Journal of the Korean
Statistical Society
Vol. 25, No. 2, 1996

On an Information Theoretic Diagnostic Measure for Detecting Influential Observations in LDA

Hea-Jung Kim [1]

ABSTRACT

This paper suggests a new diagnostic measure for detecting influential observations in two group linear discriminant analysis(LDA). It is developed from an information theoretic point of view using the minimum discrimination information(MDI) methodology. MDI estimator of symmetric divergence by Kullback(1967) is taken as a measure of the power of discrimination in LDA. It is shown that the effect of an observation over the power of discrimination is fully explained by the diagnostic measure. Asymptotic distribution of the proposed measure is derived as a function of independent chi-squared and standard normal variables. By means of the distributions, a couple of methods are suggested for detecting the influential observations in LDA. Performance of the suggested methods are examined through a simulation study.

Key Words: Power of discrimination; Influential observations; MDI methodology; Diagnostic measure; Asymptotic distribution.

[1] Department of Statistics, Dongguk University, Seoul, 100-715, Korea

1. INTRODUCTION

Recently, Critchley and Vitiello(1991) and Fung(1992) independently proposed two fundamental statistics, say d_{ij}^2 and $\hat{\psi}_{ij}$, in Fisher's linear discriminant analysis(LDA), like the residual and leverage measure in regression, on which many influence measures depend. By means of the fundamental statistics, Critchley and Vitiello(1991) examined the influence of observations upon misclassification probability estimates in LDA and Fung(1995) suggested a couple of Cook's type (Cook, 1977) diagnostic measures for detecting outliers. In addition to the studies, many articles have been published on detecting outliers and influential observations in LDA. See, for example and for further references, Campbell(1978), Radhakrishnan(1983) and Johnson(1987).

The studies mentioned above are based on either sampling theory approach or Bayesian framework. The present paper considers, however, an information theoretic approach for detecting the influential observations. Kullback(1967) and Kim(1995, 1996) discussed using the information theoretic approach in discriminant analysis. A brief review of the use of symmetric divergence by Kullback(1967) as a measure of the power of discrimination in LDA is given in Section 2. Based upon the power of discrimination, Section 3 proposes a diagnostic measure for detecting the influential observations which can be expressed in terms of the fundamental statistics, d_{ij}^2 and $\hat{\psi}_{ij}$. In Section 4 an asymptotic distribution of the diagnostic measure is derived so that one may construct critical values and expected quantiles of the measure. In Section 5 the performance of the proposed measure is examined through a simulation study. A few concluding remarks are given in Section 6.

2. POWER OF DISCRIMINATION

Suppose we have two p-variate normal populations $\Pi_i \sim N_p(\mu_i, \Sigma)$, with μ_i, $i = 1, 2$, the p mean vector and Σ, the common covariance matrix. Denoting the respective population densities by

$$f_i(x) \;=\; |2\pi\Sigma|^{-1/2} exp\{-\frac{1}{2}(x-\mu_i)'\Sigma^{-1}(x-\mu_i)\},$$

we find the Kullback-Liebler cross entropy between the two populations(cf.

Kullback, 1967), Π_1 and Π_2, as

$$\begin{aligned} I(1:2) &= \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx \\ &= \frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2). \end{aligned} \tag{2.1}$$

As suggested in Kapur and Kesavan(1992), to measure the power of discrimination between the two multivariate normal populations with the densities $f_1(x)$ and $f_2(x)$, one can use the measure of symmetric divergence by Kullback(1967):

$$J(1:2) = I(1:2) + I(2:1) = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2). \tag{2.2}$$

Definition 1. Let $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$ be distributions associated with the p vector random variable X from the populations Π_1 and Π_2, respectively. Then the power of discrimination between the two populations is

$$J(1:2) = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2).$$

Let us consider a transformation $Y = \alpha' X$. Then the power of discrimination based on Y is given by

$$\begin{aligned} J(1,2:y) &= \int \phi_1(y) \ln \frac{\phi_1(y)}{\phi_2(y)} dy + \int \phi_2(y) \ln \frac{\phi_2(y)}{\phi_1(y)} dy \\ &= \alpha'\delta\delta'\alpha / \alpha'\Sigma\alpha, \end{aligned} \tag{2.3}$$

where $\delta = (\mu_1 - \mu_2)$ and $\phi_i(y)$ is the pdf of $N(\alpha'\mu_i, \alpha'\Sigma\alpha)$ distribution, $i = 1, 2$.

