한국통계학회논문집 제3권 2호
1996년 8월 pp. 37-42

On Computing a Cholesky Decomposition

Jong Tae Park[1)]

Abstract

Maximum likelihood estimation of a Cholesky decomposition is considered under normality assumptions. It is shown that maximum likelihood estimation gives a Cholesky decomposition of the sample covariance matrix. The joint distribution of the maximum likelihood estimators is derived. The usual algorithm for a Cholesky decomposition is shown to be equivalent to a maximum likelihood estimation of a Cholesky root when the underlying distribution is a multivariate normal one.

1. Introduction

The Cholesky decomposition theorem ensures that for a p by p positive definite and symmetric matrix Σ, there exists a unique lower triangular matrix A, which is called a Cholesky root, with positive diagonal elements such that $\Sigma = AA^T$ (Anderson, 1984, p.586). In general, the Cholesky root of the sample covariance matrix is not an unbiased estimator of A. Olkin (1985), however, obtained an unbiased estimator of A by adjusting the coefficients of each column of the Cholesky root. Eaton and Olkin (1987) showed that the Cholesky root of the sample covariance matrix multiplied in the right by a diagonal matrix is a best equivariant estimator of A for a variety of loss functions.

In this work maximum likelihood estimation of the Cholesky root of the covariance matrix is considered under normality assumptions and its properties are investigated. It is also shown that the usual algorithm for finding a Cholesky root of the sample covariance matrix is equivalent to a maximum likelihood estimation of a Cholesky root of the population covariance matrix when the underlying distribution is a multivariate normal one.

1) Full-time Lecturer, Department of Computer Science & Statistics, Pyongtaek University, Pyongtaek, 450-701, Korea.

2. Maximum likelihood estimation

Let X be a p-variate random vector distributed according to a multivariate normal distribution $N(\mu,\Sigma)$ with mean vector μ and positive definite covariance matrix Σ. Let S be the usual unbiased estimator of Σ based on a sample of size N and let $n=N-1$. Then nS has a Wishart distribution $W_p(\Sigma, n)$ so that the likelihood function of Σ given S is

$$L(\Sigma)=|\Sigma|^{-n/2}\exp\{-\frac{n}{2}\mathrm{tr}(\Sigma^{-1}S)\} ,$$

when ignoring the constant terms not depending on Σ. Let A be a unique lower triangular matrix with positive diagonal elements such that $\Sigma=AA^T$. We cannot find directly the maximum likelihood estimator of A because the likelihood function includes its inverse. This difficulty can be avoided with a help of the invariance property of the maximum likelihood estimators (Zehna, 1966). Putting $B^T=A^{-1}$ gives $\Sigma^{-1}=BB^T$ and B^T is also a lower triangular matrix. Since all the products of the ith diagonal elements of A and B^T for $i=1,\dots,p$ are one, B should be an upper triangular matrix with positive diagonal elements. Let b_{ij} denote the (i,j)th element of B and b_i the ith column of B. Since

$$|\Sigma|^{-1} = |B|^2 = \prod_{i=1}^{p} b_{ii}^2$$

and

$$\mathrm{tr}(\Sigma^{-1}S) = \mathrm{tr}(B^TSB) = \sum_{i=1}^{p} \mathbf{b}_i^T S\, \mathbf{b}_i,$$

the log-likelihood function is

$$l(\Sigma) = \frac{n}{2}\{2\sum_{i=1}^{p}\log(b_{ii})-\sum_{i=1}^{p} b_i^T S\, b_i\}.$$

Let s_i denote the ith column of the sample covariance matrix S and s_{ij} the (i,j)th element of S. Taking partial derivatives of $l(\Sigma)$ with respect to b_{ii} and setting them equal to zero gives

$$\frac{1}{\hat{b}_{ii}} = s_i^T\, \hat{b}_i \quad (i=1,\dots,p). \tag{1}$$

Since $b_{ji}=0$ for $i<j$, the result obtained by taking partial derivatives of $l(\Sigma)$ with respect to b_{ji} $(i>j)$ and by setting them equal to zero can be written as

$$s_j^T\, \hat{b}_i = 0 \quad (1\le j< i\le p). \tag{2}$$

The vector $\hat{b}_i$ has zeros for the last $p-i$ elements. Hence the likelihood equations (1) and (2) can be rearranged in a simple form like

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

where S_i denotes the leading principal submatrix of S having order i, $\hat{\beta}_i$ is the column vector $\hat{b}_{ii}(\hat{b}_{1i},\dots,\hat{b}_{ii})^T$ of dimension i, and c_i is the column vector of dimension i having zeros for the first $i-1$ elements and one for the last element. Thus the maximum likelihood estimators $\hat{b}_i$ are found recursively through the $\hat{\beta}_i$.

