Statistica Sinica **11**(2001), 855-862

EXACT DISTRIBUTION OF THE MLE OF CONCENTRATION MATRICES IN DECOMPOSABLE COVARIANCE SELECTION MODELS

Jing-Long Wang

East China Normal University

Abstract: Covariance selection models were introduced by Dempster (1972). The covariance selection model with a decomposable graph is called a decomposable covariance selection model. Based on the hyper-Markov property (Dawid and Lauritzen (1993)), the exact distribution of the Maximum Likelihood Estimator (MLE) of the concentration matrix in the decomposable covariance selection model is given.

Key words and phrases: Covariance selection model, decomposable graph, inverted Wishart distribution, multivariate normal model, Wishart distribution.

1. Introduction

In a graphical Markov model, statistical variables are represented by the vertices of a graph. For a given graph D with vertex set V, let X_v denote the variable represented by the vertex $v \in V$, and $X = \times(X_v \mid v \in V)$, a p-dimensional random vector, where $p = |V|$, the cardinality of the set $|V|$. For a subset $A \subset V$, let $X_A = \times(X_v \mid v \in A)$. Two vertices α and β are said to be neighbours if they are connected by an edge. A path from α to β is a sequence $\alpha = \alpha_0, \ldots, \alpha_n = \beta$ of distinct vertices such that α_{i-1} and α_i are neighbors, $i = 1, \ldots, n$. For disjoint subsets A, B and C of V, C is said to separate A from B if every path from any $\alpha \in A$ to any $\beta \in B$ must intersect C. The random vector X is said to possess the local Markov property if each variable is conditionally independent of all other variables given its neighbours. It is said to possess the global Markov property if X_A is conditionally independent of X_B given X_C ($X_A \perp\!\!\!\perp X_B \mid X_C$) for any triple (A, B, C) of disjoint subsets of V with C separating A from B. The notation of conditional independence is due to Dawid (1979, 1980). Obviously the global Markov property implies the local Markov property. The converse is also true if the distribution of X has a positive and continuous density with respect to Lebesgue measure. The graphical Markov models are surveyed in books by Whittaker (1990), Edwards (1995), Lauritzen (1996) and Cox and Wermuth (1996).

Covariance selection models were introduced by Dempster (1972) for parsimonious estimation of Gaussian covariance matrices. The interpretation of these models in terms of conditional independence structures was given by Wermuth (1976) and studied further by Speed and Kiiveri (1986). The hyper-Markov property, or hyper-Wishart law for decomposable graphical models was introduced by Dawid and Lauritzen (1993). Bayesian model search within different model classes, including decomposable models, has been advocated by Madigan and Raftery (1994). Guidici and Green (1999) use MCMC to obtain the posterior distribution of the estimated covariance matrix. A covariance selection model is a graphical Markov model when the variables follow a multivariate normal distribution. Here, we suppose that the covariance matrix Σ is positive definite. Thus the local Markov property and global Markov property are equivalent. The covariance selection model with a decomposable graph is called a decomposable covariance selection model.

In a graph D, a subset C is complete if any two vertices of C are joined by an edge, and a complete subset that is maximal with respect to $\subseteq$ is called a clique. A graph D is said to be decomposable if all cliques of D can be numbered to form a sequence, $C_1, \ldots, C_m$, which satisfies the following condition: if $H_i = C_1 \cup \cdots \cup C_i, R_i = C_i \setminus H_{i-1}, S_i = H_{i-1} \cap C_i,\ G_i = H_{i-1} \setminus C_i$, then for all $i = 2, \ldots, m$, there is a $j < i$ such that $S_i \subseteq C_j$, and S_i separates R_i from G_i (Dawid and Lauritzen (1993)). Such a sequence is said to be perfect. Note that the separators $S_2, \ldots, S_m$ are not necessarily disjoint, and some of them may be either identical or empty. Fast algorithms exist to find the desired numbering. According to the global Markov property, we know that $X_{R_i} \perp\!\!\!\perp X_{G_i} \mid X_{S_i},\ \ i = 2, \ldots, m$.

