Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.4 pp. 1007~1012

A Cholesky Decomposition of the Inverse of Covariance Matrix

Jong Tae Park[1) · Chul Kang[2)

Abstract

A recursive procedure for finding the Cholesky root of the inverse of sample covariance matrix, leading to a direct solution for the inverse of a positive definite matrix, is developed using the likelihood equation for the maximum likelihood estimation of the Cholesky root under normality assumptions. An example of the Hilbert matrix is considered for an illustration of the procedure.

KeyWords : Cholesky decomposition; Hilbert matrix; Maximum likelihood estimation

1. Introduction

Let Σ be a symmetric and positive definite covariance matrix of dimension p. The Cholesky decomposition theorem states that there exists a unique lower triangular matrix A, called a Chlosky root, with positive diagonal elements such that

$$\Sigma = AA^T.$$

We use T to denote transpose. The precision matrix, Σ^{-1}, is then expressed as

$$\Sigma^{-1} = BB^T$$

by putting $B^T = A^{-1}$. The Cholesky root B of the precision matrix, also called a Cholesky inverse root of Σ, is an upper triangular matrix with positive diagonal elements. The usual way to compute B is to first get A and then invert it.

---

1) Assistant Professor, Department of Information Statistics, Pyongtaek University, Pyongtaek, 450-701, Korea.
E-mail : jtpark@ptuniv.ac.kr

2) Associate Professor, Department of Computational Mathematics and Informatics, Hankyong National University, Ansung, 456-749, Korea.

The Cholesky decomposition has its wide use in finding the inverse of a symmetric and positive definite matrix, in solving linear systems of equations and in multivariate analysis. In particular, the Cholesky root of the inverse of sample covariance matrix is used for, for example, finding the inverse of $X^T X$, where X is a design matrix of full rank in multiple regression, solving the normal equations for generalized least squares problems, and analyzing a quadratic form like $x^T S^{-1} x$, where x is a random vector (see Tanabe and Sagae(1992); Hawkins and Eplett(1982); Maindonald(1977) for more details). Also, the Cholesky decomposition of the covariance matrix or correlation matrix is interesting and it also makes it easy to get some statistical parameters in a simple form and to interpret some parameters. For example, Olkin(1985) considered unbiased estimation of the Cholesky root of the covariance matrix. Hawkins and Eplett(1982) gave interpretations of the Cholesky root of the inverse correlation matrix of the dependent and independent variables in a multiple regression. Kim(1994) derived a recursive procedure for getting the influence curve for the Cholesky root of a covariance matrix with the help of the usual Cholesky algorithm. He also derived the influence curve for the inverse Cholesky root using the influence curves for the vector of regression coefficients and the residual variance in multiple regression.

When the underlying distribution is a p-variate normal one with positive definite covariance matrix Σ, Park(1996) considered a maximum likelihood estimation of the Cholesky root by estimating the upper triangular Cholesky inverse root B and found that the maximum likelihood estimator $\widehat{A}$ of the Cholesky root satisfies the identity

$$S = \widehat{A}\,\widehat{A}^T \tag{1}$$

or equivalently

$$S^{-1} = \widehat{B}\,\widehat{B}^T, \tag{2}$$

where S is the usual unbiased estimator of Σ. The hat notation is used to denote the maximum likelihood estimator. The corresponding likelihood equation is given by

$$S_i \widehat{\beta}_i = c_i \quad (i = 1, \cdots, p), \tag{3}$$

where S_i denotes the leading principal submatrix of S having order i, $\widehat{\beta}_i$ is the column vector $\widehat{b}_{ii}(\widehat{b}_{1i}, \cdots, \widehat{b}_{ii})^T$ of dimension i, $\widehat{b}_{ji}$ denotes the (j, i)th entry of $\widehat{B}$, and c_i is the column vector of dimension i having zeroes for the first $i-1$ elements and one for the last element. The form in (3) is one of the three forms of the likelihood equations that Park(1996) derived. In the article, he also showed that the likelihood procedure for finding $\widehat{A}$ in (1) is equivalent to the usual algorithm for the Cholesky decomposition in p.88 of Golub and Van

Loan(1983), and therefore conclude that the likelihood procedure above is applicable to (2) for any symmetric and positive definite matrix instead of the sample covariance matrix and that the procedure is independent of normality assumptions.

In this work we develop a procedure for finding the Cholesky root $\widehat{B}$ in (2) based on the entries of a symmetric and positive definite matrix S, using the likelihood equation (3).