Lemma 1. Fisher's linear discriminant function(LDF), *i.e.* the transformation $Y = (\mu_1 - \mu_2)'\Sigma^{-1}X$, is sufficient in a sense that the transformation has no effect on the power of discrimination, i.e.

$$J(1,2:y) = J(1:2) = \delta'\Sigma^{-1}\delta.$$

Proof. It can be shown from the Cauchy-Schwarz inequality that

$$J(1,2:y) = \frac{\alpha'\delta\delta'\alpha}{\alpha\Sigma\alpha} \leq J(1:2) = \delta'\Sigma^{-1}\delta,$$

where the equality holds for $\alpha = \Sigma^{-1}\delta$. Therefore the transformation $Y = \delta'\Sigma^{-1}X$ leads to $J(1,2:y) = J(1:2) = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$.

Above lemma notes that, with regard to the power of discrimination, LDF is the best linear discriminant function of random vector X associated with Π_1 and Π_2. Since $J(1,2:y)$ is additive for independent random variables, we have for a random sample of n observations, O_n,

$$J(1,2:y,O_n) = nJ(1,2:y) = n\delta'\Sigma^{-1}\delta. \tag{2.4}$$

When the parameters μ_1, μ_2 and Σ are unknown, we take the conjugate distributions of Π_1 and Π_2 with parameters the same as the respective observed best unbiased sample estimates of the unknown parameters(cf. Kullback, 1967). This leads to so called minimum discrimination information(MDI) estimator of LDF and that of the power of discrimination. Following definition summarizes the MDI estimators.

Definition 2. Suppose we have n_1 observations, $\{X_{1j}\}$, $j = 1,\ldots,n_1$, of the multivariate normal variate $X' = (x_1,\ldots,x_p)$ from $\Pi_1 \sim N_p(\mu_1,\Sigma)$ and n_2 measurements, $\{X_{2j}\}$, $j = 1,\ldots,n_2$, of this quantity from $\Pi_2 \sim N_p(\mu_2,\Sigma)$ with $n \geq p+2$; $n = n_1 + n_2$. Then MDI estimator of LDF is

$$Y = \hat{\alpha}'X = (\bar{X}_1 - \bar{X}_2)'S^{-1}X,$$

and that of the power of discrimination obtained from LDF is

$$n\hat{J}(1,2:y) = n(\bar{X}_1 - \bar{X}_2)'S^{-1}(\bar{X}_1 - \bar{X}_2), \tag{2.5}$$

where $\bar{X}_1$ and $\bar{X}_2$ are respective sample mean vectors of Π_1 and Π_2, and S denotes the pooled sample covariance matrix.

It is shown that the estimated linear discriminant function obtained by maximizing the power of discrimination is equivalent to that of Fisher's LDF. Therefore, a diagnostic measure for the linear discriminant analysis suggested in the sequel via information theoretic approach(using the power of discrimination) may as well apply for Fisher's LDF.

3. DIAGNOSTIC MEASURE

If $X_{ij} \in R^p$ denotes an observation from Π_i, the well-known Fisher's(sample) LDF is to allocate an observation X_{ij} to Π_1 if the prior quantity $\log(q_1/q_2)$, say, plus the linear discriminant score is

$$(\bar{X}_1 - \bar{X}_2)'S^{-1}\{X_{ij} - \frac{1}{2}(\bar{X}_1 + \bar{X}_2)\} = \frac{1}{2}\hat{d}^2_{2j} - \frac{1}{2}\hat{d}^2_{1j} > 0,$$

