Duality between faithfulness assumptions in Graphical models

Dhafer Malouche, Bala Rajaratnam

Duality between faithfulness assumptions in Graphical models.

Dhafer Malouche & Bala Rajaratnam

U2S-ENIT-Ecole Supérieure de la Statistique et de l'Analyse de l'Information, Tunisia
&
Stanford University, USA

Résumé : On s'intéresse à la dualité entre deux hypothèses de *fidélité* qui peuvent être satisfaites par une distribution de probabilité d'un vecteur aléatoire. La première concerne la *fidélité* au graphe de concentration et la seconde concerne la *fidélité* au graphe de covariance. Dans chacun de ces graphes, un sommet correspond exactement à une variable. Par contre, l'absence d'une arête entre une paire de variables, dans le graphe de concentration, indique une indépendance conditionnelle entre ces deux variables sachant le reste des variables. L'absence d'une arête, dans le graphe de covariance, indique une indépendance marginale entre ces deux variables. Sur chaque graphe il a été définit un critère de séparation et la lecture d'une séparation dans le graphe indique la présence d'une indépendance conditionnelle dans la distribution de probabilité : c'est la propriété de Markov Globale. Une distribution de probabilité sera dite *fidèle* au graphe si toutes les indépendances conditionnelles existantes dans la distribution de probabilité correspondent à celles visibles dans le graphe. Dans ce papier, on s'intéresse à l'étude de ces deux hypothèses de *fidélité* et à la dualité qui peut exister entre elles. On s'intéresse ensuite aux distributions de probabilités, dites *bi-fidèles*, satisfaisant simultanément les deux hypothèses de *fidélité*. On montre, dans ce cas, que les graphes de concentration et covariance ne doivent contenir que des composantes connexes qui sont soit complètes, soit contenant que des séparateurs de taille égales à $|V| - 2$ où $|V|$ est le nombre de variables.

Abstract : In this paper we analyze the duality between two faithfulness assumptions that can be defined on a given multivariate probability distribution of a set of random variables. The first pertains to faithfulness to its concentration graph and the second pertains to faithfulness to its covariance graph. The vertices in both these graphs are in a one-to-one correspondence with the set of variables in the random vector. The concentration graph is an undirected graph constructed by looking through conditional independences between each pair of variables given the remaining variables and the covariance graph is constructed by looking through marginal independences between each pair of variables. The absence of an edge in the graph corresponds to conditional or marginal independences respectively. On each graph a separation criteria is defined which implies conditional independences present in the probability distribution: this is termed the Global Markov property. Furthermore, the faithfulness assumption is said to be satisfied when all the independence conditional statements in the probability distributions are represented in the graph. In this paper we analyze the duality between these two faithfulness hypothesis. We also prove that when the both assumptions are simultaneously satisfied, i.e., the bi-faithfulness property, all the connected components in the concentration and in the covariance graphs are either complete or all their separators have a cardinality equal to $|V| - 2$ where V is the number of variables.

Key Words : Graphical Models, Markov properties, faithfulness, concentration graphs, covariance graphs

Graphical Models and Bayesian Networks have found widespread use in statistics - especially in high dimensional settings. One important application of these models is in the area of analyzing gene expression level data (see Friedman *et al.* (2000), Magwene and Kim (2004), Castelo and Roverato (2006), Wille and Bühlman (2006), Malouche and Sevestre (2008)...). One of the main objectives in this active area of research is to re-construct or recover the graph, $G = (V, E)$, representing the interactions between genes from observed data. This graph is often termed as a *Gene Network Interaction* model (see Toh and Horimoto (2002)). Under the assumption of *Gaussianity*, this gene network coincides with the *concentration* graph (see Lauritzen (1996)) associated with an unknown probability distribution generating the data. Data sets in high dimensional setting typically have a low number of observations compared to the number of variables, i.e., $n << |V|$, and hence classical estimation procedures are no longer applicable. These classical procedures include those based on maximum likelihood estimation(mle) of the covariance matrix or its inverse (see for example Lauritzen (1996) or Edwards (2000)). Existence of the mle is not even guaranteed in high dimensions, let alone obtaining a stable estimator with good properties. Buhl (1993) showed that the maximum likelihood

estimator exists with probability one if the number of observations, n, is greater than the number of variables, V. This probability can however be smaller than one in the case when $n < |V|$.

As a result many authors have proposed estimation procedures for concentration graphs by checking for low order conditioning (for example Magwene and Kim (2004), Castelo and Roverato (2006), Wille and Bühlman (2006), Malouche and Sevestre (2008)). These approaches aim to discover conditional independences given a certain fixed number of variables. Generally this fixed number is very low. Wille and Bühlman (2006) consider one variable, Friedman *et al.* (2000) consider two variables... The assumption in these procedures that allows estimation of the true, but unknown graph, is to impose the *faithfulness* assumption. Indeed, the *faithfulness* assumption is an hypothesis which is commonly used in estimation procedures for graphical models.

