T_2} = 2 \sum_{k=1}^{p} (d_k - 1)^2.$$

Thus, $\int_{R_2} f(\boldsymbol{z} \mid \Pi_1) d\boldsymbol{z}$ would be approximated by

$$\int_{R_2} f(\boldsymbol{z} \mid \Pi_1) d\boldsymbol{z} \approx Pr(\chi^2_{f_1} \gtrless \pm K_{BAL}/\alpha_1) \tag{24}$$

and similar approximation gives

$$\int_{R_1} f(\boldsymbol{z} \mid \Pi_2) d\boldsymbol{z} \approx Pr(\chi^2_{f_2} \lessgtr \pm K_{BAL}/\alpha_2), \tag{25}$$

where the upper symbols are for all $d_k > 1$ while the lower symbols are for all $d_k < 1$, $k = 1, \cdots, p$. Using the same notations as above, we have the following result from (20), (24) and (25).

Corollary 2. In the case $d_k > 1$ or $d_k < 1$ for all $k = 1, \cdots, p$, the regions of classification, R_1 and $R_2 (R_2 = \overline{R}_1)$, defined by the balanced classification rule in Theorem 4 approximately yield the total probability of misclassification:

$$TRM \approx \pi_1 Pr(\chi^2_{f_1} \gtrless \pm K_{BAL}/\alpha_1) + \pi_2 Pr(\chi^2_{f_2} \lessgtr \pm K_{BAL}/\alpha_2). \tag{26}$$

where

$$\alpha_1 = \frac{\Sigma_{k=1}^{p}((d_k-1)/d_k)^2}{\Sigma_{k=1}^{p}(|d_k-1|/d_k)}, \quad \alpha_2 = \frac{\Sigma_{k=1}^{p}(d_k-1)^2}{\Sigma_{k=1}^{p}|d_k-1|},$$

$$f_1 = \frac{(\Sigma_{k=1}^{p}|d_k-1|/d_k)^2}{\Sigma_{k=1}^{p}((d_k-1)/d_k)^2}, \quad f_2 = \frac{(\Sigma_{k=1}^{p}|d_k-1|)^2}{\Sigma_{k=1}^{p}(d_k-1)^2}.$$

When $C_{BAL}=0$ (*i.e.* $\pi_1=\pi_2=1/2$) in the expression of K_{BAL}, (26) will give approximate TPM for the classification rule (3) which uses the usual little knowledge estimates (LKE) for π_i. In the following Section, this rule is compared to that of the balanced classification rule in Theorem 4.

4. Numerical Results

The goal of this section is to study the overall effectiveness of the balanced classification rule (BCR) suggested in Theorem 4 and to identify some situations where one would (and would not) expect substantial improvement with BCR. The performance of BCR is compared to the little knowledge optimal classification rule (LCR) which estimates the unknown prior probabilities appeared in (3) with LKE (*i.e.* $\pi_1=\pi_2=1/2$). The comparison between the two rules is conducted in terms of TPM (equivalently, optimal error rate). To put the comparison into a canonical form, we again made the transformation (18) so that $\mu=0$, $\Sigma_1=I_p$, and $\Sigma_2=D$, a diagonal matrix. Choice of parameters in the comparison are some combination of values of D and dimension p:

$$Case\ \mathrm{I}: D = diag(d_1, \cdots\cdots, d_p),\ p = 1, 2, 4, 6, 8, 10.$$

$$Case\ \mathrm{II}: D = diag(d, \cdots\cdots, d),\ p = 1, 2, 4, 6, 8, 10.$$

For Case I, to change the value of D, we set $d_k=1/(k+1/2)^m$ for the case when all $d_k<1$, while $d_k=1+(k+1/2)^m/2$ for the case when all $d_k>1$; $m=.2, .4, .6, .8, 1, 1.2, 1.4, 1.6, 1.8$. For each combination of D and dimension p, respective TPM's of BCR and LCR and their ratio (TPM_{LCR}/TPM_{BCR}) were calculated and tabulated in (Table 1). The quantities in parentheses are TPM_{BCR}'s over various situations.