3. Some properties of a Cholesky decomposition

Let Σ_i [$\hat{\Sigma}_i$] , A_i [$\hat{A}_i$] , and B_i [$\hat{B}_i$] be the leading principal submatrices of Σ [$\hat{\Sigma}$] , A [$\hat{A}$] , B [$\hat{B}$] , respectively, each having order i. A matrix with subscript $-i$ indicates a submatrix obtained by deleting the first $(p-i)$ rows and columns.

Lemma 1. If Σ [$\hat{\Sigma}$] has a Cholesky decomposition as

$$\Sigma = AA^T \quad [\ \hat{\Sigma} = \hat{A}\hat{A}^T]$$

or equivalently

$$\Sigma^{-1} = BB^T \quad [\ \hat{\Sigma}^{-1} = \hat{B}\hat{B}^T] \ ,$$

then we have for $i=1,\dots,p$

$$\Sigma_i = A_i A_i^T \quad [\ \hat{\Sigma}_i = \hat{A}_i \ \hat{A}_i^T]$$

or equivalently

$$(\Sigma^{-1})_{-i} = B_{-i}B_{-i}^T \quad [\ (\hat{\Sigma}^{-1})_{-i} = \hat{B}_{-i} \ \hat{B}_{-i}^T] \ .$$

Furthermore $B_i^T = A_i^{-1}$ [$\hat{B}_i^T = \hat{A}_i^{-1}$] and $B_i^{-1} = (B^{-1})_i$ [$\hat{B}_i^{-1} = (\hat{B}^{-1})_i$] for $i=1,\dots,p$.

Lemma 1 is easily proved by partitioning the corresponding matrices and using the invariance property of the maximum likelihood estimators. We are sometimes interested in a marginal distribution and in such a case Lemma 1 will be useful. A choice of a subset of variables needs consecutive reordering so that selected variables occupy labels in the first part.

Let $S_{i,j}$ be the $(i-1)$ by $(i-1)$ submatrix obtained by deleting the ith row and the jth column of S_i for $1 \le j \le i \le p$. Note that $S_{i,i} = S_{i-1}$. From the likelihood equations (1) and (2) together with Cramer's rule, we get relationships among the maximum likelihood estimators and cofactors of the leading principal submatrices as in the following lemma.

Lemma 2. For $1 \le j \le i \le p$, the likelihood equations (1) and (2) imply

$$\hat{b}_{11}^{\,2} = \frac{1}{s_{11}} \quad (i=j=1) \tag{4}$$

and

$$\hat{b}_{ji}\ \hat{b}_{ii} = (-1)^{i+j} \frac{|S_{i,j}|}{|S_i|} \quad (i>1). \tag{5}$$

The likelihood equations (1) and (2) yield directly that

$$\hat{B}^T S \hat{B} = I_p$$

and therefore

$$S = \hat{A}\,\hat{A}^T. \tag{6}$$

Hence the maximum likelihood estimation under normality assumptions gives the usual Cholesky decomposition of the sample covariance matrix.

Now we know that the sample covariance matrix has a unique Cholesky root consisting of the maximum likelihood estimators. Since $S = \hat{A}\,\hat{A}^T$, the joint distribution of the $\hat{a}_{ji}$ $(i \le j)$ is that of the usual Cholesky root of the sample covariance matrix and which is given by

$$\frac{n^{\frac{pn}{2}} \left(\prod_{i=1}^{p} \hat{a}_{ii}^{\,n-i}\right) \exp\{-\frac{n}{2} \operatorname{tr}(\mathrm{A}^{-1}\, \hat{\mathrm{A}}\, \hat{\mathrm{A}}^{\mathrm{T}} \mathrm{A}^{-\mathrm{T}})\}}{2^{\frac{p(n-2)}{2}}\, \Pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \{a_{ii}^{n}\, \Gamma(\frac{n-i+1}{2})\}}$$

after a simple algebra of calculations(see Anderson, 1984, Theorem 7.2.1).

Thus the joint distribution of the $\hat{b}_{hg}$ $(h \le g)$ can be obtained using that of the $\hat{a}_{ji}$. The partial derivatives of $\hat{a}_{ji}$ $(i \le j)$ with respect to $\hat{b}_{hg}$ $(h \le g)$ are

$$\frac{\partial\, \hat{a}_{ji}}{\partial\, \hat{b}_{hg}} = \begin{cases} -\ \hat{a}_{hi}\ \hat{a}_{jg} & \text{for } i \le h \le g \le j \\ 0 & \text{otherwise} \end{cases}$$

