Gaussian Faithful Markov Trees

Dhafer Malouche, Bala Rajaratnam

Gaussian Faithful Markov Trees

Dhafer Malouche [1] & Bala Rajaratnam [2]

[1] *Ecole Supérieure de la Statistique et de l'Analyse de l'Information.*
`dhafer.malouche@essai.rnu.tn`

[2] *Stanford University*
`brajarat@stanford.edu`

Résumé : Dans ce travail on s'intéresse à deux types de modèles graphiques : les modèles de concentration et les modèles de covariance. Un modèle graphique peut être défini comme un graphe associé à la distribution de probabilité d'un vecteur aléatoire. Chaque sommet dans ce graphe correspond à une variable du vecteur aléatoire. Dans un modèle de concentration l'absence d'une arête liant deux variables indique que celles-ci sont indépendantes sachant toutes les autres variables. Par contre, dans le modèle de covariance, l'absence d'une arête indique une indépendance marginale entre ces deux variables. Ces deux types de modèles permettent aussi de lire beaucoup d'autres relations d'indépendances conditionnelles entre les variables. Mais certaines d'entre elles peuvent être omises par le graphe. Dans le cas où le graphe permet de lire toutes les indépendances conditionnelles existantes dans la distribution de probabilité, on dira que cette distribution de probabilité est *fidèle* à son graphe. On présente ici deux résultats théoriques concernant les modèles graphiques représentés par un arbre qui est un graphe où chaque paire de sommets est liée exactement par une seule trajectoire. Il a été montré dans les deux cas et dans le cadre des modèles graphiques gaussiens que l'hypothèse de fidélité est nécessairement satisfaite. Ainsi, si le graphe de concentration associé à une loi gaussienne est un arbre, alors ce graphe va permettre de lire toutes les indépendances conditionnelles existantes dans la loi probabilité (voir [1]) et le même résultat a été aussi observé dans le cas de modèles de covariances (voir [2]). Cependant, les méthodes utilisées pour la démonstration des deux résultats sont complètement différentes.

Abstract: We study two types of graphical models in this paper: concentration and covariance graphical models. Graphical models use graphs to encode or capture the multivariate dependencies that are present in a given multivariate distribution. A concentration graph associated with a multivariate probability distribution of a given random vector is an undirected graph where each vertex represents each of the different components of the random vector, and where the absence of an edge between any pair of variables implies conditional independence between these two variables given the remaining ones. Similarly, a covariance graph reflects marginal independences in the sense that the absence of an edge between any pair of variables implies marginal independence between these two variables. These two graphical models do not only encode pairwise relationship between variables, but they also allow us to read many other conditional independence statements present in the probability distribution through separation criteria in this graph. In general, the graph may not reflect some of these conditional independence statements. When the graph encodes all of these conditional independences we say that the probability distribution is faithful to its corresponding graphical model. We present in this paper two mathematical results concerning Markov trees : graphical models corresponding to trees. Gaussian Markov trees are necessarily faithful to their concentration and covariance graphs. More formally this means that Gaussian distributions that have trees as concentration graphs are necessarily faithful (see [1]). Similarly an equivalent result can be proved for covariance graphs (see [2]). However the methods of proofs used for these two results are completely different.

Graphical models are mathematical tools used to represent conditional independences in a given multivariate probability distribution (see [3,4]). Many different types of graphical models have been studied in the literature. For example, directed acyclic graphs or DAGs are commonly referred to as "Bayesian networks" (see [5]). When the graph is undirected and when such graphs are constructed using marginal independence relationships between pairs of random variables in a given random vector, these graphical models are called "covariance graph" models (see [6,7,8,9]). Covariance graph models are commonly represented by graphs with exclusively bi-directed or dashed edges (see [6]). This representation is used in order to distinguish them from the traditional and widely used concentration graph models. Concentration graphs encode conditional independence between pairs of variables given the remaining ones. Formally, consider a random vector $\mathbf{X} = (X_v, v \in V)'$ with a probability distribution P where V is a finite set representing the random variables in $\mathbf{X}$. The concentration graph associated with P is an undirected graph $G = (V, E)$ where

- V is the set of vertices.
- Each vertex represents one variable in $\mathbf{X}$.
- E is the set of edges (between the verices in V) constructed using the pairwise rule: for any pair $(u, v) \in V \times V$, $u \neq v$

$$(u, v) \notin E \iff X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_{V \setminus \{u,v\}} \tag{1}$$

where $\mathbf{X}_{V \setminus \{u,v\}} := (X_w,\ w \neq u \text{ and } w \neq v)'$.