Since $TPM_{LCR}/TPM_{BCR}>1$ means the better performance of BCR relative to LCR, the table shows uniformly better performance of BCR over LCR for all

combinations of the parameter values considered in our study. Moreover, the performance of BCR becomes better as m gets larger values for both criteria(the ratio and TPM_{BCR}), and this tendency increases with the number of parameters (or the dimension p). Thus, it can be deduced that the BCR can better utilize discrepancies in variance than LCR to decrease the probability of misclassification for the equal mean classification analysis.

〈 Table 1 〉 The Ratio TPM_{LCR}/TPM_{BCR} and TPM_{BCR} for Case Ⅰ

m	$p=1$	$p=2$	$p=4$	$p=6$	$p=8$	$p=10$
			$d_k>1$			
0.2	1.1030(.4060)	1.0885(.3836)	1.0766(.3462)	1.0712(.3146)	1.0682(.2872)	1.0663(.2628)
0.4	1.1114(.3998)	1.1031(.3703)	1.0995(.3191)	1.1000(.2753)	1.1018(.2373)	1.1040(.2041)
0.6	1.1204(.3933)	1.1215(.3554)	1.1316(.2884)	1.1429(.2317)	1.1542(.1840)	1.1651(.1445)
0.8	1.1302(.3865)	1.1444(.3389)	1.1753(.2551)	1.2047(.1866)	1.2326(.1328)	1.2595(.0921)
1.0	1.1407(.3794)	1.1724(.3211)	1.2336(.2204)	1.2911(.1434)	1.3463(.0888)	1.4003(.0525)
1.2	1.1520(.3720)	1.2064(.3021)	1.3099(.1858)	1.4092(.1050)	1.5073(.0550)	1.6051(.0268)
1.4	1.1641(.3643)	1.2471(.2821)	1.4080(.1527)	1.5682(.0734)	1.7314(.0318)	1.8986(.0125)
1.6	1.1772(.3564)	1.2957(.2613)	1.5330(.1226)	1.7799(.0492)	2.0406(.0173)	2.3163(.0054)
1.8	1.1912(.3482)	1.3530(.2403)	1.6909(.0962)	2.0600(.0318)	2.4657(.0090)	2.9103(.0022)
			$d_k<1$			
0.2	1.0165(.4822)	1.0335(.4598)	1.0420(.4243)	1.0454(.3923)	1.0476(.3628)	1.0492(.3353)
0.4	1.0343(.4675)	1.0726(.4201)	1.0973(.3516)	1.1110(.2930)	1.1216(.2423)	1.1308(.1988)
0.6	1.0532(.4468)	1.1180(.3814)	1.1685(.2845)	1.2016(.2083)	1.2295(.1491)	1.2547(.1043)
0.8	1.0736(.4293)	1.1705(.3440)	1.2587(.2250)	1.3235(.1413)	1.3809(.0849)	1.4345(.0490)
1.0	1.0953(.4119)	1.2311(.3083)	1.3719(.1741)	1.4848(.0919)	1.5893(.0452)	1.6897(.0209)
1.2	1.1184(.3947)	1.3006(.2745)	1.5131(.1320)	1.6965(.0575)	1.8735(.0228)	2.0486(.0083)
1.4	1.1432(.3777)	1.3804(.2430)	1.6883(.0983)	1.9733(.0349)	2.2601(.0110)	2.5538(.0031)
1.6	1.1696(.3610)	1.4717(.2139)	1.9056(.0720)	2.3350(.0207)	2.7873(.0052)	3.2677(.0011)
1.8	1.1977(.3445)	1.5762(.1872)	2.1748(.0520)	2.8087(.0120)	3.5093(.0024)	4.2424(.0004)

For Case Ⅱ, TPM_{OCR}/TPM_{BCR} were calculated for all combinations of the following parameter values: $d=.2, .4, .6, .8, 2, 4, 6, 8$; $\pi_1=1/6, 1/3, 1/2, 2/3, 5/6$. In this case, we changed the value of π_1 to see the relative performance of BCR over LCR and the optimal rule(OCR) in (3) which takes a value π_1 other than 1/2. Note that OCR with $\pi_1=1/2$ is equivalent to LCR. The results of the calculation

are given in 〈Table 2〉.