First let us briefly state this hypothesis. Assume that the graph associated with a given multivariate probability distribution has a set of vertices that corresponds to the random variables in the random vector. Each vertex corresponds to one such random variable. A distribution is called *faithful* to a graph if all conditional independence statements in the distribution are represented by a *separation* statement in the corresponding graph.

This faithfulness assumption means that no other conditional independences exist in the distribution than the ones given by the separation statement appearing on the graph. For example, if the *faithfulness* hypothesis is not satisfied this implies that we can potentially find conditional independences that do not correspond to any separation statement in the graph. When the *faithfulness* assumption is satisfied and if all the separation statements in the graph represent a conditional independence statement in the probability distribution then there is a one-to-one association between separations in the graph and conditional independences in the probability distribution.

We present in this paper two families of undirected graphs that are associated with a given multivariate probability distribution P of a given random vector $\mathbf{X}_V = (X_v,\, v \in V)'$. The first family is named *covariance* graphs. A covariance graph associated with P is an undirected graph $G_0 = G_0(P) = (V, E_0(P))$ with a set vertices V and where the set of edges is constructed as follows

$$E_0(P) = \{\, (u,v) \in V \times V \;\text{ such that } u \neq v \;\text{ and }\; X_u \not\perp\!\!\!\perp X_v \}$$

This type of graphical models were formally defined by Cox and Wermuth (1996) and studied by others (Kauermann (1996), Chaudhuri *et al.* (2007), Khare and Rajaratnam (2008)).

Covariance graphs not only encode marginal independences between variables, but they also represent conditional independences statements between subsets of variables. This can be specified by defining a separation criteria on these undirected graphs : let A, B and S be three pairwise disjoint subsets of V with A and B non-empty, if in G_0 the set $V \setminus (A \cup B \cup S)$ separates A and B in G_0, which means that any path connecting a vertex in A and an other in B intersects $V \setminus (A \cup B \cup S)$, then the sub-random vector $\mathbf{X}_A$

is independent of $\mathbf{X}_B$ given $\mathbf{X}_S$, i.e., $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$. This property is called the *global covariance* Markov property of P with respect to G_0. We will say that P is g_0-Markov with respect to G_0. We can also state an equivalent version of this global Markov property : let A, B and S be three pairwise disjoint subsets of V with A and B non-empty, if the subset S separates A and B in G_0, then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V\setminus(A\cup B\cup S)}$.

Let us note also that this g_0-Markov property is satisfied when P satisfies the following property : for any A, B and C three pairwise disjoint subsets of V if $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B$ and $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_C$ then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B\cup C}$ (see Banerjee and Richardson (2003) for a proof). We note that this latter property is satisfied by multivariate Gaussian distributions.

Let us review the second family of graphical models associated with a given probability distribution P, namely, *concentration* graphs. These are represented by an undirected graph $G = G(P) = (V, E(P))$ where V represents the set of vertices, and the set of edges $E(P)$ is constructed as follows

$$E(P) = \{\, (u,v) \in V \times V \ \text{ such that } u \neq v \ \text{ and } \ X_u \not\perp\!\!\!\perp X_v \mid \mathbf{X}_{V\setminus\{u,v\}}\}$$

Yet an other global Markov property can be defined in this context and is satisfied when P verifies the following property, called the *intersection* property : for any pairwise disjoints subsets A, B, C and D of V

$$\text{if} \quad \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{C\cup D} \quad \text{and} \quad \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_C \mid \mathbf{X}_{B\cup D} \quad \text{then} \quad \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B\cup C} \mid \mathbf{X}_D$$

We will say that P is *global concentration* Markov or $g-$Markov with respect to G if for any pairwise disjoints subsets A, B, and S such that S separates A and B then $\mathbf{X}_A$ is independent of $\mathbf{X}_B$ given $\mathbf{X}_S$, i.e., $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$.

We now define two faithfulness assumption for a given probability distribution P. We say that P is *covariance faithful* to G_0 if G_0 is the covariance graph associated to P and if all the conditional independences present in P are represented by G_0. Equivalently this means that $G_0 = G_0(P)$ and for any triplet (A, B, S) of pairwise disjoint subsets of V, we have

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V\setminus(A\cup B\cup S)} \ \Rightarrow \ A \text{ and } B \text{ are separated by } S \text{ in } G_0 \tag{1}$$

Similarly, we say that P *concentration faithful* to G if G is the concentration graph associated to P and if all the conditional independences present in P are represented by G. Equivalently this means that $G = G(P)$ and for any triplet (A, B, S) of pairwise disjoint subsets of V, we have

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S \ \Rightarrow \ A \text{ and } B \text{ are separated by } S \text{ in } G_0 \tag{2}$$

We prove first in this paper that when the *concentration faithfulness* assumption is satisfied by a given probability distribution P with respect to its concentration graph G, the associated covariance graph G_0 is a degenerate version of G. This means that G and G_0 have the same connected components, but all the connected components of G_0

are complete. The same result holds true when the *covariance faithfulness* assumption is satisfied by a given probability distribution P with respect to its covariance graph G_0. In this case, the associated concentration graph G is a degenerate version of G_0. Hence G and G_0 have the same connected components, but all the connected components of G are complete.

