Algorithmique pour les Réseaux Bayésiens et leurs extensions

Linda Smail

UNIVERSITÉ DE MARNE-LA-VALLÉE

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE MARNE-LA-VALLÉE

Spécialité : Mathématiques Appliquées et Applications des Mathématiques

présentée et soutenue publiquement par

Linda Smail

le 30 Avril 2004

Titre

Algorithmique pour les Réseaux Bayésiens et leurs Extensions

Directeur de thèse : Jean Pierre Raoult

Jury :

Rapporteurs :	Marc Bouissou, EDF. R et D
	Evelyne Flandrin, Univ. Paris 5
	Eric Moulines, ENST
Examinateurs :	Wenceslas Fernandez de La Vega, Univ. Paris-Sud
	Kamel Mekhnacha, INRIA Rhône Alpes
	Paul Munteanu, ESIEA
	Georges Oppenheim, Univ. Marne-la-Vallée
	Jean Pierre Raoult, Univ. Marne-la-Vallée

A l'esprit de mon cher père.

Remerciements.

Je tiens particulièrement à exprimer ma reconnaissance à Jean-Pierre Raoult, mon directeur de thèse, pour sa disponibilité, son écoute et ses conseils, qui m'ont été toujours précieux, sa confiance, son investissement scientifique et humain qui ont été essentiels à la réalisation de ce travail.

Je voudrais également exprimer toute ma gratitude aux professeurs Eric Moulines et Evelyne Flandrin qui, en leur qualité de rapporteurs, m'ont fait l'honneur d'examiner minutieusement ce travail.

Les discussions que j'ai eues avec Georges Oppenheim et Marc Bouissou tout au long de ma thèse ont été particulièrement enrichissantes, je leur en suis reconnaissante et les remercie d'avoir accepté de faire partie du jury de ma thèse.

A Kamel Mekhnacha j'adresse des remerciements particuliers pour toutes les questions pertinentes qu'il a su me poser et les discussions fructueuses que nous avons eues, et qui ont participé à l'enrichissement de ce travail. Sa présence dans mon jury de thèse me fait très plaisir, tout comme celle de Paul Munteanu, et aussi Wenceslas Fernandez de La Vega, que je remercie sincèrement pour le regard attentif qu'ils ont porté sur mon travail.

Je tiens aussi à remercier pour son accueil toute l'équipe du laboratoire d'Analyse et Mathématiques Appliquées de l'université de Marne-la-Vallée, où j'ai effectué cette thèse. En particulier j'exprime ma gratitude à tous les membres du groupe de travail de fiabilité : Christiane Cocozza, Sophie Mercier et Michel Roussignol, pour leurs conseils et encouragements.

Je remercie également Emmanuel Mazer et Kamel Mekhnacha de l'équipe Laplace, localisée à INRIA Rhône Alpes, pour leur accueil, les longues journées de travail et les multiples discussions qui m'ont aidée dans mes travaux.

Tous ceux qui ont pris de leur temps pour une lecture attentive et m'ont apporté leurs remarques et judicieux conseils : Karim Mennas, Yannick Lefebvre, Wahid Boukraa, Margot Desgrouas. A chacun, toute ma gratitude.

Un grand merci également à Mireille Morvan, pour sa gentillesse et son aide.

Enfin, je remercie toute ma famille et tout particulièrement ma mère, de m'avoir soutenue et encouragée, et ma sœur Nabila pour son aide dans les moments difficiles.

Table des matières

Table des figures

Liste des tableaux

Introduction

Au carrefour de plusieurs disciplines comme les probabilités et statistiques, l'informatique, les sciences cognitives, l'analyse décisionnelle (en particulier pour le diagnostic médical), la théorie des jeux, les télécommunications (en particulier pour la transmission des informations), etc, les réseaux bayésiens constituent aujourd'hui une possibilité efficace dans la résolution de nombreux problèmes. Ils sont le résultat d'une convergence entre deux disciplines : en premier lieu les méthodes statistiques, parce qu'elles sont précisément conçues pour permettre le passage de l'observation à la loi de probabilité, ensuite les technologies de l'intelligence artificielle parce que leur vocation est de permettre aux ordinateurs de traiter de la connaissance [6].

Les réseaux bayésiens constituent l'un des formalismes les plus complets et les plus cohérents pour l'acquisition, la représentation et l'utilisation de connaissances par des ordinateurs. Encore du domaine de la recherche au début des années 90, cette technologie connaît de plus en plus d'applications pratiques depuis le diagnostic de panne à la détection de fraudes, en passant par la localisation des gènes [23], [11], [45], [27]. L'intérêt que lui portent certains industriels, qui ont investi fortement sur les réseaux bayésiens, est un signe du caractère stratégique de cette technologie.

Les réseaux bayésiens, qui doivent leur nom aux travaux de Thomas Bayes au dix-huitième siècle sur "la probabilité des causes", sont le résultat de recherches effectuées dans les années 80, par Judea Pearl [33], [34], [35], [36]. Ils ont fait l'objet de nombreuses recherches lors de ces dernières années et sont présents dans de nombreux articles ; on peut citer par exemple [31], [53], [54], [52], [12], [24], [9], [18], [25], [26] et [50].

Une des caractéristiques les plus importantes des réseaux bayésiens est leur capacité à coder les distributions jointes de familles de variables aléatoires ainsi que leurs distributions conditionnelles relativement à d'autres familles de variables aléatoires (la fixation de valeurs conditionnantes est appelée par les praticiens des réseaux bayésiens une "instantiation").

La structure graphique d'un réseau bayésien traduit les relations d'indépendance conditionnelle entre variables aléatoires (voir Annexe B) à l'aide de liens munissant l'ensemble des ces variables de la structure de graphe orienté sans circuits (voir Annexe A) ; elle améliore la compréhension du modèle du fait que les différentes relations de dépendance entre les variables aléatoires en jeu dans le modèle peuvent être lues directement à partir de cette structure. Initialement c'est cette structure graphique qui a rendu les réseaux

bayésiens très attractifs car elle pouvait coder la structure du domaine à modéliser. La facilité de compréhension des réseaux bayésiens est l'un des plus grands avantages des représentations qu'ils permettent, comparées par exemple à celles obtenues par les réseaux neuronaux qui sont plus difficiles à comprendre.

L'utilisation essentielle des réseaux bayésiens est de calculer des probabilités conditionnelles d'événements reliés les uns aux autres ; cette utilisation a été appelée par les utilisateurs anglophones "inference", d'où l'emploi par les utilisateurs français des réseaux bayésiens du terme "inférence", quoique son sens soit ici différent de celui qu'il a en statistique. Elle consiste à propager une ou plusieurs informations certaines au sens du réseau (c'est à dire les valeurs établies par certaines variables), pour en déduire comment ceci intervient sur les probabilités d'événements liés à d'autres variables. Certains types de calculs ont été particulièrement étudiés et implémentés dans des logiciels, parce qu'ils peuvent correspondre à des utilisations pratiques courantes ; il s'agit du calcul de la loi marginale d'une variable ou de sa loi conditionnelle relativement à un ensemble d'observations ; ce type d'inférence, appelé "mise à jour" des probabilités est essentiel en particulier dans les applications à l'élaboration de diagnostic, où l'on doit reconsidérer son avis sur la situation en fonction d'une ou plusieurs nouvelles observations. S'agissant en général de variables aléatoires à nombre fini de valeurs, il s'agit ici essentiellement d'un problème d'optimisation de calcul, puisque celui-ci devient de plus en plus lourd suivant la complexité du graphe relativement à la fois au nombre de variables et au nombre de valeurs prises par ces variables.

La correspondance qui existe entre la structure graphique et la structure probabilisite associée va permettre de ramener pour une bonne part l'ensemble des problèmes d'inférence à des problèmes de graphes. Cependant ces problèmes restent relativement complexes et donnent lieu à de nombreuses recherches [60], [7], [19], [51].

L'inférence probabiliste sur les réseaux bayésiens est une tâche intense et difficile ; Cooper [10] montre que ce problème est NP-difficile, quoique plusieurs recherches aient été dirigées pour développer des algorithmes efficaces pour ce genre de problèmes [8], [57] [44].

Nous présentons maintenant un bref historique des principales méthodes d'inférence "exactes" (par opposition à "approchées") existantes.

Pearl a édité dans le début des années 80 un algorithme d'inférence efficace dans des réseaux simples [35], dit polyarbres (voir Annexe A.3) pour le calcul de la distribution marginale d'une variable. Cet algorithme est exact et a une complexité polynomiale dans le nombre de nœuds, mais fonctionne seulement pour les polyarbres.

Pearl a présenté également un autre algorithme dit "loop cutset conditioning" (voir [35]) d'inférence exacte pour des graphes plus généraux (orientés sans circuits). Cet algorithme change le réseau en un polyarbre en le restreignant au complémentaire d'un sous-ensemble choisi de variables aléatoires désigné par le nom "loop cutset". Le calcul dans le réseau résultant est effectué par l'algorithme des polyarbres de Pearl et ensuite les résultats des calculs associés à chaque instantiation dans le "loop cutset" sont combinés en étant pondérées par la loi restreinte au "loop cutset". La complexité de l'algorithme augmente

exponentiellement avec la taille du sous-ensemble de variables fixées. Il est ainsi important de réduire au minimum la taille de ce sous-ensemble de variables de conditionnement ; malheureusement, le problème de minimisation de cet ensemble est NP- difficile.

Lauritzen et Spiegelhalter ont présenté ([39]) un algorithme dit dans la littérature "algorithme LS" et qui fait usage d'un "arbre de cliques" (les cliques sont des parties remarquables de l'ensemble des variables aléatoires du réseau, dont on trouvera la définition en chapitre 3). Cet algorithme est une généralisation de l'algorithme original de Pearl. L'idée de base en est de transformer le réseau bayésien en un arbre de cliques appelé "arbre de jonction" (en d'autres termes on établit entre certaines cliques des liens privilégiés). Associant à ces cliques des fonctions appelées potentiels, l'algorithme modifie progressivement ces potentiels, en passant d'une clique à une clique qui lui est liée (ce qu'on appelle "transmission de messages") de sorte qu'en fin de son déroulement le potentiel de chaque clique soit la loi conjointe des variables aléatoires constituant cette clique. L'algorithme de propagation fonctionne efficacement pour les réseaux clairsemés, mais peut être extrêmement lent pour les réseaux denses. Sa complexité est exponentielle par rapport à la taille de la plus grande clique de l'arbre de jonction [13].

Un autre algorithme est apparu pour la première fois dans Zhang et Poole [65]. C'est essentiellement l'algorithme "d'élimination de variables" de Dechter (voir [15]) ainsi appelé parce qu'il élimine par marginalisation (c'est à dire intégration) les variables les une après les autres. Un ordre dans lequel les variables doivent être marginalisées est exigé comme entrée de cet algorithme ; on l'appelle l'ordre d'élimination. Le calcul dépend de cet ordre. La complexité de l'algorithme d'élimination de variables peut être mesurée par le nombre d'opérations d'additions et de multiplications numériques qu'il exécute. Trouver un ordre d'élimination optimal est un problème NP-difficile [55].

Ils existe plusieurs autres algorithmes pour l'inférence exacte dans les réseaux bayésiens. Ils sont souvent basés sur les méthodes dites de regroupements ; nous citons par exemple l'algorithme de Shenoy-Shafer [49], Hugin [16] et l'algorithme de la propagation paresseuse "lazy propagation" [43], [63], [64]. (Voir aussi [59], [58], [56], etc). Tous ces algorithmes d'inférence, basés sur le regroupement de variables, partagent une complexité exponentielle par rapport à la taille de la plus grande clique de l'arbre de jontion associé [39], [62], [61].

Indiquons maintenant comment notre travail se situe dans ce déroulement de recherches.

Notre contribution dans cette thèse se situe à la fois au niveau de la formalisation des outils et au niveau algorithmique. Notre résultat essentiel est la proposition d'un algorithme de calcul de probabilités et de probabilités conditionnelles, dans un réseau bayésien, pour n'importe quelle sous famille de variables du réseau considéré. Cet algorithme dit "algorithme des restrictions successives" est basé sur deux nouvelles notions que nous introduisons également : la première notion est celle de réseau bayésien de niveau deux, la seconde est celle de descendance proche. Ces deux notions seront présentées en détail dans cette thèse. Nous proposons aussi, compte tenu des propriétés graphiques du réseau bayésien, et des propriétés d'indépendance conditionnelle dans le réseau bayésien, des résultats qui

permettent l'écriture de la loi d'une sous famille de variables d'un réseau bayésien sous la forme d'un produit de lois. Cette factorisation est très importante pour le calcul des lois dans des réseaux bayésiens de grande taille puisqu'elle permet de mener plusieurs petits calculs en parallèle au lieu d'un seul grand calcul.

Ce document se compose de quatre chapitres et trois annexes.

Le chapitre 1 présente en détails la théorie des réseaux bayésiens, ainsi que leurs propriétés, notamment la représentation de l'indépendance conditionnelle, nécessaire pour la compréhension de la suite de cette thèse. Nous introduisons également une nouvelle notion, celle de réseau bayésien de niveau deux, utile pour l'introduction de notre algorithme de calcul sur les réseaux bayésiens ; nous donnons également quelques résultats fondamentaux et nous situons dans notre formalisme un exemple d'école de réseau bayésien dit "Visite en Asie".

Dans le second chapitre, nous exposons une propriété graphique appelée "d-séparation" [47], [37] grâce à laquelle on peut déterminer, pour tout couple de variables aléatoires ou de groupes de variables, et tout ensemble de conditionnement, s'il y a nécessairement, ou non, indépendance conditionnelle. De nouvelles relations deviennent ainsi lisibles sur le graphe. Cette notion est très importante dans l'étude de l'indépendance conditionnelle dans les réseaux bayésiens, elle permet de préciser dans quelles conditions une information peut être traitée localement sans perturber l'ensemble du graphe. Il est important également de signaler que, si la d-séparation est une propriété purement graphique, c'est-à-dire uniquement liée au graphe, son utilisation est liée au schéma de causalité que l'on attache à ce graphe. Nous présentons également dans ce chapitre des résultats concernant le calcul de probabilités ou probabilités conditionnelles dans les réseaux bayésiens en utilisant les propriétés de la d-séparation. Ces résultats, qui concernent des écritures à notre connaissance originales de la factorisation de la loi jointe et de la loi conditionnée d'une famille de variables aléatoires du réseau bayésien (en liaison avec la notion de réseau bayésien de niveau deux) doivent trouver leur utilité pour les réseaux bayésiens de grande taille.

Le troisième chapitre donne la présentation détaillée et la justification d'un des algorithmes connus de calcul dans les réseaux bayésiens : il s'agit de l'algorithme LS (Lauritzen and Spigelhalter) [39], [22], [14], basé sur la méthode de l'arbre de jonction. Pour notre part, après avoir présenté la notion d'arbre de jonction, ou de façon équivalente, de suite recouvrante propre possédant la propriété d'intersection courante, nous proposons un algorithme en deux versions (dont l'une est originale) qui permet de construire une suite de parties d'un réseau bayésien possédant cette propriété, sans conditions restrictives sur les parties comme ceci est le cas pour la méthode des arbres de jonction. Cette présentation est accompagnée d'exemples.