Note that $(u, v) \notin E$ means that the vertices u and v are not adjacent in G.

An undirected graph $G_0 = (V, E_0)$ is called the covariance graph associated with the probability distribution P if the set of edges E_0 is constructed as follows:

$$(u, v) \notin E \iff X_u \perp\!\!\!\perp X_v \tag{2}$$

The subscript zero is invoked for covariance graphs (i.e., G_0 vs G) as the definition of covariance graphs does not involve conditional independences.

Both concentration and covariance graphs are not only used to encode pairwise relationships between pairs of variables in the random vector $\mathbf{X}$, but as we will see below, these graphs can be used to encode conditional independences that exist between subsets of variables of $\mathbf{X}$. First we introduce some definitions:

The multivariate distribution P is said to satisfy the "intersection property" if for any subsets A, B C and D of V which are pairwise disjoint,

$$\begin{cases} \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{C\cup D} \\ \text{and} \\ \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_C \mid \mathbf{X}_{B\cup D} \end{cases} \quad \text{then} \quad \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B\cup C} \mid \mathbf{X}_D \tag{3}$$

We will call the *intersection* property (see [3]) in (3) above the *concentration intersection* property in order to differentiate it from another property that is satisfied by P when studying covariance graph models.

Let P satisfy the *concentration intersection* property. Then for any triplet (A, B, S) of subsets of V pairwise disjoint, if S separates[1] A and B in the concentration graph G associated with P then the random vector $\mathbf{X}_A = (X_v,\, v \in A)'$ is independent of $\mathbf{X}_B = (X_v,\, v \in B)'$ given $\mathbf{X}_S = (X_v,\, v \in S)'$. This latter property is called *concentration global Markov* property and is formally defined as,

$$A \perp_G B \mid S \;\Rightarrow\; \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S. \tag{4}$$

In [8] it is shown that if P satisfies the following property: for any triplet (A, B, S) of subsets of V pairwise disjoint,

$$\text{if } \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \text{ and } \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_C \;\text{ then } \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B\cup C}, \tag{5}$$

then for any triplet (A, B, S) of subsets of V pairwise disjoint, if $V \setminus (A \cup B \cup S)$ separates A and B in the covariance graph G_0 associated with P then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$. This latter property is called the *covariance global Markov property* and can be written formally as follows:

$$A \perp_{G_0} B \mid V \setminus (A \cup B \cup S) \;\Rightarrow\; \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S. \tag{6}$$

In parallel to the concentration graph case, property (5) will be called the *covariance intersection* property.

Even if P satisfies both intersection properties, the covariance and concentration graphs may not be able to capture or reflect all the conditional independences present in

[1]We say that S separates A and B if any path connecting A and B in G intersects S, i.e., $A \perp_G B \mid S$, and is not to be confused with stochastic independence, which is denoted by $\perp\!\!\!\perp$ as compared to $\perp_G$.

the distribution, i.e., there may exist one or more conditional independences present in the probability distribution that does not correspond to any separation statement in either G or G_0. Equivalently, a lack of a separation statement in the graph does not necessarily imply conditional dependence. On the contrary, when no other conditional independences exist in P except the ones encoded by the graph, we classify P as a *faithful* probability distribution to its graphical model. More precisely we say that P is *concentration faithful* to its concentration graph if for any triplet (A, B, S) of subsets of V pairwise disjoint, the following statement holds:

$$S \text{ separates } A \text{ and } B \iff \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S. \tag{7}$$

