Journal of the Faculty of Environmental Science and Technology, Okayama University
Vol.3, No.1, pp.37-46, January 1998

Assessing Local Influence in Multivariate Analyses of Incomplete Data

Hyun-Jeong KIM [1] , Tomoyuki TARUMI [2] and Yutaka TANAKA [2]

(Received October 31 , 1997)

The present paper deals with multivariate analyses applied to the maximum likelihood estimate(s) for (the mean vector and) the covariance matrix based on incomplete data, and derives influence functions for the mean vector, the covariance matrix and some statistics in multivariate analyses. Influential directions in the sense of Cook's local influence are also derived. A numerical example is given to show the usefulness of the proposed method.

1 Introduction

Suppose we wish to analyze a set of multivariate (or rectangular) data, where we can assume that the observation vector follows a p-variate normal distribution. Such a data set can be analyzed with standard multivariate statistical methods when all elements are observed. In practice, however, we sometimes meet situations where some parts of rectangular data are missing. To deal with such incomplete data there are some conventional methods such as i) the method of using only complete cases, ii) the method based on covariances computed by using all available pairs of observations, and iii) the method of imputing missing observations. It is known, however, that we can obtain the maximum likelihood (ML) estimates for means, variances and covariances based on all available observations by using the EM algorithm, and that it gives a better result than the above conventional methods, when we can assume the missing observations occur at random (see, e.g., Little & Rubin, 1987).

Results of multivariate analyses sometimes depend heavily upon a small number of observations, and to detect such influential observations, methods of influence or sensitivity analysis have been studied in various multivariate methods (see, e.g., Tanaka, 1994). Similar phenomena may occur in the case of incomplete data. In the present paper we try to develop a method of sensitivity analysis in multivariate analyses of incomplete data, focusing on the ML estimates for the mean vector and covariance matrix and multivariate analyses based on them.

Most multivariate methods including principal component analysis (PCA), canonical correlation analysis (CCA), factor analysis (FA), covariance structure analysis (CSA) and discriminant analysis (DA) can be applied in two steps: The first step estimates the mean vector ($\tilde{\mu}$) and covariance matrix ($\widetilde{\Sigma}$), and then using the estimated $\tilde{\mu}$ and $\widetilde{\Sigma}$ the second step computes their major results such as eigenvalues and eigenvectors, which are given by differentiable functions of $\tilde{\mu}$ and $\widetilde{\Sigma}$. So it is basically important to evaluate the influence on the mean vector and covariance matrix to derive the influence on various statistics in multivariate methods. In section 2 we explain the EM algorithm for maximum likelihood estimation of μ and Σ based on incomplete data, and derive the influence functions for $\tilde{\mu}$ and $\widetilde{\Sigma}$ in section 3. Then, by using the chain rule we illustrate how to derive the influence functions for some statistics related to PCA in section 4. In section 5, after introducing the basic idea of Cook's local influence, we derive maximum

[1] Graduate School of Natural Science and Technology, Okayama University, Okayama, 700 Japan.
[2] Faculty of Environmental Science and Technology, Okayama University, Okayama, 700 Japan.

curvature directions by using the asymptotic covariances of $\tilde{\mu}$ and $\tilde{\Sigma}$. A numerical example is given to show the usefulness of the proposed method.

2 Maximum likelihood estimates of μ and Σ based on incomplete data using EM algorithm

Suppose that the observations $X_1, \cdots, X_n$ are obtained as a random sample from a p-variate normal distribution and that parts of the data are missing at random (MAR) or, in other words, the missing probability does not depend on the missing value of the variable. A general approach for computing maximum likelihood (ML) estimates from incomplete data is given by Dempster, Laird, and Rubin (1977). Their technique called the EM algorithm consists of an iterative calculation involving two steps which are called as the estimation and maximization steps. In the case of multivariate normal distribution it is known that the ML estimates of μ and Σ are based on the complete data sufficient statistics $T_1 = \sum X_\alpha$, $T_2 = \sum X_\alpha X_\alpha^T$ and the algorithm proceeds as follows.

Step 1. Start from appropriate initial values $\tilde{\mu} = \tilde{T}_1/n$, $\tilde{\Sigma} = \tilde{T}_2/n - \tilde{\mu}\tilde{\mu}^T$.

Step 2. (Estimation step)

The E step calculates the conditional expectations given $\tilde{\mu}$ and $\tilde{\Sigma}$ to estimate the missing values and then estimates the contributions of $\tilde{x}_\alpha^{(1)}$ to T_1 and T_2;

$$\begin{aligned}
&\tilde{x}_\alpha^{(1)} = \tilde{\mu}^{(1)} + \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}(x_\alpha^{(2)} - \tilde{\mu}_\alpha^{(2)}) \\
&\widetilde{x_\alpha^{(1)}(x_\alpha^{(1)})^T} = \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + \tilde{x}_\alpha^{(1)}(\tilde{x}_\alpha^{(1)})^T \\
&\widetilde{x_\alpha^{(1)}(x_\alpha^{(2)})^T} = \tilde{x}_\alpha^{(1)}(x_\alpha^{(2)})^T, \quad \widetilde{x_\alpha^{(2)}x_\alpha^{(2)}} = x_\alpha^{(2)}(x_\alpha^{(2)})^T, \qquad \alpha = 1, 2, \cdots, n.
\end{aligned}$$

Note. Superscripts (1) and (2) indicate the groups of variables which are missing and not missing for the α-th observation, respectively.

