<Recearch Paper>

Journal of the Korean
Statistical Society
Vol. 16, No. 2, 1987

Empirical Bayes Posterior Odds Ratio for Heteroscedastic Classification

Hea-Jung Kim*

ABSTRACT

Our interest is to access in some way the relative odds or probability that a multivariate observation Z belongs to one of k multivariate normal populations with unequal covariance matrices. We derived the empirical Bayes posterior odds ratio for the classification rule when population parameters are unknown. It is a generalization of the posterior odds ratio suggested by Geisser(1964). The classification rule does not have complicated distribution theory which a large variety of techniques from the sampling viewpoint have. The proposed posterior odds ratio is compared to the Geisser's posterior odds ratio through a Monte Carlo study. The results show that the empirical Bayes posterior odds ratio, in general, performs better than the Geisser's. Especially, for large dimension of Z and small training sample, the performance is prominent.

1. Introduction

It is sometimes of interest for an investigator to access in some way the relative odds or probability that a multivariate obserbation Z belongs to one of k multivariate normal populations Π_i, $i=1,2,\cdots,k$. If the parameters of Π_i are known and covariance matrices are unequal, the quadratic classification rule is optimal in the sense of minimizing the overall probability of misclassification. When the parameters of Π_i are unknown but the estimates of them are available, based on independent samples of size N, a large variety of other techniques have been suggested from the sampling theory viewpoint-see, for example, Gilbert(1969), Lachenbruch, Sneeringer, and Revo(1973) and Dunn

* Department of Statistics, Dongguk University, Seoul 100, Korea.

(1971). Asymptotic relative performance of prominant techniques(the quadratic, best linear, and Fisher's linear discriminant functions) is given by Marks and Dunn(1974).

Although these may found easily, the distributions required to use these discriminant functions in small samples are quite complicated - see, for instance, Anderson(1951) and Sitgreaves(1952). However, the Bayesian approach provides a useful and simple alternative. Bayesian approach to the classification problem in the case of normally distributed observations with unknown parameters were discussed by Geisser(1964) and Dunsmore (1966). The results are extremely simple to apply and there is not complicated distribution theory. But their interest was primarily on a statement concerning the relative likelihood or probability that an observation belongs to one or another of the populations, as a basis for assignment, so that they used a particular convenient prior density to reflect an initial diffuseness or vaguness about the unknown parameters.

In this paper we derive a classification method using more customary Bayesian application of making a probability statement about where a parameter lies such that the classificalion method is often just as easily applicable for a broad range of Bayesian assumptions and flexible enough that it can be tailored to a temporal imperative. Apart from these advantages, we find that it has smaller overall probability of misclassification than the conventional method by Geisser(1964). This improvement comes from the increased precision in the estimation of the relative likelihood by utilizing the data we have.

General posterior odds ratio that Z helongs to Π_i, as compared with Π_j, is derived in Section 2 under the further assumption that population parameters have a natural conjugate prior distrbution. In Section 3 several related posterior odds ratios, each appropriate for special circumstances, are described and are evaluated by the empirical Bayesian scheme. In Section 4 a study compares by Monte Carlo methods the perfomance of two posterior odds ratios in classifying individuals into two multivariate normally distributed populations - Geisser's posterior odds ratio and empirical Bayes posterior odds ratio classification methods.

2. General Posterior Odds Ratio

In this section, we shall restrict ourselves to the case of normally distributed observations with unknown parameters and assume prior probabilities $q_1, q_2, \cdots, q_k$ are known.

The extension to the case where the prior probabilities are unknown, on which Geisser (1964) have already commented, is easily managed.

Let $X_{i\alpha}'=(x_{1i\alpha},\cdots,x_{pi\alpha})$ be the α-th observation obtained from Π_i, which is $N(\theta_i,\Sigma_i)$, $i=1,2,\cdots,k$, $\alpha=1,2,\cdots,N_i$, and let Z : p×1 be an observation from one of the populations. The natural conjugate prior densities here in essence are, for unknown p component mean vector θ_i and $p\times p$ covariance matrix Σ_i,

$$P(\theta_i,\Sigma_i)\propto\frac{\exp\{-1/2\cdot\mathrm{tr}\Sigma_i^{-1}[b_i(\theta_i-\phi_i)(\theta_i-\phi_i)'+\Omega_i]\}}{|\Sigma_i|^{(m_i+1)/2}}, \tag{2.1}$$

where ϕ_i,Ω_i,b_i and m_i are parameters of the prior distribution, $m>2p$.