Dans le chapitre 4, nous donnons une présentation détaillée de l'algorithme des restrictions successives que nous proposons pour le calcul de lois (dans sa première version), et de lois conditionnelles (dans sa deuxième version). Cela est présenté après l'introduction d'une nouvelle notion : il s'agit de la descendance proche ; c'est une notion purement graphique et qui rentre dans la justification de l'algorithme des restrictions successives. Nous présentons également une application de l'algorithme des restrictions successives sur l'exemple "Visite

en Asie" présenté en chapitre 1, et nous comparons le nombre d'opérations élémentaires effectuées avec celui qui intervient dans l'application de l'algorithme LS sur le même exemple. Le gain de calcul qui, à la faveur de cet exemple, apparaît au profit de l'algorithme des restrictions successives, sera comme toujours, d'autant plus marqué que la taille des réseaux et le nombre de valeurs prises par les variables seront plus élevés. C'est ce qui justifie l'insertion de notre algorithme au sein de "ProBT", un logiciel d'inférence probabiliste, réalisé et diffusé par Kamel Mekhnacha, Juan-Manuel Ahuactzin sous la direction de Emmanuel Mazer et Pierre Bessière, de l'équipe Laplace localisée dans le laboratoire Gravir à INRIA Rhône Alpes.

En annexes nous rappelons les propriétés des graphes orientés sans circuits nécessaires pour la compréhension de la thèse (Annexe A), ainsi que les notions de base sur l'indépendance conditionnelle (Annexe B). Nous terminons en Annexe C, en donnant l'équivalence de plusieurs définitions des réseaux bayésiens, dont certaines ne sont pas évoquées au chapitre 1.

Chapitre 1

Présentation des Réseaux Bayésiens et Réseaux Bayésiens de Niveau Deux

La partie 1 de ce chapitre a pour objectif de rappeler les notions de base sur les réseaux bayésiens. Nous présentons ces modèles ainsi que les relations d'indépendance conditionnelle et les représentations graphiques associées. Nous introduisons dans la partie 2 un autre modèle graphique et probabiliste qui est le réseau bayésien de niveau deux ainsi que ses propriétés.

Les notions de théorie des graphes et de probabilités conditionnelles utilisées sont rappelées respectivement dans les Annexes A et B de la thèse.

1.1 Réseaux bayésiens

1.1.1 Notations

Soit (I, G) **un graphe orienté sans circuits** (voir Annexe A) dont l'ensemble des **nœuds** (aussi appelés **sommets**) I est fini.

Pour tout i, on note :

1. $p_G(i)$ (ou $p(i)$) l'ensemble des **parents** (ou prédécesseurs) de i : j est un parent de i si et seulement si $(j, i) \in G$; un nœud dont l'ensemble des parents est vide est appelé **racine** ;
2. $e_G(i)$ (ou $e(i)$) l'ensemble des **enfants** (ou successeurs) de $i \in I$: j est un enfant de i si et seulement si $(i, j) \in G$ (voir Annexe A) ; un nœud dont l'ensemble des enfants est vide est appelé **feuille** ;
3. $d_G(i)$ (ou $d(i)$) l'ensemble des **descendants** de $i \in I$: j est un descendant de i si et seulement s'il existe un chemin de i à j, i.e. une suite $(i_0, \ldots, i_s)$ telle que

$i_0 = i$, $i_s = j$ et $\forall\, m \in \{1, \ldots, s\}$ $(i_{m-1}, i_m) \in G$.

Pour chaque $i \in I$, soit $\mathcal{M}_i = (\Omega_i, \mathcal{F}_i, \mu_i)$ un espace mesuré, où μ_i est une mesure positive $\sigma-$finie sur l'espace mesurable $(\Omega_i, \mathcal{F}_i)$[1] ; en particulier si Ω_i est fini, μ_i est la mesure de dénombrement.

Pour toute partie non vide J de I, on note $\mathcal{M}_J = (\Omega_J, \mathcal{F}_J, \mu_J)$ l'espace mesuré produit des $\mathcal{M}_j = (\Omega_j, \mathcal{F}_j, \mu_j)$ où $j \in J$. Si cela ne prête pas à une confusion, on notera $(\Omega, \mathcal{F}, \mu)$ pour $(\Omega_I, \mathcal{F}_I, \mu_I)$. Si $x_I \in \Omega_I$, on note x_J l'élément de Ω_J obtenu on ne conservant dans $x_I = (x_i)_{i\in I}$ que les composantes appartenant à J ; autrement dit x_J est la projection de x_I sur Ω_J ; plus généralement, si $J \subseteq J' \subseteq I$, et si on considère $x_{J'} \in \Omega_{J'}$, on note x_J la projection de $x_{J'}$ sur Ω_J.

1.1.2 Définition

Nous serons fréquemment amenés à considérer des applications à valeurs réelles, définies sur des espaces du types $\Omega_J = \prod_{j\in J} \Omega_j$, où $J \subseteq I$. Si f est une telle fonction, il nous sera utile et non gênant de considérer la même notation pour une extension aux espaces $\Omega_{J'}$, où $J \subseteq J' \subseteq I$; en particulier, en procédant à une extension à $\Omega(= \Omega_I)$ tout entier, on notera, pour tout $x \in \Omega$, $f(x_I) = f(x_J)$. Grâce à cette convention, étant donné $k+1$ parties de I, soit $J, J_1, \ldots, J_k$, telles que, pour tout j $(1 \leq j \leq k)$, $J_j \subseteq J$ et, pour tout j, une application f_j de Ω_{J_j} dans $\mathbb{R}$, on pourra définir sur Ω_J l'application $\prod_{j=1}^k f_j \ : \ x_J \rightsquigarrow \prod_{j=1}^k f_j(x_{J_j})$.

Définition 1 *On appelle* **réseau bayésien** *sur* $(I, G, (\mu_i)_{i\in I})$ *toute famille* $(f_{i/p(i)})_{i\in I}$ *où, pour tout* i, $f_{i/p(i)}$ *est une application mesurable de* $(\Omega_{i\cup p(i)}, \mathcal{F}_{i\cup p(i)})$ *dans* $(\mathbb{R}_+, \mathbb{B}_+)$ *(où* $\mathbb{B}_+$ *est la tribu borélienne de* $\mathbb{R}_+$*), notée*

$$(x_i \ , \ x_{p(i)}) \quad \rightsquigarrow \quad f_{i/p(i)}(x_i/x_{p(i)})$$

telle que : pour tout $x_{p(i)} \in \Omega_{p(i)}$, $f_{i/p(i)}(\cdot\ /x_{p(i)})$ *est une densité de probabilité par rapport à la mesure* μ_i *(i.e.* $\int_{\Omega_i} f_{i/p(i)}(x_i/x_{p(i)}) d\mu_i(x_i) = 1$*) ;* μ_i *est dite "dominante".*
Si $p(i) = \emptyset$*, la notation* $f_{i/p(i)}$ *désigne une application* f_i *de* Ω_i *dans* $\mathbb{R}_+$ *et la condition ci-dessus exprime que c'est une densité de probabilité par rapport à* μ_i

Notation :

On parlera du réseau bayésien (RB) $(I, G, (\mu_i)_{i\in I}, (f_{i/p(i)})_{i\in I})$. On dira que (I, G) (ou en bref G) est le graphe (orienté sans circuits) sous-jacent au réseau bayésien. Chaque fois que cela ne prêtera pas à ambiguïté, on notera : le RB $(I, G, (\Omega_i)_{i\in I}, (f_{i/p(i)})_{i\in I})$ ou le RB $(I, G, (f_{i/p(i)})_{i\in I})$.

1.1.3 Probabilité définie par un réseau bayésien

L'intérêt de la notion de réseau bayésien, telle que nous venons de l'introduire, tient tout d'abord au fait qu'elle définit une probabilité sur Ω_I, relativement à laquelle toutes les

[1] S'il n'y a pas de risque de confusion, on omettra de rappeler $\mathcal{F}_i$

interprétations probabilistes prendront leurs sens. On a en effet la proposition suivante :

Proposition 1 *Soit* $(I, G, (\mu_i)_{i\in I}, (f_{i/p(i)})_{i\in I})$ *un réseau bayésien ; l'application, de* Ω_I $(= \prod_{i\in I} \Omega_i)$ *dans* $\mathbb{R}_+$,

$$x_I = (x_i)_{i\in I} \quad \rightsquigarrow \quad \prod_{i\in I} f_{i/p(i)}(x_i/x_{p(i)}) \;=\; f_I(x_I) \tag{1.1}$$

est une densité, par rapport à $\mu_I \;=\; \prod_{i\in I}\; \mu_i$, *d'une probabilité notée* P_I.

Démonstration

Soit $n = card(I)$ et soit $(i_1, \ldots, i_n)$ un ordre hiérarchique sur I, on rappelle (voir Annexe A.2) qu'un **ordre hiérarchique** vérifie que, pour tout $s \in \{1, \ldots, n\}$, les parents de i_s sont inclus dans $\{i_1, \ldots, i_{s-1}\}$. Pour montrer que f_I est une densité de probabilité, on doit établir que

$$\int_{\Omega_I} f_I(x_{i_1}, \ldots, x_{i_n})\; d\mu_{i_1}(x_{i_1}) \;\cdots\; d\mu_{i_n}(x_{i_n}) \;=\; 1\;;$$

on effectuera ce calcul en intégrant successivement par rapport à $i_n, i_{n-1}, \ldots, i_1$.
On va établir, par récurrence descendante sur s, que, après intégrations relativement à $i_n, \ldots, i_{s+1}$, on a obtenu la fonction :

$$h_s(x_{i_1}, \ldots, x_{i_s}) \;=\; \prod_{t\leq s} f_{i_t/p(i_t)}(x_{i_t}/x_{p(i_t)}),$$

que l'on doit intégrer relativement à x_{i_s} ; or il résulte immédiatement du fait que l'ordre est hiérarchique que, dans l'expression de h_s, x_{i_s} ne figure que dans le facteur $f_{i_s/p(i_s)}(x_{i_s}/x_{p(i_s)})$. Donc, comme les variables $x_{p(i_s)}$ sont fixées à ce stade d'intégration, l'intégrale sur Ω_{i_s} vaut 1, de sorte que la fonction à intégrer au pas suivant est bien :

$$h_{s-1}(x_{i_1}, \ldots, x_{i_{s-1}}) \;=\; \prod_{t\leq s-1} f_{i_t/p(i_t)}(x_{i_t}/x_{p(i_t)}).$$

Ceci implique en particulier que la fonction h_s est une densité, par rapport à $\mu_{\{i_1,\ldots,i_s\}}$, de $P_{\{i_1,\ldots,i_s\}}$ projection de P_I sur $\prod_{t=1}^{s}\; \Omega_{i_t}$.

■

Terminologie : On pourra parler du réseau bayésien (I, G, P_I) (en sous entendant les espaces mesurés $(\Omega_i, \mathcal{F}_i, \mu_i)$).

Notation :

Pour tout $J \subset I$, on note P_J la projection de P_I sur Ω_J ; en particulier si $J = \{i\}$, on note P_i la projection de P_I sur Ω_i.

On a alors la proposition suivante qui s'applique aux **parties commençantes** pour le graphe G : $J \subseteq I$ est commençante si les parents de tout élément de J appartiennent aussi à J.

Proposition 2 *Soit J une partie commençante de I ; alors l'application, de $\Omega_J (= \prod_{j\in J} \Omega_j)$ dans $\mathbb{R}_+$,*

$$x_J = (x_j)_{j\in J} \quad \rightsquigarrow \quad \prod_{j\in J} f_{j/p(j)}(x_j/x_{p(j)}) \;=\; f_J(x_J) \tag{1.2}$$

est une densité de P_J par rapport à μ_J ; en particulier, si i est une racine de I, c'est-à-dire si $p(i) = \emptyset$, f_i, qui est une application de Ω_i dans $\mathbb{R}_+$, est une densité de P_i par rapport à μ_i.

Démonstration :

Soit $s = card(J)$. J étant une partie commençante, il est possible de choisir un ordre hiérarchique tel que $J = \{i_1, \ldots, i_s\}$; le résultat a déja été fourni en fin de démonstration de la proposition (1).

■

Un problème fondamental relatif aux réseaux bayésiens est de caractériser, étant donné une partie J de I, la probabilité P_J, autrement dit d'en calculer une densité f_J ; la propriété précédente montre que, si J est commençante, la densité f_J s'écrit par simple produit à partir des données. Sinon, on a intérêt à considérer la partie commençante engendrée par J, soit J^+, c'est-à-dire l'union de J et de tous les nœuds ayant un descendant dans J ; J étant ici supposée non commençante elle-même, J^+ contient strictement J, et alors

$$f_J(x_J) \;=\; \int_{\Omega_{J^+-J}} f_{J^+}(x_J, x_{J^+-J})\; d\mu_{J^+-J}(x_{J^+-J})\;.$$

La difficulté technique est alors de gérer l'intégration sur Ω_{J^+-J} ; dans les chapitres suivants on rappellera des techniques connues et on introduira des techniques nouvelles à cet effet.

1.1.4 Interprétation en termes de variables aléatoires et de lois conditionnelles

Considérons maintenant que P_I est la loi d'une famille de variables aléatoires $X_I = (X_i)_{i\in I}$ définies sur un même espace mesurable. Alors P_J est la loi de la sous famille $X_J = (X_j)_{j\in J}$ et pour tout i, P_i est la loi de la variable aléatoire X_i.

P_I étant définie par sa densité (voir équation (1.1)), on dit alors que X_I est **compatible** avec le réseau bayésien $(I, G, (f_{i/p(i)})_{i\in I})$; on pourra parler du réseau bayésien $(I, G, (X_i)_{i\in I})$.

Remarque : Par abus de langage, si $j \in p(i)$ (respectivement $j \in e(i)$, $j \in d(i)$) nous disons que la variable aléatoire X_j est un parent de X_i (respectivement un enfant de X_i, un descendant de X_i).

Le théorème suivant est à la base de l'usage pratique des réseaux bayésiens. On y fait usage de l'indépendance conditionnelle : étant donné trois variables aléatoires X_1, X_2, X_3,

l'indépendance de X_1 et X_2 conditionnellement à X_3 est notée $X_1 \perp\!\!\!\perp X_2 \ / \ X_3$ (voir Annexe B où sont rappelées les conventions sur la manipulation d'égalités presque sûres).

Théorème 1 *Soit $(I, G, (\Omega_i, \mathcal{F}_i, \mu_i)_{i\in I}, (f_{i/p(i)})_{i\in I})$ un réseau bayésien et soit $(X_i)_{i\in I}$ une famille de variables aléatoires définies sur un même espace telle que, pour tout i, X_i soit à valeurs dans $(\Omega_i, \mathcal{F}_i)$ et de loi absolument continue par rapport à la mesure σ-finie μ_i.*

Pour que $(X_i)_{i\in I}$ soit compatible avec le réseau bayésien il faut et il suffit que :

1. *pour tout i, $f_{i/p(i)}$ soit une densité conditionnelle de X_i, conditionnellement à $X_{p(i)}$ (en particulier une densité de X_i si $p(i) = \emptyset$),*
2. *pour tout i, X_i soit, conditionnellement à $X_{p(i)}$, indépendante de $X_{I-\{i\cup p(i)\cup d(i)\}}$.*

Avant de rappeler quelle est la preuve (classique) de cette équivalence fondamentale, précisons que la propriété (2) (dite de Markov) exprime que toute variable aléatoire est, conditionnellement à ses parents, indépendante de toutes les variables aléatoires du réseau autres que elle-même, ses parents, et ses descendants. Dans le cas particulier d'une racine i du réseau, où $p(i) = \emptyset$, elle exprime que X_i est indépendante de toute variable autre que elle-même et ses descendants ; en particulier les racines sont indépendantes entre elles.

C'est cette propriété qui intervient le plus pour les applications à des situations aléatoires concrètes, lors de la phase de construction d'un réseau bayésien à fin de modélisation d'enchaînement de causalités ; elle exprime, intuitivement, que les seules "causes immédiates" d'une variable aléatoire sont représentées par ses variables aléatoires parentes, et que les "causes initiales" sont indépendantes entre elles et représentées par les variables aléatoires racines.