Step 3. Calculate the sufficient statistics $\tilde{T}_1$ and $\tilde{T}_2$ using the results of Step 2;

$$\begin{aligned}
\tilde{T}_1 &= \sum_\alpha X_\alpha^+ \\
\tilde{T}_2 &= \sum_\alpha (X_\alpha X_\alpha^T)^+,
\end{aligned}$$

where

$$X_\alpha^+ = \begin{cases} x_\alpha, & \text{observed} \\ (\tilde{x}_\alpha^{(1)},\ x_\alpha^{(2)})^T, & \text{missing} \end{cases}$$

$$(X_\alpha X_\alpha^T)^+ = \begin{cases} x_\alpha x_\alpha^T, & \text{observed} \\ \begin{bmatrix} \widetilde{x_\alpha x_\alpha^T} & \tilde{x}_\alpha^{(1)}(x_\alpha^{(2)})^T \\ x_\alpha^{(2)}(\tilde{x}_\alpha^{(1)})^T & x_\alpha^{(2)}(x_\alpha^{(2)})^T \end{bmatrix}, & \text{missing} \end{cases}$$

Step 4. (Maximization step)

The M step calculates the revised estimates of μ and Σ from those filled-in sufficient statistics;

$$\tilde{\mu} = \tilde{T}_1/n, \quad \tilde{\Sigma} = \tilde{T}_2/n - \tilde{\mu}\tilde{\mu}^T.$$

Step 5. Go back to Step 2 if not converged.

3 Influence functions for ML estimators of μ and Σ

Let us consider to place weight $w_\alpha^* = nw_\alpha / \sum w_\beta$ to observation α for $\alpha = 1, \cdots, n$, and assume that observations x_α are independently distributed as $\mathrm{N}(\mu, {w_\alpha^*}^{-1}\Sigma)$. Obviously the case with $w_\alpha = 1$ or $w_\alpha^* = 1$ for all α, which we call unperturbed case, is just the same with the unweighted model in the previous section, and we introduce small perturbations to the case weights. When there are no missing values, the ML estimators of μ and Σ are obtained as

$$\hat{\mu} = \sum_\alpha w_\alpha x_\alpha / \sum_\alpha w_\alpha, \quad \hat{\Sigma} = (\sum_\alpha w_\alpha)^{-1} \sum_\alpha w_\alpha (x_\alpha - \hat{\mu})(x_\alpha - \hat{\mu})^T.$$

The partial derivatives of $\hat{\mu}$ and $\hat{\Sigma}$ with respect to w_α at $\boldsymbol{w}_0 = (1, \cdots, 1)^T$ are

$$\begin{aligned} \partial\hat{\mu}/\partial w_\alpha|_{w_0} &= n^{-1}(x_\alpha - \hat{\mu}) \\ \partial\hat{\Sigma}/\partial w_\alpha|_{w_0} &= n^{-1}\left\{(x_\alpha - \hat{\mu})(x_\alpha - \hat{\mu})^T - \hat{\Sigma}\right\}. \end{aligned}$$

These partial derivatives are essentially equivalent to the empirical influence functions (EIF) of the mean vector and covariance matrix, or strictly they are $1/n$ times the corresponding EIF. It can be verified easily that similar relations hold with respect to other statistics which are derived as differentiable functions of $\hat{\mu}$ and $\hat{\Sigma}$. So, we can obtain the influence functions by multiplying n to the corresponding partial derivatives.

Now let us try to derive partial derivatives of $\hat{\mu}$ and $\hat{\Sigma}$ in the case of incomplete data.

In the above weighted model $\tilde{T}_1 = \sum_\alpha w_\alpha^* X_\alpha$ and $\tilde{T}_2 = \sum_\alpha w_\alpha^* X_\alpha X_\alpha^T$ are joint sufficient statistics, and we can obtain the estimates $\hat{\mu}_w$ and $\hat{\Sigma}_w$ using the procedure in the previous section by replacing the sufficient statistics by the corresponding quantities and by replacing $\tilde{T}_1$ and $\tilde{T}_2$ in Step 3 by $\widetilde{T}_1 = \sum_\alpha w_\alpha^* X_\alpha^+$ and $\widetilde{T}_2 = \sum_\alpha w_\alpha^* (\widetilde{X_\alpha X_\alpha^T})^+$, respectively. Substituting $\tilde{\mu} + (\partial\tilde{\mu}/\partial w_j)\Delta w_j$ and $\tilde{\Sigma} + (\partial\tilde{\Sigma}/\partial w_j)\Delta w_j$ into $\hat{\mu}$ and $\hat{\Sigma}$, respectively and comparing the coefficients of Δw_j, we obtain the following system of equations. The partial derivatives $\partial\tilde{\mu}_j (= \partial\tilde{\mu}/\partial w_j)$ and $\partial\tilde{\Sigma}_j = (\partial\tilde{\Sigma}/\partial w_j)$ at the converged solution, are obtained by solving this system of equations :

