

INFÉRENCE DE STRUCTURE DE MODÈLE GRAPHIQUE À L'AIDE D'ARBRES COUVRANTS

Loïc Schwaller [‡] & Stéphane Robin [‡]

[‡] *AgroParisTech/INRA, UMR 518 MIA, F-75005 Paris, France*
{loic.schwaller,stephane.robin}@agroparistech.fr

Résumé. On se propose d'apprendre la structure d'un modèle graphique non-orienté (aussi appelé champ aléatoire de Markov) en calculant la probabilité d'apparition a posteriori de certaines structures locales telles que les arêtes. Cette tâche serait impossible à réaliser de manière exacte sans restreindre l'espace des graphes explorés. Nous nous limitons ici à celui des arbres couvrants. Les distributions a priori sur les structures et les distributions sont choisies de telle manière à permettre une inférence rapide et exacte des probabilités a posteriori sur les arêtes, l'intégration sur l'espace des arbres étant réalisée à l'aide d'un résultat d'algèbre appelé théorème Arbre-Matrice. Enfin, on montre que l'hypothèse d'arborescence n'empêche pas notre approche d'exhiber de bon résultats sur des données simulées selon un modèle non-arborescent ainsi que sur des données de cytométrie de flux.

Mots-clés. Arbre, Arbre-Matrice, hyper Markov, inférence de réseaux, modèle graphique.

Abstract. We propose to learn the structure of a discrete undirected graphical model by computing exact posterior probabilities for local structures in a Bayesian framework. This task would be untractable without any restriction on the considered graphs. We limit our exploration to the spanning trees. We define priors on tree structures and parameters that allow fast and exact computation of the posterior probability for an edge to belong to a random tree thanks to an algebra result called the Matrix-Tree theorem. We show that the assumption we have made does not prevent our approach to perform well on synthetic and flow cytometry data.

Keywords. Graphical models, hyper Markov, Matrix-Tree, network inference, tree.

1 Introduction

Les modèles statistiques deviennent de plus en plus complexes, impliquant des structures de dépendances entre les variables de plus en plus intriquées. Les modèles graphiques sont un moyen à la fois naturel et puissant pour décrire de telles structures. Ainsi, dans de nombreux domaines d'application, il est intéressant d'être capable d'inférer un modèle graphique à partir d'observations. D'un point de vue statistique, parler d'inférence de

modèle graphique nécessite de considérer le modèle lui-même comme un paramètre. Dans un cadre bayésien, une telle inférence suppose un modèle global et en particulier, une distribution a priori sur les modèles graphiques. Ici, nous ne considérerons que des graphes non-orientés. Un modèle graphique m_G est un couple (G, \mathcal{F}_G) où G est un graphe (que l'on supposera décomposable) qui décrit les relations de dépendances conditionnelles entre les variables représentées par les nœuds et \mathcal{F}_G est une famille de distributions respectant ces relations. Dans la littérature, inférer le modèle graphique doit souvent être compris comme inférer le graphe, c'est-à-dire la structure supportant le modèle. Toujours dans un cadre bayésien, cela signifie être capable de calculer la distribution a posteriori $p(G|\mathbf{X})$ sur les graphes à la vue d'observations \mathbf{X} . Nous nous limitons ici aux modèles graphiques dont la structure est un arbre couvrant. Dans la section 2, nous décrivons le modèle global dans lequel sera réalisée l'inférence, et en particulier les distributions a priori utilisées. La section 3 se concentre sur l'inférence du modèle. Nous nous intéresserons particulièrement à la probabilité a posteriori qu'une arête appartienne au graphe. L'idée est qu'un graphe peut être localement arborescent sans être un arbre et que cette quantité reste exploitable si tel est le cas. La section 4 présente une étude simulatoire ainsi qu'une application à des données de cytométrie de flux.

2 Modèle

Soit $\mathbf{X} = (X_1, \dots, X_p)$ un vecteur aléatoire à valeurs dans un espace produit $\mathcal{X} = \bigotimes_{i=1}^p \mathcal{X}_i$ et \mathcal{F} l'ensemble des distributions sur \mathcal{X} . Soit $V = \{1, \dots, p\}$ et $E \subset \mathcal{P}_2(V)$ où $\mathcal{P}_2(V)$ est l'ensemble des parties de V de taille 2. $G = (V, E)$ est un graphe non-orienté dont les sommets sont indexés par V et les arêtes données par E . Tous les graphes considérés ici auront V pour ensemble de sommets. Un graphe connecté mais ne possédant pas de cycle est appelé un arbre couvrant et on notera \mathcal{T} l'ensemble des arbres. Un modèle graphique m_G est donné par un couple (G, \mathcal{F}_G) où G est un graphe (supposé décomposable) et $\mathcal{F}_G \subset \mathcal{F}$ est une famille de distributions vérifiant la propriété de Markov par rapport à G . Le lecteur peut se référer à l'article de Dawid et Lauritzen (1993) pour un rappel sur les modèles graphiques non-orienté (ainsi que pour les notions développées en Section 2.2).

