Journal of the Korean Statistical Society (2001), **30**: *3, pp 499–509*

Influence Analysis in Selecting Discriminant Variables

Kang-Mo Jung[1] and Myung Geun Kim[2]

ABSTRACT

We investigate the influence of observations on a test of additional information about discrimination using the influence function and the derivative influence measures. The influence function for the test statistic is derived and its sample versions are used for influence analysis. The derivative influence measures for the test statistic under a perturbation scheme are derived. It will be seen that the influence function method and the derivative influence measures yield the same result. Furthermore, we will derive the relationships between the influence function and the derivative influence measures when the sample size is large. An illustrative example is given and we will compare the results provided by the influence function method and the derivative influence measures.

Keywords: Derivative influence; Diagnostics; Influence function; Perturbation; Selecting discriminant variables.

1. INTRODUCTION

The detection of outliers and influential observations has a long history. However, many diagnostic measures have been proposed in the context of estimation. Influence analysis in hypothesis testing problems is important because even a single observation can dominate our conclusion about hypothesis as can be seen in Section 5. In multivariate analysis a few works investigated the influence of observations on test statistics. Among others, Kim (1995) used the influence function to assess the influence of observations in comparing covariance matrices. Jung and Kim (1999) investigated the influence in multivariate Behrens-Fisher problem using the influence function and the derivative influence.

The influence function (Hampel, 1974) for a parameter at a point measures the effect of an infinitesimal contamination at that point on the estimator of

[1]Department of Informatics and Statistics, Kunsan National University, Kunsan, 573-701, Korea.

[2]Department of Applied Statistics, Seowon University, Chongju, 361-742, Korea.

the parameter. Hence the influence function can serve as a diagnostic method of detecting influential observations in performing a test of hypothesis. The derivative influence (De Gruttola et al., 1987) measures the differential change in an estimated parameter resulting from a slight perturbation in the weight assigned to a given observation.

In this work we will investigate the influence of observations on a test statistic for the additional information about discrimination in linear discriminant analysis using the influence function and the derivative influence. In Section 2 a test for selecting discriminant variables is reviewed. In Section 3 we will derive the influence function for the test statistic and consider three sample versions of the influence function that will be used for investigating the influence of observations on the test statistic. The derivative influence measures for the test statistic are derived in Section 4, under the perturbation scheme in which a weight is put on the covariance matrix for an observation. Some relationships between the influence function and the derivative influence measure are found when the sample size is large. In Section 5 an illustrative example is given and we will compare the results provided by the influence function and the derivative influence.

2. TEST FOR ADDITIONAL INFORMATION

Two independent random samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n_2}\}$ are drawn from p-variate normal distributions $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, respectively. Let $\mathbf{z}$ be a p-variate generic random vector having $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ or $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$. The theoretical maximum likelihood discriminant function is $\boldsymbol{\alpha}^T \boldsymbol{z}$, where $\boldsymbol{\alpha} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$ and $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. We partition $\boldsymbol{\alpha}^T = (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T)$ and $\boldsymbol{\delta}^T = (\boldsymbol{\delta}_1^T, \boldsymbol{\delta}_2^T)$, where $\boldsymbol{\alpha}_1$ and $\boldsymbol{\delta}_1$ have the first $k(< p)$ components.

The hypothesis $\boldsymbol{\alpha}_2 = \mathbf{0}$ implies that the last $p - k$ variables in $\boldsymbol{z}$ do not provide additional information about discrimination. Whenever the hypothesis $\boldsymbol{\alpha}_2 = \mathbf{0}$ holds, the last $p - k$ variables in $\boldsymbol{z}$ can be discarded. We partition $\boldsymbol{\Sigma}$ according to the partition of $\boldsymbol{\delta}$ and denote by $\boldsymbol{\Sigma}_{11}$ the leading principal submatrix of $\boldsymbol{\Sigma}$ corresponding to $\boldsymbol{\delta}_1$. Let $\Delta_p^2 = \boldsymbol{\delta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$ and $\Delta_k^2 = \boldsymbol{\delta}_1^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\delta}_1$. Then the hypothesis $\boldsymbol{\alpha}_2 = \mathbf{0}$ is equivalent to $\Delta_p^2 = \Delta_k^2$ (Mardia et al., 1979).