$$\begin{aligned} n(\partial\tilde{\mu}_j) &- \sum_\alpha \{M_\alpha(\partial\tilde{\mu}_j) + \sum_\alpha M_\alpha(\partial\tilde{\Sigma}_j)\overline{M}_\alpha S_{22}^-(\alpha)\overline{M}_\alpha(X_\alpha^+ - \tilde{\mu}) \\ &- S_{12}(\alpha)S_{22}^-(\alpha)\overline{M}_\alpha(\partial\tilde{\Sigma}_j)\overline{M}_\alpha S_{22}^-(\alpha)\overline{M}_\alpha(X_\alpha^+ - \tilde{\mu}) - S_{12}(\alpha)S_{22}^-(\alpha)\overline{M}_\alpha(\partial\tilde{\mu}_j)\} \\ &= X_j^+ - \tilde{\mu} \end{aligned}$$

$$\begin{aligned} (\partial\tilde{\Sigma}_j) &- n^{-1}\sum_\alpha \{M_\alpha(\partial\tilde{\Sigma}_j)M_\alpha - M_\alpha(\partial\tilde{\Sigma}_j)\overline{M}_\alpha S_{22}^-(\alpha)S_{21}(\alpha) \\ &- S_{12}(\alpha)S_{22}^-(\alpha)S_{21}(\alpha) - S_{12}(\alpha)S_{22}^-(\alpha)\overline{M}_\alpha(\partial\tilde{\Sigma}_j)\overline{M}_\alpha S_{22}^-(\alpha)S_{21}(\alpha)\} \\ &- n^{-1}\sum_\alpha AX_\alpha^{+T}M_\alpha - n^{-1}\sum_\alpha M_\alpha X_\alpha^+ A^T - n^{-1}\sum_\alpha \left\{AX_\alpha^{+T}\overline{M}_\alpha + \overline{M}_\alpha X_\alpha^+ A^T\right\} \\ &+ (\partial\tilde{\mu}_j)\tilde{\mu}^T + \tilde{\mu}(\partial\tilde{\mu}_j)^T \\ &= n^{-1}\left\{(X_j X_j^T)^+ - n^{-1}T_2\right\}, \end{aligned}$$

where

$$\begin{aligned} A = M_\alpha(\partial\tilde{\mu}_j) &+ \sum_\alpha M_\alpha(\partial\tilde{\Sigma}_j)\overline{M}_\alpha S_{22}^-(\alpha)\overline{M}_\alpha(X_\alpha^+ - \tilde{\mu}) \\ &- S_{12}(\alpha)S_{22}^-(\alpha)\overline{M}_\alpha(\partial\tilde{\Sigma}_j)\overline{M}_\alpha S_{22}^-(\alpha)\overline{M}_\alpha(X_\alpha^+ - \tilde{\mu}) \\ &- S_{12}(\alpha)S_{22}^-(\alpha)\overline{M}_\alpha(\partial\tilde{\mu}_j). \end{aligned}$$

M_α and $\overline{M}_\alpha$ being defined as $M_\alpha = \mathrm{diag}(\delta_\alpha(1), \cdots, \delta_\alpha(p))$ and $\overline{M}_\alpha = I - M_\alpha$, where

$$\delta_\alpha(i) = \begin{cases} 1, & \text{if variate } i \text{ is missing for individual } \alpha, \\ 0, & \text{otherwise,} \end{cases}$$

and using the definition of M_α and $\overline{M}_\alpha$,

$$S_{11}(\alpha) = M_\alpha S M_\alpha,\ S_{12}(\alpha) = M_\alpha S \overline{M}_\alpha,\ S_{22}(\alpha) = \overline{M}_\alpha S \overline{M}_\alpha.$$

Since $S_{22}(\alpha)$ consists of a part of S which corresponds to the variables not missing for individual α and the other part filled with 0's, it is singular excepting the case where all variables are observed for individual α. $S_{22}^-(\alpha)$ denotes its Moore-Penrose inverse, which is equal to the ordinary inverse matrix when it is non-singular and can be obtain by inverting a matrix made by extracting only the non-missing part and then by replacing all elements of the non-missing part of S by the elements of the inverse matrix.

The number of unknowns in the above system of equations are p for $\partial\tilde{\mu}_j$ plus $p^* = p(p+1)/2$ for $\partial\tilde{\Sigma}_j$. It is just equal to the number of linearly indepentdent equations. Note that in the above system of equations the coefficients for the unknowns in the left-hand side do not depend on individual number j, while the quantities in the right-hand side depend on it. Therefore, we have to compute the right-hand side for each j but have to compute the left-hand side only once.

4 Influence functions for statistics in multivariate analyses

As discussed in section 1, many multivariate analyses including PCA, CCA, FA, CSA, and DA can be applied in two steps. In the first step the mean vector ($\tilde{\mu}$) and covariance matrix ($\tilde{\Sigma}$) are estimated, and then in the second step major resulting statistics are obtained in the forms of differentiable functions of $\tilde{\mu}$ and $\tilde{\Sigma}$. In the previous section it is shown that the influence functions can be obtained for $\tilde{\mu}$ and $\tilde{\Sigma}$ based on incomplete data by solving a system of equations. Once the influence functions for $\tilde{\mu}$ and $\tilde{\Sigma}$ are obtained it is easy to derive the influence functions for the statistics using the so-called chain rule.