On considère le modèle hiérarchique suivant. On commence par tirer un arbre T selon une certaine loi, puis une distribution π dans \mathcal{F}_T et enfin \mathbf{X} selon π (Figure 2). Afin qu'il soit possible d'effectuer l'inférence de la structure du modèle graphique au sein de

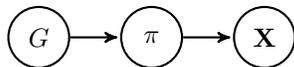


FIGURE 1 – Modèle global.

ce modèle de manière exacte, il est nécessaire de définir les lois a priori sur les arbres et les distributions soigneusement.

2.1 Distribution a priori sur les arbres

On considère une distribution a priori sur l'espace des arbres factorisée sur les arêtes de ceux-ci, à savoir de la forme

$$\forall T \in \mathcal{T}, p(T) = \frac{1}{Z_0} \prod_{\{i,j\} \in E_T} \beta_{ij}, \quad Z_0 = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} \beta_{ij}, \quad (1)$$

avec $\beta = (\beta_{ij})_{(i,j) \in V^2}$ matrice symétrique dont les termes sont positifs et choisie telle que le graphe support $G_\beta = (V, E_\beta)$ où $E_\beta = \{\{i, j\} \in \mathcal{P}_2(V), \beta_{ij} > 0\}$ est connecté. La condition sur le graphe support permet d'assurer que β induit réellement une distribution sur \mathcal{T} . Typiquement, on prendra une distribution a priori uniforme sur les arbres.

2.2 Distribution a priori sur les distributions

Dans la suite, une distribution sur un espace de distributions sera appelée hyperdistribution afin d'éviter les confusions. Il est souhaitable que la distribution a posteriori sur les arbres $p(T|\mathbf{X})$ reste factorisée sur les arêtes, comme expliqué dans la Section 3. Or, $p(T|\mathbf{X}) \propto p(T)p(\mathbf{X}|T)$ où $p(\mathbf{X}|T)$ est la vraisemblance intégrée par rapport aux paramètres de \mathbf{X} . Pour un arbre donné T , la distribution $p(\pi|T)$ doit donc être choisie de telle manière que $p(\mathbf{X}|T)$ reste factorisable sur les arêtes. Ce choix doit de plus être effectué de manière cohérente entre les différents arbres afin d'obtenir une écriture similaire à (1) pour $p(T|\mathbf{X})$. Ceci est effectué en construisant une famille compatible d'hyperdistributions fortement Markov (Dawid et Lauritzen, 1993).

Soit T un arbre couvrant. Si $\pi \in \mathcal{F}_T$ et ρ^T est une hyperdistribution sur \mathcal{F}_T , pour tous $A, B \subset V$, on note π_A la distribution marginale obtenue à partir de π sur les variables \mathbf{X}_A et $\pi_{B|A}$ la collection de distributions conditionnelles de $\mathbf{X}_B|\mathbf{X}_A$ sous θ . On note aussi ρ_A^T l'hyperdistribution marginale induite par ρ^T sur π_A et $\rho_{B|A}^T$ la collection d'hyperdistributions induites par ρ^T sur $\pi_{B|A}$.

Definition ρ^T est dite fortement hyper Markov par rapport à T si, pour tous $A, B \subset V$ tels que $A \cap B$ est complet et sépare A de B , $\{\pi_{B|A}, \pi_{A|B}, \pi_{A \cap B}\}$ sont mutuellement indépendants sous ρ^T .

De telles hyperdistributions ont des propriétés intéressantes vis-à-vis de $p(\mathbf{X}|T)$.

Proposition 1. (Dawid et Lauritzen, 1993) Si ρ^T est fortement hyper Markov par rapport à T , alors la vraisemblance intégrée $p(\mathbf{X}|T)$ est Markov par rapport à T .

