On a Mixture of Gaussian Copula Graphical Model

Mike K. P. So and Raymond W. M. Li

Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology
Corresponding Author: Raymond W. M. Li, e-mail: imwmli@ust.hk

Abstract

Identifying dependence between variables has always been an interesting question for researchers. Understanding the dependencies provide insight in many fields of applications including biostatistics, social science, etc. One of the classical approaches to target the problem is the use of a graphical model: A graph $G = (V, E)$ with each vertex $v \in V$ representing a variable and edge $e \in E$ governing the conditional dependency between the connected vertices. Researches in Graphical Model are almost exclusively based on Gaussian assumption or more general, Gaussian copula: Prior choices of a single precision matrix Σ^{-1} including Wishart for complete graphs, Hyper Wishart for decomposable graphs and G-Wishart in general are developed over past decades. Nevertheless, using a single covariance structure can still be restrictive for certain applications. In this paper, we extend existing approaches by considering a mixture of Gaussian copula in graphical models with the use of a set of latent indicator variables. Bayesian inference of the indicators and precision matrices will be discussed. Simulation study will also be conducted to examine the efficiency and accuracy of our model.

1 Introduction

Conditional dependence among variables of a dataset X can be modeled with the use of graph $G = (V, E)$ (Edwards, 2000). Presence of edge $e \in E$ between two variables X_i, X_j, which corresponds to the vertices $v_i, v_j \in V$ in graph G respectively, indicates conditional dependence between X_i and X_j. The joint density of X is then constructed given the dependence structure represented by the graph G. Traditionally graphical model assumes the joint density to be Gaussian with precision matrix Σ^{-1} constrainted by the graph G. This is called the Gaussian graphical model (Lauritzen, 1996). Under normal assumption, conditional independence between X_i and X_j is zero conditional correlation between the two, reflected by the $(i, j) - t$h entry of Σ^{-1} being zero. Dobra and Lenkoski (2011) generalizes the model using Gaussian copula, which assumes the underlying dependence is Gaussian and marginals can be modeled with other density family.

In estimation of the graph G and the dependence of data, Bayesian framework can be readily applied. Assuming the data follows a Gaussian distribution, the traditional choice of conjugate prior of precision matrix is Inverse Wishart distribution. Dawid and Lauritzen (1993) develops the Hyper Inverse Wishart distribution $HIW(D, \delta)$, which is the conjugate prior of Σ^{-1} given the constraints of the graph G:

$$p(\Sigma^{-1}|G) = \frac{1}{I_G(\delta, D)}(\det \Sigma^{-1})^{(\delta-2)/2} \exp\{-\frac{1}{2}tr(\Sigma^{-1}, D)\} \tag{1.0.1}$$

where $I_G(\delta, D)$ is the normalizing constant, D is a positive definite matrix and δ is degree of freedom. The posterior distribution of Σ^{-1} is $HIW(D + X^T X, \delta + n)$. The normalizing constant I_G is an integral which cannot be calculated exactly for most cases, numerical approximation is normally required. With the additional constraint of the graph being decomposable, the posterior can be estimated with Gibbs Sampler on a perfect sequence of subgraph, eliminating the need of numerical approximation (Letac and Massam, 2007). For non-decomposable graph, Roverato (2002) developed the algorithm in estimating the posterior.

Nevertheless, usage of Gaussian copula in graphical model still limits the variable dependence governed by a single precision matrix. We propose using a mixture of Gaussian copulae, which further generalizes the Gaussian dependence assumption. In section 2 we will formulate the proposed mixture model under a hierarchical Bayesian model with estimation procedure. In section 3 we will provide a short discussion on the model and future work.

2 Methodology

The main methodology can be divided into two parts: Estimation of the Gaussian copula mixture and the precision matrix for each component. Similar to ordinary Gaussian mixture (Diebolt and Robert, 1994), we construct the following hierarchical model for the mixture parameter:

$$\begin{aligned} \alpha &: \text{hyperparameter} \\ p|\alpha &\sim Dirichlet(\alpha) \\ z|p &\sim Categorical(p) \end{aligned}$$

With the use of hidden variable z, joint density of X can be represented as $f(X|\{z, \Sigma_1, \Sigma_2, \ldots, \Sigma_K\}) = \sum_{i=1}^{K} z_i c_i(X|\Sigma_i)$ with c_i being the i-th Gaussian copula component. Under this hierarchical model, Gibbs Sampler can be employed to sample the mixture component:

1. Sample $p^{(h+1)}$ from Dirichet distribution
2. Sample $z^{(h+1)}$ given $p^{(h+1)}$
3. Based on the hidden variable $z^{(h+1)}$, observations are partitioned into K
4. Data partitioned into the i-th set will be used in estimating the precision matrix $\Sigma_i^{(-1)}$

The next step is to sample the precision matrix according to the graph G_i for each mixture component. Assuming the graph G_i is known, given the hidden variable z, the sampling follows similarly in Gaussian graphical model

1. Given all observations $x_n \in X$ of the i-th mixture component, transform the variables by their marginals: $u_n = F(x_n)$
2. $U|\Sigma_i^{-1} \sim N(0, \Sigma_i)$
3. Given the constraint of graph G_i, sample Σ_i^{-1} from the Hyper Inverse Wishart distribution
4. Sample the transition of G_i into a new graph based on the sampled Σ_i^{-1}

Assuming the graph G_i is constrained to be decomposable, we can efficiently sample the graphical component similarly as in paper by Wang and Li (2012).

3 Conclusion

The development of mixture of Gaussian copula graphical model extends the range of application of graphical model in dependence modeling. A full Bayesian framework can be employed in estimation of the mixture components and the dependence structure of the graph. In the future, we are looking to conduct simulation study on the methodology. Discussion on how the number of mixture components K can be chosen, the identifiability of the mixture model and graph transition will be conducted in the future to ensure effective implementation of the suggested methodology.

References

Dawid, A. and Lauritzen, S. (1993). Hyper markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21(3):1272–1317.

Diebolt, J. and Robert, C. (1994). Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 363–375.

Dobra, A. and Lenkoski, A. (2011). Copula gaussian graphical models and their application to modeling functional disability data. *The Annals of Applied Statistics*, 5(2A):969–993.

Edwards, D. (2000). *Introduction to graphical modelling, Chapter 1.* Springer.

Lauritzen, S. (1996). *Graphical models*, volume 17. Oxford University Press, USA.

Letac, G. and Massam, H. (2007). Wishart distributions for decomposable graphs. *The Annals of Statistics*, 35(3):1278–1323.

Roverato, A. (2002). Hyper inverse wishart distribution for non-decomposable graphs and its application to bayesian inference for gaussian graphical models. *Scandinavian Journal of Statistics*, 29(3):391–411.

Wang, H. and Li, S. (2012). Efficient gaussian graphical model determination under g-wishart prior distributions. *Electronic Journal of Statistics*, 6:168–198.