For instance we discuss the case of PCA. The major results of PCA consist of the statistics such as the dominant eigenvalues ν_s, the associated eigenvectors $\boldsymbol{v}_s$, the orthogonal projector onto the subspace spanned by the q principal components $P = \sum_{s=1}^{q} \boldsymbol{v}_s \boldsymbol{v}_s^T$, and a part of the spectral decomposition corresponding to the largest q eigenvalues $T = \sum_{s=1}^{k} \nu_s \boldsymbol{v}_s \boldsymbol{v}_s^T$.

The empirical influence functions (or the partial derivatives) with respect to w_α at $\boldsymbol{w}_0 = (1, \cdots, 1)^T$, for ν_s, $\boldsymbol{v}_s$, P and T can be evaluated by

$$\begin{aligned}
\partial\nu_s(\alpha) &= \boldsymbol{v}_s^T(\partial\tilde{\Sigma}(\alpha))\boldsymbol{v}_s \\
\partial\boldsymbol{v}_s(\alpha) &= \sum_{r\neq s}(\nu_s - \nu_r)^{-1}(\boldsymbol{v}_r^T(\partial\tilde{\Sigma}(\alpha))\boldsymbol{v}_s) \\
\partial P(\alpha) &= \sum_{s=1}^{q}\sum_{r=q+1}^{p}(\nu_s - \nu_r)^{-1}(\boldsymbol{v}_s^T(\partial\tilde{\Sigma}(\alpha))\boldsymbol{v}_r)(\boldsymbol{v}_s\boldsymbol{v}_r^T + \boldsymbol{v}_r\boldsymbol{v}_s^T) \\
\partial T(\alpha) &= \sum_{s=1}^{q}\sum_{r=1}^{q}(\boldsymbol{v}_s^T(\partial\tilde{\Sigma}(\alpha))\boldsymbol{v}_r)\boldsymbol{v}_s\boldsymbol{v}_r^T \\
&\quad + \sum_{s=1}^{q}\sum_{r=q+1}^{p}\nu_s(\nu_s - \nu_r)^{-1}(\boldsymbol{v}_s^T(\partial\tilde{\Sigma}(\alpha))\boldsymbol{v}_r)(\boldsymbol{v}_s\boldsymbol{v}_r^T + \boldsymbol{v}_r\boldsymbol{v}_s^T)
\end{aligned}$$

respectively, by using the empirical influence function for $\hat{\Sigma}$ (Tanaka, 1988) . Thus the influence functions for statistics in major multivariate methods can be evaluated in the case of incomplete data.

5 Cook's local influence

A general method has been proposed by Cook (1986) for assessing the local influence of minor perturbations of the model.

Suppose we have a set of n observations and a statistical model with m parameters θ. Denote the unperturbed weights for n observations by $\boldsymbol{w}_0 = (1, \cdots, 1)^T$, and consider a perturbation from $\boldsymbol{w}_0$ to $\boldsymbol{w}$. Also denote the log likelihood functions for the unperturbed and perturbed cases by $L(\boldsymbol{\theta}|\boldsymbol{w}_0)$ and $L(\boldsymbol{\theta}|\boldsymbol{w})$, and the ML estimates for $\boldsymbol{\theta}$ in both cases by $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$, respectively.

In Cook's local influence the change from $\hat{\boldsymbol{\theta}}$ to $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ is measured with a criterion function called likelihood displacement defined as $D(\boldsymbol{w}) = 2[L(\hat{\boldsymbol{\theta}}|\boldsymbol{w}_0) - L(\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}|\boldsymbol{w}_0)]$, and the effect of the perturbation is represented by a graph called influence graph $(\boldsymbol{w}, D(\boldsymbol{w}))$. In particular, the change of $D(\boldsymbol{w})$ along a straight line $\boldsymbol{w} = \boldsymbol{w}_0 + t\boldsymbol{h}$ plays an important role, where $||\boldsymbol{h}|| = 1$, and Cook (1986) searches for the direction which has the largest curvature at $\boldsymbol{w}_0$.

When we consider the change of $D(\boldsymbol{w})$ along $\boldsymbol{w} = \boldsymbol{w}_0 + t\boldsymbol{h}$, $D(\boldsymbol{w})$ is expressed as a function of t, i.e., $D = D(t)$, and it is expanded around $t = 0$ as

$$D(t) = -2\boldsymbol{h}^T \left[\frac{\partial^2 L}{\partial \boldsymbol{w} \partial \boldsymbol{w}^T}\right] \boldsymbol{h} \cdot \frac{t^2}{2} + O(t^3).$$

The absolute value of the coefficient of $t^2/2$ in the right-hand side gives the normal curvature along $\boldsymbol{h}$ of the influence graph $(\boldsymbol{w}, D(\boldsymbol{w}))$, and it is rewritten as

$$C_{\boldsymbol{h}} = -2\boldsymbol{h}^T \left[\frac{\partial \hat{\boldsymbol{\theta}}^T}{\partial \boldsymbol{w}}\right] \left[\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right] \left[\frac{\partial \hat{\boldsymbol{\theta}}}{\partial \boldsymbol{w}^T}\right] h,$$

where $[\partial \hat{\boldsymbol{\theta}}^T / \partial \boldsymbol{w}]$ is an $n \times m$ matrix and $[\partial^2 L / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T]$ is an $m \times m$ matrix evaluted at $\boldsymbol{w} = \boldsymbol{w}_0$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, respectively. Thus, to find the most influential direction we need to maximize the quadratic form $C_{\boldsymbol{h}}$ under the condition of $||\boldsymbol{h}|| = 1$ and it is found as the eigenvector associated with the largest eigenvalue of the eigenvalue problem of the coefficient matrix of $C_{\boldsymbol{h}}$. If a small proportion of the observations have elements much larger than the rest in this eigenvector, these observations are regarded as the influential subset of observations.