2. Computing a Cholesky root of the inverse of sample covariance matrix

Let $A(k, l)$ and $\widehat{A}(k, l)$ $(k < l)$ denote the principal matrices formed by common entries in kth, $(k+1)$th, $\cdots$, lth rows and columns of A and $\widehat{A}$, respectively. The same notations are used for B and $\widehat{B}$. Then we get the following lemma.

Lemma 1. The principal matrix $A(k, l)$ is also a lower triangular matrix with positive diagonal elements and therefore nonsingular, and its maximum likelihood estimator is $\widehat{A}(k, l)$

Lemma 1 is true for $B(k, l)$ and $\widehat{B}(k, l)$ except for the upper triangularity of $B(k, l)$ and $\widehat{B}(k, l)$

Let a_i and b_i be the ith columns of A and B, respectively. Note that $a_{ji}=0$ for $j < i$ and $b_{ji}=0$ for $j > i$, where a_{ji} and b_{ji} denote the (j, i)th entries of A and B, respectively. Let I_p be the unit matrix of dimension p.

Lemma 2. The inverses of $A(k, l)$ and $\widehat{A}(k, l)$ $(k < l)$ are the transposes of $B(k, l)$ and $\widehat{B}(k, l)$ respectively.

Proof. Since $A^T B = I_p$, we have $a_i^T b_i = \delta_{ij}$, where δ_{ij} is the Kronecker delta. Then the result for parameter matrices comes from the fact that the inner product of the ith column of A and the jth column of B is equivalent to that of the $(i-k+1)$th column of $A(k, l)$ and the $(j-k+1)$th column of $B(k, l)$ for $k \le i,. \ j \le l$. We get the same result for the maximum likelihood estimators since $\widehat{A}^T \widehat{B} = I_p$.

Now we derive a procedure for finding the Cholesky root $\widehat{B}$ of a symmetric and positive definite matrix S^{-1} based on the entries of S from the maximum likelihood procedure (3). Let s_i be the ith column of S and let s_{ji} denote the (j, i)th element of S.

When $i=1$,

$$\hat{b}_{11}=s_{11}^{-1/2}.$$

For $2\leq i$, partitioning S_i into submatrices with S_{i-1} and s_{ii} as the block diagonal elements gives

$$|S_i|=|S_{i-1}|(s_{ii}-w_{i.i-1}^T S_{i-1}^{-1} w_{i.i-1}),$$

where $w_{j.i}=(s_{1j},\cdots,s_{ij})^T$. Let $\hat{B}_i$ be the leading principal submatrix of $\hat{B}$ having order i. Since $S_{i-1}^{-1}=\hat{B}_{i-1}\hat{B}_{i-1}^T$, we get

$$\hat{b}_{ii}=\frac{|S_{i-1}|^{1/2}}{|S_i|^{1/2}}=\{s_{ii}-\sum_{k=1}^{i-1}(\sum_{j=1}^{k}s_{ji}\hat{b}_{jk})^2\}^{-1/2}.$$

Since $\hat{A}=S\hat{B}$, we have the identity

$$\hat{a}_{ji}=s_j^T\hat{b}_i \quad (i\leq j). \tag{4}$$

Let $\hat{a}_{j(i)}=(\hat{a}_{j+1,j},\cdots,\hat{a}_{i-1,j})^T$ and $\hat{a}_{(j)i}=(\hat{a}_{i,j+1},\cdots,\hat{a}_{i,i-1})^T$. The column vectors $\hat{a}_{j(i)}$ and $\hat{a}_{(j)i}$ should be read as zeroes for $i=j+1$. For $j<i$, the likelihood equation (3) with Cramer's rule and (4) gives

$$\hat{b}_{ji}=-\hat{b}_{jj}\hat{b}_{ii}\sum_{k=1}^{j}s_{ki}\hat{b}_{kj}$$

when $j=i-1$, and otherwise

$$\begin{aligned}\hat{b}_{ji}&=-\hat{b}_{jj}\hat{b}_{ii}(\hat{a}_{ij}-\hat{a}_{j(i)}^T\hat{B}(j+1,i-1)\hat{a}_{(j)i})\\&=-\hat{b}_{jj}\hat{b}_{ii}\{\hat{a}_{ij}-\sum_{l=j+1}^{i-1}(\sum_{k=j+1}^{l}\hat{a}_{kj}\hat{b}_{kl})\hat{a}_{il}\}\\&=-\hat{b}_{jj}\hat{b}_{ii}[\sum_{k=1}^{j}s_{ki}\hat{b}_{kj}-\sum_{l=j+1}^{i-1}\{\sum_{k=j+1}^{l}(\sum_{m=1}^{l}s_{mi}\hat{b}_{ml})(\sum_{m=1}^{j}s_{mk}\hat{b}_{mj})\hat{b}_{kl}\}].\end{aligned}$$

