Journal of Multivariate Analysis 145 (2016) 58–72

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A skew Gaussian decomposable graphical model

Hamid Zareifard [a,*], Håvard Rue [b], Majid Jafari Khaledi [c], Finn Lindgren [d]

[a] Department of Statistics, Jahrom University, Jahrom 74137-66171, Iran
[b] Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway
[c] Department of Statistics, Tarbiat Modares University, P.O. Box 14115-134, Tehran, Iran
[d] Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom

ARTICLE INFO

Article history:
Received 1 August 2014
Available online 28 August 2015

AMS subject classifications:
62H12
62C10
62F15

Keywords:
Decomposable graphical models
Multivariate closed skew normal distribution
Conditional independence
Noninformative prior

ABSTRACT

This paper proposes a novel decomposable graphical model to accommodate skew Gaussian graphical models. We encode conditional independence structure among the components of the multivariate closed skew normal random vector by means of a decomposable graph so that the pattern of zero off-diagonal elements in the precision matrix corresponds to the missing edges of the given graph. We present conditions that guarantee the propriety of the posterior distributions under the standard noninformative priors for mean vector and precision matrix, and a proper prior for skewness parameter. The identifiability of the parameters is investigated by a simulation study. Finally, we apply our methodology to two data sets.

1. Introduction

In recent years, there have been many developments in multivariate statistical models. Making sense of all the many complex relationships and multivariate dependencies present in the data, formulating correct models and developing inferential procedures is an important challenge in modern statistics. In this context, graphical models currently represent an active area of statistical research which have served as tools to discover structure in data. More specifically, graphical models are multivariate statistical models in which the corresponding joint distribution of a family of random variables is restricted by a set of conditional independence assumptions, and the conditional relationships between random variables are encoded by means of a graph. In the Gaussian case, these models induce the conditional independence assumptions by zeros in the precision matrix. An important reason for working with this class of distributions is important properties like closure under marginalization, conditioning and linear combinations which are seldom preserved outside the class of multivariate normal distributions. However, in spite of substantial advances, the Gaussian distributional assumption might be overly restrictive to represent the data. The real data could be highly non-Gaussian and may show features like skewness.

In this article, we study skew distributions in graphical models with the aim of mimicking the success of Gaussian graphical models as much as possible. The last decade has witnessed major developments in models whose finite dimensional marginal distributions are multivariate skew-normal. Azzalini and Capitanio [3] introduced multivariate skew-normal (SN) distribution which enjoys some of the useful properties of normal distribution, such as property of closure

* Corresponding author.
E-mail addresses: zareifard@jahrom.ac.ir (H. Zareifard), hrue@math.ntnu.no (H. Rue), jafari-m@modares.ac.ir (M.J. Khaledi), f.lindgren@bath.ac.uk (F. Lindgren).

http://dx.doi.org/10.1016/j.jmva.2015.08.011

under marginalization and conditioning. Accordingly, an n-dimensional random vector $\mathbf{Y}$ is said to have a SN distribution if its density is

$$\phi_n(\mathbf{y}; \boldsymbol{\mu}, \Omega)\Phi(\alpha_0 + \boldsymbol{\alpha}' D_{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu}))/\Phi(\tau),$$

where $\phi_n(\cdot; \boldsymbol{\mu}, \Omega)$ is the probability density function of the n-dimensional $N_n(\boldsymbol{\mu}, \Omega)$ variable, $\Phi(\cdot)$ is cumulative distribution function of $N(0, 1)$, $\boldsymbol{\mu} \in \Re^n$, $\tau \in \Re$, $\Omega \in \Re^{n\times n}$ is a full rank covariance matrix, $D_{\Omega} = diag(\Omega_{11}, \ldots, \Omega_{nn})^{1/2}$, $\boldsymbol{\alpha} \in \Re^n$ is shape parameter and $\alpha_0 = \tau(1 + \boldsymbol{\alpha}' D_{\Omega}^{-1}\Omega D_{\Omega}^{-1}\boldsymbol{\alpha})^{1/2}$. When $\boldsymbol{\alpha} = \mathbf{0}$ we are back to the multivariate normal distribution. Capitanio et al. [6] used the SN family in graphical models examining in particular the construction of conditional independence graphs. Their results show that if $\mathbf{Y}$ be an n-variate SN distribution with covariance matrix Ω and skewness vector $\boldsymbol{\alpha}$, then

$$Y_i \perp Y_j | \mathbf{Y}_{-ij} \Leftrightarrow \Omega^{ij} = 0 \quad \text{and} \quad \alpha_i\alpha_j = 0$$