where $\hat{d}^2_{kj} = (X_{ij} - \bar{X}_k)'S^{-1}(Z - \bar{X}_k)$, $k = 1, 2$. When $i = k$, $\hat{d}^2_{ij}$ is interpreted as the atypicality of observation X_{ij} from population Π_i. Another statistic that is the difference between the discriminant score of X_{ij} and that of $\bar{X}_i$, say residual of X_{ij} is given by $\hat{\psi}_{ij} = (\bar{X}_1 - \bar{X}_2)'S^{-1}(X_{ij} - \bar{X}_i)$. These two statistics, $\hat{\psi}_{ij}$ and $\hat{d}^2_{ij}$, are said to be the two fundamental statistics in LDF(see, Fung(1992) and Critchley and Vitiello(1991)). In this section we will derive an information theoretic diagnostic measure that can be written in terms of the two fundamental statistics.

Suppose we are interested in the effect of the omission of an observation j from Π_1(or Π_2) on the power of discrimination. One may study it through the estimated loss in powers of discrimination due to the omission of an observation j from Π_i, $i = 1, 2$:

$$Loss_{ij} = n\hat{J}(1, 2 : y) - (n-1)\hat{J}_{ij}(1, 2 : y), \; i = 1, 2, \; j = 1, \ldots, n_i, \quad (3.1)$$

where $(n-1)\hat{J}_{ij}(1, 2 : y)$ is the estimated power of discrimination when an observation j is omitted from the training sample of Π_i.

Theorem 1. In two group discriminant analysis, effect in $Loss_{ij}$ due to omission of an observation j, say X_{ij}, from $\Pi_i \sim N_p(\mu_i, \Sigma)$, $i = 1, 2$; $j = 1, \ldots n_i$, is determined mainly by

$$M_{ij} = \frac{\left(\hat{\psi}_{ij} - \lambda_i^{-1}\right)^2}{(n_i - 1)\lambda_i^{-1} - \hat{d}^2_{ij}}, \quad (3.2)$$

where $\lambda_i = n_i/(n_1 + n_2 - 2)$, and $\hat{\psi}_{ij} = (\bar{X}_1 - \bar{X}_2)'S^{-1}(X_{ij} - \bar{X}_i)$ and $\hat{d}^2_{ij} = (X_{ij} - \bar{X}_i)'S^{-1}(X_{ij} - \bar{X}_i)$ are the two fundamental statistics.

Proof. Definition 2 notes that, when we use LDF, the estimated power of discrimination between Π_1 and Π_2 is given by

$$n\hat{J}(1, 2 : y) = n(\bar{X}_1 - \bar{X}_2)'S^{-1}(\bar{X}_1 - \bar{X}_2). \quad (3.3)$$

Assume, for convenience, that observation j is deleted from training sample of Π_1. Then $\bar{X}_2$ and V_2 are unaffected while $\bar{X}_1$ and V_1 change to $\bar{X}_1 - (n_1 - 1)^{-1}\Delta_1$ and $V_1 - \{1 + (n_1 - 1)^{-1}\}\Delta_1\Delta_1'$, where $\Delta_1 = (X_{1j} - \bar{X}_1)$, and V_1 and V_2 are sample corrected sums of squares and cross-product matrices so that $S = V/(n-2)$; $V = V_1 + V_2$, $n = n_1 + n_2$. Let $(n-1)\hat{J}_{1j}(1,2:y)$ be the estimated power of discrimination when an observation j is omitted from the training sample of Π_1, then

$$\begin{aligned}
\hat{J}_{1j}(1,2:y) &= (n-3)\hat{\delta}_{1j}'(V - n_1\Delta_1\Delta_1'/(n_1-1))^{-1}\hat{\delta}_{1j} \\
&= \frac{n-3}{n-2}\hat{\delta}_{1j}'\left(S^{-1} + \frac{\lambda_1 S^{-1}\Delta_1\Delta_1' S^{-1}/(n_1-1)}{1 - \lambda_1\Delta_1' S^{-1}\Delta_1/(n_1-1)}\right)\hat{\delta}_{1j} \\
&= \frac{n-3}{n-2}\left\{\hat{J}(1,2:y) - \frac{1}{\lambda_1(n_1-1)} + \frac{(\hat{\psi}_{1j} - \lambda_1^{-1})^2}{(n_1-1)\lambda_1^{-1} - \hat{d}_{1j}^2}\right\},
\end{aligned}$$