Let $\bar{\mathbf{x}}$ and $\mathbf{S}_1$ be the maximum likelihood estimators of $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}$ based on the first sample, and let $\bar{\mathbf{y}}$ and $\mathbf{S}_2$ be those of $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ based on the second sample. Let $\mathbf{d} = \bar{\mathbf{x}} - \bar{\mathbf{y}}$ and $\mathbf{S} = (n_1\mathbf{S}_1 + n_2\mathbf{S}_2)/m$, where $m = n_1 + n_2 - 2$. We partition $\mathbf{d}^T = (\mathbf{d}_1^T, \mathbf{d}_2^T)$ and $\mathbf{S}$ according to the partition of $\mathbf{d}$. Let $\mathbf{S}_{11}$ be the leading principal submatrix of $\mathbf{S}$ corresponding to $\boldsymbol{\Sigma}_{11}$. The squared sample

Mahalanobis distances are written as $D_p^2 = \mathbf{d}^T\mathbf{S}^{-1}\mathbf{d}$ and $D_k^2 = \mathbf{d}_1^T\mathbf{S}_{11}^{-1}\mathbf{d}_1$. Rao (1973) suggested a test of the hypothesis $H_0 : \boldsymbol{\alpha}_2 = \mathbf{0}$ based on the test statistic $T_* = (m-p+1)/(p-k)T$, where

$$T = \frac{D_p^2 - D_k^2}{c + D_k^2},$$

where $c = m(m-2)/(n_1 n_2)$. Under the null hypothesis, the test statistic T_* has the F distribution with degrees of freedom $p-k$ and $m-p+1$. A large value of T_* leads to the rejection of the null hypothesis.

3. INFLUENCE FUNCTION

In this section we will derive the influence function for the statistic T and three sample versions of the influence function that will be used for investigating the influence of observations on T.

Let $\theta = \theta(F_1, F_2)$ be a parameter of interest which is a functional of the two distributions F_1 and F_2. It is understood that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1(F_1)$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_2(F_2)$ $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(F_1, F_2)$. Assume that only the first distribution is perturbed as $\tilde{F}_1 = (1-\varepsilon)F_1 + \varepsilon\delta_{\mathbf{x}}$ for $0 \le \varepsilon \le 1$, where $\delta_{\mathbf{x}}$ denotes the distribution having unit mass at $\mathbf{x}$. The perturbation of θ at $\mathbf{x}$ is $\theta(\tilde{F}_1, F_2)$. The influence function for θ at $\mathbf{x}$ (Hampel, 1974) is defined by

$$IF(\theta) = \lim_{\varepsilon \to 0} \frac{\theta(\tilde{F}_1, F_2) - \theta}{\varepsilon}. \tag{3.1}$$

The influence function for a parameter at $\mathbf{x}$ measures the effect of an infinitesimal contamination at $\mathbf{x}$ on the estimator of the parameter.

The performance of the linear discriminant analysis is determined by the squared Mahalanobis distance $\boldsymbol{\Delta}_p^2$. Campbell (1978) derived the influence function for $\boldsymbol{\Delta}_p^2$ as

$$IF(\boldsymbol{\Delta}_p^2) = \lambda_1 \boldsymbol{\Delta}_p^2 + 2\phi_p - \lambda_1 \phi_p^2, \tag{3.2}$$

where $\phi_p = \boldsymbol{\alpha}^T(\mathbf{x} - \boldsymbol{\mu}_1)$ and λ_i $(i = 1, 2)$ is a weight for population i in forming $\boldsymbol{\Sigma}$ with $\lambda_1 + \lambda_2 = 1$.

