Réseaux Bayésiens de Niveau Deux et D-Séparation

Linda Smail, Jean-Pierre Raoult

Laboratoire d'Analyse et de Mathématiques Appliquées (CNRS UMR 8050)
Université de Marne-la-Vallée
5 boulevard Descartes, Champs sur Marne 77454 Marne-la-Vallée Cedex 2
linda.smail@univ-mlv.fr, raoult@math.univ-mlv.fr

Résumé. Etant donné une famille de variables aléatoires $(X_i)_{i\in I}$, munie de la structure de réseau bayésien et un sous-ensemble S de I, nous considérons le problème de calcul de la loi de la sous-famille $(X_a)_{a\in S}$ (resp. la loi de $(X_b)_{b\in \bar{S}}$, où $\bar{S} = I - S$, conditionnellement à $(X_a)_{a\in S}$). Nous mettons en évidence la possibilité de décomposer cette tâche en plusieurs calculs parallèles dont chacun est associé à une partie de S (resp. de $\bar{S}$) ; ces résultats partiels sont ensuite regroupés dans un produit. Dans le cas du calcul de $(X_a)_{a\in S}$, ceci revient à la mise en place sur S d'une structure de réseau bayésien de niveau deux.

1 Introduction

Etant donné un réseau bayésien $(X_i)_{i\in I}$, nous nous intéressons, étant donné une partie non vide S de I, à la loi P_S de la sous-famille $X_S = (X_i)_{i\in S}$ et à la loi $P_{\bar{S}/S}$, de la sous-famille $X_{\bar{S}} = (X_j)_{j\in \bar{S}}$ conditionnellement à X_S.

Dans les réseaux bayésiens possédant de nombreaux nœuds et fortement connectés, le calcul de lois ou de lois conditionnelles peut faire intervenir des sommations relatives à de très gros sous-ensembles de l'ensemble des indices I. Il y a donc intérêt à s'efforcer, au préalable, de décomposer, s'il est possible, ces calculs en plusieurs calculs moins lourds et pouvant être menés en parallèle. Cette décomposition est liée à des propriétés du graphe définissant le réseau.

Les formules donnant P_S et $P_{\bar{S}/S}$ apparaissent alors comme des produits de facteurs dépendant isolément des atomes pour des partitions appropriées.

La construction de ces partitions fait intervenir deux relations d'équivalence dans $\bar{S}$, toutes deux du type : x et y sont équivalents si et seulement s'ils sont reliés, dans un graphe non orienté (GNO) convenablement déduit du graphe orienté (GO) définissant le réseau bayésien, par une chaîne ne passant pas par S. Deux tels GNO sont considérés ; l'un, classique, est le graphe moral, pour lequel les arêtes relient les nœuds joints par un arc ou ceux ayant un enfant en commun ; l'autre, à notre connaissance original, est le graphe hyper-moral, pour lequel les arêtes relient les nœuds joints par un arc ou ceux dont les descendances proches (voir définition dans (Smail 2004) et en 2 ci-dessous) ont une intersection non vide .

La formule de calcul de P_S est liée à une structure de réseau bayésien dont les nœuds ne sont pas les éléments de I mais les atomes d'une partition de I (notion introduite en (Smail 2003) sous le nom de réseau bayésien de niveau 2).

2 Des partitions remarquables

Etant donné S, partie non vide de I, on note S^+ la partie commençante engendrée par S, c'est à dire l'ensemble des éléments de I ayant au moins un descendant dans S. Nous allons introduire deux partitions de $S^+ - S$; pour cela nous ferons usage des notions suivantes

ensemble de descendance proche d'un nœud i : $dp(i)$ est l'ensemble des enfants de i et de tous les nœuds intermédiaires sur les chemins reliant i à chacun de ses enfants.

ensemble de descendance proche de C : $T(C)$ est l'ensemble des éléments appartenant aux descendances proches des nœuds dans C autre que ceux dans C lui même.

ensemble des racines extérieures de C : $R(C)$ est l'ensemble des parents des éléments de $C \cup T(C)$ autres que ceux dans $C \cup T(C)$ lui même,

2.1 Graphes moral et hyper-moral

Etant donné un graphe orienté sans circuits (I, G), son graphe moral (resp. hyper-moral) associé est le graphe non orienté (I, G_m) (resp (I, G_{hm}) où G_m (resp G_{hm}) est l'ensemble :

- des paires (i, j) telles que $(i, j) \in G$ ou $(j, i) \in G$,
- des paires (i, j) telles que i et j ont un enfant (resp. un descendant proche) en commun.

2.2 Partitions S-morale et S-hyper-morale de $S^+ - S$

On appelle partition S -morale (resp. S-hyper-morale) la partition de $S^+ - S$ notée $\mathcal{P}^m_{S^+-S}$ (resp. $\mathcal{P}^{hm}_{S^+-S}$), définie par la relation d'équivalence $\sim_{S^+-S}$ (resp. $\approx_{S^+-S}$), pour laquelle x et y sont équivalents si et seulement s'ils sont reliés par une chaîne dans le graphe moral G_m (resp. hyper-moral G_{hm}).

