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).