As discussed by Tanaka (1994) and more precisely by Tanaka et al. (1997) the above method of Cook's local influence has close relationship with their general procedure based on influence functions.

From the theory of ML estimation, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is given by $E[-\partial^2 L / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T]$. If we estimate the asymptotic covariance matrix by $[\widehat{\text{acov}}(\hat{\boldsymbol{\theta}})] = [-\ddot{L}]^{-1}_{\boldsymbol{w}=\boldsymbol{w}_0}$, and use the relation $[\partial \boldsymbol{\theta}^T / \partial \boldsymbol{w}] = n^{-1}[EIF]$, the most influential direction $\boldsymbol{h}_{\max}$ is obtained by

$$\boldsymbol{h}_{\max} = \text{argmax} \left\{ 2n^{-2} \boldsymbol{h}^T [EIF][\widehat{\text{acov}}(\hat{\boldsymbol{\theta}})]^{-1} [EIF]^T \boldsymbol{h} \right\},$$

where $[EIF]$ is an $n \times m$ matrix of $\{EIF(\boldsymbol{x}_i; \hat{\boldsymbol{\theta}})\}$. Then this $\boldsymbol{h}_{\max}$ is obtained as the eigenvector associated with the largest eigenvalue of an $n \times n$ eigenvalue problem

$$\{2n^{-2}[EIF][\widehat{\text{acov}}(\hat{\boldsymbol{\theta}})]^{-1}[EIF]^T - \lambda I\}\boldsymbol{h} = \boldsymbol{0},$$

which can be transformed into an $m \times m$ eigenvlaue problem

$$\{2n^{-2}[EIF]^T[EIF] - \lambda[\widehat{\text{acov}}(\hat{\boldsymbol{\theta}})]\}\boldsymbol{a} = \boldsymbol{0}. \tag{1}$$

Therefore, $\boldsymbol{h}_{\max}$ can be obtained by $\boldsymbol{h}_{\max} = [EIF]\boldsymbol{a}_{\max}$, where $\boldsymbol{a}_{\max}$ is the eigenvector associated with the largest eigenvalue of the above $m \times m$ eigenvalue problem.

To search for influential subsets Tanaka et al. (1990) suggest to apply "canonical variate analysis" to $\{EIF(\boldsymbol{x}_i;\hat{\boldsymbol{\theta}})\}$, namely, to solve an eigenvalue problem as

$$\{n^{-1}[EIF]^T[EIF] - \lambda[\widehat{\mathrm{acov}}(\hat{\boldsymbol{\theta}})]\}\boldsymbol{a} = \boldsymbol{0},$$

in their general procedure (also see, Tanaka, 1994). Obviously their first canonical variate gives essentially the same information as the most influential direction in the sense of Cook's local influence.

6 Asymptotic covariance matrix of the estimated means and covariances

To apply the analysis of Cook's local influence we need information about $\partial^2 L/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ or the asymptotic covariance matrix of the estimated parameters.

When the data are missing completely at random, it is known that the expected information matrix of $\theta = (\mu, \Sigma)$ represented as a vector has the form

$$J(\theta) = \begin{bmatrix} J(\mu) & 0 \\ 0 & J(\Sigma) \end{bmatrix}$$

and the (j,k)th element of $J(\mu)$ is

$$\sum_{i=1}^{n} \psi_{jki},$$

where $\psi_{jki} = (j,k)$th element of $\Sigma_{obs,i}^{-1}$, if both x_{ij} and x_{ik} are present; $= 0$, otherwise, and $\Sigma_{obs,i}$ is the covariance matrix of the variables present in observation i. The (lm, rs)th element of $J(\Sigma)$ is

$$\frac{1}{4}(2-\delta_{lm})(2-\delta_{rs})\sum_{i=1}^{n}(\psi_{lri}\psi_{msi} + \psi_{lsi}\psi_{mri})$$

where $\delta_{lm} = 1$, if $l = m$, 0 if $l \neq m$ (see, Little and Rubin, 1987). Then, subsituting the elements in $\tilde{\mu}$ and $\widetilde{\Sigma}$ into the corresponding parameters we can evaluate the asymptotic covariance matrix $\mathrm{acov}(\tilde{\theta})$.

7 Example

A numerical example will be discussed in detail. The set of data, which is taken from Johnson and Wichern (1992), p.183, consists of three measurements (X_1: sweat rate, X_2: sodium content, X_3: potassium content) of perspiration from 20 healthy females. The data are reproduced in Table 1. To illustrate our procedure for incomplete data we have introduced five missing values artificially. Values with asterisk are regarded as missing.