Similarly, P is said to be *covariance faithful* to its covariance graph G_0 if for any triplet (A, B, S) of subsets of V pairwise disjoint, the following statement holds:

$$V \setminus (A \cup B \cup S) \text{ separates } A \text{ and } B \iff \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S. \tag{8}$$

A natural question of both theoretical and applied interest in probability theory and statistics is to understand the implications of the *faithfulness* assumption. This assumption is fundamental since it yields a bijection between the probability distribution P and the graph G in terms of the independences that are present in the distribution. In this paper we present two results concerning the faithfulness assumption. The first result concerns covariance graphs and was proved by Malouche and Rajaratnam 2009 (see [2]). This result is formally written in Theorem 1. It was shown that when P is a multivariate Gaussian distribution, whose covariance graph is a tree, the probability distribution P is necessarily covariance faithful, i.e., such probability distributions satisfy property (8). Equivalently, the associated covariance graph G is fully able to capture all the conditional independences present in the multivariate distribution P.

Theorem 1 *Malouche and Rajaratnam (2009)*
Let $\boldsymbol{X}_V = (X_v,\, v \in V)'$ be a random vector with Gaussian distribution $P = \mathcal{N}_{|V|}(\mu, \Sigma)$. Let $G_0 = (V, E_0)$ be the covariance graph associated with P. If G_0 is a tree or more generally a union of connected components each of which are trees (or a union of "tree connected components"), then P is covariance faithful to G_0.

The proof of Theorem 1 requires among others a result proved by Jones and West 2005 (see [11]). This result gives a method that can be used to compute the covariance matrix Σ from the precision matrix K using the paths in the concentration graph G. The result can also be easily extended to show that the precision matrix K can be computed from the covariance matrix Σ using the paths in the covariance graph G_0. We now state the result by Jones and West 2005. Let us first recall that we denote by $\mathcal{P}(u, v, G)$ the set of paths connecting the vertices u and v in an undirected graph G.

Theorem 2 *Jones and West (2005) (modified).*
Let $\boldsymbol{X}_V = (X_v,\, v \in V)'$ be a random vector with Gaussian distribution $P = \mathcal{N}_{|V|}(\mu, \Sigma)$ where Σ and $K = \Sigma^{-1}$ are positive definite matrices. Let $G = (V, E)$ and $G_0 = (V, E_0)$ denote respectively the concentration and covariance graph associated with the probability distribution of $\boldsymbol{X}_V$.

For all (u, v) *in* $V \times V$

$$k_{uv} = \sum_{p \in \mathcal{P}(u,v,G_0)} (-1)^{|p|+1} |\sigma|_p \frac{|\Sigma \setminus p|}{|\Sigma|}$$

and

$$\sigma_{uv} = \sum_{p \in \mathcal{P}(u,v,G)} (-1)^{|p|+1} |k|_p \frac{|K \setminus p|}{|K|}$$

where, if $p = (u_0, \ldots, u_n)$,

$$|\sigma|_p = \sigma_{u_0 u_1} \sigma_{u_1 u_2} \ldots \sigma_{u_{n-1} u_n}, \quad |k|_p = k_{u_0 u_1} k_{u_1 u_2} \ldots k_{u_{n-1} u_n},$$

$K \setminus p = (k_{uv}, (u, v) \in (V \setminus p) \times (V \setminus p))$ *and* $\Sigma \setminus p = (\sigma_{uv}, (u, v) \in (V \setminus p) \times (V \setminus p))$ *denote respectively* K *and* Σ *with rows and columns corresponding to variables in path* p *omitted. The determinant of a zero-dimensional matrix is defined to be* 1.