Define

$$\bar{X}_i=\frac{1}{N_i}\sum_{\alpha=1}^{N_i}X_{i\alpha},\quad V_i=\sum_{\alpha=1}^{N_i}(X_{i\alpha}-\bar{X}_i)(X_{i\alpha}-\bar{X}_i)'.$$

Since $\bar{X}_i$ and V_i are independent and since

$$\bar{X}_i|\theta_i,\Sigma_i\sim N(\theta_i,\Sigma_i/N_i)\ \text{and}\ V_i|\Sigma_i\sim W(\Sigma_i,p,N_i-1),$$

the likelihood is given by

$$P(\bar{X}_i,V_i|\theta_i,\Sigma_i)\propto\frac{|V_i|^{(N_i-p-2)/2}}{|\Sigma_i|^{N_i/2}}\exp\{-1/2\cdot\mathrm{tr}\ \Sigma_i^{-1}[N_i(\theta_i-\bar{X}_i)(\theta_i-\bar{X}_i)'+V_i]\}. \tag{2.2}$$

Multiplying (2.2) by (2.1) yields the posterior density of the parameters in the i-th population which is

$$P(\theta_i,\Sigma_i|\bar{X}_i,V_i,\Pi_i)\propto\frac{\exp\{-1/2\cdot\mathrm{tr}\Sigma_i^{-1}[V_i+\Omega_i+N_i(\theta_i-\bar{X}_i)(\theta_i-\bar{X}_i)'+b_i(\theta_i-\phi_i)(\theta_i-\phi_i)']\}}{|\Sigma_i|^{(N_i+m_i+1)/2}}. \tag{2.3}$$

Hence, from the posterior density (2.3), we may obtain the predictive density of Z on the hypothesis that it was obtained from Π_i, which results in

$$\begin{aligned}P(Z|\bar{X}_i,V_i,\Pi_i)&=\iint P(Z|\theta_i,\Sigma_i,\Pi_i)P(\theta_i,\Sigma_i|\bar{X}_i,V_i,\Pi_i)d\theta_i d\Sigma_i\\&\propto 1/|B_i|^{(N_i+m_i-p)/2},\end{aligned} \tag{2.4}$$

where

$$\begin{aligned}(N_i+b_i+1)B_i=&V_i+\Omega_i+\frac{N_ib_i}{N_i+b_i}(\bar{X}_i-\phi_i)(\bar{X}_i-\phi_i)'\\&+\frac{N_i+b_i}{N_i+b_i+1}\left(Z-\frac{N_i\bar{X}_i+b_i\phi_i}{N_i+b_i}\right)\left(Z-\frac{N_i\bar{X}_i+b_i\phi_i}{N_i+b_i}\right)'.\end{aligned}$$

This is the kernal of a mutivariate Student t-density. Using the definition of B_i and completing the form of (2.4), the predictive density of Z is

$$P(Z|\bar{X}_i, V_i, \Pi_i) = \frac{|Q_i|^{-1/2}[\Gamma\{N_i+m_i-p)/2\}/\Gamma\{(N_i+m_i-2p)/2\}]\{(N_i+b_i)/[\pi(N_i+b_i+1)]\}^{p/2}}{\left[1+\frac{N_i+b_i}{N_i+b_i+1}\left(Z-\frac{N_i\bar{X}_i+b_i\phi_i}{N_i+b_i}\right)'Q_i^{-1}\left(Z-\frac{N_i\bar{X}_i+b_i\phi_i}{N_i+b_i}\right)\right]^{(N_i+m_i-p)/2}}, \tag{2.5}$$

where $Q_i=V_i+\Omega_i+\frac{N_ib_i}{N_i+b_i}(\bar{X}_i-\phi_i)(\bar{X}_i-\phi_i)'$.

Now apply Bayes Theorem to the predictive probability density (2.5) to compute the posterior probability that a new observation Z belongs to population Π_i. Then we get

$$P(\Pi_i|Z,q) = \frac{q_iP(Z|\bar{X}_i, V_i, \Pi_i)}{\sum_{j=1}^{k} q_jP(Z|\bar{X}_j, V_j, \Pi_j)}, \tag{2.6}$$

where q implies that $p(\Pi_i|Z,q)$ is conditioned on the known q_i 's.