Démonstration :

1. Condition nécessaire :

Soit $X_I = (X_i)_{i\in I}$ une famille de variables aléatoires compatible avec le réseau bayésien $(I, G, (f_{i/p(i)})_{i\in I})$. Montrons successivement les propriétés 1 et 2.

1. Soit $g_{i/p(i)}$ la densité de X_i conditionnellement à $X_{p(i)}$ définie par,

$$g_{i/p(i)}(x_i/x_{p(i)}) \ = \ \frac{f_{i \,\cup\, p(i)}(x_i \cup p(i))}{f_{p(i)}(x_{p(i)})} \ .$$

Si les parents de i sont des racines, $p(i)$ est une partie commençante ainsi que $i \cup p(i)$ et on a, d'après la proposition 2,

$$f_{i \,\cup\, p(i)}(x_i \cup p(i)) \ = \ f_{i/p(i)}(x_i/x_{p(i)}) \prod_{j\in p(i)} f_j(x_j) \ = \ f_{i/p(i)}(x_i/x_{p(i)}) f_{p(i)}(x_{p(i)}) \ ;$$

donc il est immédiat que $g_{i/p(i)} \ = \ f_{i/p(i)}$.

Sinon, soit $(p(i))^+$ la partie commençante engendrée par $p(i)$ (et alors $i \cup (p(i))^+$ est i^+, partie commençante engendrée par i) ; posons $a(i) = (p(i))^+ - p(i)$; $a(i)$, ensemble des **aïeux** de i, est non vide et i^+ est partitionné en $i \cup p(i) \cup a(i)$.

Il résulte de la proposition 2 que

$$f_{i^+}(x_{i^+}) = \prod_{j\in i^+} f_{j/p(j)}(x_j/x_{p(j)}) = f_{i/p(i)}(x_i/x_{p(i)}) \prod_{j\in p(i)} f_{j/p(j)}(x_j/x_{p(j)}) \prod_{k\in a(i)} f_{k/p(k)}(x_k/x_{p(k)})$$

et

$$f_{(p(i))^+}(x_{(p(i))^+}) = \prod_{j\in (p(i))^+} f_{j/p(j)}(x_j/x_{p(j)}) = \prod_{j\in p(i)} f_{j/p(j)}(x_j/x_{p(j)}) \prod_{k\in a(i)} f_{k/p(k)}(x_k/x_{p(k)})$$

Alors

$$\begin{aligned} f_{i\cup p(i)}(x_{i\cup p(i)}) &= \int_{\Omega_{a(i)}} f_{i^+}(x_i, x_{p(i)}, x_{a(i)}) d\mu_{a(i)}(x_{a(i)}) \\ &= \int_{\Omega_{a(i)}} f_{i/p(i)}(x_i/x_{p(i)}) \prod_{j\in p(i)} f_{j/p(j)}(x_j/x_{p(j)}) \prod_{k\in a(i)} f_{k/p(k)}(x_k/x_{p(k)})\ d\mu_{a(i)}(x_{a(i)}) \\ &= f_{i/p(i)}(x_i/x_{p(i)}) \int_{\Omega_{a(i)}} \prod_{j\in p(i)} f_{j/p(j)}(x_j/x_{p(j)}) \prod_{k\in a(i)} f_{k/p(k)}(x_k/x_{p(k)})\ d\mu_{a(i)}(x_{a(i)}) \\ &= f_{i/p(i)}(x_i/x_{p(i)})\ f_{p(i)}(x_{p(i)})\ , \end{aligned}$$

d'où, ici encore, $g_{i/p(i)} = f_{i/p(i)}$.

2. Montrons que pour tout $i \in I$, $X_i \perp\!\!\!\perp X_{I-\{i\cup p(i)\cup d(i)\}} \ / \ X_{p(i)}$, c'est à dire que

$$f_{i/I-\{i\cup d(i)\}} = f_{i/p(i)}.$$

On va calculer pour tout i :

$$f_{i/I-\{i\cup d(i)\}}(x_i/x_{I-\{i\cup d(i)\}}) = f_{i/I-\{i\cup d(i)\}}(x_i/x_{I-\{i\cup p(i)\cup d(i)\}}, x_{p(i)}).$$

D'une part $I - d(i)$ est une partie commençante, donc d'après la proposition 2 on a le résultat suivant :

$$f_{I-d(i)}(x_{I-d(i)}) \ = \prod_{j\in\ I-d(i)} f_{j/P(j)}(x_j/x_{P(j)})\ ;$$

d'autre part, il en est de même pour $I - \{i \cup d(i)\}$ et alors on a,

$$f_{I-\{i\cup d(i)\}}(x_{I-\{i\cup d(i)\}}) \ = \prod_{j\in\ I-\{i\cup d(i)\}} f_{j/P(j)}(x_j/x_{P(j)}),$$

Alors

$$\begin{aligned} f_{i/I-\{i\cup d(i)\}}(x_i/x_{I-\{i\cup p(i)\cup d(i)\}}, x_{p(i)}) &= \frac{f_{I-d(i)}(x_i, x_{I-\{i\cup p(i)\cup d(i)\}}, x_{p(i)})}{f_{I-\{i\cup d(i)\}}(x_{I-\{i\cup p(i)\cup d(i)\}}, x_{p(i)})} \\ &= \frac{\prod_{j\in I-d(i)}\ f_{j/p(j)}(x_j/x_{p(j)})}{\prod_{j\in\ I-\{i\cup d(i)\}}\ f_{j/(j)}(x_j/x_{p(j)})} \\ &= \frac{f_{i/p(i)}(x_i/x_{p(i)}) \prod_{j\in\ I-\{i\cup d(i)\}}\ f_{j/p(j)}(x_j/x_{p(j)})}{\prod_{j\in\ I-\{i\cup d(i)\}}\ f_{j/p(j)}(x_j/x_{p(j)})} \\ &= f_{i/p(i)}(x_i/x_{p(i)}) \end{aligned}$$

d'où le résultat.

2. Condition suffisante :

Soit P_I la loi de la famille de variables aléatoires $X_I = (X_i)_{i\in I}$. Par hypothèse sa loi est absolument continue par rapport à la mesure μ_I sur $(\Omega_I, \mathcal{F}_I)$; soit g_I une densité.

Pour tout i et toute partie J de I telle que $i \notin J$, notons $g_{i/J}$ une densité conditionnelle de X_i, conditionnellement à X_J ; par hypothèse (1), si $J = p(i)$, $g_{i/p(i)}$ coïncide (p.s) avec la fonction $f_{i/p(i)}$ du réseau bayésien $(I, G, (f_{i/p(i)})_{i\in I})$.

Montrer que X_I est compatible avec le réseau bayésien revient à établir que g_I n'est autre (p.s) que $f_I = \prod_{i\in I} f_{i/p(i)}$.

Soit $(i_1, \ldots, i_n)$ un ordre hiérarchique sur I. D'après la propriété générale d'emboîtement des conditionnements, on a

$$g_I(x_{i_1}, \ldots, x_{i_n}) \;=\; \prod_{i_s=1}^{n} g_{i_s/i_1,\ldots,i_{s-1}}(x_{i_s}/x_{i_1}, \ldots, x_{i_{s-1}}) \;.$$

Comme l'ordre considéré est hiérarchique, on a , pour tout $s \in \{1, \ldots, n\}$

$$p(i_s) \;\subset\; \{i_1, \ldots, i_{s-1}\} \;,$$

et de plus les éléments de $\{i_1, \ldots, i_{s-1}\} - p(i_s)$ ne sont pas des descendants de i_s ; or, par hypothèse (2), on a pour tout s

$$X_{i_s} \quad \perp\!\!\!\perp \quad X_{I-\{i_s\cup p(i_s)\cup d(i_s)\}} \;/\; X_{p(i_s)} \;,$$

donc en particulier

$$X_{i_s} \quad \perp\!\!\!\perp \quad X_{\{i_1,\ldots,i_{s-1}\}} \;/\; X_{p(i_s)} \;,$$

d'où

$$g_{i_s/i_1,\ldots,i_{s-1}}(x_{i_s}/x_{i_1}, \ldots, x_{i_{s-1}}) \;=\; g_{i_s/p(i_s)}(x_{i_s}/x_{p(i_s)}) \;=\; f_{i_s/p(i_s)}(x_{i_s}/x_{p(i_s)}).$$

Donc on a bien

$$g_I(x_I) \;=\; \prod_{s=1}^{n} f_{i_s/p(i_s)}(x_{i_s}/x_{p(i_s)}) \;=\; f_I(x_I) \;.$$

■

1.1.5 Commentaire et généralisation

La propriété (2) du théorème (1) peut être posée comme définition d'une famille de variables aléatoires $(X_i)_{i\in I}$ possédant la structure de réseau bayésien sur le graphe (I, G). Elle permet de généraliser la notion de réseau bayésien, car la notion d'indépendance conditionnelle a un sens en termes de probabilités sur les espaces mesurables les plus généraux (et pas nécessairement en termes de densités comme nous venons de le faire). Rappelons la (voir aussi Annexe B).

Définition 2 *Soient données* 3 *variables aléatoires définies sur un même espace,* X_1, X_2 *et* X_3. X_1 *et* X_2 *sont indépendantes conditionnellement à* X_3 *si et seulement si il existe des versions de probabilités conditionnelles, relativement à* X_3 , *de* X_1, X_2 *et* (X_1, X_2), *soit* $P_{1/3}$, $P_{2/3}$ *et* $P_{1,2/3}$, *telles que, pour tout couple de parties mesurables* $A_1(\subseteq \Omega_1)$ *et* $A_2(\subseteq \Omega_2)$ *on ait, pour* P_3-*presque tout* $x_3 \in \Omega_3$,

$$P_{1,2/3}(A_1 \cap A_2/x_3) \;=\; P_{1/3}(A_1/x_3)\; P_{2/3}(A_2/x_3)\;.$$

Il est donc possible de définir en toute généralité la notion de réseau bayésien sur le graphe orienté sans circuits (I, G) par :

Définition 3 *Etant donné un graphe orienté sans circuits* (I, G) *et, pour tout* $i \in I$, *une variable aléatoire* X_i *à valeurs dans* $(\Omega_i, \mathcal{F}_i)$, *on dit que* (I, G, X_I) *(où* $X_I = (X_i)_{i\in I}$*) est un réseau bayésien si, pour tout* i, X_i *est, conditionnellement à* $X_{p_G(i)}$, *indépendante de* $X_{I-\{i\cup p_G(i)\cup d_G(i)\}}$.

Mais dans ce travail, nous nous limiterons au cas dit "dominé", c'est à dire où, pour tout i, la loi conditionnelle $P_{i/p(i)}$ est définie par sa densité relativement à la mesure σ-finie μ_i. Dans les chapitres suivants, pour rester proches des applications informatiques, nous présenterons nos résultats dans le cadre des variables aléatoires discrètes (plus précisement les espaces Ω_i seront finis) et noterons $P_{i/p(i)}$ ou P_J au lieu de $f_{i/p(i)}$ ou f_J et $\sum$ au lieu de $\int$ (les mesures μ_i étant les mesures de dénombrement) :

$$\left|\begin{array}{lcl} P_{i/p(i)}(x_i/x_{p(i)}) & = & P[X_i = x_i/X_{p(i)} = x_{p(i)}] \\ \\ P_J(x_J) & = & P[X_J = x_J] \end{array}\right.$$

Remarque

On trouve parfois la terminologie "Réseaux Bayésiens Mixtes" pour désigner des réseaux où certaines variables aléatoires sont discrètes, et d'autres "continues", c'est-à-dire absolument continues par rapport à la mesure de lebesgue sur $\mathbb{R}$. Ils rentrent bien sûr dans la catégorie générale des réseaux bayésiens définis par des densités introduits en 1.1.2 ci-dessus. Mais les problèmes de calcul qu'ils posent nécessitent des outils spécifiques (voir [14], [40]).

1.1.6 Complément sur l'indépendance conditionnelle

L'indépendance conditionnelle est à la base de l'interprétation pratique des réseaux bayésiens, à tel point que certains auteurs ont développé une théorie "abstraite" de graphoïde ([41]) où ils manipulent des triplets (i, j, k) vérifiant des axiomes qui se trouvent être des propriétés de la relation $X_i \perp\!\!\!\perp X_j/X_k$. Une problématique intéressante est alors, étant donnée la loi de $(X_i)_{i\in I}$, la recherche de tous les triplets (i, j, k) tels que $X_i \perp\!\!\!\perp X_j/X_k$.

Nous ne nous intéressons pas dans cette thèse à ces points de vue, mais nous rappelons que, étant donné un réseau bayésien, il peut bien sûr se produire que certaines relations d'indépendance conditionnelle existent qui ne sont pas traduites par cette structure de réseau. Un exemple élémentaire classique est celui du triplet (X_1, X_2, X_3) tel que ces 3 variables aléatoires soient indépendantes deux à deux mais pas dans leur ensemble ; il est impossible de définir sur $\{1, 2, 3\}$ un graphe orienté sans circuits tel que le réseau bayésien associé justifie simultanément les 3 indépendances mutuelles ; si on considère le graphe orienté $1 \rightarrow 3 \leftarrow 2$ qui rend compte de $X_1 \perp\!\!\!\perp X_2$ et laisse ouverte la possibilité que (X_1, X_2, X_3) ne soient pas globalement indépendantes, on doit par ailleurs indiquer que $X_2 \perp\!\!\!\perp X_3$ et $X_1 \perp\!\!\!\perp X_3$, autrement dit que le triplet (X_1, X_2, X_3) serait aussi compatible avec les graphes orientés $2 \rightarrow 1 \leftarrow 3$ et $1 \rightarrow 2 \leftarrow 3$.

Nous donnons maintenant une propriété d'indépendance conditionnelle qui résulte de la structure de réseau bayésien.

Proposition 3 *Soit $(I, G, (X_i)_{i \in I})$ un réseau bayésien et soit J une partie de I telle que aucun élément de J ne soit descendant d'un autre élément de J. Alors les X_j, où $j \in J$, sont indépendantes conditionnellement à l'ensemble $X_{p(J)}$ de leurs parents et, pour tout $j \in J$, on a*

$$P_{j/p(J)} \;=\; P_{j/p(j)}.$$

Démonstration

Soit $J \subset I$, tel qu'aucun élément de J ne soit un descendant d'un autre élément de J ; cela signifie que $a(J) \cap J \;=\; \emptyset$. Pour montrer que les variables aléatoires $(X_j)_{j \in J}$ sont indépendantes conditionnellement à $X_{p(J)}$, on doit calculer $f_{J/p(J)}(x_J / x_{p(J)})$.