The result presented in Theorem 1 can be considered as a dual of a previous probabilistic result proved by Becker *et al.* 2005 (see [1]) for concentration graphs that demonstrates that Gaussian distributions having concentration trees[2], are necessarily *concentration faithful* to its concentration graph (implying property (7) is satisfied). This result was proved by showing that Gaussian distributions satisfy the following two properties called respectively "weak transitivity" (see eqn (9)) and "decomposable transitivity" (see eqn (10)). For any triplet (A, B, C) pairwise disjoint subsets of V such that $V \setminus (A \cup B \cup C) \neq \emptyset$ and for any $w \in V \setminus (A \cup B \cup C)$

$$\left\{ \begin{array}{l} \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C \\ \text{and} \\ \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{C \cup \{w\}} \end{array} \right. \Longrightarrow \mathbf{X}_A \perp\!\!\!\perp X_w \mid \mathbf{X}_C \text{ or } X_w \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C \tag{9}$$

For any any pair of disjoint subsets (A, B) of V such that $V \setminus (A \cup B) \neq \emptyset$ and for any triplet (u, v, w) of vertices in $V \setminus (A \cup B)$ we have

$$\left\{ \begin{array}{l} \mathbf{X}_{A \cup \{u\}} \perp\!\!\!\perp \mathbf{X}_{B \cup \{v\}} \mid X_{\{w\}} \\ \text{and} \\ \mathbf{X}_u \perp\!\!\!\perp \mathbf{X}_v \mid \mathbf{X}_{A \cup B} \end{array} \right. \Longrightarrow X_u \perp\!\!\!\perp X_w \mid \mathbf{X}_B \text{ or } X_w \perp\!\!\!\perp X_v \mid \mathbf{X}_A \tag{10}$$

We now formally present in Theorem 3 the result proved by Becker *et al.* 2005 pertaining to concentration trees.

Theorem 3 *Becker, Geiger and Meek (2005)*
Let $\boldsymbol{X}_V = (X_v, v \in V)'$ *be a random vector with Gaussian distribution* $P = \mathcal{N}_{|V|}(\mu, \Sigma)$. *Let* $G = (V, E)$ *be the concentration graph associated with* P. *If* G *is a tree or more generally a union of connected components each of which are trees (or a union of "tree connected components"), then* P *is concentration faithful to* G.

[2]i.e., the concentration graph is a tree.

Acknowledgments: Bala Rajaratnam was supported in part by NSF grant DMS 0505303, DMS 0906392, SUFSC2008-SUSHSTF2009-SMSCVISG200906. Dhafer Malouche was supported in part by the Stanford University France-Stanford Center grant 2008. Both authors gratefully acknowledge the facilities provided by the Department of Mathematics, Laboratoire Dieudonné, Université de Nice, Sophia-Antipolis, Nice, France.

Bibliographie

[1] Becker, A., Geiger, D., Meek, C., 2005. Perfect tree-like markovian distributions. Probability and Mathematical Statistics 25 (2), 231–239.

[2] Malouche, D., Rajaratnam, B., 2009. Gaussian Covariance Faithful Markov Trees, Technical report, Department of Statistics, Stanford University.

[3] Kindermann, R., & Snell, J. L. 1980.*Markov Random Fields and Their Applications.* American Mathematical Society, Providence, Rhode Island.

[4] Lauritzen, S. L., 1996. *Graphical Models.* New York : Oxford University Press.

[5] Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems.* Tech. rept. Morgan Kaufman.

[6] Cox, D.R., & Wermuth, M. 1993. Linear dependencies represented by chain graphs (with Discussion). *Statist. Sci.*, **8**, 204–218, 247–277.

[7] Cox, D. R., & Wermuth, N. 1996. *Multivariate Depencies : Models, Analysis and Interpretations.* Chapman and Hall.

[8] Kauermann, G. 1996. On a dualization of graphical Gaussian models. *Scand. J. Statist.*, **23**, 105–116.

[9] Khare, K., & Rajaratnam, B. 2009.Wishart distributions for decomposable covariance graph models. Technical report, Department of Statistics, Stanford University.

[10] Banerjee, M., & Richardson, T. 2003. On a Dualization of Graphical Gaussian Models: A Correction Note. *Scand. J. Statist.*, **Vol 30**, 817–820.

[11] Jones, B., & West, M. 2005. Covariance decomposition in undirected Gaussian graphical models.*Biometrika*, **92**, 770–786.