Consider the functional $\tau = \tau(F_1, F_2)$ defined by

$$\tau(F_1, F_2) = \frac{\boldsymbol{\Delta}_p^2(F_1, F_2) - \boldsymbol{\Delta}_k^2(F_1, F_2)}{c + \boldsymbol{\Delta}_k^2(F_1, F_2)}.$$

Let $\hat{F}_1$ and $\hat{F}_2$ be the empirical distribution functions based on the samples $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_{n_2}\}$, respectively. Then it is clear that $\tau(\hat{F}_1, \hat{F}_2) = T$,

because $\mu_1(\hat{F}_1) = \bar{\mathbf{x}}$, $\mu_2(\hat{F}_2) = \bar{\mathbf{y}}$ and $\mathbf{\Sigma}(\hat{F}_1, \hat{F}_2) = \mathbf{S}$. Thus the functional $\tau(F_1, F_2)$ is well defined. For a functional θ, we have $\theta(\tilde{F}_1, F_2) = \theta + IF(\theta)\varepsilon + O(\varepsilon^2)$. By the Taylor expansion we obtain

$$\frac{1}{c + \mathbf{\Delta}_k^2(\tilde{F}_1, F_2)} = \frac{1}{c + \mathbf{\Delta}_k^2} \left[1 - \frac{1}{c + \mathbf{\Delta}_k^2} IF(\mathbf{\Delta}_k^2)\varepsilon + O(\varepsilon^2)\right].$$

From the above equation we obtain

$$\begin{aligned} \tau(\tilde{F}_1, F_2) &= \frac{\mathbf{\Delta}_p^2(\tilde{F}_1, F_2) - \mathbf{\Delta}_k^2(\tilde{F}_1, F_2)}{c + \mathbf{\Delta}_k^2(\tilde{F}_1, F_2)} \\ &= \tau + (c + \mathbf{\Delta}_k^2)^{-2} \left[(c + \mathbf{\Delta}_k^2) IF(\mathbf{\Delta}_p^2) - (c + \mathbf{\Delta}_p^2) IF(\mathbf{\Delta}_k^2)\right] \varepsilon + O(\varepsilon^2). \end{aligned}$$

Therefore the influence function for τ is given by

$$IF(\tau) = (c + \mathbf{\Delta}_k^2)^{-2} \left[(c + \mathbf{\Delta}_k^2) IF(\mathbf{\Delta}_p^2) - (c + \mathbf{\Delta}_p^2) IF(\mathbf{\Delta}_k^2)\right], \tag{3.3}$$

where $IF(\mathbf{\Delta}_k^2) = \lambda_1 \mathbf{\Delta}_k^2 + 2\phi_k - \lambda_1 \phi_k^2$ and ϕ_k is similarly defined as ϕ_p in (3.2), according to the partition of $\mathbf{\Sigma}$.

Next we will consider three sample versions as in Critchley (1985) : the empirical influence function (EIF), the sample influence function (SIF) and the deleted empirical influence function (DIF). A large absolute value of each sample version indicates that the corresponding observation is influential.

3.1. Empirical Influence Function

The EIF for D_p^2 and T are obtained by substituting the empirical distribution functions $\hat{F}_1, \hat{F}_2$ for F_1, F_2, respectively, in (3.2) and (3.3). It is easily seen that the EIF for D_p^2 at $\mathbf{x}_j$ becomes

$$EIF_j(D_p^2) = \lambda_1 D_p^2 + 2a_p - \lambda_1 a_p^2, \tag{3.4}$$

where $a_p = \hat{\phi}_p = \mathbf{d}^T \mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$ and $\lambda_1 = (n_1 - 1)/m$. The EIF for T at $\mathbf{x}_j$ can be written as

$$EIF_j(T) = (c + D_k^2)^{-2} \left[(c + D_k^2) EIF_j(D_p^2) - (c + D_p^2) EIF_j(D_k^2)\right], \tag{3.5}$$

where $EIF_j(D_k^2)$ is obtained similarly to (3.4).