It follows, from (2.6), that the posterior odds ratio for classifying Z into Π_i, as compared with Π_j, becomes proportional to the ratio of the associated multivariate Student t-densities,

$$C_{ij}=\frac{P(\Pi_i|Z,q)}{P(\Pi_j|Z,q)}$$

$$=L_{ij}\frac{\left[1+\frac{N_j+b_j}{N_j+b_j+1}\left(Z-\frac{N_j\bar{X}_j+b_j\phi_j}{N_j+b_j}\right)'Q_j^{-1}\left(Z-\frac{N_j\bar{X}_j+b_j\phi_j}{N_j+b_j}\right)\right]^{(N_j+m_j-p)/2}}{\left[1+\frac{N_i+b_i}{N_i+b_i+1}\left(Z-\frac{N_i\bar{X}_i+b_i\phi_i}{N_i+b_i}\right)'Q_i^{-1}\left(Z-\frac{N_i\bar{X}_i+b_i\phi_i}{N_i+b_i}\right)\right]^{(N_i+m_i-p)/2}}, \tag{2.7}$$

where L_{ij} is constant given by

$$L_{ij}=\frac{q_i}{q_j}\left\{\frac{|Q_j|}{|Q_i|}\right\}^{1/2}\frac{\Gamma\{(N_i+m_i-p)/2\}\Gamma\{(N_j+m_j-2p)/2\}}{\Gamma\{(N_j+m_j-p)/2\}\Gamma\{(N_i+m_i-2p)/2\}}\left\{\frac{(N_j+b_j+1)(N_i+b_i)}{(N_i+b_i+1)(N_j+b_j)}\right\}^{p/2}.$$

It is seen that the sample sizes need not be large for (2.7) to be applicable, as was case in the sampling theory approach, and (2.7) is flexible enough that a classification problem can be tailored to a broad range of prior assumptions. Some of them will be presented in Section 3. But the number of hyperparameters in (2.7) is large and unknown so that we can not apply conventional introspection techniques to assess them. Alternatively, we adopt an empirical Bayesian approach and use the data to estimate the hyperparameters.

3. Empirical Bayes Posterior Odds Ratios

We have used the Bayesian assumption(2.1), $\theta_i|\Sigma_i\sim N(\phi_i,\Sigma_i/b_i)$ and $\Sigma_i\sim W^{-1}(\Omega_i,p,m_i)$, and derived a posterior odds ratio(2.7), as a convenient mathematical tool to classify a new observation Z into Π_i, $i=1,\cdots,k$. One can take the assumption about $\theta_i|\Sigma_i$ and Σ_i seriously, but stop short of full Bayesianhood by assuming that the hyperparameters are unknown and must be estimated from the data. That is, one can be an empirical Bayesian. Berger(1985) makes a cogent argument for this point of view. In this section we shall discuss three different empirical Bayes posterior odds ratios for varying Bayesian assumptions, each appropriate for temporal imperative, on $\theta_i|\Sigma_i$ and Σ_i.

Case A. General, but the coordinates of θ are similar

This is a popular Bayesian assumption(Press and Rolph 1980 ; Efron and Morris 1982) such that

$$\theta_i|\Sigma_i\sim N(\tau_i e, \Sigma_i/b_i),\quad \Sigma_i\sim W^{-1}(\Omega_i, p, m_i),\quad i=1,2,\cdots,k,$$

where $e=(1,1,\cdots,1)'$.

Under this assumption the posterior odds ratio C_{ij} in(2.7) becomes

$$\frac{P_A(\Pi_i|Z,q)}{P_A(\Pi_j|Z,q)}=C_{ij}\Big|_{\Phi_\beta=\tau_\beta e},\quad \beta=i,j. \tag{3.1}$$

Press and Rolph(1980) suggests, adopting an empirical Bayesian approach, moment estimators for the hyperparameters

$$\hat{\tau}_i=e'\bar{X}_i/p,\quad \hat{b}_i=\{(\bar{X}_i-\hat{\tau}_i e)'\ (\bar{X}_i-\hat{\tau}_i e)/\mathrm{tr}(V_i/N_i)\ -1/N_i\}^{-1}$$

and

$$\hat{\Omega}_i=(m_i-2p-2)V_i/(N_i-1).$$

Finally an empirical Bayes posterior odds ratio can be obtained by substituting these estimators into (3.1) to give

$$\frac{\hat{P}_A(\Pi_i|Z,q)}{\hat{P}_A(\Pi_j|Z,q)}=C_{ij}\Big|_{\phi_\beta=\hat{\tau}_\beta e,\ b_\beta=\hat{b}_\beta,\ \Omega_\beta=\hat{\Omega}_\beta},\quad \beta=i,j. \tag{3.2}$$

