J. Appl. Math. & Informatics Vol. 28(2010), No. 3 - 4, pp. 913 - 921
Website: http://www.kcam.biz

INFLUENCE ANALYSIS OF CHOLESKY DECOMPOSITION

MYUNG GEUN KIM

ABSTRACT. The derivative influence measure is adapted to the Cholesky decomposition of a covariance matrix. Formulas for the derivative influence of observations on the Cholesky root and the inverse Cholesky root of a sample covariance matrix are derived. It is easy to implement this influence diagnostic method for practical use. A numerical example is given for illustration.

AMS Mathematics Subject Classification : 62J20.
Key words and phrases : Covariance matrix, derivative influence, influence function, outliers.

1. Introduction

Cholesky decomposition has been used in many areas of statistics, for example in multiple regression[5], in linear mixed-effects model[9], in longitudinal data[11], etc. A modified Cholesky decomposition was considered by [12] in which more references can be found.

Diagnostic methods have been suggested in wide areas of statistics and some of them are included in [2] and [1]. A sample covariance matrix is very sensitive to outliers, and so is the Cholesky root of it. Thus it is necessary to devise diagnostic methods of measuring the influence of observations on the Cholesky root. The influence of observations on the Cholesky root of a sample covariance matrix was investigated by [7] using the influence function. However, no other method is available to the best of my knowledge.

In this work the derivative influence measure suggested by [3] is adapted to the Cholesky decomposition of a covariance matrix. This influence measure needs a perturbation scheme in which all the other observations have the same distribution except for only one observation which is distributed as another distribution. The slope of the perturbed estimator under this perturbation scheme is considered in order to investigate the influence of observations on the estimator of

Received October 20, 2009. Accepted January 8, 2010.

interest. In Section 2, we consider an appropriate perturbation scheme under which the perturbed maximum likelihood estimators of the model parameters under the multivariate normality are found. We derive formulas for the derivative influence on the Cholesky root of a sample covariance matrix in Section 3. The inverse Cholesky root is considered in Section 4. The formulas derived in Sections 3 and 4 are similar to those based on the influence function considered by [7]. A numerical example is provided for illustration in Section 5.

2. Preliminaries

A random sample of size n, $\{x_1, x_2, \cdots, x_n\}$, from a p-variate normal distribution, $N(\mu, \Sigma)$ with mean vector μ and positive definite covariance matrix Σ is drawn. The maximum likelihood estimators (MLEs) of μ and Σ based on the sample of size n are denoted by $\bar{x}$ and S, respectively. It is well known that the sample covariance matrix S is positive definite with probability one if and only if $n > p$ ([4]). Thus it is assumed that the sample size is sufficiently large, in order to ensure that the sample covariance matrix S is positive definite.

Under normality assumption, it can be easily checked that the MLE of the Cholesky root of the covariance matrix Σ is just the Cholesky root of the MLE of Σ. First, we reparametrize the covarianve matrix Σ as a product of the Cholesky root and its transpose. Then we find the likelihood equations of the Cholesky root, which become equivalent to the usual Cholesky algorithm.

For each $s = 1, 2, \cdots, n$, we consider a perturbation scheme in which all observations have $N(\mu, \Sigma)$ as their common distribution, except for the sth observation x_s which is distributed as $N(\mu, \Sigma/w_s)$. When $w_s = 1$, it reduces to the unperturbed scheme. Under this perturbation scheme, the perturbed MLEs of μ and Σ will be derived here for easy understanding. The log-likelihood function up to unimportant terms is

$$l(\mu, \Sigma|w_s) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{r\neq s}(x_r-\mu)^T\Sigma^{-1}(x_r-\mu) - \frac{w_s}{2}(x_s-\mu)^T\Sigma^{-1}(x_s-\mu).$$

First, $\partial l(\mu, \Sigma|w_s)/\partial\mu = 0$ yields the perturbed MLE of μ as

$$\hat{\mu}(w_s) = \frac{n}{w_s+n-1}\bar{x} + \frac{w_s-1}{w_s+n-1}x_s.$$

Let σ_{ij} and σ^{ij} be the elements in the ith row and jth column of Σ and Σ^{-1}, respectively. Then we have