where $\hat{\delta}_{1j} = (\bar{X}_1 - \bar{X}_2 - \Delta_1/(n_1-1))$. Thus $Loss_{1j}$ is given by

$$\frac{(n-3)(n-1)}{(n-2)}\left\{\frac{D^2(2n-3)}{(n-3)(n-1)} + \frac{1}{\lambda_1(n_1-1)} - \frac{\left(\hat{\psi}_{1j} - \lambda_1^{-1}\right)^2}{(n_1-1)\lambda_1^{-1} - \hat{d}_{1j}^2}\right\}, \quad (3.4)$$

where $D^2 = (\bar{X}_1 - \bar{X}_2)'S^{-1}(\bar{X}_1 - \bar{X}_2)$.

Therefore, the deletion of X_{1j} effects $Loss_{1j}$ only through the last term of (3.4). Similar proof holds for the deletion of an observation j from Π_2.

Diagnostic measure M_{ij} in (3.2) shows that $O(n^{-1})$ effect of $\hat{\psi}_{ij}$ dominates the $O(n^{-2})$ effect of $\hat{d}_{ij}^2$, and that $\hat{\psi}_{ij}$ can vary independently of $\hat{d}_{ij}^2$. Therefore, Theorem 1 notes that, to detect influential observations over the power of discrimination, we need examination of M_{ij} rather than separate examinations of the two fundamental statistics. Theorem 1 also tells us that, other things being equal, the improvement in the ratio upon deleting an observation increases (i) with larger value of $|\hat{\psi}_{ij} - \lambda_i^{-1}|$, and (ii) with increasing $\hat{d}_{ij}^2$.

Noticing that the difference in the power of discrimination between before and after omitting X_{ij} is $n\hat{J}(1,2:y) - (n-1)\hat{J}_{ij}(1,2:y)$, we have following result.

Corollary 1. $M_{ij} > 0$ for all $i = 1,2,\ j = 1,\ldots,n_i$.

Proof. It can be easily seen that positive definiteness of the pooled sample covariance matrix under such perturbation leads to $\hat{d}_{ij}^2 < \lambda_i^{-1}(n_i - 1),\ i =$

$1, 2;\ j = 1, \ldots n_i$, because $|S - \lambda_i \Delta_i \Delta_i'/(n_i - 1)| = |S|(1 - \lambda_i \Delta_i' S^{-1} \Delta_i/(n_i - 1)) > 0$. This gives the result.

Corollary 2. Gain in the power of discrimination occur when the deletion of X_{ij} results in

$$M_{ij} > \frac{1}{\lambda_i(n_i - 1)} + \frac{D^2(2n - 3)}{(n - 1)(n - 3)}, \quad i = 1, 2,\ j = 1, \ldots, n_i. \tag{3.5}$$

Proof. The condition $n\hat{J}(1, 2 : y) - (n - 1)\hat{J}_{ij}(1, 2 : y) < 0$ and Corollary 1 give the result.

The inequality (3.5) indicates that the lower bound of M_{ij} can be used as a criterion for detecting influential observations upon power of discrimination in LDF.

Corollary 3. The diagnostic measure M_{ij} is invariant with respect to linear transformations.

Proof. Let $W_{ij} = AX_{ij} + b, ; i = 1, 2; j = 1, \ldots, n_i$, for some nonsingular $A : p \times p,\ b : p \times 1$. Then $W_{ij} \sim N_p(A\mu_i + b, A\Sigma A')$. Hence, if $\hat{d}_{ij}^2$ and $\hat{\psi}_{ij}$ based on $W_{ij}'s$ are respectively denoted by

$$\hat{d}_{ij}^2(W) = (W_{ij} - \bar{W}_i)' S_W^{-1} (W_{ij} - \bar{W}_i) = (AX_{ij} - A\bar{X}_i)'(ASA')^{-1}(AX_{ij} - A\bar{X}_i) = \hat{d}_{ij}^2$$

and

$$\hat{\psi}_{ij}(W) = (\bar{W}_1 - \bar{W}_2)' S_W^{-1} (W_{ij} - \bar{W}_i) = (A\bar{X}_1 - A\bar{X}_2)'(ASA')^{-1}(AX_{ij} - A\bar{X}_i) = \hat{\psi}_{ij}.$$

Thus M_{ij} which is a function of the two invariant statistics is invariant.