3. An example

Consider the Hilbert matrix H of dimension p whose (i, j)th element is

$$h_{ij}=\frac{1}{i+j-1} \quad (1\leq i,\ \ j\leq p)$$

(see Forsythe and Moler(1967) for more details). It is well known that the Hilbert matrix is symmetric and positive definite, and that it is nearly singular for large dimension. The inverse of the Hilbert matrix of dimension 6 is found using the procedure described in Section 2. The computations below are performed on IBM

PC using Microsoft FORTRAN and double-precision floating-point arithmetic. The Cholesky root of H^{-1} is

$$\begin{pmatrix} 1 & -1.73 & 2.24 & -2.65 & 3 & -3.32 \\ 0 & 3.46 & -13.42 & 31.75 & -60 & 99.50 \\ 0 & 0 & 13.42 & -79.37 & 270 & -696.50 \\ 0 & 0 & 0 & 52.92 & -420 & 1857.31 \\ 0 & 0 & 0 & 0 & 210 & -2089.47 \\ 0 & 0 & 0 & 0 & 0 & 836.79 \end{pmatrix}.$$

Hence the inverse of the Hilbert matrix is

$$\begin{pmatrix} 36 & -630.0 & 3360.0 & -7560.0 & 7560.0 & -2772.0 \\ -630.0 & 14700.0 & -88200.0 & 211680.0 & -220500.0 & 83160.0 \\ 3360.0 & -88200.0 & 564480.0 & -1411200.0 & 1512000.0 & -582200.0 \\ -7560.0 & 211680.0 & -1411200.0 & 3628800.0 & -3969000.0 & 1552320.0 \\ 7560.0 & -220500.0 & 1512000.0 & -3969000.0 & 4410000.0 & -1746360.0 \\ -2772.0 & 83160.0 & -582120.0 & 1552320.0 & -1746360.0 & 698544.0 \end{pmatrix}.$$

We have used the same FORTRAN code depicted in p.83 of Forsythe and Moler(1967) to get an input data for the Hilbert matrix. The result above is much better than that by Forsythe and Moler(1967) and is almost equivalent to the matrix (19.5) in p.82 of their book. Hence this example shows that the procedure in Section 2 overcomes much of rounding errors caused by a machine.

4. Concluding remarks

The procedure developed in Section 2 is based on maximum likelihood estimation of a positive definite covariance matrix under normality assumptions. The likelihood procedure for finding the inverse Cholesky root of the sample covariance matrix is equivalent to the usual algorithm for the Cholesky decomposition, and therefore conclude that the procedure is applicable to any symmetric and positive definite matrix instead of the sample covariance matrix and that it is actually independent of normality assumptions. The procedure performs computations of the entries in the $(i+1)$th column of $\widehat{B}$ when the first i columns of $\widehat{B}$ have been completely formed. It does not need any pivoting and has numerical stability.

For further study we will generalize the proposed procedure for the UL decomposition of the inverse of a nonsingular matrix.

References

1. Forsythe, G. E. and Moler C. B.(1967). *Computer solution of linear algebraic systems*, Prentice-Hall, Englehood Cliffs, N.J.

2. Golub, G. H. and C. Van Loan(1983). *Matrix computations*, The Johns Hopkins University Press, Baltimore, Maryland.
3. Hawkins, D. M. and W. J. R. Eplett(1982). The Cholesky factorization of the inverse correlation or covariance matrix in multiple regression, *Technometrics*, 24, 191-198.
4. Kim, M. G.(1994). Influence curve for the Cholesky root of a covariance matrix, *Communications in Statistics-Theory and Methods*, 23(5), 1399-1412.
5. Maindonald, J. H.(1977). Least squares computations based on the Cholesky decomposition of the correlation matrix, *Journal of Statistical Computation and Simulation*, 5, 247-258.
6. Olkin, I.(1985). Estimating a Cholesky decomposition, *Linear Algebra and Its Applications*, 67, 201-205.
7. Park, J. T.(1996). On computing a Cholesky decomposition, *The Korean Communications in Statistics*, 3(2), 37-42.
8. Tanabe, K. and M. Sagae(1992). An exact Cholesky decomposition and the generalized inverse of the variance-covariance matrix of the multinomial distribution, with applications, *Journal of Royal Statistical Society*, B, 54, 211-219.

[received date : Sep. 2003, accepted date : Nov. 2003]