Case B. General, but the coordinates of θ are thought to be close to $O: p\times 1$

This is the case where Bayesian assumption(Kim 1986 ; Efron and Morris 1976 ; Stein 1962) is

$$\theta_i|\Sigma_i\sim N(O, \Sigma_i/b_i),\quad \Sigma_i\sim W^{-1}(\Omega_i, p, m_i),\quad i=1,2,\cdots,k.$$

Hence,

$$\frac{P_B(\Pi_i|Z,q)}{P_B(\Pi_j|Z,q)}=C_{ij}\Big|_{\phi_\beta=O},\quad \beta=i,j, \tag{3.3}$$

where C_{ij} is the odds ratio in(2.7).

Using the empirical Bayes approach, Kim(1986) obtained estimators for the hyperparameters

$$\hat{b}_i=\{(N_i-1)(N_i+m_i-p-1)\bar{X}_i' V_i^{-1}\bar{X}_i/(p-2)(N_i+m_i-2p-2)\ -1/N_i\}^{-1},$$

$$\hat{\Omega}_i=(m_i-2p-2)V_i/(N_i-1).$$

It follows that, under Case B, an empirical Bayes posterior odds ratio is

$$\frac{\hat{P}_B(\Pi_i|Z,q)}{\hat{P}_B(\Pi_j|Z,q)}=C_{ij}\Big|_{\phi_\beta=O,\ b_\beta=\hat{b}_\beta,\ \Omega_\beta=\hat{\Omega}_\beta},\ \beta=i,j. \tag{3.4}$$

Case C. General, but the coordinates of θ are close to $O: p\times 1$ and $E(\Sigma)$ is intraclass

When we suspect that populations Π_i, $i=1,2,\cdots,k$, have intraclass correlation in nature, Bayesian assumption(Press and Rolph 1980) may be

$$\theta_i|\Sigma_i\sim N(O,\Sigma_i/b_i),\ \ \Sigma_i\sim W^{-1}(\Omega_i,p,m_i),$$

$$\text{where } \Omega_i=a_i^2\{\rho_i ee'+(1-\rho_i)I_p),\ \ i=1,2,\cdots,k.$$

Under this assumption the posterior odds ratio C_{ij} in(2.7) becomes

$$\frac{P_C(\Pi_i|Z,q)}{P_C(\Pi_j|Z,q)}=C_{ij}\Big|_{\phi_\beta=O,\ \Omega_\beta=a_\beta^2\{\rho_\beta ee'+(1-\rho_\beta)I_p\}},\ \beta=i,j. \tag{3.5}$$

Press and Rolph(1980)gives an empirical Bayes estimators for the hyperparameters

$$\hat{b}_i=\{(N_i+m_i-p-1)\bar{X}_i'[N_iV_i/(N_i-1)+\hat{\Omega}_i]^{-1}\bar{X}_i/(p-2)-1/N_i\}^{-1},$$

$$\hat{\Omega}_i=\hat{a}_i^2\{\hat{\rho}_i ee'+(1-\hat{\rho}_i)I_p\},$$

where

$$\hat{a}_i^2=(m_i-2p-2)\mathrm{tr}(V_i)/[p(N_i-1)],\ \ V_i=\{v_{\gamma\delta}\},$$

$$\hat{\rho}_i=\{(2p-1)\sum_{\gamma,\delta=1}^{p} v_{\gamma\delta}/p-p\mathrm{tr}(V_i)\}/\{\sum_{\gamma,\delta=1}^{p} v_{\gamma\delta}-\mathrm{tr}(V_i)\}.$$

An empirical Bayes posterior odds ratio can be obtained by substituting these estimators into(3.5) to give

$$\frac{\hat{P}_C(\Pi_i|Z,q)}{\hat{P}_C(\Pi_j|Z,q)}=C_{ij}\Big|_{\phi_\beta=O,\ b_\beta=\hat{b}_\beta,\ \Omega_\beta=\hat{\Omega}_\beta},\ \beta=i,j. \tag{3.6}$$

It would seem reasonable and desirable to use empirical Bayes posterior odds ratio which has property

$$\lim_{N_i,N_j\to\infty}\frac{\hat{P}(\Pi_i|Z,q)}{\hat{P}(\Pi_j|Z,q)}=\frac{P(\Pi_i|Z,q)}{P(\Pi_j|Z,q)}. \tag{3.7}$$

Since the empirical Bayes estimators for the hyperparameters are obtained by the method of moments, it can easily be shown that the empirical Bayes posterior odds ratios, derived for the Cases A, B, and C, exhibit this property. —see, for example, Kim (1986).