3.2. Sample Influence Function

The SIF can be obtained by setting $F_1 = \hat{F}_1, F_2 = \hat{F}_2$ and taking $\varepsilon = -1/(n_1 - 1)$ in the definition of the influence function (3.1) instead of taking a limit. Then the SIF for a parameter θ at $\mathbf{x}_j$ can be rewritten as $(n_1-1)[\theta(\hat{F}_1, \hat{F}_2)-\theta(\hat{F}_{1(j)}, \hat{F}_2)]$, where $\hat{F}_{1(j)} = (1 + (n_1 - 1)^{-1})\hat{F}_1 - (n_1 - 1)^{-1}\delta_{\mathbf{x}_j}$ is the deleted version of $\hat{F}_1$ with the jth observation $\mathbf{x}_j$ deleted.

The mean vector evaluated at $\hat{F}_{1(j)}$ and the covariance matrix evaluated at $\hat{F}_{1(j)}$ and $\hat{F}_2$ are given by

$$\begin{aligned} \boldsymbol{\mu}_1(\hat{F}_{1(j)}) &= \bar{\mathbf{x}} - (\mathbf{x}_j - \bar{\mathbf{x}})/(n_1 - 1), \\ \mathbf{S}_{(j)} = \boldsymbol{\Sigma}(\hat{F}_{1(j)}, \hat{F}_2) &= \frac{1}{m-1}\left[m\mathbf{S} - \frac{n_1}{n_1 - 1}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T\right], \end{aligned}$$

respectively. Hereafter we denote by the subscript (j) the estimator based on the reduced data set with the deletion of observation $\mathbf{x}_j$. Since $\mathbf{d}_{(j)} = \mathbf{d} - (\mathbf{x}_j - \bar{\mathbf{x}})/(n_1 - 1)$ and

$$\mathbf{S}_{(j)}^{-1} = \frac{m-1}{m}\left[\mathbf{S}^{-1} + \frac{n_1\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T\mathbf{S}^{-1}}{m(n_1 - 1) - n_1(\mathbf{x}_j - \bar{\mathbf{x}})^T\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})}\right],$$

we have

$$D_{p(j)}^2 = \frac{m-1}{m}\left[D_p^2 + \frac{(n_1 a_p - m)^2}{n_1(m(n_1 - 1) - n_1 b_p)} - \frac{m}{n_1(n_1 - 1)}\right],$$

where $b_p = (\mathbf{x}_j - \bar{\mathbf{x}})^T\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$. Thus the SIF for D_p^2 at $\mathbf{x}_j$ becomes

$$SIF_j(D_p^2) = \frac{n_1 - 1}{m}\left[D_p^2 - \frac{m-1}{n_1}\frac{(n_1 a_p - m)^2}{m(n_1 - 1) - n_1 b_p} + \frac{m(m-1)}{n_1(n_1 - 1)}\right]. \tag{3.6}$$

Also the SIF for T at $\mathbf{x}_j$ can be written as

$$SIF_j(T) = (n_1 - 1)\left[\frac{D_p^2 - D_k^2}{c + D_k^2} - \frac{D_{p(j)}^2 - D_{k(j)}^2}{c' + D_{k(j)}^2}\right], \tag{3.7}$$

where $c' = (m-1)(m-3)/((n_1 - 1)n_2)$.

3.3. Deleted Empirical Influence Function

The DIF is obtained by replacing F_1 with $\hat{F}_{1(j)}$ in (3.2) and (3.3), and it measures the effect of deleting the jth observation in the first sample on the

estimator. Using $\mathbf{d}_{(j)}$ and $\mathbf{S}_{(j)}^{-1}$ derived in the previous subsection, we obtain

$$a_{p(j)} = a_p(\hat{F}_{1(j)}, \hat{F}_2) = (m-1)\left[\frac{n_1 a_p - m}{m(n_1-1) - n_1 b_p} + \frac{1}{n_1 - 1}\right].$$