At first we have applied the EM algorithm to estimate μ and Σ. The iterative procedure is considered to be converged, when it holds that the Euclidean norm of the differences of successive two values of $\tilde{\mu}$ and $\widetilde{\Sigma}$ is smaller than $\epsilon = 0.001$. The obtained estimates are

$$\tilde{\mu} = [4.815518,\ 43.49231,\ 9.964999],\quad \widetilde{\Sigma} = \begin{bmatrix} 2.957339 & 12.050798 & -1.83225 \\ 12.050798 & 165.8363953 & -4.320019 \\ -1.83225 & -4.320019 & 3.446301 \end{bmatrix}$$

Then, to evaluate the influence on the estimates of the mean vector $\tilde{\mu}$ and the covariance matrix $\widetilde{\Sigma}$, the empirical influence functions are computed for $\tilde{\mu}$ and $\widetilde{\Sigma}$. Since major multivariate methods are based only

on $\widetilde{\Sigma}$ (not on $\tilde{\mu}$), we shall focus our interest in $\widetilde{\Sigma}$ for simplicity. To detect singly influential observation the vector-valued influence functions are summerized into the so-called generalized Cook's D:

$$D_i = (n-1)^{-2}[\text{vech}(\text{EIF}(\boldsymbol{x}_i, \widetilde{\Sigma}))]^T[\widehat{\text{acov}}(\text{vech}\widetilde{\Sigma})]^{-1}[\text{vech}(\text{EIF}(\boldsymbol{x}_i, \widetilde{\Sigma}))]$$

Table 1. Sweat Data

Individual	X1 Sweat rate	X2 Sodium	X3 Potassium
1	3.7*	48.5	9.3
2	5.7	65.1	8.0
3	3.8	47.2	10.9
4	3.2*	53.2*	12.0
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14
8	7.2	33.1	7.6
9	6.7	47.4	8.5
10	5.4	54.1	11.3
11	3.9	36.9	12.7
12	4.5	58.8	12.3
13	3.5	27.8	9.8
14	4.5	40.2	8.4
15	1.5	13.5	10.1
16	8.5	56.4	7.1
17	4.5*	71.6*	8.2
18	6.5	52.8	10.9
19	4.1	44.1	11.2
20	5.5	40.9	9.4

Source: John and Wichern (1992), p.183

The index plot of D_i is shown in Figure 1. In this figure we can find that the influence of the 15th observation is much larger than those of the other 19 observstions.

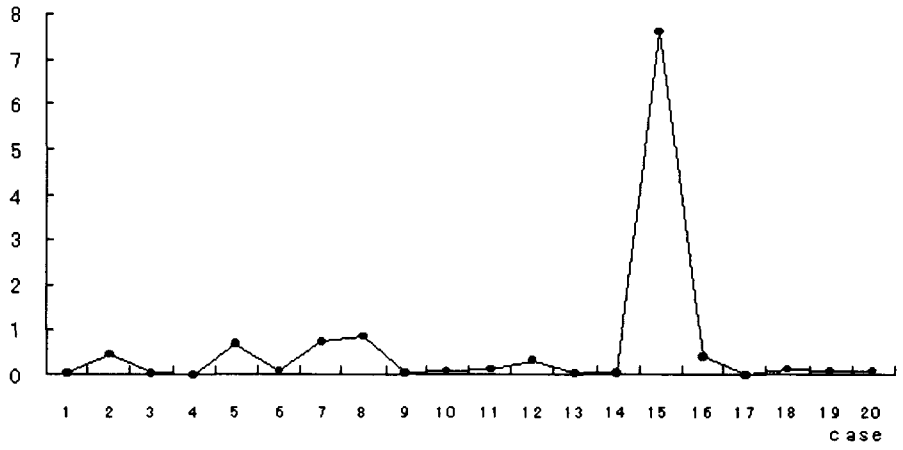

Figure 1: Index plot of D_i

To detect jointly influential observations we have solved the eigenvalue problem (1) and, by using the relation $\boldsymbol{h} = [\text{EIF}]\boldsymbol{a}$, obtained dominant eigenvalues λ_s and associated eigenvectors $\boldsymbol{h}_s$. The eigenvalues

are $17.3579 \gg 2.63429 > 1.74988 > \cdots$ in order of magnitude. The eigenvectors, normalized so that the norm of each $\boldsymbol{h}$ is equal to the corresponding eigenvalue, are displayed in the forms of index plot and scatter plot in Figures 2 and 3, respectively.

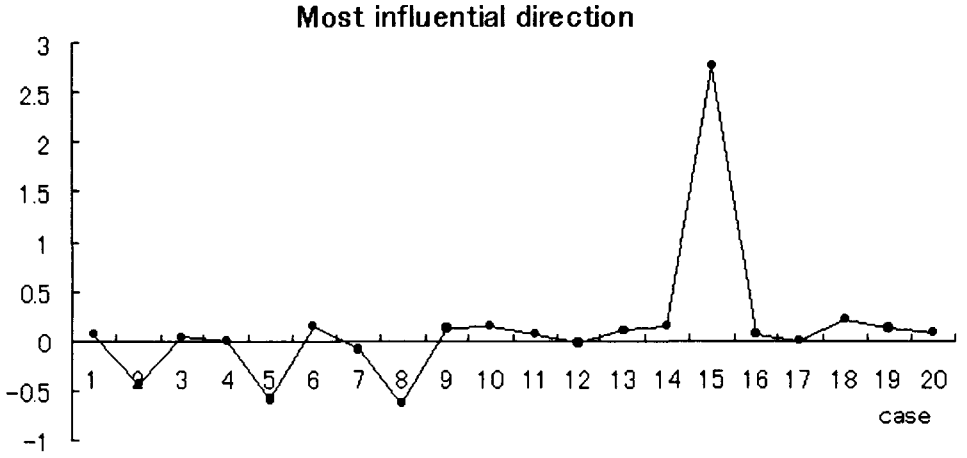

Figure 2: Index plot of the elements of $\boldsymbol{h}_1(\boldsymbol{h}_{\max})$