It is well known that Mahalanobis generalized distance D^2 is invariant with respect to the linear transformations. Thus, it is straight-forward to show the invariance of the estimated loss in powers of discrimination defined in (3.1).

4. ASYMPTOTIC DISTRIBUTION OF THE DIAGNOSTIC MEASURE

It is clear from (3.2) that the proposed measure is a function of the two fundamental statistics, $\hat{\psi}_{ij}$ and $\hat{d}_{ij}^2$. Thus the following lemma by Fung(1995) is useful for deriving the asymptotic distribution for the measure M_{ij}.

Lemma 2(Fung, 1995). When an observation X_{ij} is omitted from $\Pi_i \sim N_p(\mu_i, \Sigma)$, the statistics $\hat{d}_{ij}^2$ and $\hat{\psi}_{ij}/D$ are asymptotically distributed as

$$\hat{d}_{ij}^2 = U + Z^2 \text{ and } \hat{\psi}_{ij}/D = Z,$$

where U and Z are independent and are distributed as χ^2_{p-1} and $N(0,1)$, respectively, $i = 1, 2$, $j = 1, \ldots, n_i$, and D^2 converges almost surely to $(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$.

Based upon the asymptotic results of the above lemma, one could obtain approximate expected quantiles for quantile-quantile(Q-Q) plot for the diagnostic measure.

Theorem 2. If an observation X_{ij} in M_{ij} is from $\Pi_i \sim N_p(\mu_i, \Sigma)$, the quantiles π_k of M_{ij} $k = 1, \ldots, n_i$, $i = 1, 2$, can then be approximately obtained by solving

$$E\left[\Phi\left\{\frac{a_i}{D^2 + \pi_k} + q(n_i, U, \pi_k)^{1/2}\right\} - \Phi\left\{\frac{a_i}{D^2 + \pi_k} - q(n_i, U, \pi_k)^{1/2}\right\}\right] = \frac{(k - 1/2)}{n_i}, \tag{4.1}$$

where $a_i = \lambda_i^{-1} D$, $q(n_i, U, \pi_k) = \pi_k\{\lambda_i^{-1}(n_i - 1) - U - \lambda_i^{-2}/(D^2 + \pi_k)\}/(D^2 + \pi_k)$, the expectation is with respect to U with distribution χ^2_{p-1}, and $\Phi\{\cdot\}$ denotes the df of $N(0,1)$.

Proof. Lemma 2 notes that the asymptotic distribution of M_{ij} is the same as the distribution of function of the first and second power of U and Z, i.e.

$$M_{ij} \approx \frac{(DZ - \lambda_i^{-1})^2}{\lambda_i^{-1}(n_i - 1) - (U + Z^2)}. \tag{4.2}$$

Thus the quantiles π_k for M_{ij} are evaluated as

$$P(M_{ij} \leq \pi_k) \approx P(|Z - a_i/(D^2 + \pi_k)| \leq q(n_i, U, \pi_k)^{1/2}) = (k - 1/2)/n_i.$$

This gives the result.