Si $p(J)$ est une partie commençante (donc aussi $J \cup p(J)$) on a

$$\begin{aligned} f_{J/p(J)}(x_J / x_{p(J)}) &= \frac{f_{J \cup p(J)}(x_J, x_{p(J)})}{f_{p(J)}(x_{p(J)})} \\ &= \frac{\prod_{j \in J \cup p(J)} f_{j/p(j)}(x_j / x_{p(j)})}{\prod_{j \in p(J)} f_{j/p(j)}(x_j / x_{p(j)})} \\ &= \prod_{j \in J} f_{j/p(j)}(x_j / x_{p(j)}) \end{aligned}$$

Sinon, considérons $(p(J))^+$, partie commençante engendrée par $p(J)$ et posons $a(J) =$

$(p(J))^+ - p(J)$; il vient que $J \cup p(J) \cup a(J)$ est aussi une partie commençante ; alors

$$
\begin{aligned}
& f_{J/p(J)}(x_J/x_{p(J)}) \\
&= \frac{f_{J\cup p(J)}(x_J, x_{p(J)})}{f_{p(J)}(x_{p(J)})} \\
&= \frac{\int_{a(J)} f_{J\cup p(J)\cup a(J)}(x_J, x_{p(J)}, x_{a(J)}) d\mu_{a(J)}(x_{a(J)})}{\int_{\Omega_{a(J)}} f_{p(J)\cup a(J)}(x_{p(J)}, x_{a(J)}) d\mu_{a(J)}(x_{a(J)})} \\
&= \frac{\int_{\Omega_{a(J)}} \prod_{j\in J} f_{j/p(j)}(x_j/x_{p(j)}) \prod_{k\in p(J)} f_{k/p(k)}(x_k/x_{p(k)}) \prod_{\ell\in a(J)} f_{\ell/p(\ell)}(x_\ell/x_{p(\ell)}) d\mu_{a(J)}(x_{a(J)})}{\int_{\Omega_{a(J)}} \prod_{k\in p(J)} f_{k/p(k)}(x_k/x_{p(k)}) \prod_{\ell\in a(J)} f_{\ell/p(\ell)}(x_\ell/x_{p(\ell)}) d\mu_{a(J)}(x_{a(J)})} \\
&= \prod_{j\in J} f_{j/p(j)}(x_j/x_{p(j)})
\end{aligned}
$$

Or $a(J) \cap J = \emptyset$; pour tout $j \in J$, le facteur $f_{j/p(j)}(x_j/x_{p(j)})$ ne dépend pas de $a(J)$, donc il sort de l'intégrale.

Dans les deux cas, on obtient que les variables aléatoires $(X_j)_{i\in J}$ sont indépendantes conditionnellement à $X_{p(J)}$.
Pour tout $j \in J$ on a

$$
\begin{aligned}
f_{j/p(J)}(x_j/x_{p(J)}) &= \int_{J-\{j\}} f_{J/p(J)}(x_J/x_{p(J)}) d\mu_{J-\{j\}}(x_{J-\{j\}}) \\
&= \int_{J-\{j\}} \prod_{\substack{k\in J \\ k\neq j}} f_{k/p(k)}(x_k/x_{p(k)}) d\mu_{J-\{j\}}(x_{J-\{j\}}) \\
&= f_{j/p(j)}(x_j/x_{p(j)}) \int_{J-\{j\}} \prod_{\substack{k\in J \\ k\neq j}} f_{k/p(k)}(x_k/x_{p(k)}) d\mu_J(x_J) \ ;
\end{aligned}
$$

on intègre successivement par rapport aux variables aléatoires $(X_k)_{k\in J-\{j\}}$; or, pour tout $k \in J - \{j\}$, $f_{k/p(k)}$ est une densité de probabilité et les parents $p(k)$ sont fixés, donc l'intégrale ne fait intervenir que le facteur 1, d'où

$$
f_{j/p(J)}(x_j/x_{p(J)}) = f_{j/p(j)}(x_j/x_{p(j)}).
$$

■

1.1.7 Exemple d'un réseau bayésien

L'exemple que nous allons présenter ici est issu des travaux de Lauritzen et Spiegelhalter [39] ; il a été cité dans plusieurs références sur les réseaux bayésiens [14], [16] et [22] ; il figure aussi dans la présentation de plusieurs logiciels de calcul sur les réseaux bayésiens (Netica, Hugin, Bayesia, BKD) et il a pour domaine une base de connaissances médicales. **Cet**

exemple nous a servi de cas test pour notre propre algorithme des restrictions successives (voir chapitre 4).

La dyspnée (difficulté de respiration) peut être causée par une tuberculose, un cancer du poumon ou une bronchite (ou par aucune ou plusieurs d'entre ces causes). Un passage récent en Asie augmente les risques de tuberculose ; fumer est reconnu comme étant un facteur important de risque pour le cancer du poumon ou la bronchite. Une radiographie des poumons ne permet pas de différencier le cancer du poumon et la tuberculose, ni ne permet d'observer la présence ou non d'une dyspnée.

On voudrait que cette base de connaissances permette de résoudre des situations telles que celle d'un patient atteint de dyspnée, ayant séjourné en Asie, sans connaissance sur son passé de fumeur, ni de radiographie des poumons. Le médecin voudrait pouvoir déterminer les probabilités d'existence des 3 maladies connues ; ou encore connaître les causes après avoir donné les valeurs observées qui sont les symptômes.

Soient les variables aléatoires suivantes :

1. Fumer : Consommation du tabac, qui prend les valeurs {Oui, Non}.
2. V. en Asie : Visite en Asie, qui prend les valeurs {Oui, Non}.
3. Bronchite, qui prend les valeurs {Présent, Absent}.
4. Cancer : Existence d'une tumeur au poumon, qui prend les valeurs {Présent, Absent}.
5. Tuber. : Tuberculose, qui prend les valeurs {Présent, Absent }.
6. Dyspnée, qui prend les valeurs {Présent, Absent}.
7. Radiologie : Résultat radiologique, qui prend les valeurs {Anormal, Normal}.

A propos de ces variables, on fait les hypothèses de modélisation suivantes :

Les variables Fumer (consommation du tabac) et Visite en Asie sont les causes, supposées indépendantes ; elle constituent donc les racines du réseau bayésien.

Les variables Bronchite, Cancer et Tuberculose sont les maladies, influencées par Visite en Asie pour la Tuberculose et consommation du tabac (Fumer) pour le Cancer et la Bronchite.

Les variables Dyspnée et Radiologie constituent les symptômes ; on suppose que la Dyspnée est influencée par les 3 maladies, alors que la Radiologie est influencée par la Tuberculose et le Cancer. De plus, on n'est pas capable d'isoler les influences du Cancer et de la Tuberculose sur la Radiologie et sur la Dyspnée ; c'est pourquoi une nouvelle variable Tuber. ou Cancer (Tuberculose ou Cancer) a été introduite, elle exprime la présence de l'une au moins des maladies Cancer et Tuberculose.

Cette base de connaissances peut être modélisée comme sur la figure 1.1 ci-dessous.

Maintenant pour compléter l'information sur ce réseau bayésien on donne les lois de probabilité de chacune de ses variables aléatoires ; ces lois dans ce cas de variables discrètes sont des tables de probabilités ; pour les deux variables sans parent (Fumer) et (V. en Asie) ce sont des probabilités absolues tandis que pour les autres variables ce sont des tables de probabilités conditionnelles, conditionnellement aux parents.

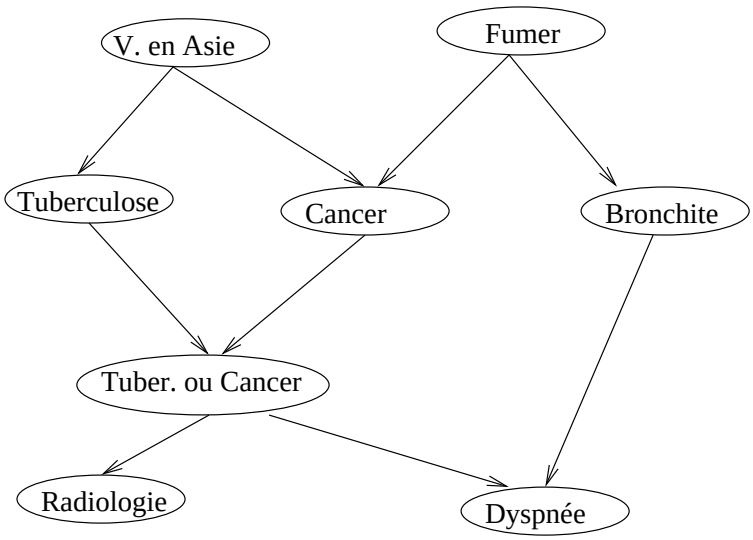


FIG. 1.1 – Exemple "Visite en Asie".

- Pour (Fumer) : $P(Fumer = Oui) = 0.5$.
- Pour (V. en Asie) : $P(V.enAsie = Oui) = 0.01$.
- Pour (Bronchite) :

$$P(Bronchite = Présent/Fumer = Oui) = 0.6$$
$$P(Bronchite = Présent/Fumer = Non) = 0.3$$

- Pour (Tuberculose)

$$P(Tuber. = Présent/V.enAsie = Oui) = 0.05$$
$$P(Tuber. = Présent/V.enAsie = Non) = 0.01$$

- Pour (Cancer) :

$$P(Cancer = Présent/Fumer = Oui) = 0.1$$
$$P(Cancer = Présent/Fumer = Non) = 0.01$$

-Pour ("Tuberculose ou Cancer") :

Tuber.	Cancer	Tuber. ou Cancer = vraie
Présent	Présent	1
Présent	Absent	1
Absent	Présent	1
Absent	Absent	0

TAB. 1.1 – Loi de "Tuber. ou Cancer" conditionnellement à Tuber. et Cancer.

-Pour (Radiologie) :

Tuber. ou Cancer	Radiologie = Anormal
Vraie	0.98
Faux	0.05

TAB. 1.2 – Loi de Radiologie conditionnellement à "Tuberculose ou Cancer" .

-Pour (Dyspnée) :

Tuber. ou Cancer	Bronchite	Dyspnée = Present
Vraie	Présent	0.9
Vraie	Absent	0.7
Faux	Présent	0.8
Faux	Absent	0.1

TAB. 1.3 – Loi de Dyspnée conditionnellement à "Tuber. ou Cancer" et Bronchite.

La probabilité jointe du réseau bayésien dans ce cas s'écrit comme produit des tables de probabilités conditionnelles associées aux variables de ce réseau.

Le calcul de la loi de chaque variable aléatoire du réseau, ainsi que des lois marginales ou lois conditionnelles, de certaines variables du réseau, pour répondre aux questions déjà posées, peut se faire par des algorithmes de calculs qu'on verra aux chapitres 3 et 4.

1.2 Réseaux Bayésiens de niveau deux

1.2.1 Besoin d'une notion nouvelle

Un problème fondamental sur les réseaux bayésiens est celui appelé en anglais "inference in bayesian networks", qui nous occupera dans la suite ; il s'agit de rechercher des probabilités P_J (où $J \subset I$) et $P_{J/K}$ (où $J \subset I$, $K \subset I$, $J \cap K = \emptyset$) autrement dit, dans le cadre dominé qui est le nôtre, de calculer des densités f_J et $f_{J/K}$. Plus précisement dans le cas des variables aléatoires discrètes, si, pour tout $J \subseteq I$ le cardinal de J est noté $|J|$ et pour tout i le cardinal de Ω_i est n_i, f_J est stocké informatiquement comme un tableau

$|J|$-dimensionnel à $\prod_{j\in I}\ n_j$ éléments (avec : $f_J((x_j)_{j\in J}) = P[\ \prod_{j\in J} X_j = x_j]\ )$ et $f_{J/K}$ est stocké informatiquement comme $\prod_{k\in K}\ n_k$ tableaux $|J|$-dimensionnels avec

$$f_{J/K}((x_j)_{j\in J}/(x_k)_{k\in K}) = P[\ \prod_{j\in J}[X_j = x_j]\ /\ \prod_{k\in K}[X_k = x_k]].$$

On a vu que f_J est élémentaire dans le cas où J est une partie commençante et que la structure de réseau bayésien reste alors totalement visible dans l'expression de f_J. Mais pour une partie J quelconque, donner f_J (en particulier, dans le cas discret, le tableau des $P(\ \prod_{j\in J}[X_j = x_j]\ )$, même si cela correspond à un besoin de calcul, fait en général perdre de l'information par rapport à la donnée du réseau bayésien initial ; on n'y hérite pas simplement de la structure qui permettait de factoriser $f_I(x_I)$ sous la forme $\prod_{i\in I}\ f_{i/p(i)}(x_i/x_{p(i)})$ et dont une certaine trace peut cependant subsituer sur f_J à condition de s'intéresser non à l'ensemble des éléments de J considérés isolément mais à une partition de J. C'est cette trace que nous traduisons par la mise en évidence (en chapitre 2, section 3) et la construction algorithmique (en chapitre 4) d'une structure de "réseau bayésien de niveau 2" sur J. Nous allons introduire ici cette notion en toute généralité (sur I).

1.2.2 Définition

Notation

Soient I un ensemble fini, une partition $\mathcal{I}$ de I, et un graphe orienté sans circuits $\mathcal{G}$ sur $\mathcal{I}$. Si $J \in \mathcal{I}$, notons $p_{\mathcal{G}}(J)$[2] l'ensemble des parents de J, c'est-à-dire l'ensemble des J' tels que $(J', J) \in \mathcal{G}$; ceci définit un sous-ensemble de parties de I (appartenant à $\mathcal{I}$), qu'on pourra, sans risque de confusion, identifier à leur union (elle-même partie de I), d'où l'écriture $p_{\mathcal{G}}(J) = \cup_{J'\in\mathcal{I},(J',J)\in\mathcal{G}} J'$.

Définition 4 *On dira que la densité de probabilité f_I est caractérisée par le* **réseau bayésien de niveau deux** *(RB2), sur I, $(\mathcal{I}, \mathcal{G}, (f_{J/p(J)})_{J\in\mathcal{I}})$, si, pour tout $J \in \mathcal{I}$, est donnée la densité conditionnelle $f_{J/p(J)}$ (qui, si $p_{\mathcal{G}}(J) = \emptyset$, est la densité marginale f_J), de telle sorte que*

$$f_I(x_I) = \prod_{J\in\mathcal{I}}\ f_{J/p(J)}(x_J/x_{p(J)}).$$

Remarque 1 : Un réseau bayésien usuel (que nous dirons de niveau 1 : RB1) est un cas particulier de RB2, avec la partition de I en singletons.

Remarque 2 : Les notations et propriétés détaillées en section 1 pour les RB1 s'étendent immédiatement aux RB2. En particulier, en termes de variables aléatoires, tout X_J est, conditionnellement à $X_{p_{\mathcal{G}}(J)}$, indépendant de $X_{I-\{J\cup p_{\mathcal{G}}(J)\cup d_{\mathcal{G}}(J)\}}$

[2] L'indice $\mathcal{G}$ est omis s'il n'y a pas de risque de confusion.

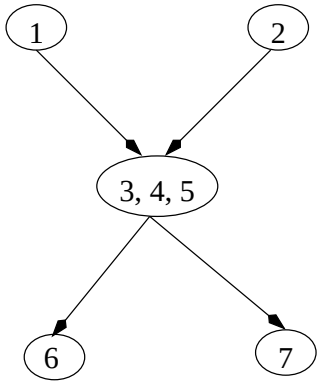


FIG. 1.2 – *Exemple d'un réseau bayésien de niveau deux.*

La loi jointe f_I associée au RB2 de la figure 1.2 s'écrit :

$$\begin{aligned} f_I(x_1,x_2,x_3,x_4,x_5,x_6,x_7) &= f_1(x_1)f_2(x_2)f_{3,4,5/1,2}(x_3,x_4,x_5/x_1,x_2)f_{6/3,4,5}(x_6/x_3,x_4,x_5). \\ & \quad f_{7/3,4,5}(x_7/x_3,x_4,x_5). \end{aligned}$$

Il peut arriver que, pour certains couples $(J^{'},J)$, où $J^{'}$ est un parent de J (i.e. $(J',J)\in\mathcal{G}$), $J^{'}$ n'étant pas un singleton dans I, on dispose de l'information selon laquelle $f_{J/p(J)}$ (la loi de J conditionnellement à ses parents $p(J)$) ne fait intervenir effectivement qu'un sous-ensemble strict, $\widetilde{J}'$, de J', parent de J ; en d'autres termes, si on note $\widetilde{p}(J)=\bigcup_{J'\in p(J)}\widetilde{J}'$, X_J et $X_{p(J)-\widetilde{p}(J)}$ sont indépendantes conditionnellement à $X_{\widetilde{p}(J)}$; on représente graphiquement cette information par un schéma du type donné en figure 1.3 :

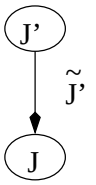


FIG. 1.3 – *Représentation d'information sur un réseau bayésien de niveau deux.*

On dira que le RB2 est renseigné.