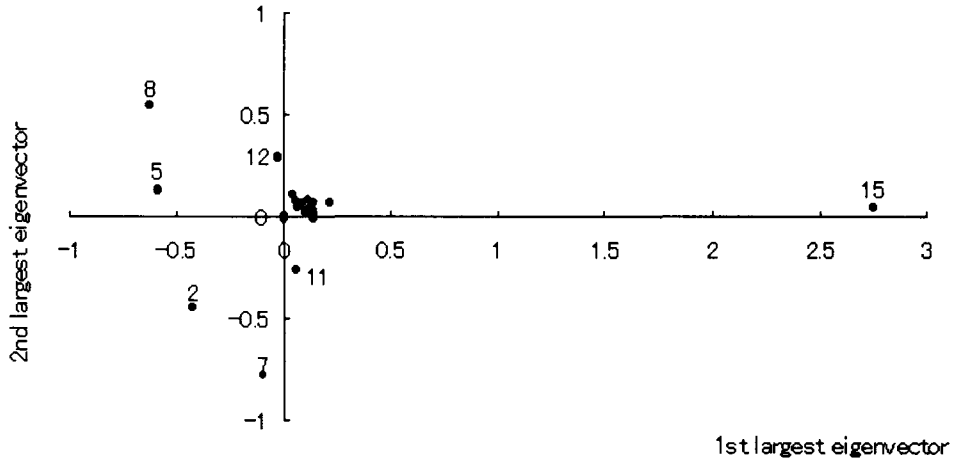

Figure 3: Scatter plot of $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$

From Figure $1 \sim 3$ it seems that there is one singly influential observation (♯15) but no jointly influential ones. If we omit observation ♯15, the estimated mean vector and covariance matrix become

$$\tilde{\mu} = [5.011782,\ 45.256798,\ 9.957895],\quad \tilde{\Sigma} = \begin{bmatrix} 2.431982 & 6.538390 & -1.898826 \\ 6.538390 & 119.461587 & -4.267498 \\ -1.898826 & -4.267498 & 3.626648 \end{bmatrix}.$$

It is noted that in particular the change is large in the parts related to variables X_1 and X_2.

Finally, as an illustration to evaluate the influence on PCA based on the estimated $\tilde{\Sigma}$, we have computed the EIF for the largest eigenvalue $\tilde{\nu}_1$, the variance of the first principal component. The results are given

in Figure 4. Here again we can find that the influence of the 15th observation is much larger than those of the other observations.

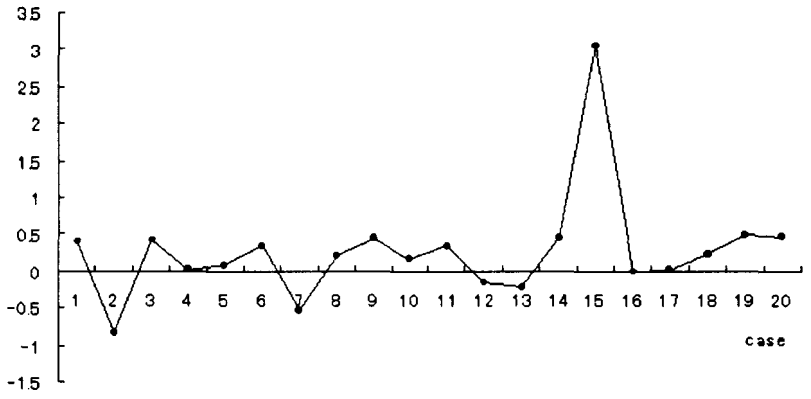

Figure 4: Index plot of the EIF for the largest eigenvalue $\tilde{\nu}_1$ in PCA

8 Discussion

In the present paper we have derived empirical influence functions (EIF) for the ML etimates of the mean vector and the covariance matrix obtained by the EM algorithm. As we discussed in section 3 the EIF is just the partial derivative with respect to weight w_α, when we place the weight $nw_\alpha / \sum w_\beta$ to the α-th observation. Of course we may use other kinds of case weight, however, if we do so, the partial derivative does not correspond to the EIF, though we can evaluate the change of estimated parameters due to minor perturbations to the weights using the techniques in sections 3 and 4.

As alternatives to the EIF we may use the sample influence function (SIF) and its one-step approximation (SIF^1). The SIF is defined as

$$SIF_i = -(n-1)(\hat{\theta}_{(i)} - \hat{\theta}),$$

where $\hat{\theta}_{(i)}$ is the estimate based on the sample without the i-th observation. To compute SIF we have to apply the iterative procedure of the EM algorithm to the data set n times by omitting each observation one by one in turn.