Let $K = (\Sigma)^{-1}$, the concentration matrix. The multivariate normal can be transformed to an exponential model. Thus $\theta = (K, h)$ can be said to be the canonical parameter, where $h = K\xi$, ξ is the mean vector. In the decomposable covariance selection model, we have (Lauritzen (1996))

$$K = \sum_{i=1}^{m} [(\Sigma_{C_iC_i})^{-1}]^p - \sum_{i=2}^{m} [(\Sigma_{S_iS_i})^{-1}]^p, \tag{1}$$

where, for any set $A \subseteq V$, Σ_{AA} is a submatrix of Σ, and $[(\Sigma_{AA})^{-1}]^p$ denotes the $p \times p$ matrix obtained from $(\Sigma_{AA})^{-1} = \{a_{\gamma\mu}\}$ as $[(\Sigma_{AA})^{-1}]^p = \{b_{\gamma\mu}\}$ with

$$b_{\gamma\mu} = \begin{cases} a_{\gamma\mu}, & \text{if } \gamma \in A,\ \mu \in A, \\ 0, & \text{otherwise.} \end{cases}$$

We have

$$\det K = \frac{\prod_{i=2}^{m} |\Sigma_{S_iS_i}|}{\prod_{i=1}^{m} |\Sigma_{C_iC_i}|}. \tag{2}$$

The decomposable covariance selection model is considered with a sample $(X^1,\ldots,X^n)$. The maximum likelihood estimate (MLE) of Σ (or equivalently, the MLE of concentration matrix $K=\Sigma^{-1}$) exists with probability one if and only if $n>\max_{1\le i\le m} p_i$, where $p_i=|C_i|$. Then

$$\hat{\Sigma}=\hat{K}^{-1},\quad \hat{K}=n\{\sum_{i=1}^{m}[(ssd_{C_iC_i})^{-1}]^p-\sum_{i=2}^{m}[(ssd_{S_iS_i})^{-1}]^p\},\tag{3}$$

where, for any set $A\subseteq V$, $ssd_{AA}=\sum_{j=1}^{n}(X_A^j-\bar{X}_A)(X_A^j-\bar{X}_A)^t$, $\bar{X}$ is the sample mean. Note that $ssd_{C_iC_i}/n$ and $ssd_{S_iS_i}/n$ are submatrices of $\hat{\Sigma}$.

We have that $ssd_{R_iR_i}\perp\!\!\!\perp ssd_{G_iG_i}\mid ssd_{S_iS_i}, i=2,\ldots,m$. This is the so called hyper-Markov property, or hyper-Wishart law for the decomposable covariance selection model (Dawid and Lauritzen (1993)).

Obviously, ssd_{AA} is distributed according to Wishart distribution $W(\Sigma_{AA}, n-1)$. Based on the hyper-Markov property, the following recursive operation can be performed: $p(ssd_{C_1C_1})=w(ssd_{C_1C_1}|\Sigma_{C_1C_1},n-1)$, and for any $j=2,\ldots,m$

$$p(ssd_{C_1C_1}\cup\cdots\cup ssd_{C_jC_j})=\frac{p(ssd_{C_1C_1}\cup\cdots\cup ssd_{C_{j-1}C_{j-1}})w(ssd_{C_jC_j}|\Sigma_{C_jC_j},n-1)}{w(ssd_{S_jS_j}|\Sigma_{S_jS_j},n-1)}.$$

Here for any j, $p(ssd_{C_1C_1}\cup\cdots\cup ssd_{C_jC_j})$ denotes the density of $ssd_{C_1C_1}\cup\cdots\cup ssd_{C_jC_j}$ and, for any set $A\subseteq V$, $w(ssd_{AA}|\Sigma_{AA},n-1)$ denotes the density of $W(\Sigma_{AA},n-1)$,

$$w(ssd_{AA}|\Sigma_{AA},n-1)=\frac{|ssd_{AA}|^{\frac{1}{2}(n-q-2)}\exp\{-\frac{1}{2}tr[(\Sigma_{AA})^{-1}ssd_{AA}]\}}{2^{\frac{1}{2}(n-1)q}|\Sigma_{AA}|^{\frac{1}{2}(n-1)}\Gamma_q(\frac{n-1}{2})},\tag{4}$$

where $q=|A|$ and $\Gamma_q(t)=\pi^{\frac{1}{4}q(q-1)}\prod_{i=1}^{q}\Gamma[t-\frac{1}{2}(i-1)]$. Then the density of $(ssd_{C_1C_1}\cup\cdots\cup ssd_{C_mC_m})$ can be obtained,

$$p(ssd_{C_1C_1}\cup\cdots\cup ssd_{C_mC_m})=\frac{\prod_{i=1}^{m}w(ssd_{C_iC_i}|\Sigma_{C_iC_i},n-1)}{\prod_{i=2}^{m}w(ssd_{S_iS_i}|\Sigma_{S_iS_i},n-1)}.\tag{5}$$