$$\begin{aligned}\frac{\partial l(\mu, \Sigma|w_s)}{\partial\sigma_{ij}} &= -\frac{n}{2}(2-\delta_{ij})\sigma^{ij} + \frac{1}{2}\sum_{r\neq s}(x_r-\mu)^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma_{ij}}\Sigma^{-1}(x_r-\mu)\\ &\quad+\frac{w_s}{2}(x_s-\mu)^T\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma_{ij}}\Sigma^{-1}(x_s-\mu)\end{aligned}$$

using the following formulas (refer to Chap. 8 of [10] for more details)

$$\frac{\partial \log|\Sigma|}{\partial \sigma_{ij}} = (2-\delta_{ij})\sigma^{ij} \text{ and } \frac{\partial \Sigma^{-1}}{\partial \sigma_{ij}} = -\Sigma^{-1}\frac{\partial \Sigma}{\partial \sigma_{ij}}\Sigma^{-1},$$

where δ_{ij} is the Kronecker delta function. In order to solve $\partial l(\mu, \Sigma|w_s)/\partial \sigma_{ij} = 0$, we need more computations and first, we can easily get

$$x_r - \hat{\mu}(w_s) = x_r - \bar{x} - \frac{w_s - 1}{w_s + n - 1}(x_s - \bar{x}) \ (r \neq s)$$

$$x_s - \hat{\mu}(w_s) = \frac{n}{w_s + n - 1}(x_s - \bar{x}).$$

Since $\sum_{r=1}^{n}(x_r - \bar{x})(x_s - \bar{x})^T = 0$, we have

$$\sum_{r \neq s}^{n}(x_r - \bar{x})(x_s - \bar{x})^T = -(x_s - \bar{x})(x_s - \bar{x})^T.$$

With the above results, a little computation shows that

$$\begin{aligned} & \sum_{r \neq s}(x_r - \hat{\mu}(w_s))(x_r - \hat{\mu}(w_s))^T \\ = \ & nS + \frac{(n-1)(w_s-1)^2 + (w_s+n-1)(w_s-1-n)}{(w_s+n-1)^2}(x_s - \bar{x})(x_s - \bar{x})^T \end{aligned}$$

and therefore the likelihood equation $\partial l(\mu, \Sigma|w_s)/\partial \sigma_{ij} = 0$ gives

$$n(2-\delta_{ij})\sigma^{ij} = \operatorname{tr}\left\{\frac{\partial \Sigma}{\partial \sigma_{ij}}\Sigma^{-1}Q(w_s)\Sigma^{-1}\right\},$$

where

$$Q(w_s) = nS + \frac{n(w_s - 1)}{w_s + n - 1}(x_s - \bar{x})(x_s - \bar{x})^T.$$

Since $\operatorname{tr}\left\{(\partial\Sigma/\partial\sigma_{ij})\Sigma^{-1}Q(w_s)\Sigma^{-1}\right\} = (2-\delta_{ij}) \times (i,j)$ element of $\Sigma^{-1}Q(w_s)\Sigma^{-1}$, the perturbed MLE of σ^{ij} is the (i,j) element of $S(w_s)^{-1}Q(w_s)S(w_s)^{-1}/n$, when it is denoted by $S(w_s)$ the perturbed MLE of Σ, so that we have $S(w_s)^{-1} = S(w_s)^{-1}Q(w_s)S(w_s)^{-1}/n$. Thus the perturbed MLE of Σ turns out to be

$$S(w_s) = \frac{1}{n}Q(w_s) = S + \frac{w_s - 1}{w_s + n - 1}(x_s - \bar{x})(x_s - \bar{x})^T. \tag{1}$$

For a parameter θ of interest, we write its MLE and the corresponding perturbed MLE under the above perturbation scheme as $\hat{\theta}$ and $\hat{\theta}(w_s)$, respectively. When $w_s = 1$, the perturbed MLE $\hat{\theta}(w_s)$ reduces to the unperturbed MLE $\hat{\theta}$. De Gruttola et al.[3] suggested an influence measure, which they called the derivative influence, defined by

$$\left.\frac{\partial \hat{\theta}(w_s)}{\partial w_s}\right|_{w_s=1}$$

in order to investigate the influence of the sth observation x_s on the estimator $\hat{\theta}$. A large absolute value of the derivative influence implies high influence. This influence measure can be easily implemented in wide area of applications.