Dans l'exemple de la figure 1.2, si on suppose que $f_{6/3,4,5}$ (respectivement $f_{7/3,4,5}$) ne fait intervenir que les indices $3,4$ (respectivement $4,5$) on pourra représenter le RB2 comme sur la figure 1.4.

A priori, il peut paraître plus lourd de stocker en mémoire d'ordinateur un RB2 (même renseigné) qu'un RB1 (car dans le RB1 on utilise plus finement des informations sur la structure de f_I). Mais il est des circonstances (et ceci apparaîtra naturellement pour les RB2 obtenus pour formaliser une restriction f_J) où on ne dispose pas d'informations

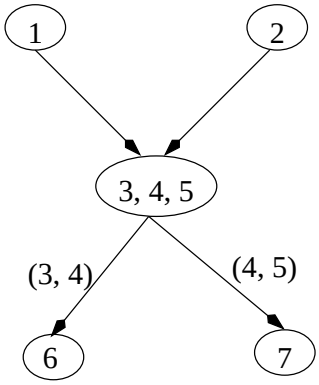


FIG. 1.4 – *Réseau bayésien de niveau deux renseigné.*

permettant de construire pour un RB2 donné une décomposition autre qu'arbitraire en RB1, qui est aussi lourde alors que le RB2 donné.

Par exemple le stockage du RB2 renseigné donné en exemple ci dessus (figure 1.4) est exactement de même poids que celui du RB1 figurant sur la figure 1.5 (qui graphiquement est moins éclairant).

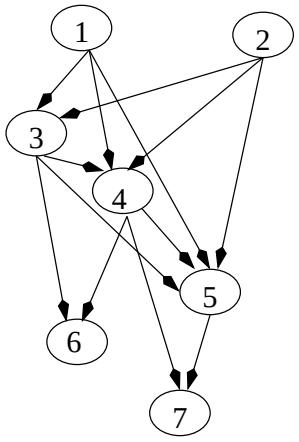


FIG. 1.5 – *RB*1 *équivalent au RB*2 *renseigné de la figure* 1.4.

1.2.3 Lien avec les réseaux bayésiens orientés objets

Les OOBN ("Object Oriented Bayesian Networks", que nous traduisons en français "Réseaux Bayésiens Orientés Objets") ont été introduits par D. Koller et A. Pfeffer [38] pour modéliser des systèmes complexes.

L'idée de départ de [38] est que, si un système (famille $(X_i)_{i\in I}$ de variables aléatoires)

est complexe, il y a souvent avantage à le découper en sous-familles, elles-même identifiées comme des sous-systèmes, de sorte que :

- les relations de causalité entre sous-systèmes soient modélisées par un réseau bayésien (dont les nœuds sont donc identifiés chacun à un sous-système) qu'on appellera ici "réseau structurant",

- chaque sous-système soit lui même doté, pour son fonctionnement interne, d'une structure de réseau bayésien (éventuellement réduit à une variable aléatoire unique) qu'on dira "fragmentaire".

Il en résulte évidement une structure de réseau bayésien sur le système tout entier (dit "global") : il est indexé par I et ses liens sont :

- entre nœuds appartenant à un même sous-système, les liens du réseau bayésien fragmentaire correspondant,

- entre nœuds appartenant à des sous-systèmes différents, les liens reliant les nœuds d'un réseau bayésien fragmentaire et ceux des réseaux bayésiens fragmentaires qui sont enfants de celui-ci dans le réseau bayésien structurant.

Une connaissance plus fine des relations entre sous-systèmes peut permettre d'éliminer certains de ces liens dans le réseau global.

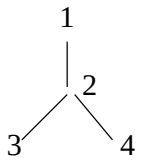
Voici des circonstances où de telles suppressions sont naturelles (mais ce n'est pas nécessairement le cas) : soit un couple (J, J') de sous-systèmes (J et J' sont des parties de I) tel que J' soit un enfant de J dans le réseau bayésien structurant :

- il peut se produire qu'il existe une partition (J_1, J_2) de J telle que $X_{J'}$ soit indépendante de X_{J_2} conditionnellement à X_{J_1} ; il n'y aura alors pas de lien reliant les nœuds J_2 à ceux de J' (ceci pourra typiquement être le cas avec J_1 ensemble des feuilles de J),

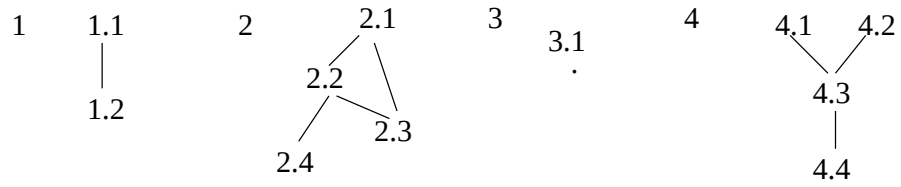
- il peut se produire qu'il existe une partition (J'_1, J'_2) de J' telle que $X_{J'_2}$ soit indépendante de X_J conditionnellement à $X_{J'_1}$; il n'y aura alors pas de lien reliant les nœuds J'_2 à ceux de J (ceci pourra typiquement être le cas avec J'_1 ensemble des racines de J').

Exemple

Réseau structurant



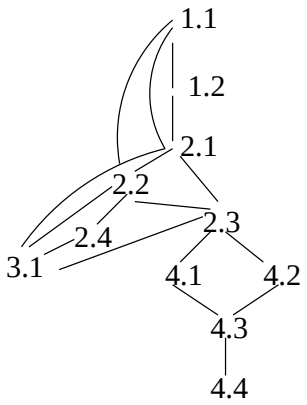
Réseaux fragmentaires



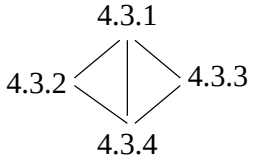
Renseignements complémentaires (on assimile les nœuds aux variables aléatoires qu'ils indexent)

(2.3) et (2.4) indépendants du fragment 1 conditionnellement à (2.1)
(4.3) et (4.4) indépendants du fragment 2 conditionnellement à (4.1) et (4.2)
(2.1), (2.2) et (2.4) indépendants du fragment 4 conditionnellement à (2.3)

Réseau global



Evidement, rien n'interdit de considérer qu'un sous-système soit lui-même structuré en sous-sous systèmes, reproduisant à leur tour, à leur échelle, la distinction entre "réseau structurant" et "réseaux fragmentaires". Dans l'exemple ci-dessus, supposons que à (4.3) soit attaché non une variable aléatoire mais le réseau suivant :



Supposons aussi qu'on dispose des renseignements complémentaires suivants :

(4.4) indépendant de (4.3.1), (4.3.2) et (4.3.3) conditionnellement à (4.3.4)
(4.3.2), (4.3.3) et (4.3.4) indépendants de (4.1) et (4.2) conditionnellement à (4.3.1).

Le réseau global serait alors

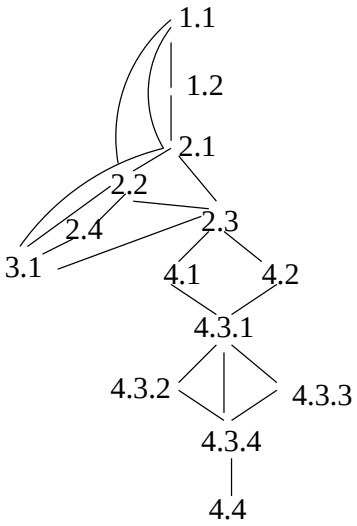


FIG. 1.6 – *Réseau global.*

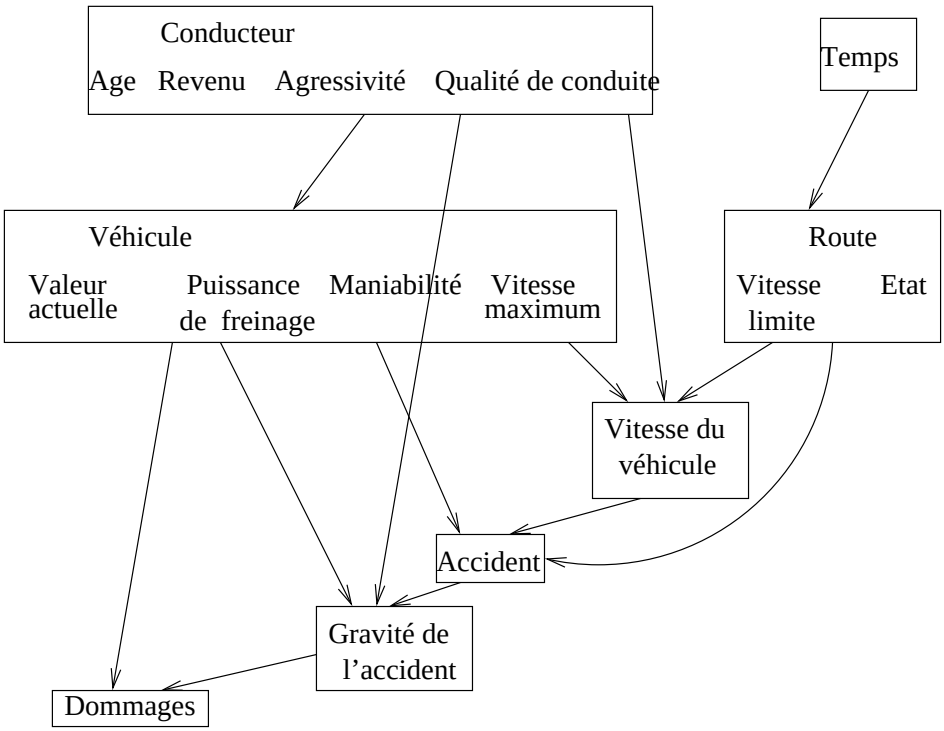
La bonne définition pour les OOBN est donc de nature récursive : un OOBN est un réseau bayésien aux nœuds duquel sont attachés soit des OOBN, soit des variables aléatoires, de sorte que toute suite $(R_1, R_2, \ldots)$ telle que, pour tout $j \geq 1$, R_j soit l'OOBN attaché à un nœud de R_{j-1}, soit finie, son dernier élément étant une variable aléatoire. Soit $(R_1, \ldots, R_k)$ une telle suite (qui représente donc des sous-systèmes emboîtés) ; R_k est réduit à une variable aléatoire et R_{k-1} est un réseau bayésien classique ; en appellera k la profondeur d'une telle suite ; la plus grande profondeur accessible dans l'OOBN est appelée son niveau (l'exemple de la figure 1.6 est celui d'un OOBN de niveau 3). Un RB (réseau bayésien "classique") est un OOBN de niveau 1 (on a vu qu'on disait aussi RB1)

Nos RB2 (Réseaux Bayésiens de niveau deux) sont des cas particuliers d'OOBN de niveau 2 pour lesquels les réseaux fragmentaires n'ont pas été munis de structure particulière ; autrement dit ce sont des réseaux bayésiens complets : on peut munir chacune des parties de la partition définissant le RB2 d'un ordre quelconque de ses éléments et choisir pour arcs les couples (i, i') tels que i soit classé avant i'.

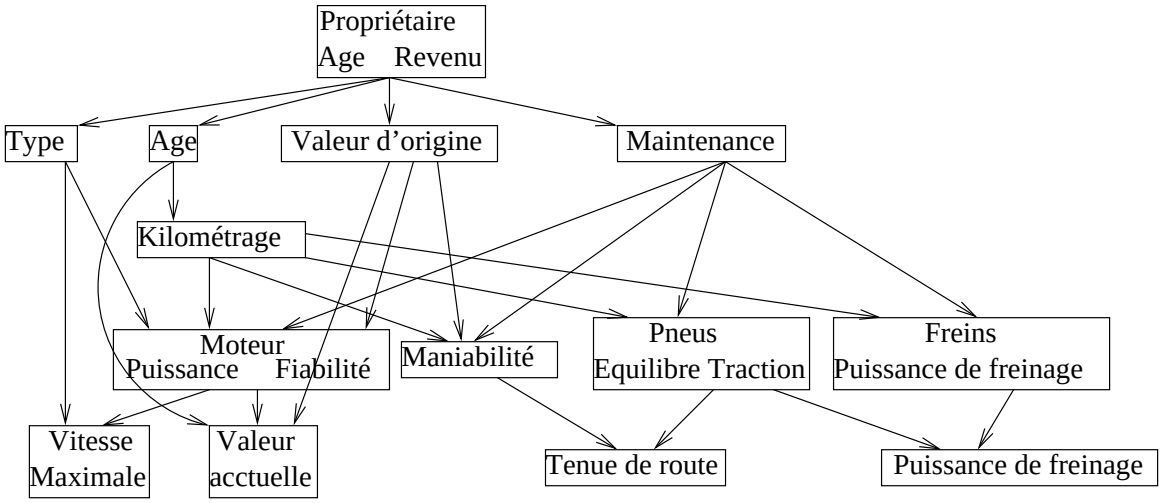
Mais notre objectif dans l'introduction des RB2 diffère de celui de [38] dans l'introduction des OOBN : ces auteurs visent à fournir pour des RB un schéma de construction souple, permettant d'utiliser des RB comme "briques" pour en élaborer d'autres ; ensuite pour les calculs d'inférence, il se réfèrent aux techniques d'arbres de jonction, et subissent donc les contraintes inhérentes à ces techniques ; pour nous les RB2 ne sont pas déterminés comme des éléments d'une modélisation mais sont des outils élaborés en fonction d'un objectif de calcul dans un réseau bayésien donné.

En conclusion de ce paragraphe sur les OOBN, donnons un exemple d'utilisation qui figure dans [38]. Il est relatif à l'analyse d'un accident de voiture.

Réseau bayésien structurant



Réseau bayésien fragmentaire (Véhicule)



Chapitre 2

Propriétés de Séparation dans les Réseaux Bayésiens

Dans les gros réseaux bayésiens, le calcul de lois ou de lois conditionnelles peut faire intervenir des sommations relatives à de très gros sous-ensembles de l'ensemble I des indices. Il y a donc intérêt à s'efforcer, au préalable, de segmenter, si on le peut, ces calculs en plusieurs calculs moins lourds et pouvant être menés en parallèle. Ces segmentations sont liées à des propriétés du graphe définissant le réseau.

Etant donné $S \subset I$, non vide, ce chapitre explicite ces segmentations pour le calcul de P_S et $P_{\bar{S}/S}$, où $\bar{S} = I - S$.

En première partie, on présente certains résultats ne faisant intervenir que la structure de graphe non orienté ; on les précise ensuite dans le cas des réseaux bayésiens, la seconde partie étant consacrée essentiellement aux propriétés graphiques et la troisième aux calculs de lois.

La plupart des résultats de ce chapitre sont classiques, mais nous en donnons des démonstrations très explicites. La présentation que nous donnons de la segmentation de calcul de la loi P_S est à notre connaissance originale, grâce à l'usage de RB2.