Expressing (4.1) through the integral with respect to the densities of U and Z, we have

$$\int_0^\infty \int_{a_i/(D^2+\pi_k)-q(n_i,u,\pi_k)^{1/2}}^{a_i/(D^2+\pi_k)+q(n_i,u,\pi_k)^{1/2}} \phi(z)dzf(u)du = \frac{(k - 1/2)}{n_i}, \tag{4.3}$$

where $q(n_i, u, \pi_k) = \pi_k\{\lambda_i^{-1}(n_i - 1) - u - \lambda_i^{-2}/(D^2 + \pi_k)\}/(D^2 + \pi_k)$, and $f(u)$ and $\phi(z)$ are respective pdf's of χ^2_{p-1} and $N(0, 1)$. The expected quantiles π_k can be evaluated through a numerical integration using Simpson's rule. The observed quantiles for M_{ij} can be plotted against the expected quantiles for detection of influential observations over the power of discrimination. Moreover, using the Bonferroni inequality, we may approximate the $100\alpha\%$-critical value $C_\alpha(i)$ for the maximum value of M_{ij} over i-th training sample points as

$$p(M_{ij} \leq C_\alpha(i)) \;=\; 1 - \alpha/n_i, \;\; i = 1, 2. \tag{4.4}$$

$C_\alpha(i)$ can be also evaluated by the Simpson's rule.

5. PERFORMANCE OF THE MEASURE

In the previous section, we have derived two devices for detecting influential observations based on M_{ij}; (i) the approximate expected quantiles (4.1) for Q-Q plot of M_{ij}; (ii) the approximate critical value for the maximum value of the measure in (4.4). It is shown that performance of the two devices are directly related to the accuracy of the approximation in (4.1). The accuracy is examined through a simulation study to see whether the use of the two devices are adequate for detecting influential observations over the power of discrimination in linear discriminant analysis.

Corollary 3 notes that M_{ij} is invariant with respect to a linear transformation. Thus, without loss of generality, our simulation generates, $n_1 = n_2 = 50$ independent observations from each population; $\Pi_1 \sim N_p(0, I)$ and $\Pi_2 \sim N_p(\delta, I)$, where $\delta = \text{diag}\{\delta_1, \ldots, \delta_p\}$. For each set of parameter values of $\{p, \delta_1, \ldots, \delta_p\}$, the M_{ij} values are obtained and ordered for each repetition. It can be easily seen that, in case $n_1 = n_2$, the asymptotic distribution of M_{ij} is the same for $i = 1, 2$. This is due to symmetry property of M_{ij}. For the examination of (i), the order statistics are averaged over the 10,000 repetitions of the simulation. For each case of the parameter values(Case 1: $p = 3, \delta_\ell = \ell, \ell = 1, 2, 3$, and Case 2: $p = 3, \delta_\ell = (-1)^\ell, \ell = 1, 2, 3$), these expected quantiles are plotted against those obtained using (4.1) in Figure 1 and Figure 2. The figures show that the approximate expected quantiles obtained by (4.1) are remarkably accurate. This confirms us that the Q-Q plot of M_{ij} is a useful exploratory tool for detecting influential observations.

The plots based on other set of parameter values give the same results as Figure 1 and Figure 2, and hence they are exempted from the figures.

For the examination of (ii), the critical values for the maximum values of M_{ij}, $\hat{d}^2_{ij}$ and $\hat{\psi}_{ij}/D$ obtained by the same simulations with 10,000 repetitions and the approximations(exploiting (4.1) and Lemma 2) are given in Table 1.

Table 1. 1%, 5% and 10% critical values for the maximum values of the measures

		Case 1			Case 2		
		M_{ij}	$\hat{d}^2_{ij}$	$(\hat{\psi}_{ij}/D)^2$	M_{ij}	$\hat{d}^2_{ij}$	$(\hat{\psi}_{ij}/D)^2$
1%	Simulation	8.53	19.03	13.98	2.42	19.03	14.21
	Approximation	8.72	19.26	14.09	2.48	19.26	14.09
5%	Simulation	3.49	16.33	11.73	0.94	16.23	11.14
	Approximation	3.31	16.45	11.44	0.92	16.45	11.44
10%	Simulation	2.51	16.62	10.53	0.68	16.03	10.53
	Approximation	2.44	16.23	10.83	0.70	16.24	10.83