3. Cholesky root

The Cholesky decompositions of S and $S(w_s)$ can be expressed as

$$S = AA^T \text{ and } S(w_s) = A(w_s)A(w_s)^T,$$

respectively, where $A = (a_{ij})$ and $A(w_s) = (a_{ij}(w_s))$ are the lower triangular matrices with positive diagonal elements. Since $S(w_s)$ is the perturbed MLE of Σ, $S(w_s)$ is the perturbed counterpart of S. Further, by the analogy with the unperturbed case stated in Section 2, the Cholesky root $A(w_s)$ of $S(w_s)$ is the perturbed MLE of the Cholesky root of Σ. Hence the Cholesky root $A(w_s)$ of $S(w_s)$ is the perturbed counterpart of the Cholesky root A of S. The assumption of positive definiteness ensures the uniqueness of the Cholesky root.

In view of (1), when we put $z = (z_1, z_2, \cdots, z_p)^T = x_s - \bar{x}$, we can write

$$A(w_s)A(w_s)^T = AA^T + \left[\frac{\sqrt{w_s - 1}}{\sqrt{w_s + n - 1}} z\right] \left[\frac{\sqrt{w_s - 1}}{\sqrt{w_s + n - 1}} z\right]^T. \qquad (2)$$

In order to express the elements of $A(w_s)$ in terms of those of A, we need a variant of the usual Cholesky algorithm ([13], p.142) which can be easily derived as in what follows

$$a_{ii}(w_s) = \left[\frac{w_s - 1}{w_s + n - 1} z_i^2 + \sum_{k=1}^{i} a_{ik}^2 - \sum_{k=1}^{i-1} a_{ik}(w_s)^2\right]^{1/2} \quad (1 \le i \le p) \qquad (3)$$

$$a_{ij}(w_s) = a_{jj}(w_s)^{-1} \left[\frac{w_s - 1}{w_s + n - 1} z_i z_j + \sum_{k=1}^{j} a_{ik}a_{jk} - \sum_{k=1}^{j-1} a_{ik}(w_s)a_{jk}(w_s)\right] \qquad (4)$$

$$(1 \le j < i \le p).$$

Note that $a_{ij} = a_{ij}(w_s)|_{w_s=1}$. Let

$$\xi_{ij} = \left.\frac{da_{ij}(w_s)}{dw_s}\right|_{w_s=1} \quad (1 \le j \le i \le p).$$

Then ξ_{ij} is the derivative influence which measures the influence of x_s on the estimator a_{ij}. We will find ξ_{ij} recursively as follows. Differentiating both sides of (3) and (4) with respect to w_s and then evaluating the resulting equations at $w_s = 1$ yield

$$\sum_{k=1}^{i} a_{ik}\xi_{ik} = \frac{1}{2n} z_i^2 \quad (1 \le i \le p) \qquad (5)$$

$$\sum_{k=1}^{j} (a_{ik}\xi_{jk} + a_{jk}\xi_{ik}) = \frac{1}{n} z_i z_j \quad (1 \le j < i \le p). \qquad (6)$$

A procedure for computing ξ_{ij} from (5) and (6) can be described as follows.

For $i = 1, 2, \cdots, p$

(i) For $j = 1, 2, \cdots, i-1$

compute ξ_{ij} using (6)
(ii) compute ξ_{ii} using (5)

Equations (5) and (6) can be expressed in a matrix form as follows. We denote by K the lower triangular matrix whose (i,k)th element is given by ξ_{ik}. Then (5) and (6) can be collected in a matrix form

$$AK^T + KA^T = \frac{1}{n} zz^T. \tag{7}$$

This equation for the derivative influence is similar to that for the influence function given in the equation above (14) of [7] which includes one more term S in the left-hand side of (7). Since $K = \partial A(w_s)/\partial w_s|_{w_s=1}$, a direct differentiation of both sides of (2) with respect to w_s can also lead to (7).

4. Inverse Cholesky root

Let $B = (b_{ij})$ and $B(w_s) = (b_{ij}(w_s))$ be the upper triangular matrices such that

$$S^{-1} = BB^T \text{ and } S(w_s)^{-1} = B(w_s)B(w_s)^T.$$

Then we have $B^T = A^{-1}$ and $B(w_s)^T = A(w_s)^{-1}$, and $B(w_s)$ is the perturbed counterpart of B. Usually, B is called the inverse Cholesky root of S.