Cela signifie que $p(\mathbf{X}|T)$ peut être factorisée sur les arêtes de T .

$$p(\mathbf{X}|T) = \int \pi(\mathbf{X}) \rho^T(\pi) d\pi = \prod_{i \in V} p(X_i|T) \prod_{\{i,j\} \in E_T} \frac{p(X_i, X_j|T)}{p(X_i|T)p(X_j|T)} \quad (2)$$

Il faut maintenant choisir toutes les hyperdistributions ρ^T de telle manière qu'elles soient hyper Markov par rapport à leur arbres respectifs et que les intégrales locales utilisées en (2) ne dépendent pas de T .

Proposition 2. (*Dawid et Lauritzen, 1993*) Soit ρ une hyperdistribution générale sur \mathcal{F} telle que, pour tout $A \subset V$, $\pi_A \perp\!\!\!\perp \pi_{V \setminus A|A}$ sous ρ . Alors, pour tout arbre $T \in \mathcal{T}$, il existe une unique hyperdistribution ρ^T sur \mathcal{F}_T fortement hyper Markov par rapport à T et telle que, pour toute arête $\{i, j\} \in E_T$,

$$\rho_{ij}^T = \rho_{ij}. \quad (3)$$

$\{\rho^T\}_{T \in \mathcal{T}}$ est une famille compatible d'hyperdistributions.

On requiert en fait que l'hyperdistribution ρ servant de base à la famille compatible soit fortement hyper Markov par rapport au graphe complet. Si $\{\rho^T\}_{T \in \mathcal{T}}$ est construite ainsi, la Proposition 2 garantit que ρ^T est fortement hyper Markov par rapport à T pour tout $T \in \mathcal{T}$ et que les vraisemblances locales intégrées $p(X_i, X_j|T)$ et $p(X_i|T)$ ne dépendent en fait pas de T . Ces termes peuvent être calculés une fois pour toute et utilisés pour tous les arbres. Ainsi, on a

$$p(T|\mathbf{X}) = \frac{1}{Z(\mathbf{X})} \prod_{\{i,j\} \in E_T} \omega_{ij}(X_i, X_j), \quad \omega_{ij}(X_i, X_j) = \beta_{ij} \frac{p(X_i, X_j)}{p(X_i)p(X_j)}, \quad (4)$$

avec $Z(\mathbf{X})$ constante de normalisation. Le modèle complet est décrit en Figure 2. Dans la pratique, les intégrales locales permettant d'obtenir ω peuvent être calculées simplement dans un cadre conjugué, mais peuvent aussi être obtenues par des méthodes d'intégration numérique, de type Monte Carlo par exemple.

$$\begin{aligned} T &\sim \mathcal{U}(\mathcal{T}); \\ \pi &\sim \rho^T; \\ \mathbf{X} &\sim \pi. \end{aligned}$$

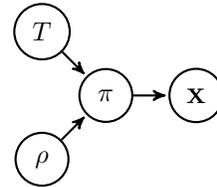


FIGURE 2 – Modèle basé sur une famille compatible.

3 Inférence

On cherche à retrouver la structure du modèle graphique dont est issue la distribution de \mathbf{X} . La distribution a posteriori sur les arbres est connue à constante de normalisation

près. On pourrait donc choisir de déterminer l'arbre le plus probable à posteriori (MAP). Ici, on s'intéresse à la probabilité a posteriori qu'une arête donnée appartienne à l'arbre décrivant le modèle graphique. Cette probabilité est donnée par

$$p(\{k, l\} \in E_T | D) = \sum_{\substack{T \in \mathcal{T} \\ E_T \ni \{k, l\}}} p(T | D) = 1 - \left(\sum_{\substack{T \in \mathcal{T} \\ E_T \not\ni \{k, l\}}} \prod_{\{i, j\} \in E_T} \omega_{ij}(X_i, X_j) \right) / Z(\mathbf{X}). \quad (5)$$

Il s'avère que les deux sommes impliquées en (5) peuvent être calculées exactement grâce à la forme factorisée de la distribution a posteriori sur les arbres. On note Δ le laplacien de ω défini par $\Delta_{ij} = -\omega_{ij}$ si $i \neq j$ et $\sum_j \omega_{ij}$ sinon. On note également $|\Delta^{(u,v)}|$ le mineur de Δ obtenu en enlevant la ligne u et la colonne v .