Replacing F_1 and F_2 with $\hat{F}_{1(j)}$ and $\hat{F}_2$ in (3.2) respectively yields the DIF for D_p^2 at $\mathbf{x}_j$ as

$$DIF_j(D_p^2) = \lambda_1 D_{p(j)}^2 + 2a_{p(j)} - \lambda_1 a_{p(j)}^2. \tag{3.8}$$

Similarly the DIF for T at $\mathbf{x}_j$ is computed as

$$DIF_j(T) = (c + D_{k(j)}^2)^{-2}\left[(c + D_{k(j)}^2)DIF_j(D_p^2) - (c + D_{p(j)}^2)DIF_j(D_k^2)\right]. \tag{3.9}$$

When only the second distribution function F_2 is perturbed, the influence function for a parameter is defined similarly to (3.1) and three sample versions of the influence functions for Δ_p^2 and τ can be similarly obtained by taking $a_p = -\mathbf{d}^T\mathbf{S}^{-1}(\mathbf{y}_j - \bar{\mathbf{y}})$ and $b_p = (\mathbf{y}_j - \bar{\mathbf{y}})^T\mathbf{S}^{-1}(\mathbf{y}_j - \bar{\mathbf{y}})$ with n_1 and n_2 interchanged.

Three sample versions of the influence function for τ have similar values and as the sample sizes n_1 and n_2 tend to be large, their values give the same degree of influence. Hence the use of only one version is enough to investigate the influence of observations on the test statistic T_*.

4. DERIVATIVE INFLUENCE MEASURE

Let $\hat{\theta}$ be an estimator of the parameter θ. The derivative influence is derived by considering a perturbation scheme represented by ω in which the distribution for only one observation is perturbed. When we denote by $\hat{\theta}(\omega)$ the estimator $\hat{\theta}$ under this perturbation scheme, we assume that $\hat{\theta} = \hat{\theta}(\omega_0)$ for some ω_0 called the null point. The derivative influence (De Gruttola et al., 1987) is defined by

$$\nabla\hat{\theta}(\omega_0) = \left.\frac{d\hat{\theta}(\omega)}{d\omega}\right|_{\omega=\omega_0}.$$

Assume that only one observation, say $\mathbf{x}_j (1 \le j \le n_1)$, is drawn from a perturbed distribution $N(\boldsymbol{\mu}_1, \Sigma/\omega)$ for $0 \le \omega \le 1$ and the others are the same as assumed in Section 2. Then under this perturbation scheme, the maximum likelihood estimators of $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}$ become

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_1(\omega) &= (n_1 + \omega - 1)^{-1}\left[n_1\bar{\mathbf{x}} + (\omega - 1)\mathbf{x}_j\right], \\
\hat{\boldsymbol{\Sigma}}(\omega) &= \frac{m}{n_1 + n_2}\mathbf{S} + \frac{n_1(\omega - 1)}{(n_1 + n_2)(n_1 + \omega - 1)}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T,
\end{aligned}$$

respectively, with $\hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{y}}$. Since under the perturbation the expectation of $\hat{\boldsymbol{\Sigma}}(\omega)$ yields $m/(n_1+n_2)\boldsymbol{\Sigma}$, we consider the perturbed covariance matrix

$$\mathbf{S}(\omega) = \mathbf{S} + \frac{n_1(\omega-1)}{m(n_1+\omega-1)}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T$$

for unbiasedness. The inverse of $\mathbf{S}(\omega)$ is easily computed as

$$\mathbf{S}(\omega)^{-1} = \mathbf{S}^{-1} - \frac{n_1(\omega-1)\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})^T\mathbf{S}^{-1}}{m(n_1+\omega-1) + n_1(\omega-1)(\mathbf{x}_j - \bar{\mathbf{x}})^T\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})}.$$

After a simple algebra, $D_p^2(\omega)$ is obtained as

$$D_p^2(\omega) = D_p^2 - \frac{\omega-1}{n_1}\left[\frac{(n_1a_p - m)^2}{m(n_1+\omega-1) + n_1(\omega-1)b_p} - \frac{m}{n_1+\omega-1}\right]. \tag{4.1}$$

The value of $D_p^2(1)$ reduces to the unperturbed case D_p^2, that is, $\omega = 1$ is the null point. Accordingly the perturbed statistic $T(\omega)$ is defined by

$$T(\omega) = \frac{D_p^2(\omega) - D_k^2(\omega)}{c + D_k^2(\omega)},$$

where $D_k^2(\omega)$ can be obtained similarly to $D_p^2(\omega)$.