In order to derive the influence measure for the inverse Cholesky root of S, we will use the following equation

$$A(w_s)^T B(w_s) = I. \tag{8}$$

When we solve this equation with respect to $b_{ij}(w_s)$, we get

$$b_{kk}(w_s)a_{kk}(w_s) = 1 \; (1 \le k \le p) \tag{9}$$

$$b_{ik}(w_s)a_{ii}(w_s) = -\sum_{j=i+1}^{k} a_{ji}(w_s)b_{jk}(w_s) \; (1 \le i \le k-1). \tag{10}$$

Let

$$\eta_{ij} = \frac{db_{ij}(w_s)}{dw_s}\Big|_{w_s=1} \; (1 \le i \le j \le p).$$

Then η_{ij} is the derivative influence which measures the influence of x_s on the estimator b_{ij}. We will find η_{ij} recursively as follows. Since $b_{ij} = b_{ij}(w_s)|_{w_s=1}$, differentiating both sides of (9) and (10) with respect to w_s and then evaluating the resulting equations at $w_s = 1$ give

$$\eta_{kk} = -\frac{b_{kk}}{a_{kk}}\xi_{kk} \; (1 \le k \le p) \tag{11}$$

$$\eta_{ik} = -\frac{1}{a_{ii}}\left(\sum_{j=i}^{k} \xi_{ji}b_{jk} + \sum_{j=i+1}^{k} a_{ji}\eta_{jk}\right) \; (1 \le i \le k-1). \tag{12}$$

Using the identities $a_{kk}b_{kk} = 1$ $(k = 1, 2, \cdots, p)$, alternative expressions for (11) and (12) can be obtained. A procedure for computing η_{ik} from (11) and (12) can be described as follows.

For $k = p, p-1, \cdots, 1$
(i) compute η_{kk}
(ii) For $i = k-1, k-2, \cdots, 1$
compute η_{ik}

When we write as E the upper triangular matrix whose (i,j)th element is given by η_{ij}, (11) and (12) can be expressed in a matrix form

$$A^T E + K^T B = 0$$

which gives

$$E = -BK^T B.$$

This equation provides a relationship between the two influence measures ξ_{ij} and η_{ij}, and incidentally, this relationship has the same form as that for the case of the influence function given in (17) of [7]. Also, it can be derived directly from (8) by differentiating both sides of (8) with respect to w_s and then by evaluating the resulting equations at $w_s = 1$ since

$$K = \frac{\partial A(w_s)}{\partial w_s}\Big|_{w_s=1} \text{ and } E = \frac{\partial B(w_s)}{\partial w_s}\Big|_{w_s=1}.$$

Also, since $B^T A = A^T B = I$, the matrix form (7) can be expressed in terms of the inverse Cholesky root B instead of the Cholesky root A as follows

$$K^T B + B^T K = \frac{1}{n} B^T z z^T B.$$

5. A numerical example

For illustration, we consider the cost data set ([6], p.276) which consists of 36 measurements on the per-mile cost of three variables: fuel, repair and capital in this order.

For each $i (1 \le i \le p)$, we denote by A_i and B_i the leading principal submatrices of the Cholesky root A and the inverse Cholesky root B, respectively. Each A_i is the Cholesky root of the leading principal submatrix, having the same order i, of S, and $B_i^T = A_i^{-1}$. Hence, as noted in [7], the influence of observations on the $a_{ij} (1 \le j \le i)$ or the $b_{ji} (1 \le j \le i)$ depends only on the first i variables.

Figures 1 and 2 show index plots of the derivative influence of observations on the Cholesky root $A = (a_{ij})$ and the inverse Cholesky root $B = (b_{ij})$ of the sample covariance matrix S, respectively. Observations 9 and 21 have large influence on a_{11} and b_{11}. For the marginal distribution of the first variable, observations 9 and 21 are possible outliers based on the values of the standardized variable. Influential observations are 21, 9 and 20 for a_{21}; 20 for a_{22}, b_{12} and b_{22}. These observations are possible outliers based on Mahalanobis distances for