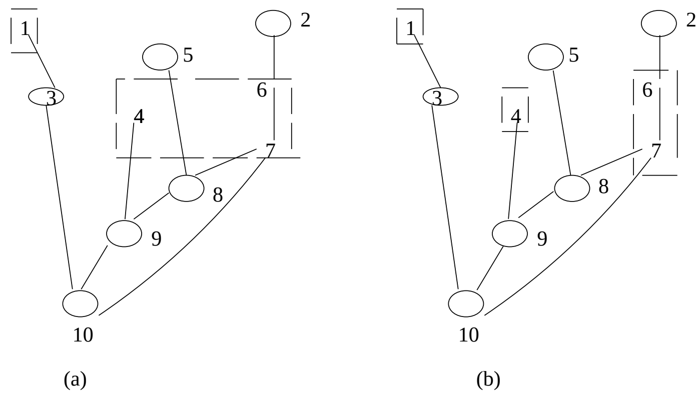
2.3 Partition de S

Soit K l'ensemble des éléments de S qui ne sont dans la descendance proche d'aucun élément de $S^+ - S$, (dans l'exemple ci dessus, $K = \{2, 5\}$). Alors (voir démonstration dans (Smail 2004)), L'ensemble des singletons $\{k\}$, où $k \in K$ (si $K \neq \emptyset$) et des parties $T(C) \cap S^+$, où $C \in \mathcal{P}^{hm}_{S^+-S}$, constitue une partition de S.

3 Factorisation du calcul de $P_{\bar{S}/S}$ et P_S

Théorème 1 *Soit $(I, G, (X_i)_{i \in I})$ un réseau bayésien et soit S une partie non vide de I. Soit $\mathcal{P}^m_{S^+-S}$ la partition S-morale de $S^+ - S$. Alors*

$$P_{\bar{S}/S} = \prod_{i \notin S^+} P_{i/p(i)} \prod_{C \in \mathcal{P}^m_{S^+-S}} P_{C/F(C) \cap S^+} \tag{1}$$



$$S = \{2, 3, 5, 8, 9, 10\}$$

FIG. 1 – *(a) : $\mathcal{P}_{\bar{S}}^{hm}$ la partition S-hyper-morale. (b) : $\mathcal{P}_{\bar{S}}^{m}$ la partition S-morale.*

où

$$P_{C/F(C)\cap S^+} = \frac{\prod_{i\in C\cup(e(C)\cap S^+)} P_{i/p(i)}}{\sum_C \prod_{i\in C\cup(e(C)\cap S^+)} P_{i/p(i)}} \ . \qquad (2)$$

Théorème 2 *Soit (I, G, P_I) un réseau bayésien et soit S une partie de I. Soit $\mathcal{P}_{S^+-S}^{hm}$ la partition S-hyper-morale de $S^+ - S$ et soit K l'ensemble des éléments de S qui ne sont dans la descendance proche d'aucun élément de $S^+ - S$. Alors*

$$P_S = \prod_{k\in K} P_{k/p(k)} \prod_{C\in\mathcal{P}_{S^+-S}^{hm}} P_{T(C)/R(C)}$$

où

$$P_{T(C)/R(C)} = \sum_C \prod_{i\in C\cup T(C)} P_{i/p(i)}(x_i/x_{p(i)})$$

Ce théorème (dont on trouve la démonstration dans (Smail 2004)) s'interprète naturellement dans le cadre des RB2 :

Corollaire 1 *Sous les hypothèses du théorème 2, la partition Q_s $(= \{\{k\}; k \in K\} \cup \{T(C); C \in \mathcal{P}_{\bar{S}}^{hm}\})$ munit S de la structure de réseau bayésien de niveau deux, où la structure de graphe ordonné, G', est définie par :*

$$\left| \begin{array}{l} si\ k \in K\ \ et\ \ k' \in K, (k,k') \in G'\ \ si\ (k,k') \in G \\ si\ k \in K\ \ et\ C \in \mathcal{P}_{\bar{S}}^{hm}, (k, T(C)) \in G'\ \ si\ k \in R(C) \\ si\ C \in \mathcal{P}_{\bar{S}}^{hm}\ \ et\ \ C' \in \mathcal{P}_{\bar{S}}^{hm}, (T(C), T(C')) \in G'\ \ si\ \ T(C) \cap R(C') \neq \emptyset \end{array} \right.$$

(mais il n'existe pas de couple $(T(C), k)$ appartenant à G').

On remarque que le théorème 2 fournit en fait une structure de RB2 renseigné puisqu'il précise, pour tout atome A de Q_s, quels sont exactement les éléments de I, appartenant aux parents de A (pour G'), qui interviennent comme conditionnement de cet atome. Dans l'exemple de la figure 1, voici la représentation du RB2 sur S

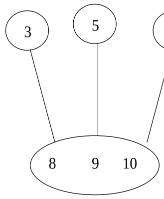


FIG. 2 – *Réseau bayésien de niveau deux représentant P_S.*

Références

Finn, V. Jensen (1996), An Introduction to Bayesian Networks, Springer Verlag, New York.

Judea, Pearl (1986), Fusion propagation and structuring in belief networks, Artificial Intelligence 29 (3) 241–288

Judea, Pearl (1988), Probabilistic Reasoning in Intelligent Systems. Morgan-Kaufmann, San Mateo.

Petr Hájek, Tomás Havránek and Radim Jiroušek (1992), Information Processing in Expert Systems, CRC Press.

Smail Linda (2003), Calcul des lois conditionnelles dans les réseaux bayésiens, 35èmes Journées de Statistique, Lyon.

Smail Linda (2004), Algorithmique pour les réseaux bayésiens et leurs extensions, Thèse en Mathématiques Appliquées soutenue en 2004, Université de Marne-la-Vallée.

Summary

Given a family of random variables, say $(X_i)_{i \in I}$, endowed with the structure of bayesian network, and a subset S of I, we consider the problem of computing the distribution of the subfamily $(X_a)_{a \in S}$ (resp. the distribution of $(X_b)_{b \in \bar{S}}$, where $\bar{S} = I - S$, conditionnally to $(X_a)_{a \in S}$). We present how it is possible to decompose this task into several parallel computations, each of them related to a subset of S (resp. of $\bar{S}$); these partial results are then put together as a product. Concerning the computation of $(X_a)_{a \in S}$, this amounts to the exhibition of a structure of bayesian network of level 2 on S.