where $\mathbf{Y}_{-ij}$ is $\mathbf{Y}$ with the ith and jth elements deleted and Ω^{ij} denotes the (i, j)th entry of Ω^{-1}. Comparing with the Gaussian graphical model, an extra constraint $\alpha_i\alpha_j = 0$ is necessary to capture conditional independence property. It means if we believe Y_i and Y_j are conditionally independent then at least one of α_i and α_j must be zero. Hence, applying this constraint in practical issues is challenging. Alternatively, Dominguez-Molina et al. [7] and Gonzalez-Farias et al. [11] proposed the multivariate closed skew normal (CSN) distribution which includes the property of SN family. Also unlike the SN family, the CSN family enjoys this property that the joint distribution of i.i.d. CSN random variables is the multivariate CSN distribution.

Although much progress has been made in the context of skew normal distributions, the achieved successes in graphical models are limited. The preservation of conditional independence property for skew normal variables has shown that extending the class of skew-normal distributions to graphical models is challenging. The aim of this paper is to develop a multivariate closed skew normal graphical model. We encode conditional independence structure among the components of the multivariate closed skew normal random vector with respect to a decomposable graph G. The main motivation to use the decomposable graphs for encoding the conditional independence is that for this type of graph there exists an ordering of the vertices such that the zero elements in precision matrix are reflected in its Cholesky decomposition [21]. Under decomposable graphs, the conditional independence property is maintained for our skewed graphical model, and simplification occurs in both the interpretation of data and the estimation procedure. Models can be specified in terms of conditional and marginal probability distributions, leading to a simplified analysis based on lower dimensional components [10,16,14].

A fully Bayesian approach to inference is adopted that leads to a coherent treatment of all parameter uncertainty given the model as well as model uncertainty. Additionally, our modeling framework easily handles latent unobserved variables. In fact, the latent variables can naturally be incorporated as additional parameters in the model using data augmentation [25]. The precision matrix is the fundamental object that evaluates conditional dependence between random variables. Estimating a sparse precision matrix is crucial specially in high-dimensional problems. In this context, a family of conjugate prior distributions for the precision matrix is developed. Development of this class of distributions enables Bayesian inference, which allows for the appropriate estimation of the precision matrix, even in the case when the sample size n is small, something which is otherwise not generally possible in the maximum likelihood framework [4,14]. Conditions for propriety of the posterior are given under the standard noninformative priors on mean vector and precision matrix as well as a proper prior for the skewness parameter. We also develop and implement a Markov chain Monte Carlo (MCMC) sampling approach for inference.

The organization of the paper is as follows. Section 2 introduces the required preliminaries and notation. In Section 3, a novel skew Gaussian decomposable graphical model is constructed using a multivariate closed skew normal distribution and its properties are established. Section 4 discusses Bayesian analysis using Gibbs sampling to sample from the posterior distribution. A simulation study is reported in Section 5. Section 6 illustrates the use of proposed methodology in two real data sets: an analysis of student marks from Mardia et al. [17] and an analysis of the carcass data from gRbase package of R. Finally, conclusions and discussion are given in Section 7. Appendix contains proofs of some of the results in the main text.

2. Preliminaries

2.1. Multivariate closed skew-normal distribution

An n-dimensional random vector $\mathbf{Y}$ is said to have a multivariate closed skew-normal distribution, denoted by $CSN_{n,m}(\boldsymbol{\mu}, \Sigma, \Gamma, \boldsymbol{\nu}, \Delta)$, if its density function is of the form

$$f(\mathbf{y}) = \phi_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma)\Phi_m(\Gamma(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta)/\Phi_m(0; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma'), \tag{1}$$

where $\boldsymbol{\mu} \in \Re^n$, $\boldsymbol{\nu} \in \Re^m$, and $\Sigma \in \Re^{n\times n}$ and $\Delta \in \Re^{m\times m}$ are both covariance matrices, $\Gamma \in \Re^{m\times n}$, and $\Phi_n(\cdot; \boldsymbol{\mu}, \Sigma)$ is the cumulative distribution function of the n-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. To derive this distribution, Gonzalez-Farias et al. [11] consider a $(n + m)$-dimensional normal random vector

$$(W_{0_1}, \ldots, W_{0_m}, W_1, \ldots, W_n)' = \begin{pmatrix} \mathbf{W_0} \\ \mathbf{W} \end{pmatrix} \sim N_{n+m}\left(\mathbf{0}, \begin{pmatrix} \Delta + \Gamma\Sigma\Gamma' & \Gamma\Sigma \\ \Sigma\Gamma' & \Sigma \end{pmatrix}\right).$$