4. A Monte Carlo Study

In this section we compare by Monte Carlo methods for performance of two posterior odds ratios in classifiying individuals into two multivariate normally distributed populations when covariance matrices are unequal—Geisser's posterior odds ratio and Empirical posterior odds ratio for the case B.

The comparision is carried out using generated pairs of samples from two multivariate normal populations. We split the total sample into a training sample and a validation sample. Parameters that are varied in the study include the number of dimension(p), mean vectors, covariance matrices, and training sample size(N_i) under the assumption that prior probabilities of origin from the populations are same and equal to 0.5.

As a criterion for evaluating the performance of the two posterior odds ratios, we have chosen the relative estimated mean of log posterior odds ratio which is, for moderate training sample size, a nearly unbiased estimate of the expected relative posterior odds ratio($RPOR$).

$$RPOR=\frac{\log(\text{empirical Bayes posterior odds ratio of } Z \text{ classified into correct population})}{\log(\text{Geisser's posterior odds ratio of } Z \text{ classified into correct population})}, \quad (4.1)$$

$$\hat{E}[RPOR]=\log(M1)/\log(M2), \quad (4.2)$$

where $M1$=mean of empirical Bayes posterior odds of validation sample to classify into correct population.

$M2$=mean of Geisser's posterior odds ratios of validation sample to classify into correct population.

We used a parameter D as a measure of the degree of separation of two populations, where

$$D=(\theta_1-\theta_2)'\left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1}(\theta_1-\theta_2). \quad (4.3)$$

When $\Sigma_1=\Sigma_2$, D is the Maharanobis distance between the two populations.

For each set of values of $p, N_1, N_2, \theta_1, \theta_2, \Sigma_1$, and Σ_2 our program generated pairs of training samples from $N(\theta_1, \Sigma_1)$ and $N(\theta_2, \Sigma_2)$, and formed the two posterior odds ratios. In addition, 100 "new" observations(validation sample) from the two populations were used to calculate the desired odds ratios for the classification. Here we took $m=2p+3$ to reflect minimal prior knowledge but to permit $E[V]$ to exist. Relative mean of

Table 1. Estimated Expected Relative Log Posterior Odds Ratio

training sample size(N) / distance(D)		log posterior odds ratio								[$RPOR$]			
		Empirical				Geisser							
dimension ($p<N$)		5	15	30	50	5	15	30	50	5	15	30	50
3	.239	9.01	26.15	34.08	54.44	5.89	27.30	38.76	54.08	1.592	0.956*	0.879*	1.006
	2.640	8.94	25.64	46.63	61.69	8.08	22.60	35.56	56.57	1.106	1.134	1.275	1.09
	7.900	10.58	32.02	57.78	68.01	9.83	45.15	49.76	62.61	1.076	0.709*	1.160	1.09
5	14.880		58.38	99.78	145.86		43.52	86.88	136.43		1.341	1.148	1.069
	280.822		83.32	146.85	208.91		60.22	124.05	144.37		1.357	1.183	1.447
	885.922		107.66	173.10	251.68		70.95	144.37	226.83		1.517	1.198	1.110
7	2.944		56.27	72.49	112.51		50.92	80.88	122.65		1.104	0.896*	0.917*
	42.400		84.66	124.85	186.51		63.13	114.76	177.09		1.341	1.092	1.053
	157.939		100.32	148.63	252.28		71.77	131.19	223.00		1.397	1.132	1.131
10	.818		73.94	207.69	302.22		57.76	166.99	274.60		1.280	1.243	1.100
	36.084		80.46	227.78	332.55		66.88	167.66	278.92		1.201	1.358	1.190
	128.780		86.57	230.34	336.56		73.47	168.32	290.00		1.178	1.368	1.160

max $\hat{E}[AER]=.09$ (AER : Actual error rate)

Note "*" denotes the case where $\hat{E}$ [$PROR$]<1

these two ratios was used to estimate expected $RPOR$. Each estimate of expected $RPOR$s was based on 2×100 posterior odds ratios for each set of values of parameters. These results are summerized in Table 1.