We know that $(ssd_{AA})^{-1}$ is distributed according to the inverted Wishart distribution $W^{-1}\left((\Sigma_{AA})^{-1},n-1\right)$, and has density

$$w^{-1}((ssd_{AA})^{-1}|(\Sigma_{AA})^{-1},n-1)=\frac{|ssd_{AA}|^{\frac{1}{2}(n+q)}\exp\{-\frac{1}{2}tr[(\Sigma_{AA})^{-1}ssd_{AA}]\}}{2^{\frac{1}{2}(n-1)q}|\Sigma_{AA}|^{\frac{1}{2}(n-1)}\Gamma_q(\frac{n-1}{2})}.\tag{6}$$

In Section 2, the density of the MLE $\hat{K}$ of the concentration matrix K in decomposable covariance selection models will be given, along with a discussion on the conjugate prior distribution of K and the hyper-Markov property of the distribution of $\hat{K}$.

Section 2 involves extensive matrix operations. The reader is referred to Giri (1977) and Muirhead (1982) for details.

2. Density of MLE of Concentration Matrix

Since $S_i \subset C_i$, $ssd_{S_iS_i}$ is a submatrix of $ssd_{C_iC_i}$. Therefore (3) implies that $\hat{K}$ is a function of $(ssd_{C_1C_1} \cup \cdots \cup ssd_{C_mC_m})$, $ssd_{C_1C_1} \cup \cdots \cup ssd_{C_mC_m} \longrightarrow \hat{K}$. For any j, $ssd_{C_jC_j}$ is a submatrix of $n(\hat{K})^{-1}$ so this transformation is invertible. The number of variables in both $\hat{K}$ and $ssd_{C_1C_1} \cup \cdots \cup ssd_{C_mC_m}$ is

$$r = \sum_{i=1}^{m} \frac{p_i(p_i+1)}{2} - \sum_{i=2}^{m} \frac{q_i(q_i+1)}{2},$$

where $p_i = |C_i|$ and $q_i = |S_i|$.

The density of $ssd_{C_1C_1} \cup \cdots \cup ssd_{C_mC_m}$ is given by (5). Hence the key to finding the density of $\hat{K}$ is to get the Jacobian $J = \frac{\partial \hat{K}}{\partial(ssd_{C_1C_1} \cup \cdots \cup ssd_{C_mC_m})}$. From (3), we have $d\hat{K} = n\{\sum\limits_{i=1}^{m} [d\{(ssd_{C_iC_i})^{-1}\}]^p - \sum_{i=2}^{m} [d\{(ssd_{S_iS_i})^{-1}\}]^p\}$. Since $dA^{-1} = -A^{-1} \cdot dA \cdot A^{-1}$, we have

$$\begin{aligned} d\hat{K} = n\{&\sum_{i=1}^{m} [-(ssd_{C_iC_i})^{-1} \cdot d(ssd_{C_iC_i}) \cdot (ssd_{C_iC_i})^{-1}]^p \\ &- \sum_{i=2}^{m} [-(ssd_{S_iS_i})^{-1} \cdot d(ssd_{S_iS_i}) \cdot (ssd_{S_iS_i})^{-1}]^p\}. \end{aligned} \tag{7}$$

As $d\{ssd_{S_iS_i}\}$ is a submatrix of $d\{ssd_{C_iC_i}\}$, (7) implies that $d\{\hat{K}\}$ is a function of $d\{ssd_{C_1C_1}\} \cup \cdots \cup d\{ssd_{C_mC_m}\}$. We know that the Jacobian of the transformation, $A \longrightarrow B = B(A)$, is equal to the Jacobian of the transformation $dA \longrightarrow dB = dB(dA)$. Hence the desired Jacobian is the Jacobian of $d\{ssd_{C_1C_1}\} \cup \cdots \cup d\{ssd_{C_mC_m}\} \longrightarrow d\{\hat{K}\}$ obtained from (7). Then

$$J = -n\{\sum_{i=1}^{m} [J_{C_i}]^r - \sum_{i=2}^{m} [J_{S_i}]^r\}, \tag{8}$$

where J_{C_i} and J_{S_i} are the following Jacobians:

$$J_{C_i} = \frac{\partial\{(ssd_{C_iC_i})^{-1} \cdot d(ssd_{C_iC_i}) \cdot (ssd_{C_iC_i})^{-1}\}}{\partial\{d(ssd_{C_iC_i})\}},$$

$$J_{S_i} = \frac{\partial\{(ssd_{S_iS_i})^{-1} \cdot d(ssd_{S_iS_i}) \cdot (ssd_{S_iS_i})^{-1}\}}{\partial\{d(ssd_{S_iS_i})\}}.$$

In fact,

$$J_{C_i} = -\frac{\partial\{d(ssd_{C_iC_i})^{-1}\}}{\partial\{d(ssd_{C_iC_i})\}}, \qquad J_{S_i} = -\frac{\partial\{d(ssd_{S_iS_i})^{-1}\}}{\partial\{d(ssd_{S_iS_i})\}}.$$

Then

$$\begin{aligned} J_{C_i}^{-1} &= -\frac{\partial\{d(ssd_{C_iC_i})\}}{\partial\{d(ssd_{C_iC_i})^{-1}\}} \\ &= \frac{\partial\{ssd_{C_iC_i} \cdot d\{(ssd_{C_iC_i})^{-1}\} \cdot ssd_{C_iC_i}\}}{\partial\{d(ssd_{C_iC_i})^{-1}\}}, \qquad (9) \\ J_{S_i}^{-1} &= -\frac{\partial\{d(ssd_{S_iS_i})\}}{\partial\{d(ssd_{S_iS_i})^{-1}\}} \\ &= \frac{\partial\{ssd_{S_iS_i} \cdot d\{(ssd_{S_iS_i})^{-1}\} \cdot ssd_{S_iS_i}\}}{\partial\{d(ssd_{S_iS_i})^{-1}\}}. \qquad (10) \end{aligned}$$

Because the sequence $C_1, \ldots, C_m$ is perfect, (9) and (10) imply that $J_{S_i}^{-1}$ is a submatrix of $J_{C_i}^{-1}$, and there is a $j < i$ such that $J_{S_i}^{-1}$ is a submatrix of $J_{C_j}^{-1}$. Obviously, the (8) can be written as

$$J = -n\{\sum_{i=1}^{m} [((J_{C_i})^{-1})^{-1}]^r - \sum_{i=2}^{m} [((J_{S_i})^{-1})^{-1}]^r\}, \qquad (11)$$

which is analogous to (1). Similarly to (2) derived from (1), we get, from (11), the absolute determinant value of Jacobian J,

$$|J|_+ = n^r \frac{\prod_{i=2}^m |(J_{S_i})^{-1}|_+}{\prod_{i=1}^m |(J_{C_i})^{-1}|_+} = n^r \frac{\prod_{i=1}^m |J_{C_i}|_+}{\prod_{i=2}^m |J_{S_i}|_+}.$$

For a $q \times q$ symmetric matrix S and a $q \times q$ non-singular matrix C, it is known that the absolute determinant value of the Jacobian of the transformation $S \longrightarrow CSC^t$ is

$$\left|\frac{\partial(CSC^t)}{\partial(S)}\right|_+ = |C|^{q+1}.$$

Hence, $|J_{C_i}|_+ = |(ssd_{C_iC_i})^{-1}|^{p_i+1}$, $|J_{S_i}|_+ = |(ssd_{S_iS_i})^{-1}|^{q_i+1}$. Then

$$|J|_+ = n^r \frac{\prod_{i=2}^m |ssd_{S_iS_i}|^{q_i+1}}{\prod_{i=1}^m |ssd_{C_iC_i}|^{p_i+1}}.$$

From (5) and (4), the density of $\hat{K}$ is

$$\begin{aligned} p(\hat{K}) &= \frac{\prod_{i=1}^m w(ssd_{C_iC_1}|\Sigma_{C_iC_i}, n-1)}{\prod_{i=2}^m w(ssd_{S_1S_i}|\Sigma_{S_iS_i})} \cdot (|J|_+)^{-1} \\ &= \frac{1}{n^r} \cdot \frac{\prod_{i=1}^m w^{-1}((ssd_{C_iC_1})^{-1}|(\Sigma_{C_iC_i})^{-1}, n-1)}{\prod_{i=2}^m w^{-1}((ssd_{S_1S_i})^{-1}|(\Sigma_{S_iS_i})^{-1}, n-1)}, \qquad (12) \end{aligned}$$

where $ssd_{C_iC_i}$ and $ssd_{C_iC_i}$ are submatrices of $n(\hat{K})^{-1}$.