〈 Table 2 〉 The Ratio TPM_{OCR}/TPM_{BCR} for Case II

		d							
π_1	p	0.2	0.4	0.6	0.8	2	4	6	8
1/6	1	1.0922	0.8918	0.8175	0.7792	1.5218	1.9901	2.3989	2.7737
	2	1.1132	0.9339	0.8709	0.8405	1.4081	1.8105	2.1662	2.4939
	4	1.1165	0.9584	0.9071	0.8850	1.3198	1.6716	1.9875	2.2802
	6	1.1144	0.9671	0.9220	0.9044	1.2797	1.6097	1.9088	2.1869
	8	1.1122	0.9715	0.9305	0.9158	1.2559	1.5735	1.8634	2.1335
	10	1.1104	0.9743	0.9361	0.9235	1.2397	1.5494	1.8334	2.0985
1/3	1	1.3691	1.0807	0.9716	0.9138	0.8477	1.0157	1.1651	1.3020
	2	1.3330	1.0791	0.9866	0.9396	0.8960	1.0438	1.1797	1.3056
	4	1.2957	1.0718	0.9944	0.9574	0.9270	1.0542	1.1769	1.2923
	6	1.2764	1.0667	0.9968	0.9650	0.9391	1.0557	1.1718	1.2821
	8	1.2645	1.0632	0.9979	0.9694	0.9458	1.0556	1.1678	1.2751
	10	1.2563	1.0606	0.9985	0.9723	0.9501	1.0552	1.1647	1.2700
1/2	1	1.6460	1.2695	1.1257	1.0484	1.1848	1.5029	1.7820	2.0378
	2	1.5529	1.2244	1.1023	1.0387	1.1521	1.4271	1.6730	1.8997
	4	1.4748	1.1853	1.0817	1.0299	1.1234	1.3629	1.5822	1.7862
	6	1.4384	1.1664	1.0716	1.0256	1.1094	1.3327	1.5403	1.7345
	8	1.4168	1.1549	1.0653	1.0229	1.1008	1.3146	1.5156	1.7043
	10	1.4023	1.1470	1.0609	1.0210	1.0949	1.3023	1.4991	1.6843
2/3	1	1.9229	1.4583	1.2798	1.1830	1.0162	1.2593	1.4735	1.6700
	2	1.7728	1.3696	1.2180	1.1377	1.0240	1.2355	1.4263	1.6027
	4	1.6539	1.2987	1.1690	1.1024	1.0252	1.2086	1.3796	1.5393
	6	1.6004	1.2661	1.1463	1.0862	1.0243	1.1942	1.3561	1.5083
	8	1.5691	1.2466	1.1326	1.0764	1.0233	1.1851	1.3417	1.4897
	10	1.5482	1.2333	1.1233	1.0697	1.0225	1.1788	1.3319	1.4771
5/6	1	2.1998	1.6472	1.4439	1.3176	1.3533	1.7465	2.0905	2.4057
	2	1.9927	1.5149	1.3338	1.2368	1.2801	1.6188	1.9196	2.1967
	4	1.8331	1.4122	1.2563	1.1749	1.2216	1.5173	1.7849	2.0332
	6	1.7624	1.3657	1.2211	1.1468	1.1946	1.4712	1.7246	1.9607
	8	1.7213	1.3382	1.2000	1.1300	1.1783	1.4440	1.6895	1.9189
	10	1.6942	1.3197	1.1857	1.1184	1.1673	1.4259	1.6663	1.8914