2.1 Séparation dans un graphe non-orienté (I, H)

2.1.1 Définition et théorème fondamental

Définition 5 *Etant donné un graphe non orienté (I, H) et 3 parties disjointes A, B et S de I, on dit que S **sépare** A et B (où que S est un **séparateur** de A et B) si toute chaîne joignant un sommet a dans A à un sommet b dans B a au moins un sommet intermédiaire dans S*

Donc A et B sont non séparés par S s'il existe au moins une chaîne reliant un sommet de A à un sommet de B sans contenir aucun sommet dans S (on dit qu'elle contourne S).

En particulier, la séparation de A et B par $\emptyset$ exprime qu'il n'existe pas de chaîne reliant un sommet de A à un sommet de B, autrement dit que A et B sont dans deux composantes connexes distinctes.

Définition 6 *Soit $\mathcal{C}$ un recouvrement de I. Nous dirons que $\mathcal{C}$ et la loi P_I de la famille $(X_i)_{i\in I}$ sont **compatibles** si P_I peut s'écrire sous la forme*

$$\forall\ x_I \in \Omega_I \quad P_I(x_I) \ = \ \frac{1}{\alpha} \prod_{C\in\mathcal{C}} f_C(x_C) \tag{2.1}$$

où, pour tout $C \in \mathcal{C}$, f_C est une application de Ω_C dans $\mathbb{R}^+$ connue.

Définition 7 *Soit (I, H) un graphe non orienté ; on dit que la loi P_I est compatible avec la structure H si elle est compatible avec le recouvrement de I par l'ensemble des cliques associées au graphe (I, H).*

Remarque : On rappelle qu'une partie J de I est dite **complète** (pour H) si toute paire d'éléments distincts de J est une arête de H ; les **cliques** sont les parties complètes maximales (pour l'inclusion). Donc pour que P_I soit compatible avec une stucture de graphe non orienté H il suffit qu'elle soit compatible avec un recouvrement par des parties complètes ; en effet toute application définie sur Ω_C, où C est complète, peut être considérée comme définie sur $\Omega_{C'}$, où C' est une clique contenant C.

Théorème 2 *Soit P compatible avec la structure de graphe non orienté H, et soient A, B et S trois parties disjointes de I, telles que S sépare A et B. Alors X_A et X_B sont indépendantes conditionnellement à X_S (en particulier, si S est vide, X_A et X_B sont indépendantes)*

Démonstration

Pour toute partie J de I, on note $\mathcal{C}_J$ l'ensemble des cliques de I contenues dans J.

Regardons tout d'abord le cas particulier où $S = \emptyset$. Soit A' (resp. B') la fermeture connexe de A (resp. B) ; alors toute clique de I est contenue soit dans A', soit dans B', soit dans $I-(A'\cup B')$ (si $A'\cup B' \neq I$). On peut donc écrire (en convenant comme d'habitude qu'un produit portant sur un ensemble vide vaut 1)

$$P(x_I) = \frac{1}{\alpha} \prod_{C\in\mathcal{C}_{A'}} f_C(x_C) \prod_{C\in\mathcal{C}_{B'}} f_C(x_C) \prod_{C\in\mathcal{C}_{(I-(A'\cup B'))}} f_C(x_C)$$

et alors

$$\begin{aligned} P_{A'\cup B'}(x_{A'\cup B'}) \ &= \ \sum_{I-(A'\cup B')} P(x'_A, x'_B, x_{I-(A'\cup B')}) \\ &= \ \frac{k}{\alpha} \prod_{C\in\mathcal{C}_{A'}} f_C(x_C) \prod_{C\in\mathcal{C}_{B'}} f_C(x_C), \end{aligned}$$

où $k = \sum_{I-(A'\cup B')} \prod_{C\in(I-(A'\cup B'))} f_C(x_C)$.

La loi de $X_{A' \cup B'}$ est donc de la forme $g(x_{A'})h(x_{B'})$, ce qui exprime (voir Annexe B) que $X_{A'}$ et$X_{B'}$ sont indépendants, et donc, puisque $A \subseteq A'$ et $B \subseteq B'$, X_A et X_B sont indépendants.

Soit maintenant S différent de l'ensemble vide. Notons A' (resp. B') l'ensemble des sommets qui soit appartiennent à A (resp. B), soit sont reliés à un sommet de B (resp. A) par une chaîne qui contourne S.

Il résulte immédiatement de la séparation de A et B par S que A' et B' sont disjointes, séparées par S ; en effet, s'il existait une chaîne reliant $x' \in A'$ à $y' \in B'$ en contournant S, en pourrait l'intégrer à une chaîne reliant, en contournant S, $x \in A$ (avec $x = x'$ si $x' \in A$) à $y \in B$ (avec $y = y'$ si $y' \in B$).

Notons D le complémentaire de $A' \cup B' \cup S$ dans I ; par définition de A' et B', si D est non vide, il est composé de sommets tels que toute chaîne les reliant à $A \cup B$ traverse S. S sépare alors A' et D (et de même B' et D) ; en effet, s'il existait une chaîne reliant z ($\in D$) à un sommet de A' en contournant S, en pourrait, par définition de A', la prolonger en une chaîne reliant z à un sommet de A en contournant S, ce qui est contraire à la définition de D.

Ces propriétés de séparation de A', B' et D par S impliquent que les cliques sont des sous-ensembles de $A' \cup S$, $B' \cup S$ ou $D \cup S$ (si D non vide).

Il résulte alors de la compatibilité de P avec H que

$$P(x_I) = \frac{1}{\alpha} \prod_{C \in \mathcal{C}_{A' \cup S}} f_C(x_C) \prod_{C \in \mathcal{C}_{B' \cup S} - \mathcal{C}_S} f_C(x_C) \prod_{C \in \mathcal{C}_{D \cup S} - \mathcal{C}_S} f_C(x_C)$$

(les cliques contenues dans S, s'il en existe, sont prises en compte dans $\mathcal{C}_{A' \cup S}$, et donc pas dans $\mathcal{C}_{B' \cup S}$ ni $\mathcal{C}_{D \cup S}$)

Si D est vide, autrement dit $I = A' \cup B' \cup S$, cette formule se simplifie en

$$P(x_I) = P_{A' \cup B' \cup S}(x_{A' \cup B' \cup S}) = \frac{1}{\alpha} \prod_{C \in \mathcal{C}_{A' \cup S}} f_C(x_C) \prod_{C \in \mathcal{C}_{B' \cup S}} f_C(x_C)$$

Si D est non vide, la loi de $X_{A' \cup B' \cup S}$ s'obtient par sommation relativement à X_D, qui porte uniquement sur $\prod_{C \in \mathcal{C}_{D \cup S} - \mathcal{C}_S} f_C(x_C)$, d'où

$$P_{A' \cup B' \cup S}(x_{A' \cup B' \cup S}) = \frac{1}{\alpha} k(x_S) \prod_{C \in \mathcal{C}_{A' \cup S}} f_C(x_C) \prod_{C \in \mathcal{C}_{B' \cup S}} f_C(x_C),$$

où $k(x_S) = \prod_{C \in \mathcal{C}_{D \cup S} - \mathcal{C}_S} f_C(x_C)$.

Dans tous les cas, la loi de $X_{A' \cup B' \cup S}$ est donc de la forme $g(x_{A' \cup S})h(x_{B' \cup S})$, ce qui exprime (voir Annexe B) que $X_{A'}$ et$X_{B'}$ sont indépendants conditionnellement à X_S ; il en est a fortiori de même pour X_A et X_B.

2.1.2 Partition S-conditionnelle pour un graphe non orienté

Le théorème 2 justifie l'introduction de la notion suivante (où $\bar{S}$ désigne le complémentaire de S, $I - S$)

Définition 8 *Etant donné une structure de graphe non orienté sur I, et une partie S de I, distincte de I, on appele partition S-conditionnelle la partition de $\bar{S}$ en les atomes pour la relation d'équivalence, notée $\sim_{\bar{S}}$, définie sur $\bar{S}$ par :*
$i \sim_{\bar{S}} i'$ si et seulement s'il existe une chaîne reliant i à i' sans sommet intermédiaire dans S.

On note $\mathcal{P}_{\bar{S}}$ cette partition.

On remarque que sont en particulier équivalents deux éléments de $\bar{S}$ qui sont reliés par une arête.

La terminologie "partition S-conditionnelle" se justifie par le fait que, si la loi P_I est compatible avec la structure H et si C et C' sont deux atomes distincts (donc disjoints) de $\mathcal{P}_{\bar{S}}$, les variables aléaoires X_C et $X_{C'}$ sont indépendantes conditionnellement à X_S. En revanche la compatibilité de P_I avec la structure H n'entraîne pas l'indépendance des variables aléatoires appartenant à un même atome de la partition $\mathcal{P}_{\bar{S}}$, ce qui ne veut pas dire qu'elles soient nécessairement dépendantes : elle peuvent se trouver être indépendantes, mais alors ceci constitue une propriété de P_I non indiquée par la compatibilité avec H (voir [17])

Un corollaire du théorème 2 est alors :

Corollaire 1 *Si P_I est compatible avec la structure de graphe non orienté H sur I, on a*

$$P_{\bar{S}/S} = \prod_{C \in \mathcal{P}_{\bar{S}}} P_{C/S}$$

Dans le cas particulier où $S = \emptyset$, et donc $\bar{S} = I$, $\mathcal{P}_I$ est la partition de I en composantes connexes pour H (souvent réduite à $\{I\}$ dans la pratique) et $P_I = \prod_{C \in \mathcal{P}_C} P_C$.

Une conséquence pratique de ce corollaire est que le calcul de $P_{\bar{S}/S}$ peut se segmenter en les calculs des $P_{C/S}$, où $C \in \mathcal{P}_{\bar{S}}$; nous allons préciser ces calculs à l'aide de la notion de frontière relative à H.

Définition 9 *Soit (I, H) un graphe non orienté et soit A une partie non vide de I ; on appelle frontière de A (pour H) et on note $F(A)$ l'ensemble des éléments de $\bar{A}$ qui sont reliés à au moins un élement de A par une arête (autrement dit $F(A)$ est l'ensemble des voisins de A relativement à H).*

Une conséquence immédiate de cette définition est la proposition suivante.

Proposition 4 *A et $I - (A \cup F(A))$ sont séparés par $F(A)$.*

Il en résulte le corollaire (qui énonce la "propriété de Markov locale")

Corollaire 2 *Soit P_I compatible avec la structure de graphe non orienté H sur I, et soit $D \subseteq I - (A \cup F(A))$; alors X_A et X_D sont indépendants conditionnellement à $F(A)$ (autrement noté $P_{A\cup D/F(A)} = P_{A/F(A)}P_{D/F(A)}$) ; donc, pour toute partie E de I, disjointe de A et contenant $F(A)$, $P_{A/E} = P_{A/F(A)}$*

Etant donné $S \subset I$, les frontières des atomes de la partition $\mathcal{P}_{\bar{S}}$ sont particulièrement utiles à considérer, en vertu de la proposition suivante.

Proposition 5 *Soit S une partie de I et C un atome de la partition $\mathcal{P}_{\bar{S}}$; alors la frontière $F(C)$ est contenue dans S.*

Démonstration : Soit i qui n'est ni dans S ni dans C ; on va démontrer qu'il n'appartient pas à $F(C)$. En effet i est dans un atome de la partition $\mathcal{P}_{\bar{S}}$ autre que C ; donc S sépare i de C, c'est-à-dire qu'il ne peut pas exister de chaîne reliant i à un élément de C alors que, par définition de la frontière, pour tout élément de $F(C)$, il existe une arête entre lui et un élément de C.

■

Il résulte alors de cette proposition et du corollaire 2 ci dessus le résultat suivant.

Corollaire 3 *Soit P_I compatible avec la structure de graphe non orienté H sur I, et soit $S \subset I$; alors, pour tout atome C de $\mathcal{P}_{\bar{S}}$, on a*

$$P_{C/S} = P_{C/F(C)}$$

Ce corollaire, joint au corollaire 2, fournit le :

Corollaire 4 *Si P_I est compatible avec la structure de graphe non orienté H sur I, on a*

$$P_{\bar{S}/S} = \prod_{C\in\mathcal{P}_{\bar{S}}} P_{C/F(C)}$$

2.2 D-séparation dans un graphe orienté sans circuits

2.2.1 Graphe moral

Définition 10 *Etant donné un graphe orienté sans circuits (I, G), son graphe moral associé est le graphe non orienté (I, G^m) ou G^m 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 en commun (il existe k tel que $(i, k) \in G$ et $(j, k) \in G$)

Un rôle important sera joué dans la suite par la partie commençante associée à une partie J de I, autrement dit l'ensemble des ancêtres de tous les éléments de J, noté J^+. Il est élémentaire que J^+ est aussi un réseau bayésien car, du fait que, pour tout $i \in J^+$, $p(i)$

est contenu dans J^+, on a

$$P_{J^+} = \sum_{\overline{J^+}} \prod_{i \in I} P_{i/p(i)} = [\prod_{i \in J^+} P_{i/p(i)}] \sum_{\overline{J^+}} (\prod_{i \in \overline{J^+}} P_{i/p(i)}) \ ;$$

or $\sum_{\overline{J^+}} \left(\prod_{i \in \overline{J^+}} P_{i/p(i)} \right) = 1$ (il suffit pour s'en assurer d'ordonner les éléments de $\overline{J^+}$ selon un ordre hiérarchique et d'effectuer les sommations selon les éléments de $\overline{J^+}$ en remontant cet ordre).

Remarquons que la restriction du graphe moral G^m à J^+ n'est pas le graphe moral associé à la restriction de G à J^+, qu'on notera $G^m_{J^+}$. Considérons, pour nous en assurer, l'exemple donné par la figure suivante, où on prend $J = \{5\}$ (sommet entouré).

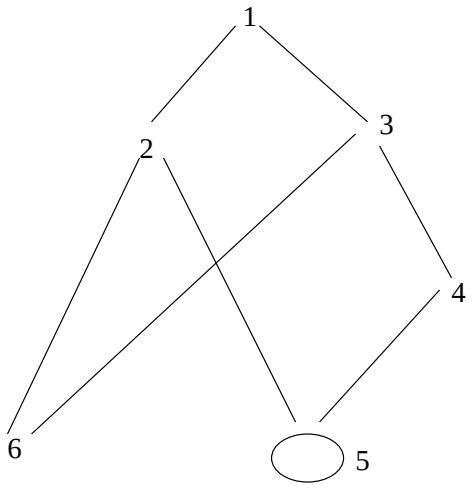


FIG. 2.1 –

Dans cet exemple, $J^+ = \{1, 2, 3, 4, 5\}$; $\{2, 3\}$ est une arête dans la restriction de G^m à J^+ (car 2 et 3 sont tous deux parents de 6) mais non dans $G^m_{J^+}$.

L'intérêt du graphe moral est que, pour tout i, la partie de I constituée de i et de ses parents relativement à G est complète relativement à G^m et donc est contenue dans une clique.

Il en résulte que, si la famille des variables aléatoires $(X_i)_{i \in I}$ est un réseau bayésien pour G, elle est compatible avec la structure de graphe non orienté G^m. Donc, si trois parties non vides de I, A, B, et S, sont telles que S sépare A et B dans le graphe moral, il en résulte que les variables aléatoires X_A et X_B sont indépendantes conditionnellement à X_S.

Plus précisément, considérons le réseau bayésien restreint à $(A \cup B \cup S)^+$; alors, si S sépare A et B dans le graphe moral $G^m_{(A \cup B \cup S)^+}$, X_A et X_B sont indépendantes conditionnellement à X_S.

Il est intéressant de caractériser la séparation directement sur le graphe orienté sans circuits G. C'est l'objet de la d-séparation ("d" pour "directed", qui signifie orienté en

anglais), introduite en 2.2.3 ci-dessous, après quelques compléments de terminologie sur les chaînes dans les graphes orientés.