Remark on conjugate priors for Bayesian inference. In the decomposable covariance selection model, a conjugate prior distribution of covariance matrix Σ, or strictly speaking, a conjugate prior distribution of $(\Sigma_{C_1C_1} \cup \cdots \cup \Sigma_{C_mC_m})$ has been given by Dawid and Lauritzen (1993). They introduce the so called hyper-inverse Wishart law: for any i $\Sigma_{C_iC_i} \sim W^{-1}(\Psi_{C_iC_i}, \nu + p_i)$, $\Sigma_{S_iS_i} \sim W^{-1}(\Psi_{S_iS_i}, \nu + q_i)$, $\Sigma_{C_iC_i} \perp\!\!\!\perp (\Sigma_{C_1C_1} \cup \cdots \cup \Sigma_{C_{i-1}C_{i-1}}) \mid \Sigma_{S_iS_i}$, where all $\Psi_{C_iC_i}$ and $\Psi_{S_iS_i}$, the submatrices of Ψ, are positive definite. Then the density of $\Sigma_{C_1C_1} \cup \cdots \cup \Sigma_{C_mC_m}$ is

$$p(\Sigma_{C_1C_1} \cup \cdots \cup \Sigma_{C_mC_m}) = \frac{\prod_{i=1}^m w^{-1}(\Sigma_{C_iC_i} | \Psi_{C_iC_i}, \nu + p_i)}{\prod_{i=2}^m w^{-1}(\Sigma_{S_1S_i} | \Psi_{S_iS_i}, \nu + q_i)}.$$

By (5), this is indeed a conjugate prior distribution of $\Sigma_{C_1C_1} \cup \cdots \cup \Sigma_{C_mC_m}$. Therefore, the density of the conjugate prior distribution of K is

$$p(K) = p(\Sigma_{C_1C_1} \cup \cdots \cup \Sigma_{C_mC_m}) \cdot |J'|_+ = \frac{\prod_{i=1}^m w((\Sigma_{C_iC_1})^{-1} | (\Psi_{C_iC_i})^{-1}, \nu + p_i)}{\prod_{i=2}^m w((\Sigma_{S_1S_i})^{-1} | (\Psi_{S_iS_i})^{-1}, \nu + q_i)},$$

where

$$J' = \frac{\partial K}{\partial\{\Sigma_{C_1C_1} \cup \cdots \cup \Sigma_{C_mC_m}\}} = -\{\sum_{i=1}^m [J'_{C_i}]^r - \sum_{i=2}^m [J'_{S_i}]^r\}.$$

J'_{C_i} and J'_{S_i} are the following Jacobians:

$$J'_{C_i} = -\frac{\partial\{(\Sigma_{C_iC_i})^{-1}\}}{\partial\{\Sigma_{C_iC_i}\}} = \frac{\partial\{(\Sigma_{C_iC_i})^{-1} \cdot d(\Sigma_{C_iC_i}) \cdot (\Sigma_{C_iC_i})^{-1}\}}{\partial\{d(\Sigma_{C_iC_i})\}},$$

$$J'_{S_i} = -\frac{\partial\{(\Sigma_{S_iS_i})^{-1}\}}{\partial\{\Sigma_{S_iS_i}\}} = \frac{\partial\{(\Sigma_{S_iS_i})^{-1} \cdot d(\Sigma_{S_iS_i}) \cdot (\Sigma_{S_iS_i})^{-1}\}}{\partial\{d(\Sigma_{S_iS_i})\}}.$$

Then $|J'_{C_i}|_+ = |(\Sigma_{C_iC_i})^{-1}|^{p_i+1}$, $|J'_{S_i}|_+ = |(\Sigma_{S_iS_i})^{-1}|^{q_i+1}$, and

$$|J'|_+ = \frac{\prod_{i=2}^m |\Sigma_{S_iS_i}|^{q_i+1}}{\prod_{i=1}^m |\Sigma_{C_iC_i}|^{p_i+1}}.$$