A couple of points are noted in constructing the table. First, as would be expected, *BCR* achieves gain in accuracy of the equal mean classification for almost all situations. Comparing to *LCR*(i.e., in cases $\pi_1 = 1/2$) we can see that the gain in accuracy is uniform, and it becomes dramatic as d gets larger value. Second, for some situations of $\pi_1 = 1/6, 1/3$, *OCR* performs better than *BCR*. This means that the balanced classification rule in Theorem 4 does not always guarantee the better performance over *OCR*. Nevertheless, *BCR* is worthy of using for the equal mean classification analysis in the following reasons:(i) It guarantees at least uniformly better performance than the usual little knowledge optimal classification rule. (ii) Unlike *OCR* which uses intuitive or trial-and-error estimate of π_1, *BCR* gives a formal and closed form estimate of π_1 as defined in (16).

5. Concluding Remarks

We have considered the problem inherent in developing an optimal classification rule with unknown prior probabilities for the two-group equal mean classification. As an alternative to the usual little knowledge optimal classification rule, a balanced classification rule is proposed. The efficacy of the suggested rule is examined through limited but informative numerical studies. This studies indicate that in many circumstances dramatic gains in classification accuracy can be achieved by use of the suggested rule. In addition to the efficacy of the suggested rule, it has the following favorable merits: (i) The balanced classification rule enables us to get formal and closed form estimates of prior probabilities of individual group membership involved in the optimal classification rule. (ii) When the rule is applied to sample based classification by replacing the population parameters with their respective sample counterparts(cf. Wald,1944 Anderson, 1984; Friedman, 1989), we can obtain the estimates of prior probabilities in a rather unified way irrespective of sampling scheme for the training samples such as the mixed sampling and the independent sampling.

References

[1] Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, 2-nd ed., Wiley & Sons, New York.

[2] Bernardo, J. M. (1979), "Expected information as expected utility," *Ann Statist.*, Vol. 7, pp. 686–690.

[3] Buehler, R. J. (1971), *Measuring information and uncertainty*, *In Foundations of Statistical Inference*, Ed. by Godambe and Sprott, Holt Rinehart Winston, Toronto.

[4] Dunn, O. J. and Holloway, L. N. (1967), "The robustness of Hotelling's T^2," *J. Amer. Statist. Assoc.*, Vol. 62, pp. 124-136.

[5] Fatti, L. P., Hawkins, D. M. and Raath, L. R. (1982), *Discriminant analysis Topics in Applied Multivariate Analysis*, Ed. by Hawkins, D. M., Cambridge University Press, Cambridge.

[6] Friedman, J. H. (1989), "Regularized discriminant analysis," *J. Amer. Statist Assoc.*, Vol. 84, pp. 165-175.

[7] Goldstein, M. and Dillon, W. R. (1978), *Discrete Discriminant Analysis*, Wiley & Sons, New York.

[8] Johnson, N. L. and Kotz, S. (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley & Sons, New York.

[9] Johnson, R. A. and Wichern, D. W. (1992), *Applied Multivariate Statistical Analysis*, 3-rd ed., Prentice Hall, New Jersey.

[10] Kendall, M. C. and Stuart, A. (1966), *The Advanced Theory of Statistics*, Vol 2, Hafner Publishing Company, New York.

[11] Lindley, D. V. (1965), *Introduction to Probability and Statistics*, Vol. 1 Cambridge University Press, Cambridge.

[12] Patnaik, P. B. (1949), "The noncentral χ^2 and F-distributions and their approximations," Biometrika, Vol. 36, pp. 202-232.

[13] Press, S. J. (1982), *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, Robert E. Krieger, Florida.

[14] Wald, A. (1944), "On a statistical problem arising in the classification of an individual into one of two groups," *Ann. Math. Statist.*, Vol. 15, pp. 145-162