so that the corresponding Jacobian is easily found to be $\prod_{i=1}^{p} \hat{a}_{ii}^{\,p+1}$. Since $|\Sigma|^{\frac{1}{2}} = \prod_{i=1}^{p} \hat{a}_{ii} = 1/\prod_{i=1}^{p} \hat{b}_{ii}$, the joint distribution of the $\hat{b}_{ji}$ $(j \le i)$ is given by

$$\frac{n^{\frac{pn}{2}} \left(\prod_{i=1}^{p} b_{ii}^{n}\right) \exp\{-\frac{n}{2}\,\mathrm{tr}(\mathrm{B}^{\mathrm{T}}\,\hat{\mathrm{B}}^{-\mathrm{T}}\,\hat{\mathrm{B}}^{-1}\mathrm{B})\}}{2^{\frac{p(n-2)}{2}}\,\pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p}\{\,\hat{b}_{ii}^{\,n-i+p+1}\,\Gamma(\frac{n-i+1}{2})\}}.$$

4. Equivalence of the likelihood procedure and the usual Cholesky decomposition

The usual algorithm for finding the lower triangular Cholesky root of S is based on a direct comparison of the entries in the equation (6) (Golub and Van Loan, 1983, p.88). Now we will show that the usual algorithm is equivalent to a maximum likelihood estimation approach. The equation (6) can be expressed as

$$\hat{A} = S\hat{B}$$

which gives

$$\hat{a}_{ji} = s_j^{T}\,\hat{\boldsymbol{b}}_i \quad (i \le j). \tag{7}$$

This complements the likelihood equation (2). Let $\hat{B}_j$ be the leading principal submatrix of $\hat{B}$ having order j and let $\boldsymbol{w}_{j.i} = (s_{1j}, \cdots, s_{ij})^{T}$. Since $\hat{a}_{jj}\,\hat{b}_{jj} = 1$ $(1 \le j \le p)$, the equation (4) gives

$$\hat{a}_{11} = s_{11}^{1/2}$$

and from (5) together with the symmetry of S, we have for $j \le 2$

$$\begin{aligned}\hat{a}_{jj} &= \frac{|S_j|^{1/2}}{|S_{j-1}|^{1/2}} \\ &= (s_{jj} - \boldsymbol{w}_{j.j-1}^{T} S_{j-1}^{-1}\,\boldsymbol{w}_{j.j-1})^{1/2} \\ &= (s_{jj} - \boldsymbol{w}_{j.j-1}^{T}\,\hat{B}_{j-1}\,\hat{B}_{j-1}^{T}\,\boldsymbol{w}_{j.j-1})^{1/2} \\ &= (s_{jj} - \sum_{i=1}^{j-1} \hat{a}_{ji}^{2})^{1/2} \quad \text{by (7).}\end{aligned}$$

Let $\hat{\boldsymbol{\beta}}_{i.j} = \hat{b}_{ii}(\hat{b}_{1i}, \cdots, \hat{b}_{ji})^T$. For $j > i$, we have

$$\begin{aligned}(\boldsymbol{s}_j^T \hat{\boldsymbol{b}}_i)\hat{b}_{ii} &= \boldsymbol{w}_{j.i-1}^T \hat{\boldsymbol{\beta}}_{i.i-1} + s_{ij} \hat{b}_{ii}^2 \\ &= (-\boldsymbol{w}_{j.i-1}^T S_{i-1}^{-1} \boldsymbol{w}_{i.i-1} + s_{ij}) \hat{b}_{ii}^2 \quad \text{by (3)} \\ &= \hat{b}_{ii}^2 (s_{ji} - \sum_{k=1}^{i-1} \hat{a}_{ik} \hat{a}_{jk}).\end{aligned}$$

Hence we get

$$\hat{a}_{ji} = \frac{1}{\hat{a}_{ii}} (s_{ji} - \sum_{k=1}^{i-1} \hat{a}_{ik} \hat{a}_{jk}) \quad (j > i).$$

Thus we can conclude that the usual algorithm for finding a Cholesky root of the sample covariance matrix is equivalent to a maximum likelihood estimation of a Cholesky root of the population covariance matrix under normality assumptions.

References

[1] Anderson, T.W. (1984). *An introduction to multivariate statistical analysis*, 2nd Ed. New York: Wiley.

[2] Eaton, M.L. and Olkin, I. (1987). Best equivariant estimators of a Cholesky decomposition, *Annals of Statistics*, Vol. 15, 1639-1650.

[3] Golub, G.H. and Van Loan, C. (1983). *Matrix computations*, The Johns Hopkins University Press, Baltimore, Maryland.

[4] Olkin, I. (1985). Estimating a Cholesky decomposition, *Linear Algebra and Its Applications*, Vol. 67, 201-205.

[5] Zehna, P.W. (1966). Invariance of maximum likelihood estimation, *Annals of Mathematical Statistics*, Vol. 37, 744.