Table 1 is the list of estimated expected relative log posterior odds ratio for each set of values of $p, N_1=N_2=N$, and D. The reader should keep in mind that the larger $\hat{E}[RPOR](>1)$, the better the performance of the Empirical Bayes posterior odds ratio relative to the Geisser's posterior odds ratio. Some of $\hat{E}[RPOR]$'s have been starred(*), indicating opposite performance. For the starred cases their conditions have small distance measure(D) than the others, and so we can conjecture that the starred cases occur mainly due to indistinguishable generated pairs of samples. Hence, they can be ignored.

Table 1 shows that, for large $p(=5,7,10)$ and small $N(=15)$, the empirical Bayes posterior odds ratios are at their best performance. This consists with a characteristic of the empirical Bayes estimator.

5. Conclusion

The area of heteroscadastic classification analysis still contains many issues that are not yet fully resolved; these include the question of whether one classification rule will produce better classifications than the others. To solve this problem we need to derive the distribution required to use a discriminant function. From the sampling theory viewpoint it is quite complicated in small samples. Bayesian approaches to the classification problem in the case of normally distributed observations are extremely simple to apply and there is not complicated distribution theory.

In this article we have derived a general posterior odds ratio for classification rule using the natural conjugate prior of the population parameters. To assess the hyperparameters involved in the classification rule, we suggest empirical Bayes estimators which fit to several assumptions about natural conjugate priors. The resulting posterior odds ratio is called the empirical Bayes posterior odds ratio. We compared by a Monte Carlo study the performance of it with that of the other Bayesian approach(the Geisser's posterior odds ratio). It is not possible to make specific recommendation about the accurate use for classification analysis from the limited Monte Carlo study. Some generalizations are evident, however. The performance of the empirical Bayes posterior odds ratio, in general, is better than the Geisser's posterior odds ratio. Especially, for large dimension(p) and small training sample size(N), the performance of the empirical Bayes posterior odds ratio relative to the Geisser's posterior odds ratio is prominent.

Acknowledgement

The detailed and constructive comments of referees were very useful in making the paper more readable and in correcting some errors. Their efforts are greatly appreciated.

References

(1) Anderson, T.W. (1951), Classification by Multivariate Analysis, *Psychometrika*, Vol.16, 31~50.

(2) Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Varlag, Inc.

(3) Dunn, O.J. (1971), Some Expected Value for Probabilities of Correct Classification in Dircriminant Analysis, *Technometrics*, 13, 345~53.

(4) Dunn, O.J. and Mark, S. (1974), Discriminant Functions When Covariance Matrices are Unequal, *Journal of the American Statistical Association*, Vol.69, No.346, 555~59.

(5) Dunsmore, I.R. (1966), A Bayesian Approach to Classification, *Journal of the Royal Statisti-*

cal, B, Vol. 28, 568~77.

(6) Efron, B and Morris, C. (1972), Empirical Bayes on Vector observations; An Extension of Steins Method, *Biometrika*, 59, 2, 335~347.

(7) Efron, B and Morris, C. (1976), Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution, *The Annals of Statistics*, Vol. 4, No. 1, 33~50.

(8) Geisser, S. (1964), Posterior Odds for Multivariate Normal Classifications, *Journal of the Royal Statistical*, Vol. 36, 150~59.

(9) Gilbert, E. S. (1969), The Effect of Unequal Variance Covariance Matrices on Fisher's Linear Discriminant Function, *Biometrika*, 25, 505~15.

(10) Kim, H. J. (1986), An Empirical Bayes Estimation of Multivariate Normal Mean Vector, *Journal of the Korean Statistical Society*, Vol. 15, No. 2, 97~106.

(11) Lachenbruch, P. A., Sneeringer, C. and Revo, L. T. (1973), Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Nonnormality, *Communications in Statistics*, 1, No. 1, 39~56.

(12) Press, S. J. and Rolph, J. E. (1980), Empirical Bayes Estimation of the Mean in a Multivariate Distribution, *Technical Report*, No. 70, University of California, Riverside.

(13) Sitgreaves, R, (1952), On the Distribution of Two Random Matrices Used in Classification Procedures, *The Annals of Statistics*, Vol. 23, 74~81.

(14) Stein, C. (1962), Confidence Sets for the Mean of a Multivariate Normal Distribution, *Journal of the Royal Statistical Society, Series*, B, Vol. 24, 265~85.