Remark on the hyper Markov property. Let

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

and suppose A is distributed according to an inverted Wishart distribution. It is known that A_{11} is independent of $(A_{22} - A_{21}A_{11}^{-1}A_{12}, A_{21}A_{11}^{-1})$. It is interesting to know whether the analogous property holds for $\hat{K}$. Consider $m = 2$ and let $\{C_1, C_2\}$ be a perfect sequence with $S_2 = C_1 \cap C_2$, $R_2 = C_2 \backslash C_1$ and $G_2 = C_1 \backslash C_2$. Then

$$\hat{K} = \begin{pmatrix} \hat{K}_{G_2G_2} & \hat{K}_{G_2S_2} & 0 \\ \hat{K}_{S_2G_2} & \hat{K}_{S_2S_2} & \hat{K}_{S_2R_2} \\ 0 & \hat{K}_{R_2S_2} & \hat{K}_{R_2R_2} \end{pmatrix},$$

$$n(\hat{K})^{-1} = \begin{pmatrix} ssd_{G_2G_2} & ssd_{G_2S_2} & ssd_{G_2R_2} \\ ssd_{S_2G_2} & ssd_{S_2S_2} & ssd_{S_2R_2} \\ ssd_{R_2G_2} & ssd_{R_2S_2} & ssd_{R_2R_2} \end{pmatrix},$$

where $ssd_{G_2R_2} = -ssd_{G_2S_2}(ssd_{S_2S_2})^{-1} ssd_{S_2R_2}$, and

$$\begin{pmatrix} ssd_{G_2G_2} & ssd_{G_2S_2} \\ ssd_{S_2G_2} & ssd_{S_2S_2} \end{pmatrix} = ssd_{C_1C_1},$$

$$\begin{pmatrix} ssd_{S_2S_2} & ssd_{S_2R_2} \\ ssd_{R_2S_2} & ssd_{R_2R_2} \end{pmatrix} = ssd_{C_2C_2}.$$

According to (12) and (6), we have that

$$p(\hat{K}) \propto \frac{\prod_{i=1}^{2} |ssd_{C_iC_i}|^{\frac{1}{2}(n+p_i)}}{|ssd_{S_2S_2}|^{\frac{1}{2}(n+q_2)}} \exp\{-\frac{1}{2} tr[\sum_{i=1}^{2} (\Sigma_{C_iC_i})^{-1} ssd_{C_iC_i} - ((\Sigma_{S_2S_2})^{-1} ssd_{S_2S_2})]\}.$$

Obviously,

$$ssd_{C_1C_1} = \begin{pmatrix} \hat{K}_{G_2G_2} & \hat{K}_{G_2S_2} \\ \hat{K}_{S_2G_2} & \hat{K}_{S_2S_2} - \hat{K}_{S_2R_2}(\hat{K}_{R_2R_2})^{-1}\hat{K}_{R_2S_2} \end{pmatrix}^{-1},$$

$$ssd_{C_2C_2} = \begin{pmatrix} \hat{K}_{S_2S_2} - \hat{K}_{S_2G_2}(\hat{K}_{G_2G_2})^{-1}\hat{K}_{G_2S_2} & \hat{K}_{S_2R_2} \\ \hat{K}_{R_2S_2} & \hat{K}_{R_2R_2} \end{pmatrix}^{-1},$$

$$ssd_{S_2S_2} = \left(\hat{K}_{S_2S_2} - \hat{K}_{S_2G_2}(\hat{K}_{G_2G_2})^{-1}\hat{K}_{G_2S_2} - \hat{K}_{S_2R_2}(\hat{K}_{R_2R_2})^{-1}\hat{K}_{R_2S_2}\right)^{-1}.$$

Therefore

$$|ssd_{C_1C_1}| = n^{p_1}\left(|\hat{K}_{S_2S_2} - \hat{K}_{S_2G_2}(\hat{K}_{G_2G_2})^{-1}\hat{K}_{G_2S_2} - \hat{K}_{S_2R_2}(\hat{K}_{R_2R_2})^{-1}\hat{K}_{R_2S_2}| \cdot |\hat{K}_{G_2G_2}|\right)^{-1},$$

$$|ssd_{C_2C_2}| = n^{p_2}\left(|\hat{K}_{S_2S_2} - \hat{K}_{S_2G_2}(\hat{K}_{G_2G_2})^{-1}\hat{K}_{G_2S_2} - \hat{K}_{S_2R_2}(\hat{K}_{R_2R_2})^{-1}\hat{K}_{R_2S_2}| \cdot |\hat{K}_{R_2R_2}|\right)^{-1}.$$