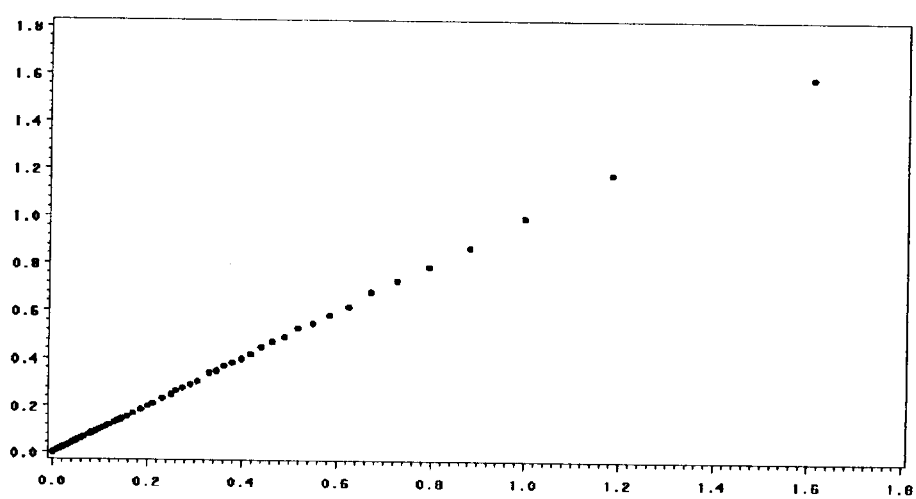

Figure 1. Q-Q plots of M_{ij} for Case 1 {asymptotic(horizontal axis) versus simulation(vertical axis)}.

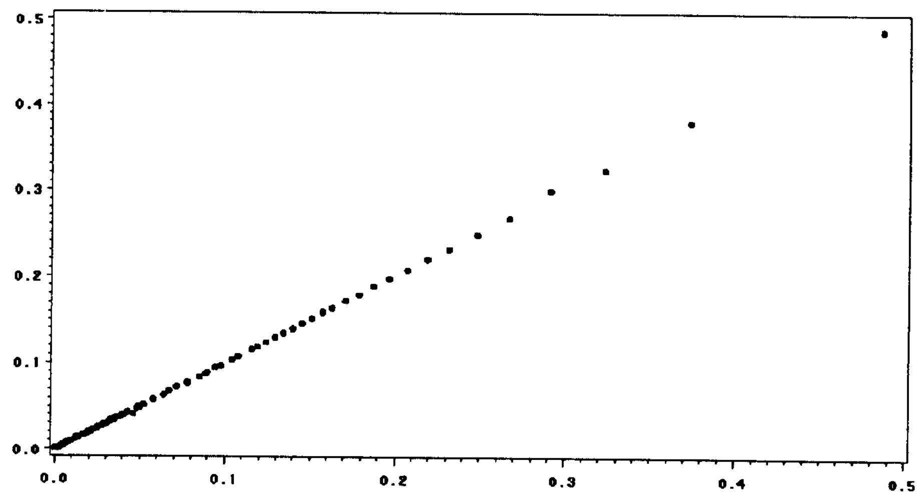

Figure 2. Q-Q plots of M_{ij} for Case 2 {asymptotic(horizontal axis) versus simulation(vertical axis)}.