The derivatives for $D_p^2(\omega)$ and $T(\omega)$ with respect to ω become

$$\nabla D_p^2(\omega) = -\frac{(n_1a_p - m)^2 m}{[m(n_1+\omega-1) + n_1(\omega-1)b_p]^2} + \frac{m}{(n_1+\omega-1)^2}, \tag{4.2}$$

$$\nabla T(\omega) = (c + D_k^2(\omega))^{-2}\left[(c + D_k^2(\omega))\nabla D_p^2(\omega) - (c + D_p^2(\omega))\nabla D_k^2(\omega)\right], \tag{4.3}$$

respectively. We are interested in the derivatives of $D_p^2(\omega)$ and $T(\omega)$ evaluated at two points $\omega = 0$ and 1. At $\omega = 1$, the perturbed estimators $\hat{\boldsymbol{\mu}}_1(\omega)$ and $\mathbf{S}(\omega)$ reduce to the unperturbed case, respectively. Thus the derivatives at $\omega = 1$, that is, the derivative influence, measure the local change in the estimators caused by the jth observation. The larger value of $\nabla T(1)$ indicates that the slope of $T(\omega)$ at the null point changes rapidly. In the other case $\omega = 0$, we obtain $\hat{\boldsymbol{\mu}}_1(0) = \bar{\mathbf{x}}_{(j)}$, $\mathbf{S}(0) = m/(m-1)\mathbf{S}_{(j)}$ and $D_p^2(0) = m/(m-1)D_{p(j)}^2$. Furthermore, $T(0)$ is equal to $T_{(j)}$ apart from the factor c in the statistic T. It implies that the perturbed estimators at $\omega = 0$ are the estimators based on the data set with the deletion of $\mathbf{x}_j$. Thus the derivatives at $\omega = 0$ measure the rate of change in the estimators when $\mathbf{x}_j$ is omitted.

At $\omega = 1$ the derivative influence for D_p^2 can be rewritten as

$$\nabla D_p^2(1) = -\frac{a_p^2}{m} + \frac{2a_p}{n_1}.$$

Comparing $\nabla D_p^2(1)$ and (3.4) yields the identity $\nabla D_p^2(1) \approx EIF_j(D_p^2)/(n_1 - 1) - D_p^2/m$ if n_1 is sufficiently large such that $1/n_1 \approx 1/(n_1 - 1)$. It implies that $\nabla D_p^2(1)$ and $EIF_j(D_p^2)$ have the similar influence information, and so do $\nabla T(1)$ and $EIF_j(T)$ from (3.5) and (4.3).

From (4.2) it is straightforward to show

$$\nabla D_p^2(0) = -\frac{(n_1 a_p - m)^2 m}{[m(n_1 - 1) - n_1 b_p]^2} + \frac{m}{(n_1 - 1)^2}.$$

In (3.8) if the term having the order $O(1/n_1)$ can be cancelled, we get $\nabla D_p^2(0) \approx DIF_j(D_p^2)/(n_1 - 1) - D_p^2/m$ with the condition of $(m-1)/m \approx 1$. Thus the influence measures using $\nabla D_p^2(0)$ and $DIF_j(D_p^2)$ are equivalent, and so are $\nabla T(0)$ and $DIF_j(T)$.