The one-step approximation (SIF^1) to the SIF is the approximate estimate which is obtained by applying the EM algorithm to the perturbed data set only one cycle starting from the solution for the unperturbed data set. Figure 5 shows the scatter matrix of EIF, SIF and SIF^1 for $\tilde{\sigma}_{11}$ and $\tilde{\sigma}_{12}$, where $\tilde{\sigma}_{11}$ and $\tilde{\sigma}_{12}$ are selected for illustration because variables X_1 and X_2 contain missing values. It is noticed that those three correspond to each other quite well and that any of them can be used for detecting influential observations. The easiness of computation is SIF^1 $\succ$ EIF $\succ$ SIF, but the interpretability is SIF $\succ$ SIF^1 $\simeq$ EIF. The EIF has the advantage that it is the only one which can be used for computing Cook's local influence.

In section 6, we obtained the estimated asymptotic covariance matrix $\widehat{\text{acov}}(\hat{\theta})$ by putting the estimated parameter to the formula of the expected information matrix under the assumption of "missing completely at random (MCAR)". The MCAR means that missing probability does not depend on not only the missing value but also the observed values. If we wish to obtain the estimate for acov$(\hat{\theta})$ applicable to the case of missing at random (MAR), we have to obtain the estimated observed information matrix. For this purpose

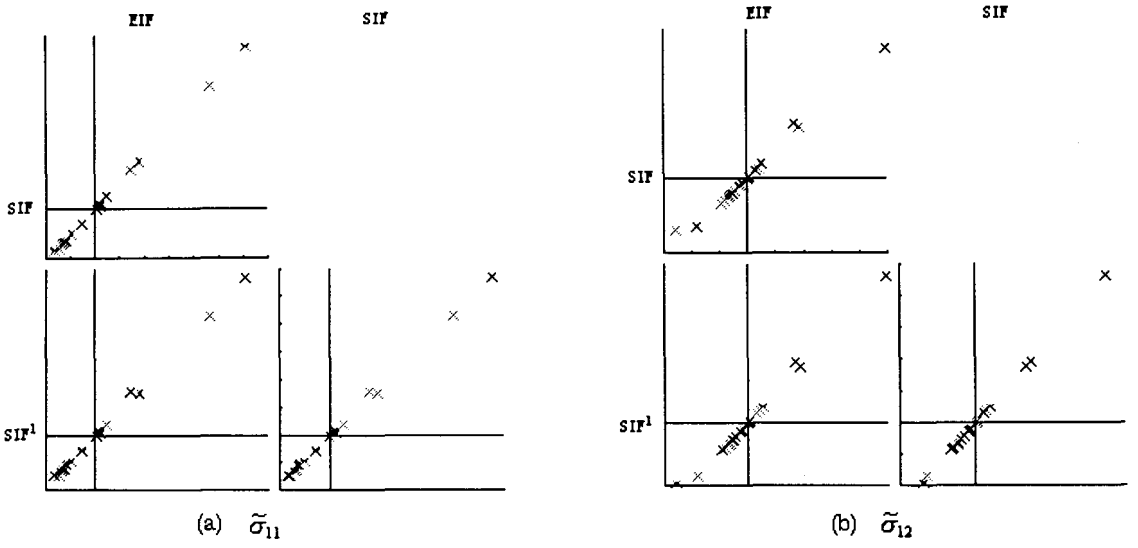

Figure 5: Comparison of EIF, SIF, and SIF^1 for $\tilde{\sigma}_{11}$ and $\tilde{\sigma}_{12}$

we can use one of several methods discussed in McLahran and Krishnan (1997), Chap. 4. It will be our future work to develop a method for the case of MAR and compare its performance with the method in the present paper.

References

[1] Cook, R. D.(1986). Assessment of local influence, *J. Roy. Statist. Soc.* **B48**, 133–69.

[2] Dempster, A. P., Laird, N. M. and Rubin, D. B.(1977). Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc.* **B39**, 1–38.

[3] Johnson, R. A. and Wichern, D. W.(1992). Applied Multivariate Statistical Ananlysis 3rd Edition, *Prentice Hall.*

[4] Little, R. J. A. and Rubin, D. B.(1987). Statistical Analysis with Missing Data, *John Wiley and Sons.*

[5] McLachlan, G. J. and Krishnan, T.(1997). The EM Algorithm and Extensions, *John Wiley and Sons.*

[6] Orchard, T. and Woodbury, M. A.(1972). A missing information principle: Theory and application., *Proc. 6th Berkely Symposium on Math. Statist. and Prob.* **1**, 697–715.

[7] Tanaka, Y.(1988). Sensitivity analysis in principal component analysis: Influence on the subspace spanned by principal components. *Comm. Statist.*, **A 17**, 3157–75. (Corrections, **A 18**(1989), 4305).

[8] Tanaka, Y., Castaño-Tostado, E. and Odaka, Y.(1990). Sensitivity analysis in factor analysis. *COMPSTAT*(Edited by Momirović, K. and Mildner, V.), Physica-Verlag, 205–10.

[9] Tanaka, Y.(1994). Recent Advance in Sensitivity Analysis in Multivariate Statistical Methods. *Jour. Japan Soc. Comm. Statist.*, **7**, 1–25.

[10] Tanaka, Y., Watadani, S. and Zhang, F. H.(1997). Q-Mode and R-mode influence analyses in multivariate methods. *Technical report No.66*, Okayama Statistical Association.