2.2.2 Chaîne dans un graphe orienté.

Etant donné un graphe orienté (I, G), soit une chaîne $(x_0, \ldots, x_n)$ reliant les sommets $a(= x_0)$ et $b(= x_n)$; ceci signifie que $(x_0, \ldots, x_n)$ est une chaîne pour la structure de graphe non orienté, soit H, sous jacente à G autrement dit que, pour tout i $(1 \leq i \leq n)$, on a $\{x_{i-1}, x_i\} \in H$ c'est à dire soit $(x_{i-1}, x_i) \in G$, soit $(x_i, x_{i-1}) \in G$.
Si $n \geq 2$, les sommets $x_1, \ldots, x_{n-1}$ sont dits "intermédiaires" sur cette chaîne .
En un sommet intermédiaire x_i la chaîne est dite :
- convergente si $(x_{i-1}, x_i) \in G$ et $(x_{i+1}, x_i) \in G$

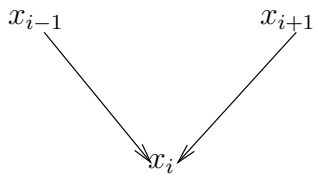


FIG. 2.2 – *Chaîne convergente en x_i.*

- divergente si $(x_i, x_{i-1}) \in G$ et $(x_i, x_{i+1}) \in G$

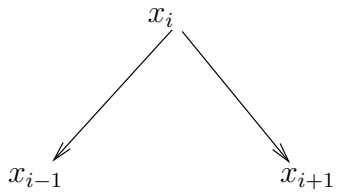


FIG. 2.3 – *Chaîne divergente en x_i.*

- en série si $(x_{i-1}, x_i) \in G$ et $(x_i, x_{i+1}) \in G$ ou si $(x_{i+1}, x_i) \in G$ et $(x_i, x_{i-1}) \in G$

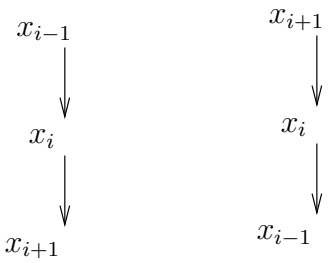


FIG. 2.4 – *Chaîne en série en x_i.*

Il est clair qu'une chaîne ne peut être convergente en deux sommets consécutifs, ni divergente en deux sommets consécutifs.

2.2.3 Blocage, d-séparation

Définition 11 *Soit donné un graphe orienté sans circuits (I, G) et une partie non vide S de I. Une chaîne reliant deux sommets de I non dans S est dite bloquée par S si elle est de longueur au moins 2 et l'un au moins des sommets intermédiaires, soit x, vérifie l'une des deux propriétés suivantes :*
- $x \in S$ et la chaîne est, en x, soit en série soit divergente
- $x \notin S^+$ et la chaîne est, en x, convergente.

Une chaîne reliant deux sommets non dans S est donc non-bloquée par S si elle est constituée d'un arc unique joignant ces deux sommets ou bien si elle est convergente en tout sommet intermédiaire appartenant à S et non convergente (i.e. en série ou divergente) en tout sommet intermédiaire qui n'a aucun descendant dans S (i.e. qui n'appartient pas à S^+) ; en remarque que le non-blocage par S n'introduit aucune condition sur le comportement de la chaîne en les sommets qui sont ancêtres de sommets de S mais non dans S lui-même.

Exemple

Considérons le graphe orienté sans circuits suivant [47]

1. La chaîne (y, x, z, s) est bloquée par $\{x\}$, puisque elle est divergente en x ; elle est aussi bloquée par $\{z\}$ puisque elle est en série en z.
2. La chaîne (w, y, r, s) est bloquée par $\{r\}$, puisque elle est en série en r.
3. La chaîne (w, y, r, z, s) est non-bloquée par $\{r\}$, puisque elle est convergente en r.

Définition 12 *Soit (I, G) un graphe orienté sans circuits et soient 3 parties non vides A, B et S de I. A et B sont dites d-séparées par S si toute chaîne reliant un sommet de A à un sommet de B est bloquée par S.*

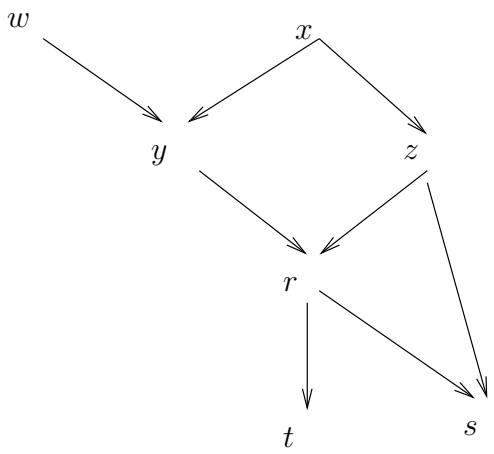


FIG. 2.5 – *Exemple de différents types de chaînes.*

Remarque :

Si A et B sont d-séparées par S il existe pas d'arc, dans G reliant un sommet de A à un sommet de B, puisque toute chaîne bloquée est de longueur au moins 2.

Exemple

Considérons le graphe orienté sans circuits précédent (figure 2.5)

1. x et r sont d-séparés par $\{y, z\}$ puisque la chaîne (x, y, r) est bloquée en y, et la chaîne (x, z, r) est bloquée en z.
2. x et t sont d-séparés par $\{y, z\}$ puisque la chaîne (x, y, r, t) est bloquée en y, et la chaîne (x, z, r, t) est bloquée en z, et la chaîne (x, z, s, r, t) est bloquée en z et en s.
3. w et t sont d-séparés par $\{r\}$ puisque les deux chaînes (w, y, r, t) et (w, y, x, z, r, t) sont bloquées en r.
4. y et z sont d-séparés par $\{x\}$ puisque la chaîne (y, x, z) est bloquée en x, la chaîne (y, r, z) est bloquée en r, et la chaîne (y, r, s, z) est bloquée en s.
5. w et s sont d-séparés par $\{r, z\}$ puisque la chaîne (w, y, r, s) et bloquée en r, les deux chaînes (w, y, r, z, s) et (w, y, x, z, s) sont bloquées en z.
6. w et s sont aussi d-séparés par $\{y, z\}$ puisque la chaîne (w, y, r, s) et bloquée en y, la chaîne (w, y, r, z, s) est bloquée en y, r et z, et la chaîne (w, y, x, z, s) est bloquée en z.
7. w et x ne sont pas d-séparés par $\{y\}$ puisque la chaîne (w, y, x) est non bloquée en y.
8. w et t ne sont pas d-séparés par $\{y\}$ puisque, malgré que la chaîne (w, y, r, t) est bloquée en y, la chaîne (w, y, x, z, r, t) est non bloquée en y ; elle est non bloquée

ailleurs parce qu'il y a pas d'autres nœuds dans $\{y\}$ et il n'y a aucune chaîne convergente en y.

L'usage de la d-séparation tient au théorème suivant :

Théorème 3 *Soit (I, G) un graphe orienté sans circuits, et A, B et S trois parties non vides de I. A et B sont d-séparées par S dans G si et seulement si elles sont séparées dans le graphe moral associé à $G^m_{(A\cup B\cup S)^+}$, restriction de G à $(A \cup B \cup S)^+$.*

Démonstration

1. Supposons les sous-ensembles A et B non d-séparés par S dans G. Nous allons établir qu'ils ne sont pas séparés par S dans $G^m_{(A\cup B\cup S)^+}$.
   Soit donc une chaîne dans G, reliant les sommets $a \in A$ et $b \in B$, qui n'est pas bloquée par S.
   On remarque tout d'abord que cette chaîne est entièrement contenue dans $(A \cup B \cup S)^+$. C'est évidemment le cas si elle est réduite à un arc joignant a à b. Sinon, admettons qu'il existe un sommet intermédiaire hors de $(A\cup B\cup S)^+$; la chaîne étant notée $(x_0, \ldots, x_n)$ (avec $x_0 = a$ et $x_n = b$) soit x_i $(1 \leq i \leq n)$ le premier tel sommet, vérifiant $x_i \notin (A\cup B\cup S)^+$; alors $(x_{i-1}, x_i) \in G$ car, s'il était vrai que $(x_i, x_{i-1}) \in G$, x_i appartiendrait à $(A \cup B \cup S)^+$ comme x_{i-1} ; ensuite on aurait $(x_i, x_{i+1}) \in G$, sans quoi la chaîne serait convergente en x_i, ce qui est interdit car $x_i \notin S^+$ et la chaîne est supposée non bloquée par S ; il en résulte que $x_{i+1} \notin (A \cup B \cup S)^+$, car si x_{i+1} appartenait à $(A\cup B\cup S)^+$ il en serait de même de x_i, parent de x_{i+1} ; on établit ainsi par récurrence que dès qu'un élément de la chaîne n'appartient pas à $(A \cup B \cup S)^+$, il en est de même pour tous les suivants, ce qui est absurde pour $x_n = b$.
   On va maintenant associer à la chaîne considérée dans G, non bloquée dans S, une chaîne, pour le graphe moral réduit à $(A \cup B \cup S)^+$, contournant S.
   Pour tout sommet intermédiaire x_i qui est dans S, il y a nécessairement convergence en x_i, mais ni en x_{i-1} ni en x_{i+1} car il ne peut jamais y avoir deux sommets consécutifs convergents ; ceci interdit donc à x_{i-1} et x_i d'appartenir à S ; de plus la paire $\{x_{i-1}, x_{i+1}\}$ appartient au graphe moral par définition de celui-ci. En substituant, ainsi, dans la chaîne donnée, pour tout $x_i \in S$, la paire $\{x_{i-1}, x_{i+1}\}$ aux deux paires initiales $\{x_{i-1}, x_i\}$ et $\{x_i, x_{i+1}\}$, on construit une chaîne, pour $G^m_{(A\cup B\cup S)^+}$, contournant S.
2. Réciproquement, supposons les sous-ensembles A et B non séparés par S dans $G^m_{(A\cup B\cup S)^+}$. Nous allons établir qu'ils ne sont pas d-séparés par S dans G.
   Soit donc une chaîne pour $G^m_{(A\cup B\cup S)^+}$, reliant $a \in A$ et $b \in B$ en contournant S. On va lui associer une chaîne pour G, non bloquée par S, reliant un sommet de A (non nécessairement a lui même) à un sommet de B (non nécessairement b lui même). Cette construction est mise en évidence sur la figure 2.6, où la chaîne donnée est (a, c, d, e, f, g, b).
   Une première modification de la chaîne donnée, $(x_0, \ldots, x_n)$, consiste à la remplacer, s'il y a lieu, par une chaîne pour G (et non dans le graphe moral) ; en effet toute arête

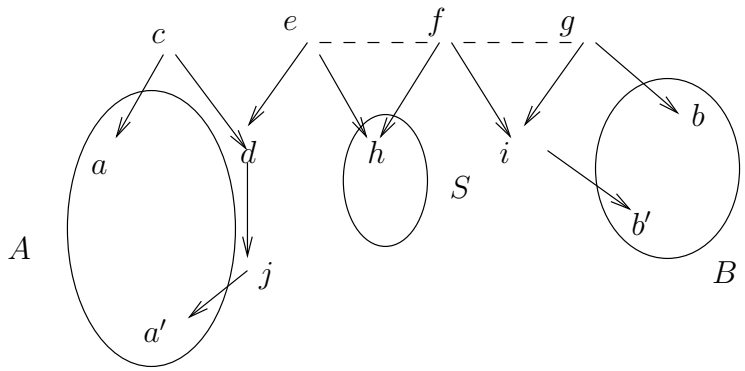


FIG. 2.6 – *schéma de construction d'une chaîne non bloquée.*

qui n'est pas dans G relie, par définition du graphe moral, deux sommets x_{i-1} et x_i qui, sont, pour G, parents d'un même sommet y ; on remplace donc l'arête $\{x_{i-1}, x_i\}$ par la succession des deux arêtes $\{x_{i-1}, y\}$ et $\{y, x_i\}$; dans cette nouvelle chaîne les seuls sommets éventuellement dans S sont les sommets nouveaux ainsi introduits et la condition de convergence en tout sommet dans S (qui fait partie de la définition d'une chaîne non bloquée par S) est donc assurée ; la chaîne ainsi construite reste dans $(A \cup B \cup S)^+$, comme la chaîne initiale ; notons la $(y_0, \ldots, y_m)$, où $y_0 = a$ et $y_m = b$; dans la figure 2.6, c'est $(a, c, d, e, h, f, i, g, b)$.

Il reste à assurer l'autre condition de non blocage par S, à savoir la non-convergence en tout sommet n'appartenant pas à S^+ ; soit donc (cas de d et i dans le figure 2.6) un sommet intermédiaire y_ℓ sur la chaîne déja construite, en lequel celle-ci est convergente et non dans S^+ ; y_ℓ appartient à $(A \cup B \cup S)^+$, donc ici à $(A \cup B)^+$, ce qui implique qu'il existe une chaîne $(z_0, \ldots, z_k)$ dans G, descendante (i.e. $(z_{j-1}, z_j) \in G$ pour tout j tel que $1 \leq j \leq k$) reliant $y_\ell (= z_0)$ à un sommet z_k qui est dans $A \cup B$, par exemple a' dans A (voir (d, j, a') dans la figure 2.6) ; aucun des sommets de cette chaîne n'est dans S^+ (sans quoi y_ℓ y serait lui-même) ; on remplace alors la portion de chaîne $(y_\ell, \ldots, y_i)$ par $(z_k, z_{k+1}, \ldots, z_0)$ et, conservant la portion de chaîne $(y_1, \ldots, y_m)$, on obtient une chaîne reliant $a' (= z_k)$ à $b (= y_m)$ et comportant au moins un sommet de moins de type interdit, c'est à dire non dans S^+ et en lequel la chaîne est convergente (éventuellement d'autres sommets de type interdit peuvent avoir disparu, qui figuraient dans la portion de chaîne $(y_\ell, \ldots, y_i)$) ; si c'était à un point de B, par exemple b', qu'avait abouti la chaîne descendante construite à partir de y_i, la construction d'une nouvelle chaîne reliant a à b' aurait été analogue (voir (i, b') dans la figure 2.6).

Par récurrence, on arrive ainsi à construire une chaîne non bloquée par S reliant A à B.

■

Le théorème suivant est une conséquence immédiate des théorèmes 2 (appliqué à $G^m_{(A \cup B \cup S)^+}$) et 3 :

Théorème 4 *Soit (I, G, P_I) un réseau bayésien et soit trois parties non vides A, B et S telles que A et B soient d-séparées par S. Alors les variables aléatoires X_A et X_B sont indépendantes conditionnellement à X_S.*

Terminologie : Les praticiens des réseaux bayésiens expriment la circonstance décrite par le théorème ci-dessus en disant que la connaissance de la valeur x_S prise par la variable aléatoire X_S (ou, ce qui est équivalent, la connaissance des valeurs x_i prises par toutes les variables aléatoires X_i telles que $i \in S$) "bloque" la transmission d'information entre les variables aléatoires X_A et X_B.

2.2.4 Couverture et frontière de Markov

Les notions et résultats introduits en 2.1.2 s'étendent aux réseaux bayésiens en prenant pour graphe non orienté le graphe moral ou le graphe moral associé à une restriction.

La frontière $F(A)$ d'une partie A non vide de I, aussi appelée "**frontière de Markov**", est l'ensemble des sommets voisins d'un élément de A dans le graphe moral ; autrement dit un élément i de $\bar{A}$ appartient à $F(A)$ s'il vérifie l'une au moins des trois propriétés suivantes :

(*i*) i a un enfant dans A (i.e. $\exists\ a \in A\ (i, a) \in G$)

(*ii*) i a un parent dans A (i.e. $\exists\ a \in A\ (a, i) \in G$)

(*iii*) i a un enfant en commun avec un élément de A (i.e. $\exists\ a \in A$, $\exists\ j \in I\ (a, j) \in G$ et $(i, j) \in G$).