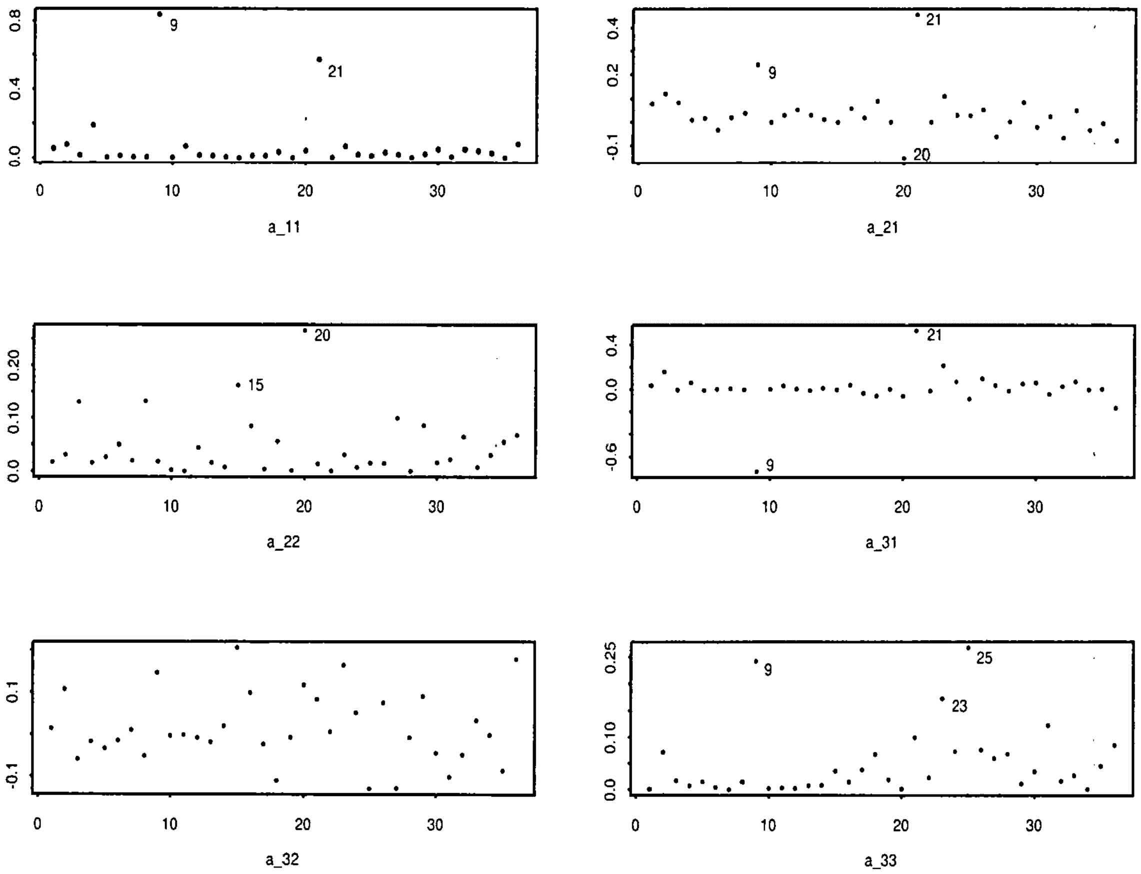

FIGURE 1. Index plots of the derivative influence on the Cholesky root

the bivariate distribution of the first two variables. Influential observations are 9 and 21 for a_{31} and b_{13}; none for a_{32} and b_{23}; 25 and 9 for a_{33} and b_{33}. For the joint distribution of all three variables, observations 9 and 21 are possible outliers based on Mahalanobis distances and more detailed analysis of the cost data can be found in [8] and references therein. Even though observations 9 and 21 are possible outliers, they do not have little influence on a_{32} and b_{23}. For a_{33} and b_{33}, influential observations are not identical to the corresponding outliers. In general, it is empirically well known that outliers need not be influential and influential observations need not be outliers ([2], p.95). But while outliers are often influential, not all influential observations need be outliers ([1], p.317).