Since $\nabla D_p^2(\omega)$ is differentiable over the range of ω, the average of $\nabla D_p^2(\omega)$ can be obtained as

$$\int_0^1 \nabla D_p^2(\omega) d\omega = D_p^2(1) - D_p^2(0).$$

The condition of $(m-1)/m \approx 1$ and (3.6) yield $D_p^2(1) - D_p^2(0) \approx SIF_j(D_p^2)/(n_1 - 1) - D_p^2/m$. Also $T(1) - T(0)$ provides the same influence information as $SIF_j(T)$, because $SIF_j(T) = (n_1 - 1)(T(1) - T(0))$, ignoring the factor c. Since $D_p^2(0) = m/(m-1)D_{p(j)}^2$ is the deletion influence on D_p^2, $T(1) - T(0)$ gives the same influence information as the deletion influence (or SIF). Thus to investigate the influence of observations on the test statistic for selecting discriminant variables we use $T(1) - T(0)$ when we consider the deletion influence.

5. NUMERICAL EXAMPLE

The influence function method and the derivative influence measures are applied to the Flea-Beetles data (Seber, 1984, p. 295) which have measurements on four variables for two species of flea beetles. Nineteen observations are obtained for the first species and twenty observations for the second species. The observations are labelled as 1 to 19 for the first species and 20 to 39 for the second species. Kim (1996) analyzed the data set for the purpose of investigating the observations that have a large influence on the misclassification probability, and concluded that observations 27 and 36 are possibly influential.

First we tested the hypothesis of additional information for all combinations of variables using the test statistic T_*. Only the hypothesis that the third variable has no additional information is not rejected, and the other hypotheses are rejected at the significance level 0.05. Here we investigate the influence of observations on the test statistic of the null hypothesis that the third variable has no additional information. In this case the value of T_* is 1.82 and the p-value is 0.19.

Next three sample versions of the influence function for the statistic T are given in Figure 5.1. In this plot, y-axis indicates EIF, SIF, DIF for T. From Figure 5.1, we may conclude that observation 27 is the most influential on the test statistic and that three sample versions yield the same result.

We conducted the influence analysis using three derivative influence measures $\nabla T(1), T(1) - T(0)$ and $\nabla T(0)$ described in Section 4. The results are depicted in Figure 5.2. In this plot we can observe that $T(1) - T(0)$ is the average of $\nabla T(1)$ and $\nabla T(0)$ as explained in Section 4. From Figures 5.1 and 5.2 we can see that the influence measures EIF, SIF and DIF are equivalent to $\nabla T(1), T(1) - T(0)$ and $\nabla T(0)$, respectively, as expected. This is also supported by observing that the sample correlation coefficients between corresponding measures are all one. It indicates that the influence function method and the derivative influence measures give the identical results.