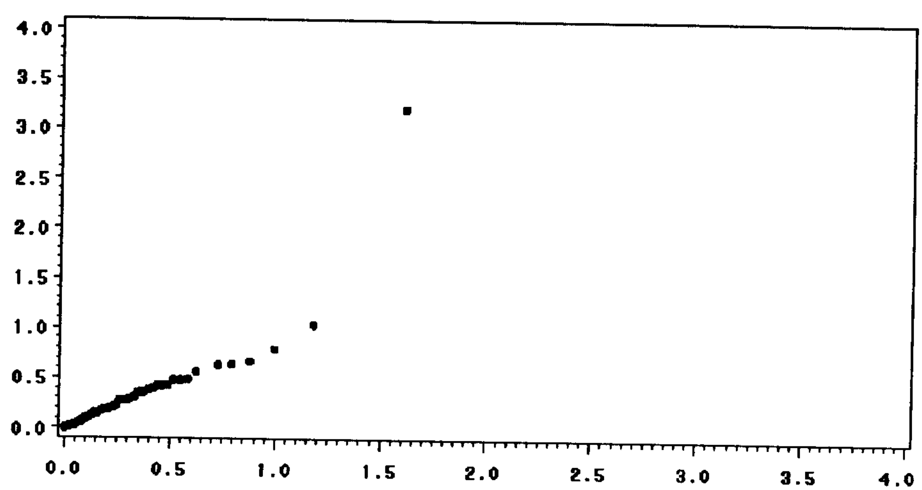

Figure 3. Q-Q plots of M_{ij} for the illustration {Expected(horizontal axis) versus observed(vertical axis)}.

As an illustration, for Case 1, we generated 49 and 51 independent observations from Π_1 and Π_2, respectively. Then we put the first observation(lebeled with index 51) generated from Π_2 into the sample from Π_1, so that we may have $n_1 = n_2 = 50$ observations to obtained the Q-Q plot for M_{ij}.

The Q-Q plot for M_{ij} is given in Figure 3. As expected the observation-lebeled 51-has $M_{1,51} = 3.19787$ and stands out from the others. It seems that the Q-Q plot successfully reveals the influential observation. It is also noted from Table 1 that the suggested measure for observation 51 is larger than the corresponding 10% critical value for maximum M_{ij}.

6. CONCLUDING REMARKS

We proposed a new diagnostic measure for detecting single influential observation in LDA. When we apply the measure sequentially, it could also be useful for identifying multiple influential observations. The measure is developed from an information theoretic point of view using MDI estimator of symmetric divergence by Kullback(1967) that can be taken as a measure of the power of discrimination in LDA. Asymptotic distribution of the proposed measure is shown to be a function of independent chi-squared and standard normal variables. Based on the asymptotic distribution, we proposed a couple of methods(Q-Q plot and critical value of maximum value of the measure) for detecting an observation that deteriorates the power of discrimination in LDA. The simulation studies in Section 5 confirm us that the suggested methods are useful tools for detecting the influential observation.

The proposed measure can be easily extended to detect multiple influential observations in blocks avoiding the masking problem(cf. Rousseeuw and Zomeren, 1990) and to detect influential observations in multiple discriminant analysis. A study pertaining to these problems are left as a further research topic of interest.

REFERENCES

(1) Campbell, N. A. (1978). The influence function as an aid in outlier detection in discriminant analysis. *Applied Statistics*, **27**, 252-258.

(2) Cook, D. R.(1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15-18.

(3) Critchley,F., and Vitiello, C. (1991). The influence of observations on misclassification probability estimates in linear discriminant analysis. *Biometrika*, **78**, 677-690.

(4) Fung, W. K. (1992). Some diagnostic measures in discriminant analysis. *Statistics and Probability Letters*, **13**, 279-285.

(5) Fung, W. K. (1995). Diagnostics in linear discriminant analysis. *Journal of the American Statistical Association*, **90**, 952-956.

(6) Johnson, W. (1987). The detection of influential observations for allocation, separation, and the determination of probabilities in a Bayesian framework. *Journal of Business and Economic Statistics*, **5**, 369-381.

(7) Kapur, J. N., and Kesavan, H. K. (1992). *Entropy Optimization Principles with Applications*, Academic Press, Inc., San Diego.

(8) Kim, H. J. (1995). On a balanced quadratic classification. *Communications in Statistics: Theory and Methods*, **24**, No. 3, 607-623.

(9) Kim, H. J. (1996). On a constrained optimal rule for classification with unknown prior individual group membership. *Journal of Multivariate Analysis*, will be published.

(10) Kullback, S. (1967). *Information Theory and Statistics*, Dover Publications, Inc., New York.

(11) Radhakrishnan, R. (1983). Influence functions for certain parameters in discriminant analysis. *Metron*, XLI-X, 1-2, **30**, 183-194.

(12) Rousseeuw, P. J., and van Zomeren. (1990). Unmasking multivariate outliers and leverage points (with discussion). *Journal of the American Statistical Association*, **85**, 633- 651.