Selection of Gaussian graphical models

Sophie Donnet[1], Jean-Michel Marin[2*]

[1] CEREMADE, Université Paris Dauphine, France
[2] INRIA Saclay, projet SELECT, Université Paris-Sud, France
[*] jean-michel.marin@inria.fr

The last decade has witnessed the apparition of applied problems typified by very high-dimensional variables, in marketing database or gene expression studies for instance. Graphical models (Lauritzen (1996)) enable concise representations of associational relations between variables. If the graph is known, the parameters of the model are easily estimated. However, a quite challenging issue is the selection of the most appropriate graph for a given dataset. We consider this problem and the case of decomposable Gaussian graphical models (Dawid and Lauritzen (1993)).

Let $\mathcal{G} = (V, E)$ be an undirected graph with vertices $V = \{1, \dots, p\}$ and set of edges E. We suppose that $\mathcal{G}$ is decomposable. To each vertex $v \in V$ of the graph, we associate a random variable y_v. Let $\mathbf{y} = (y_1, \dots, y_p)$, a graphical model is a family of distributions on $\mathbf{y}$ which are Markov with respect to a graph. A Gaussian graphical model is such that

$$\mathbf{y}|\mathcal{G}, \Sigma_\mathcal{G} \sim \mathcal{N}_p\left(0_p, \Sigma_\mathcal{G}\right) , \tag{1}$$

where Σ_G is a positive definite matrix which ensures that the distribution of $\mathbf{y}$ is Markov with respect to $\mathcal{G}$. $\Sigma_\mathcal{G}$ ensures that the distribution of $\mathbf{y}$ is Markov iff $(i,j) \notin E \iff \left(\Sigma_\mathcal{G}^{-1}\right)_{(i,j)} = 0$.
We observe a sample $\mathbf{y}^1, \dots, \mathbf{y}^n$ from (1) (the data are centered). We would like to identify the set of most relevant graphs. For the considered multivariate random phenomenon, we are interested in the set of most relevant conditional independence structures.

We consider the Bayesian paradigm. Conditionally on $\mathcal{G}$, we use a Hyper-Inverse Wishart (HIW) distribution associated to the graph $\mathcal{G}$ as prior distribution on $\Sigma_\mathcal{G}$: $\Sigma_\mathcal{G}|\mathcal{G}, \delta_\mathcal{G}, \Phi_\mathcal{G} \sim \mathrm{HIW}_\mathcal{G}\left(\delta_\mathcal{G}, \Phi_\mathcal{G}\right)$ where $\delta_\mathcal{G} > 0$ and $\Phi_\mathcal{G}$ is a $p \times p$ symmetric positive definite matrix. Conditionally on $\mathcal{G}$, the HIW distribution is conjugate

$$\Sigma_\mathcal{G}|\mathcal{G}, \mathbf{y}^1, \dots, \mathbf{y}^n, \delta_\mathcal{G}, \Phi_\mathcal{G} \sim \mathrm{HIW}_\mathcal{G}\left(\delta_\mathcal{G} + n, \Phi_\mathcal{G} + \sum_{i=1}^n \mathbf{y}^\mathbf{i}\left(\mathbf{y}^\mathbf{i}\right)^\mathrm{T}\right) . \tag{2}$$

Moreover, for such a prior, $f(\mathbf{y}|\mathcal{G}, \delta_\mathcal{G}, \Phi_\mathcal{G}) = \dfrac{h_\mathcal{G}(\delta_\mathcal{G}, \Phi_\mathcal{G})}{(2\pi)^{np/2} h_\mathcal{G}\left(\delta_\mathcal{G} + n, \Phi_\mathcal{G} + \sum_{i=1}^n \mathbf{y}^\mathbf{i}\left(\mathbf{y}^\mathbf{i}\right)^\mathrm{T}\right)}$ where $h_\mathcal{G}$ is the normalizing constant of the HIW distribution associated to the graph $\mathcal{G}$. Finally, we assume a uniform prior distribution in the space of graphs: $\pi(\mathcal{G}) \propto 1$. In that case,

$$\pi\left(\mathcal{G}|\mathbf{y}^1, \dots, \mathbf{y}^n, \delta_\mathcal{G}, \Phi_\mathcal{G}\right) \propto f(\mathbf{y}|\mathcal{G}, \delta_\mathcal{G}, \Phi_\mathcal{G}) . \tag{3}$$

It is well-known that (3) is sensible to the specification of the hyper-parameters $\delta_\mathcal{G}$ and $\Phi_\mathcal{G}$ (Giudici and Green (1999), Jones et al. (2005)). In this work, we address this problem and present different strategies. Then, we introduce a new sampling scheme to explore the space of graphs and conclude with some experiments on simulated and real datasets.

Dawid, A. and Lauritzen, S. (1993) Hyper Markov Laws in the statistical analysis of decomposable graphical models, Annals of Statistics, 21, 1272-1317.

Giudici, P. and Green, P. (1999) Decomposable graphical Gaussian model determination, Biometrika, 86, 785-801.

Jones, B., Carvalho, C., Dobra, A., Hans, C. Carter, C. and West, M. (2005) Experiments in Stochastic Computation for High-Dimensional Graphical Models, Statistical Science, 20, 388-400.

Lauritzen, S. (1996) Graphical models, Oxford Statistical Science Series, 17, Oxford Science Publications