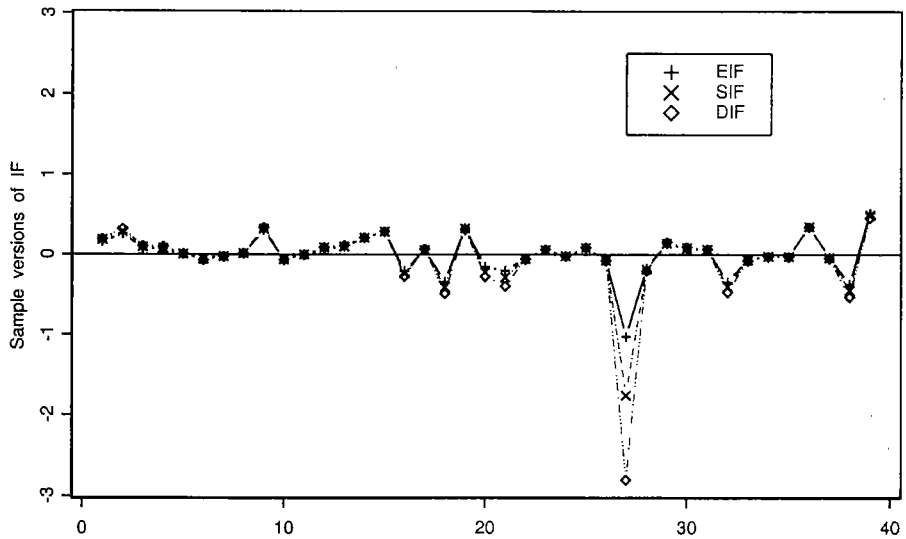

Figure 5.1: Sample versions of the influence function for T

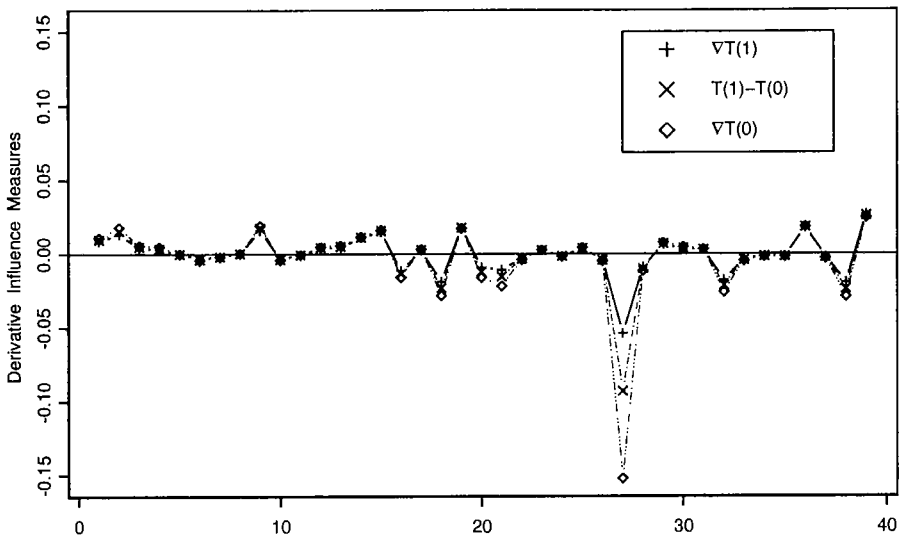

Figure 5.2: Derivative influence measures for T

Finally we performed the case deletion for the test statistic T_*. The change in the value of the test statistic T_* due to a single case deletion is the maximum for the deletion of observation 27 and the next for the deletion of observation 39. However, the effect of omitting observation 39 is relatively small to that of observation 27, because we have $T_* - T_{*(27)} = -3.0$ and $T_* - T_{*(39)} = -0.72$. Thus we may conclude that only observation 27 is influential on the test statistic T_*. This result is in parallel with the influence function method and the derivative influence measures. After the deletion of observation 27, the p-value for the test statistic based on the remaining sample becomes 0.035 and thus the null hypothesis would be rejected at the significance level 0.05. It implies that a reverse decision is made by removing only observation 27.

6. Concluding Remarks

The influence function method was applied to the investigation of the influence of observations on a test for addition information about discrimination in linear discriminant analysis. The results based on the influence function are identical to those based on the derivative influence measures. The derivative influence measure is more flexible than the influence function method because the former allows us to take diverse perturbation schemes according to the purpose of investigation.

REFERENCES

Campbell, N. A. (1978). The influence function as an aid in outlier detection in discriminant analysis, *Applied Statistics*, **27**, 251–258.

Critchley, F. (1985). Influence in principal component analysis, *Biometrika*, **72**, 627–636.

De Gruttola, V., Ware, J. H. and Louis, T. A. (1987). Influence analysis of generalized least squares estimators, *Journal of the American Statistical Association*, **82**, 911–917.

Hampel, F. R. (1974). The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, **69**, 383–393.

Jung, K.-M. and Kim, M. G. (1999). Influence analysis of the likelihood ratio test in multivariate Behrens-Fisher problem, *The Korean Communications in Statistics*, **6**, 939–946.

Kim, M. G. (1995). The influence in comparing covariance matrices, *Communications in Statistics: Theory and Methods*, **24**, 1431–1441.

Kim, M. G. (1996). Local influence on misclassification probability, *Journal of the Korean Statistical Society*, **25**, 145–151.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, Wiley.

Seber, G. A. F. (1984). *Multivariate observations*, John Wiley & Sons.