ESTIMATION OF A COVARIANCE MATRIX UNDER LATTICE CONDITIONL INDEPENDENCE RESTRICTIONS

Author Style: KONNO, Yoshihiko
Author Style: Faculty of Science, Chiba University, Chiba-shi 263-8622, Japan
email address: konno@math.s.chiba-u.ac.jp

Introduction

Conditional independence (CI) models, which given by a combination of conditional distribution and independence, has received increasing attention in recent years. Since 1980, particular attention has been directed to graphical Markov models, i.e., the multivariate CI models determined by undirected or directed graphs. See Laruritzen (1996) for an introduction to graphical models. Classes of CI models where all distributions are assumed to be multivariate normal is of special interest. Under this assumption, Andersson and Perlman(1993, AP) introduced so called lattice conditional independence (LCI) models which have pleasant feature of admitting explicit maximum likelihood estimators. This is because the likelihood function and parameter space for a LCI model can be factored into products of conditional likelihood functions and parameter spaces, where the standard multivariate techniques can be applied. In this paper we consider the problem of estimating the covariance matrices under LCI restriction in a decision theoretic set-up. The Stein loss function is used in this study and, using the factorization mentioned above, minimax estimators are obtained. Since the maximum likelihood estimator has constant risk and is different from minimax estimator, this shows that the maximum likelihood estimator under LCI restriction is inadmissible. These results extend those obtained by James and Stein (1960) and Dey and Srinivasan (1986) for estimating normal covariance matrices to the LCI models.

Setup of problem

We first introduce the lattice conditional independence (LCI) model. For the finite set $I = \{1, 2, \ldots, p\}$, let $\mathcal{K}$ be a distributive lattice on I, i.e., a ring of subsets of the finite set I, which is closed under union and intersection and includes the empty set. Following an approach due to AP, we define the LCI model $\boldsymbol{N}(\mathcal{K})$ as the set of all normal distributions $N(0, \Sigma)$ on R^I such that, for every pair $L, M \in \mathcal{K}$,

$$x_L \quad \text{and} \quad x_M \quad \text{are conditionally independent given} \quad x_{L\cap M}. \tag{1}$$

Here x_L is the coordinate projection of $x \in R^I$ onto R^L. We denote by $\boldsymbol{P}_I(\mathcal{K})$ the set of all positive definite $I \times I$ matrices such that (1) is satisfied and $x \sim N(0, \Sigma)$. Based on i.i.d. observations $x_1, x_2, \ldots, x_n$ from the LCI model $\boldsymbol{N}(\mathcal{K})$, we consider the estimation problem of Σ in $\boldsymbol{P}_I(\mathcal{K})$ in a decision theoretic way. We employ the Stein loss function

$$\boldsymbol{L}(\hat{\Sigma}, \Sigma) = \text{ tr } (\hat{\Sigma}\Sigma^{-1}) - \log\det(\hat{\Sigma}\Sigma^{-1}) - p, \tag{2}$$

where $\hat{\Sigma}$ is an estimator of Σ, and evaluate the performance of an estimator by using the risk function $\boldsymbol{R}(\hat{\Sigma}, \Sigma) = \boldsymbol{E}[\boldsymbol{L}(\hat{\Sigma}, \Sigma)]$, where the expectation is taken with respect to the joint distribution of $(x_1, x_2, \ldots, x_n)$.

Question we address in this paper is the following: For a subspace of a symmetric cone characterized by a LCI condition, we can project the sample covariance matrix onto this space and obtain the explicit form of the maximum likelihood estimator(MLE) for the covariance matrix with the LCI condition. From the Pythagoras theorem for I-divergence, or equivalently the loss function (2) due to Speed and Kiveri (1986), the MLE is better than the sample covariance matrix in terms of the loss function (2). Moreover, from the same argument, we may project minimax estimators for the full model onto the sub-cone to get estimators which are better than the original estimators, provided that the projected estimators exist and that the LCI condition is true. Under the full model, the sample covariance matrix is improved by minimax estimators. However, it is not apparent that the estimators projected from minimax estimators for the full model improve upon the MLE under the LCI condition and that these induced estimators are minimax under the LCI model. Using invariance property that the LCI model possesses, we can see that these decision-theoretic properties hold under the LCI model.

An Example. Consider trivariate normal distribution where the second and third variables are conditionally independent given the first variable. To express this as a LCI model, let $I = \{1, 2, 3\}$ and suppose that $\mathcal{K}_1 = \{\phi, 1, 12, 13, I\}$, so $\mathcal{J}(\mathcal{K}_1) = \{1, 12, 13\}$. Let $x_1, x_2, \ldots, x_n$ follow $\boldsymbol{N}(\mathcal{K}_1)$ and put $S = \sum_{i=1}^{n} x_i x_i' = (s_{ij})$. Using reconstruction algorithm in Andersson, and Perlman (1993), we can see that the maximum likelihood estimator is given by

$$n\hat{\Sigma}_{\mathcal{K}_1}^{mle} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{21}s_{11}^{-1}s_{13} \\ s_{31} & s_{31}s_{11}^{-1}s_{12} & s_{33} \end{bmatrix} \text{ or } n\hat{\Sigma}_{mle}^{\mathcal{K}_1} = T_{\mathcal{K}_1}T_{\mathcal{K}_1}' \text{with} T_{\mathcal{K}_1} = \begin{bmatrix} s_{11}^{1/2} & 0 & 0 \\ s_{21}s_{11}^{-1/2} & t_{22} & 0 \\ s_{31}s_{11}^{-1/2} & 0 & t_{33} \end{bmatrix},$$

where $t_{ii}^2 = s_{ii} - s_{i1}s_{11}^{-1}s_{1i}$ with $t_{ii} > 0$ for $i = 2, 3$. From Theorem 3.1 in Konno (2000), we can see that a minimax estimator is given by $\hat{\Sigma}_{\mathcal{K}_1}^{m} = T_{\mathcal{K}_1}D_{\mathcal{K}_1}T_{\mathcal{K}_1}'$, where

$$D_{\mathcal{K}_1} = \text{ diag } \left(\frac{1}{n+2}, \frac{1}{n-1}, \frac{1}{n-1}\right).$$

Hence it is written as

$$\hat{\Sigma}_{\mathcal{K}_1}^{m} = \begin{bmatrix} \frac{s_{11}}{n+2} & \frac{s_{12}}{n+2} & \frac{s_{13}}{n+2} \\ \frac{s_{21}}{n+2} & \frac{1}{n-1}\left(s_{22} - \frac{3}{n+2}s_{21}s_{11}^{-1}s_{12}\right) & \frac{s_{21}s_{11}^{-1}s_{13}}{n+2} \\ \frac{s_{31}}{n+2} & \frac{s_{31}s_{11}^{-1}s_{12}}{n+2} & \frac{1}{n-1}\left(s_{33} - \frac{3}{n+2}s_{31}s_{11}^{-1}s_{13}\right) \end{bmatrix}.$$

References

Andersson, S.A. and Perlman, M.D. (1993): Lattice models for conditional independence in a multivariate normal distribution. *Ann. Statist.* **21** 1318-1358.

Konno, Y. (2000): Inadmissibility of the maximum likelihood estimator of normal covariance matrices with the lattice conditional independence. To appear *J. Multivariate Anal.*

Speed, T.P. and Kiveri, H.T. (1986): Gaussian Markov distribution over finite graph. *Ann. Statist.* **14**, 117-131.