Théorème 1 (Théorème Arbre-Matrice, Chaiken (1982)). *Tous les mineurs $|\Delta^{(u,v)}|$ sont égaux et on a l'égalité suivante*

$$|\Delta^{(u,v)}| = \sum_{T \in \mathcal{T}} \prod_{\{i, j\} \in E_T} \omega_{ij}. \quad (6)$$

$Z(\mathbf{X})$ est donc calculé comme un résultat direct de ce théorème. La somme au numérateur de (5) peut également être obtenue grâce à ce théorème en mettant les termes correspondant à l'arête $\{k, l\}$ dans ω à zéro.

4 Simulations et application à des données de cytométrie de flux

Le but des simulations est de montrer que les performances de notre approche en termes d'inférence restent bonnes si le véritable modèle n'est pas arborescent. On utilise un algorithme d'inférence utilisant un échantillonnage de type MCMC dans l'espace des DAGs (Niinimäki *et al.*, 2011) comme point de comparaison. On se place dans un cadre

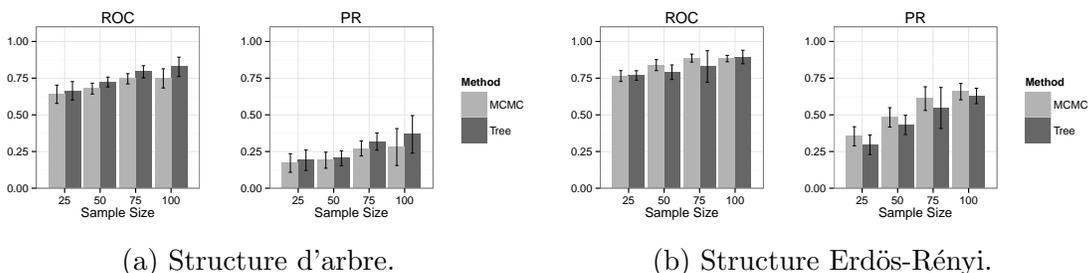
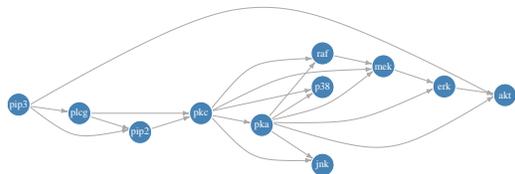


FIGURE 3 – Résultats sur données simulées. Aire sous les courbes ROC et PR pour différentes tailles d'échantillons.



(a) Voie Raf.

	MCMC	Tree
ROC	0.688 (0.062)	0.690 (0.047)
PR	0.575 (0.050)	0.570 (0.051)

(b) Résultats de l'inférence.

FIGURE 4 – Résultats sur les données de cytométrie de flux.

multinomial avec a priori Dirichlet. On observe que les performances des deux méthodes sont similaires en terme de scores ROC et PR, que le graphe soit un arbre ou non (Figure 3). La restriction de l'espace des graphes à celui des arbres ne semble donc pas affecter l'inférence en ce qui concerne les arêtes, ce qui est bien ce qu'on cherchait à vérifier ici. L'aspect exact de notre approche permet de plus d'aboutir à un algorithme rapide.

Enfin, nous avons appliqué notre approche basée sur les arbres couvrants à des données de cytométrie de flux (Werhli *et al.*, 2006). Ces données décrivent le niveau d'activation de 11 protéines et phospholipides impliqués dans la voie de signalisation Raf. La structure communément admise pour cette voie de signalisation est donnée en Figure 4a. Comme dans le cas des données simulées, les performances de notre algorithme sont équivalentes à celles de l'algorithme basé sur l'échantillonnage dans l'espace des DAGs (Figure 4b).

Références

- S. CHAIKEN : A Combinatorial Proof of the All Minors Matrix Tree Theorem. *SIAM Journal on Algebraic Discrete Methods*, 3(3):319–329, 1982.
- A. P. DAWID et S. L. LAURITZEN : Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models. *The Annals of Statistics*, 21(3):1272–1317, 1993.
- M. MEILA et T. JAAKKOLA : Tractable bayesian learning of tree belief networks. *Statistics and Computing*, 16(1):77–92, 2006.
- T. NIINIMAKI, P. PARVIAINEN et M. KOIVISTO : Partial order mcmc for structure discovery in bayesian networks. In F. G. COZMAN et A. PFEFFER, édés : *UAI*, p. 557–564. AUAI Press, 2011. ISBN 978-0-9749039-7-2.
- A. V. WERHLI, M. GRZEGORCZYK et D. HUSMEIER : Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. *Bioinformatics (Oxford, England)*, 22(20):2523–31, oct. 2006. ISSN 1367-4811.