REFERENCES

1. V. Barnett and T. Lewis, *Outliers in Statistical Data*, 3rd Ed., John Wiley & Sons, 1994

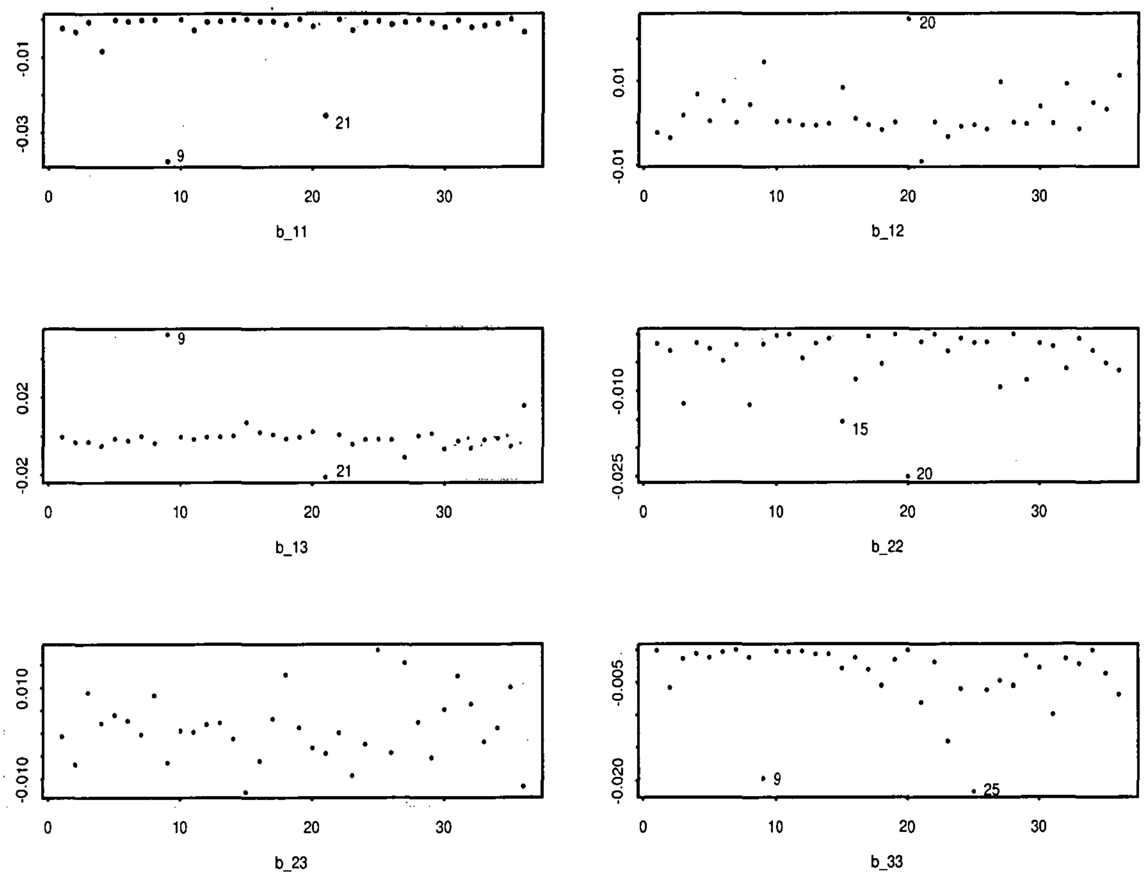

FIGURE 2. Index plots of the derivative influence on the inverse Cholesky root

2. S. Chatterjee and A. S. Hadi, *Sensitivity Analysis in Linear Regression*, Wiley, New York, 1988
3. V. De Gruttola, J. H. Ware and T. A. Louis, *Influence analysis of generalized least squares estimators*, Journal of the American Statistical Association, **82** (1987), 911–917
4. R. L. Dykstra, *Establishing the positive definiteness of the sample covariance matrix*, Annals of Mathematical Statistics, **41** (1970), 2153–2154
5. D. M. Hawkins and W. J. R. Eplett, *The Cholesky factorization of the inverse correlation or covariance matrix in multiple regression*, Technometrics, **24** (1982), 191–197
6. A. J. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Englewood Cliffs: Prentice-Hall, 1992
7. M. G. Kim, *Influence curve for the Cholesky root of a covariance matrix*, Communications in Statistics: Theory and Methods, **23** (1994), 1399–1412
8. M. G. Kim, *Multivariate outliers and decompositions of Mahalanobis distance*, Communications in Statistics: Theory and Methods, **29** (2000), 1511-1526
9. M. J. Lindstrom and D. M. Bates, *Newton-Raphson and EM algorithm for linear mixed-effects models for repeated measures data, Journal of the American Statistical Association*, **83** (1988), 1014–1022
10. J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, 1988

11. M. Pourahmadi, *Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation*, Biometrika, **86** (1999), 677-690
12. M. Pourahmadi, *Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters*, Biometrika, **94** (2007), 1006–1013
13. G. W. Stewart, *Introduction to Matrix Computations*, Academic Press, 1973

M.G. Kim received his Ph.D from Ohio State University. He is now a professor of Mathematics Education Department at Seowon University. His research interest centers on diagnostics in multivariate analysis and linear model.

Department of Mathematics Education, Seowon University, 231 Mochung-Dong, Heungduk-Gu, Cheongju, Chung-Buk, 361-742, Korea
e-mail : mgkim@seowon.ac.kr