1. Introduction

Graphical models (Lauritzen, 1996 and Whittaker, 1990) are a key technique for the analysis of the conditional independence structure of a multivariate distribution. One of the most interesting aspects is the possibility of selecting, through an appropriate procedure, the graph underlying the data. For a \( d \)-dimensional random variable \( Y \) the pairwise conditional independence relationships can be represented through a conditional independence graph, that is, an undirected graph \( G(V, E) \) where to each component of \( Y \) is associated an element of the vertices set \( V = \{1, 2, \ldots, d\} \) and a missing edge in \( E \subseteq V \times V \) represents conditional independence between the corresponding components of \( Y \). When \( Y \) is supposed to follow a \( N_d(\mu, \Sigma) \) distribution, the selection procedure for a graphical model involves identifying the zero entries of the concentration matrix \( \Sigma^{-1} \).

In fact a conditional independence constraint between the pair of variables, \( Y_i \) and \( Y_j \), is equivalent to specify a zero in the \((i, j)\)th entry of \( \Sigma^{-1} \) which corresponds to a missing edge between vertices \( \{i\} \) and \( \{j\} \) in the graph.

The standard approach to model selection is the well-known stepwise backward selection (Edwards, 2000). This procedure is based on partial correlations, and by starting from a saturated model with a complete graph, tests each of the edges individually at a fixed significance level and proceeds until no further edge is removed.

In this paper we propose the use of the extended Skew Normal distribution (ESN) defined in Capitanio et al. (2003) for model selection in conditional independence graphs, in order to take into account departure from normality in terms of asymmetry. The authors discuss some issues related to the conditional independence graph of an ESN variate; here we partially extend their work by proposing a test for a single edge exclusion/inclusion, which is outlined in the next section.

2. Methodology

When the deviation from normality is due to a lack of symmetry the class of Skew Normal (SN) distributions defined by Azzalini and Dalla Valle (1996) provides a model to represent the data. This family extends the Gaussian distribution by adding a skewness parameter \( \alpha \); the Normal model is obtained as a special case when \( \alpha = 0 \). The SN
distribution possesses some suitable properties, such as closure under marginalization and linear transformations. The extended Skew Normal distribution is a slight extension of the $SN$, which achieves also closure under conditioning.

The density function of a $d$-dimensional $ESN$ variate is:

$$f(y) = \frac{1}{\Phi(\tau)} \phi_d(y - \xi; \Omega) \Phi(\tau (1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T \omega^{-1}(y - \xi))$$  \hspace{1cm} (1)$$

where $\phi_d(y; \Omega)$ is the density of a $d$-dimensional $N_d(0, \Omega)$ variate, $\Phi$ is the distribution function of $N(0, 1)$, $\Omega$ is a full rank covariance matrix, $\omega$ is a diagonal matrix such that $\Omega = \omega^{-1} \Omega \omega^{-1}$ is a correlation matrix, $\alpha$ is a parameter regulating skewness, $\xi$ is the location parameter and $\tau \in \Re$ is an additional shape parameter. The log-likelihood function of the parameters cannot be maximized in closed form, so that the maximization must be performed using numerical routines.

If the $Y$ variate has density (1), pairwise conditional independence between two components $Y_i$ and $Y_j$ occurs if and only if the following conditions hold simultaneously:

(a) $\Omega_{ij} = 0$ and (b) $\alpha_i \alpha_j = 0$, where $\Omega_{ij}$ denotes the $(i,j)$th entry of $\Omega^{-1}$.

For this model a partial correlation equal to zero does not imply the fulfillment of conditions (a) and (b), and consequently the edge selection procedure mentioned in the introduction is not appropriate. Taking into account conditions (a) and (b) for pairwise conditional independence, we propose a selection procedure based on tests for a single edge exclusion/inclusion whose null hypothesis is:

$$H_0 : g(\theta) = \begin{pmatrix} \Omega_{ij} \\ \alpha_i \alpha_j \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$  \hspace{1cm} (2)$$

Let $\hat{\theta}$ be the maximum likelihood estimator of $\theta = (\xi, \Omega^{-1}, \alpha, \tau)$ and $I(\theta)$ the information matrix. By applying the delta method the covariance matrix of $g(\theta)$ is:

$$\Sigma_g(\theta) = \begin{bmatrix} Var(\hat{\Omega}_{ij}) & \alpha_j Cov(\hat{\Omega}_{ij}, \hat{\alpha}_i) + \alpha_i Cov(\hat{\Omega}_{ij}, \hat{\alpha}_j) \\ \alpha_i^2 Var(\hat{\alpha}_j) + \alpha^2 Cov(\hat{\alpha}_i, \hat{\alpha}_j) + 2\alpha_i \alpha_j Cov(\hat{\alpha}_i, \hat{\alpha}_j) \end{bmatrix}.$$  

The test for (2) can be based on the Wald-type statistic $W_n(Y)$ where:

$$W_n(Y) = \left( g(\hat{\theta}) - g(\theta) \right)^T \Sigma_g(\hat{\theta})^{-1} \left( g(\hat{\theta}) - g(\theta) \right) \overset{d}{\rightarrow} \chi^2_d$$

Numerical experiments, on samples of size 200, 500 and 1000, whose results are not reported for brevity but are available upon request from the authors, show that the actual level of the test is pretty close to the nominal one and the power is satisfactory.

**References**