Let $Z_{S_2G_2} = \hat{K}_{S_2G_2}(\hat{K}_{G_2G_2})^{-\frac{1}{2}}$, $W_{S_2R_2} = \hat{K}_{S_2R_2}(\hat{K}_{R_2R_2})^{-\frac{1}{2}}$. Then $\hat{K}_{S_2G_2} = Z_{S_2G_2}(\hat{K}_{G_2G_2})^{\frac{1}{2}}$, $\hat{K}_{S_2R_2} = W_{S_2R_2}(\hat{K}_{R_2R_2})^{\frac{1}{2}}$. Hence, the density of $\hat{K}$ is of the form $p(\hat{K}) = g\left(\hat{K}_{G_2G_2}, \hat{K}_{S_2S_2}, Z_{S_2G_2}, W_{S_2R_2}\right) \cdot h\left(\hat{K}_{R_2R_2}, \hat{K}_{S_2S_2}, Z_{S_2G_2}, W_{S_2R_2}\right)$, where $Z_{S_2G_2} = \hat{K}_{S_2G_2}(\hat{K}_{G_2G_2})^{-\frac{1}{2}}$, $W_{S_2R_2} = \hat{K}_{S_2R_2}(\hat{K}_{R_2R_2})^{-\frac{1}{2}}$. It is obvious that

$$\left|\frac{\partial(\hat{K}_{G_2G_2}, \hat{K}_{R_2R_2}, \hat{K}_{S_2S_2}, Z_{S_2G_2}, W_{S_2R_2})}{\partial(\hat{K}_{G_2G_2}, \hat{K}_{R_2R_2}, \hat{K}_{S_2S_2}, \hat{K}_{S_2G_2}, \hat{K}_{S_2R_2})}\right|_+ = \left(|\hat{K}_{G_2G_2}| \cdot |\hat{K}_{R_2R_2}|\right)^{-\frac{q_2}{2}}.$$

Thereby $\hat{K}_{G_2G_2} \perp\!\!\!\perp \hat{K}_{R_2R_2} \mid \hat{K}_{S_2S_2}, \hat{K}_{S_2G_2}(\hat{K}_{G_2G_2})^{-\frac{1}{2}}, \hat{K}_{S_2R_2}(\hat{K}_{R_2R_2})^{-\frac{1}{2}}$.

When $m > 2$, we have an analogous result:

$$\hat{K}_{G_mG_m} \perp\!\!\!\perp \hat{K}_{R_mR_m} \mid \hat{K}_{S_mS_m}, \hat{K}_{S_mG_m}(\hat{K}_{G_mG_m})^{-\frac{1}{2}}, \hat{K}_{S_mR_m}(\hat{K}_{R_mR_m})^{-\frac{1}{2}}.$$

Acknowledgements

I wish to thank the editors and referees for their valuable comments, corrections and suggestions, which resulted in considerable improvement of my original paper.

References

Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis, and Interpretation.* Chapman and Hall, London.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B* **41**, 1-13.

Dawid, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8**, 598-617.

Dawid, A. P. and Lauritzen, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21**, 1272-1317.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28**, 157-175.

Edwards, D. (1995). *Introduction to Graphical Modelling.* Springer, New York.

Giri, N. C. (1977). *Multivariate Statistical Inference.* Academic Press, New York.

Guidici, P. and Green, P. J. (1999). Decomposable graphical Gaussian model determination. *Biometrica* **86**, 785-801.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford University Press, Oxford.

Madigan, D. and Raftery, A. (1994). Model selection and accounting for model uncertainty using Occam's window. *J. Amer. Statist. Assoc.* **89**, 1535-1546.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Speed, T. P. and Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14**, 138-150.

Wermuth, N. (1976). Analogous between multiplicative models in contingency tables and covariance selection. *Biometrics* **32**, 95-108.

Whittaker, J. L. (1990). *Graphical Models in Applied Multivariate Statistics.* Wiley, New York.

Department of Statistics, East China Normal University, Zhonshan Road (Northern), Shanghai 200062, China.
E-mail: jwang@stat.ecnu.edu.cn

(Received January 2000; accepted February 2001)