We proceed to define a less restrictive and more natural *faithfulness* assumption by not allowing for the conditioning subset S to be empty. In particular in implications (1) and (2) S is not allowed to be an empty set, i.e., $S \neq \emptyset$. We call this less restrictive condition the $g-$*faithfulness* assumption in the case of concentration graphs and g_0-*faithfulness* assumption in the case of covariance graphs. We prove that there is a duality between these later two less restrictive faithfulness hypothesis (see Theorem 1). We consider probability distributions satisfying both *faithfulness* assumptions. We call this the *bi-faithful* assumption. Hence a probability distribution P is *bi-faithful* if P is simultaneously $g-$faithful to its concentration graph G and g_0-faithful to its covariance graph G_0. In theorem 2 (see below) we deduce consequences of the *bi-faithfulness* assumption.

Theorem 1. Let $\mathbf{X}_V = (X_u,\, u \in V)'$ be a random vector with probability distribution P. Let $G = (V, E)$ and $G_0 = (V, E_0)$ denote respectively the concentration graph and the covariance graph constructed from P.

i. If P is $g-$faithful to G, then G and G_0 have the same connected components and if all the separators in any connected component $(G_0)_U$ of G_0 have cardinality smaller then $|U| - 2$, then $E_U \subseteq (E_0)_U$.

ii. If P is g_0-faithful to G_0, then G and G_0 have the same connected components and if all the separators in any connected component G_U of G have cardinality smaller then $|U| - 2$, then $(E_0)_U \subseteq E_U$.

Theorem 2. Let $\mathbf{X}_V = (X_v,\, v \in V)'$ be a random vector with probability distribution P. Let $G = (V, E)$ and $G_0 = (V, E_0)$ denote respectively the concentration and the covariance graphs associated with P. Assume that P is bi-faithful to (G, G_0), i.e. P is simultaneously $g-$faithful to G and g_0-faithful to G_0 then

i. G and G_0 have the same connected components,

ii. for any $U \subseteq V$ that is a connected component in G or G_0, then G_U and $(G_0)_U$ are belonging to $\mathcal{G}$ where

$$\begin{aligned} \mathcal{G} \;=\; & \{G_W = (W, E_W) \text{ where } W \subseteq V,\; G_W \text{ is either complete} \\ & \text{or } \forall\, S \text{ separator in } G_W,\; |S| = W - 2\} \end{aligned}$$

Bibliography

[1] Banerjee, M., Richardson, T., 2003. On a dualization of graphical gaussian models: A correction note. Scand. J. Statist. Vol 30, 817–820.

[2] Buhl, S. L., 1993. On the existence of maximum likelihood estimators for graphical gaussian models. Scan. Journal of Statistic. 20, 263–270.

[3] Castelo, R., Roverato, A., 2006. A robust procedure for gaussian graphical models search for microarray data with p larger than n. Journal of Machine Learning Research 57, 2621–2650.

[4] Edwards, D., 2000. Introduction to graphical modelling. Springer texts in statistics.

[5] Friedman, N., Linial, M., Nachman, I., Pe'er, D., 2000. Using bayesian networks to analyse expression data. J. Comput. Biol. 7(3-4), 601–620.

[6] Khare, K. and Rajaratnam, B., 2008. Conjugate Wishart distributions for covariance graph models, Technical report, Department of Statistics, Stanford University.

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

[8] Magwene, P., Kim, J., 2004. Estimating genomic coexpression networks using first-order conditional independence. Genom Biol. 5(12).

[9] Malouche, D., Sevestre-Ghalila, S., 2008. Estimating high dimensional faithful gaussian graphical models by low-order conditioning. Proceeding, of 26th IASTED International Multi-Conference on Applied Informatics, Artificial Intelligence and Applications 595-025, 1–6.

[10] Toh, H., Horimoto, K., 2002. Inference of genetic network by combined approach of cluster analysis and graphical gaussian modelling. Bioinformatics 18(2), 287–297.

[11] Wille, A., Bühlman, P., 2006. Low-order conditional independence graphs for inferring genetic network. Statistical Applications in Genetics and Molecular Biology 5, 1–32.