On introduit les notions et notations suivantes, toutes associées à une partie C de I, non vide :

couverture de Markov de $C = M(C) = C \cup F(C)$ (où $F(C)$ est la frontière de Markov)

ensemble des racines extérieures de C : $R(C)$ est l'ensemble des éléments de $F(C)$ qui n'ont pas de parents dans C (soit qu'il n'aient pas de parents, et donc sont des racines de I, soit que leurs parents soient hors de C)

ensemble des voisins non racines de C : $T(C) = F(C) - R(C)$: les éléments de $T(C)$ appartiennent à la frontière de Markov de C et ont au moins un parent dans C ; autrement dit ce sont les enfants, non dans C, d'éléments de C.

2.2.5 Partition S-conditionnelle pour un réseau bayésien

Alors que la relation $\sim_{\bar{S}}$ sur $\bar{S}$ (définition 8, en 2.1.2) est une relation d'équivalence, il n'en est pas de même de la relation $\approx_{\bar{S}}$ définie sur $\bar{S}$ par :
$x \approx_{\bar{S}} y$ si et seulement s'il existe une chaîne reliant x à y et non bloquée par S.

En effet, la relation $\approx_{\bar{S}}$ n'est pas transitive : soit z ($\in \bar{S}$) tel que $x \approx_{\bar{S}} z$ et $z \approx_{\bar{S}} y$; la concaténée des chaînes, non bloquées par S, reliant x à z et z à y n'est pas nécessairement non bloquée par S car il peut se produire que $z \notin S^+$ et que la chaîne concaténée soit

convergente en z, d'où le blocage.

En revanche, si en se limite à S^+, on dispose sur $S^+ - S$ de la relation d'équivalence $\sim_{S^+-S}$, où $x \sim_{S^+-S} y$ peut être défini de manière équivalente par "il existe une chaîne de x à y, dans $S^+ - S$, non bloquée par S" ou "il existe une chaîne, pour le graphe moral $G^m_{S^+-S}$, reliant x à y sans sommet intermédiaire dans S". Cette relation d'équivalence définit la partition $\mathcal{P}_{S^+-S}$, appelée **partition S-conditionnelle** de $S^+ - S$.

Le lecteur est invité à utiliser, pour la compréhension des notions, résultats et démonstrations de la section suivante 2.3, l'exemple ci-dessous, où $S^+ = I$ et donc $S^+ - S = \bar{S}$ (les sommets appartenant à S sont entourés et les atomes de la partition $\mathcal{P}_{\bar{S}}$ sont délimités par des traits en tiretés).

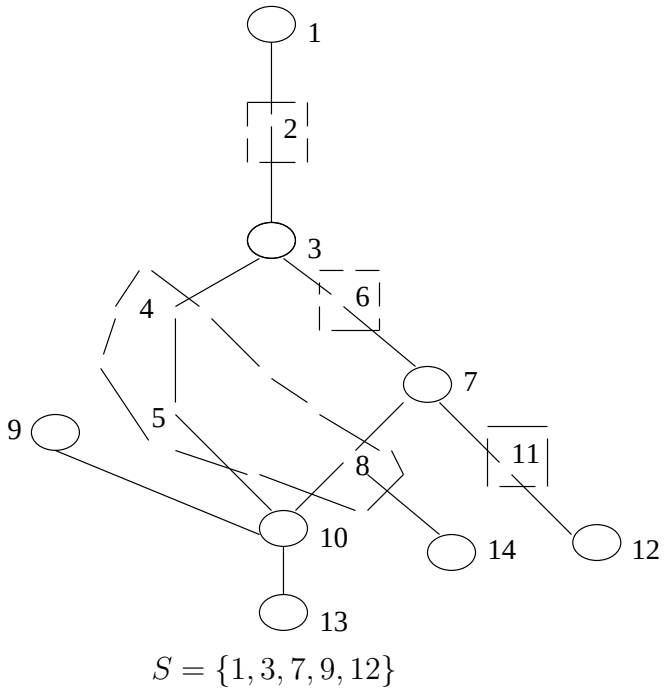


FIG. 2.7 – *Exemple de partition S-conditionnelle.*

$C_1 = \{2\}$	$F(C_1) = \{1, 3\}$	$M(C_1) = \{1, 2, 3\}$	$R(C_1) = \{1\}$	$T(C_1) = \{3\}$
$C_2 = \{4, 5, 8\}$	$F(C_2) = \{3, 7, 9, 10, 14\}$	$M(C_2) = \{3, 4, 5, 7, 8, 9, 10, 14\}$	$R(C_2) = \{3, 7, 9\}$	$T(C_2) = \{10, 14\}$
$C_3 = \{6\}$	$F(C_3) = \{3, 7\}$	$M(C_3) = \{3, 6, 7\}$	$R(C_3) = \{3\}$	$T(C_3) = \{7\}$
$C_4 = \{11\}$	$F(C_4) = \{7, 12\}$	$M(C_4) = \{7, 11, 12\}$	$R(C_4) = \{7\}$	$T(C_4) = \{12\}$

TAB. 2.1 – Différentes parties associées aux 4 atomes de la partition $P_{\bar{S}}$.

La proposition suivante rassemble des propriétés, qui nous serviront dans la suite, des

objets que nous venons d'introduire, dans le cas particulier où toutes les feuilles de I appartiennent à S, autrement dit $I = S^+$ (et donc $S^+ - S = \bar{S}$)

Proposition 6 *Soit S une partie non vide de I, telle que $I = S^+$; soit C un atome de $\mathcal{P}_{\bar{S}}$, la partition S-conditionnelle de $\bar{S}$; alors :*

1. *$T(C) \neq \emptyset$*
2. *$M(C) \subseteq (T(C))^+$*
3. *Les enfants des éléments de C appartiennent à $C \cup T(C)$.*
4. *Les parents des éléments de $C \cup T(C)$ appartiennent à $M(C)$.*
5. *Les enfants des éléments de $\overline{M(C)}$ $(I - M(C))$ appartiennent à $R(C) \cup \overline{M(C)}$.*
6. *Les parents des éléments de $R(C) \cup [(T(C))^+ - M(C)]$ appartiennent aussi à $R(C) \cup [(T(C))^+ - M(C)]$.*

Démonstration :

1. Par hypothèse $I = S^+$; soit $i \in C$, il est ancêtre d'au moins un élément de S, et donc il existe un chemin $(i_1, i_2, \ldots, i_k)$ tel que $i_1 = i$ et $i_k \in S$, donc $i_k \notin C$. Soit i_ℓ $(2 \leq \ell \leq k)$ le premier élément non dans C sur ce chemin ; $i_\ell \in T(C)$ car il est enfant de $i_{\ell-1}$ qui appartient à C, et donc $i \in (T(C))^+$.
2. On vient d'établir que $C \subset (T(C))^+$; pour achever la démonstration que $M(C) \subseteq (T(C))^+$, il faut établir que $R(C)$ (éventuellement vide) est aussi contenu dans $(T(C))^+$. Soit donc $r \in R(C)$; comme il appartient à la frontière de Markov de C, $F(C)$, sans être dans $T(C)$, il est dans l'un des cas éventuellement non exclusifs suivants :
   - ou il a un enfant dans C (et alors r est ancêtre d'un élément de $T(C)$ puisque son enfant l'est)
   - ou il a avec un élément de C un enfant en commun non dans C ; cet enfant appartient à $F(C)$ mais non à $R(C)$ (puisqu'il a un parent dans C) et donc est un élément de $T(C)$; on a là encore établi que $r \in (T(C))^+$.
3. $T(C)$ a été introduit comme ensemble des enfants, non dans C, d'éléments de C.
4. Il résulte immédiatement de la défnition de la couverture de Markov que tout parent d'un élément de C est dans $M(C)$; d'autre part, comme tout élément de $T(C)$ a au moins un parent dans C, tous ses parents qui ne seraient pas dans C sont dans $F(C)$ et donc dans $M(C)$.
5. Pour prouver que tout enfant d'un élément non dans $M(C)$ appartient à $R(C) \cup \overline{M(C)}$, nous allons établir que tout parent d'un élément du complémentaire de $R(C) \cup \overline{M(C)}$, c'est à dire $C \cup T(C)$, appartient à $M(C)$. C'est le cas par définition de $M(C)$, pour tout élément de C ; soit par ailleurs $i \in T(C)$ et soit j un parent de i qui ne serait pas dans $M(C)$; mais i a aussi au moins un parent, soit k, dans C ; k et j ayant un enfant en commun, j doit appartenir à $M(C)$.
6. Posons $D = R(C) \cup ((T(C))^+ - M(C))$; D est contenu dans $(T(C))^+$ et donc il en est de même de l'ensemble des parents des éléments de D. Pour démontrer que

ces parents appartiennent à D, nous allons établir qu'ils ne peuvent pas appartenir au complémentaire de D dans $(T(C))^+$, c'est à dire $C \cup T(C)$. C'est évident, par définition même de $M(C)$, pour les éléments de $R(C)$; quant aux éléments de $(T(C))^+ - M(C)$, il ne peuvent avoir de parent ni dans C, par définition de $M(C)$, ni dans $T(C)$, par définition de $(T(C))^+$.

■

Soit K l'ensemble des éléments de S qui n'ont aucun parent dans $S^+ - S$ (soit qu'il n'aient aucun parent, et donc soient des racines de S^+, soit que leurs parents soient tous dans S).

(Dans l'exemple ci dessus, $K = \{1, 9, 13\}$)

Proposition 7 *L'ensemble Q_s des singletons $\{k\}$, où $k \in K$ (si $K \neq \emptyset$), et des parties $T(C) \cap S^+$, où $C \in \mathcal{P}_{S^+-S}$, constitue une partition de S.*

Démonstration

Q_s est un recouvrement de S, car tout élément i de S qui n'appartient pas à K a au moins un parent, soit j, dans $\bar{S}$; j appartient donc à l'un des atomes C de la partition $\mathcal{P}_{S^+-S}$, et donc i est dans $T(C)$.

Etablissons par ailleurs que les éléments de Q_s sont disjoints ; on doit distinguer deux cas :

- si C et C' sont deux atomes distincts de $\mathcal{P}_{S^+-S}$, $T(C) \cap T(C') \cap S^+ = \emptyset$; en effet, s'il existait $i \in T(C) \cap T(C') \cap S^+$, on disposerait d'une chaîne (j, i, j'), où $j \in C$, $j' \in C'$ et $i \in S$, convergente en i ; cette chaîne relierait dans S^+, sans blocage par S, un élément de C à un élément de C', ce qui est contraire à la définition de la partition $\mathcal{P}_{S^+-S}$,

- si $k \in K$ et $C \in \mathcal{P}_{S^+-S}$, k n'appartient pas à $T(C)$, puisque k n'a aucun parent dans $S^+ - S$, alors que chaque élément de $T(C)$ a moins un parent dans $C \subseteq S^+ - S$.

■

2.3 Factorisation des calculs de lois et lois conditionnelles

L'usage de la partition $\mathcal{P}_{S^+-S}$ permet d'ordonner les calculs de $P_{\bar{S}/S}$ et P_S.

2.3.1 Factorisation du calcul de $P_{\bar{S}/S}$

Rappelons que, S^+ étant une partie commençante, on a

$$P_{S^+} = \prod_{i \in S^+} P_{i/p(i)}$$

et donc

$$P_{\bar{S}/S} = \frac{P_I}{P_S} = [\prod_{i \notin S^+} P_{i/p(i)}] \frac{P_{S^+}}{P_S}$$

d'où, comme $S \subseteq S^+$,

$$P_{\bar{S}/S} = [\prod_{i \notin S^+} P_{i/p(i)}] P_{S^+-S/S}.$$

Or le corollaire 4 (en 2.1.2), exprimé dans le cadre du graphe moral restreint à $G^m_{S^+}$, nous apprend que

$$P_{S^+-S/S} = \prod_{C \in \mathcal{P}_{S^+-S}} P_{C/F(C)\cap S^+}.$$

Plus précisément, on a le théorème suivant :

Théorème 5 *Soit (I, G, P_I) un réseau bayésien et soit S une partie non vide de I. Soit $\mathcal{P}_{S^+-S}$ la partition S-conditionnelle de $S^+ - S$. Alors*

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

où

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

Démonstration :

La seule propriété qui reste à établir est l'égalité, pour $C \in \mathcal{P}_{S^+-S}$,

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

.

On peut donc sans perte de généralité, se limiter, dans la suite de cette preuve, au cas où $S^+ = I$.

On va calculer successivement $P_{M(C)}$ (on rappelle que $C \cup F(C) = M(C)$, couverture de Markov de C) puis $P_{F(C)}$ et enfin $P_{C/F(C)} = \frac{P_{M(C)}}{P_{F(C)}}$.

1 Calcul de $P_{M(C)}$.

On sait (proposition 6 (2)) que $M(C) \subseteq (T(C)^+$. On peut donc, pour l'obtention de $P_{M(C)}$ se limiter à $(T(C))^+$, c'est-à-dire effectuer le calcul :

$$P_{M(C)} = \sum_{(T(C))^+ - M(C)} \prod_{i \in (T(C))^+} P_{i/p(i)}.$$

Considérons la partition de $(T(C))^+$ comme union de $R(C) \cup ((T(C))^+ - M(C))$ et de $C \cup T(C)$. Alors

$$P_{M(C)} = \sum_{((T(C))^+ - M(C))} [\prod_{i \in C \cup T(C)} P_{i/p(i)} \prod_{i \in R(C) \cup ((T(C))^+ - M(C))} P_{i/p(i)}].$$

Il résulte de la proposition 6(4) que les éléments de $i \cup p(i)$, où $i \in C \cup T(C)$ ne sont pas dans $(T(C))^+ - M(C)$.

On a donc

$$P_{M(C)} = [\prod_{i \in C \cup T(C)} P_{i/p(i)}][\sum_{(T(C))^+ - M(C)} \prod_{i \in R(C) \cup ((T(C))^+ - M(C))} P_{i/p(i)}]. \qquad (2.2)$$

On remarque que, dans ce produit, le deuxième facteur ne dépend que de $R(C)$ (voir proposition 6(6)).

2. Calcul de $P_{F(C)}$.

Comme $F(C) = M(C) - C$, on va effectuer le calcul

$$P_{F(C)} = \sum_C P_{M(C)}.$$

Utilisant l'expression (2.2) de $P_{M(C)}$ et la remarque qui la suit, il vient

$$P_{F(C)} = [\sum_C \prod_{i \in C \cup T(C)} P_{i/p(i)}][\sum_{(T(C))^+ - M(C)} \prod_{i \in R(C) \cup ((T(C))^+ - M(C))} P_{i/p(i)}],$$

3 Calcul de $P_{C/F(C)}$.

Il résulte des expressions ci-dessus de $P_{M(C)}$ et $P_{F(C)}$ que

$$P_{C/F(C)} = \frac{\prod_{i \in C \cup T(C)} P_{i/p(i)}}{\sum_C \prod_{i \in C \cup T(C)} P_{i/p(i)}}.$$

■

2.3.2 Factorisation de calcul de P_S

Le théorème ci-dessous fournit la possibilité de segmenter le calcul de P_S en plusieurs calculs menés indépendamment.

Théorème 6 *Soit (I, G, P_I) un réseau bayésien et soit S une partie de I. Soit $\mathcal{P}_{S^+ - S}$ la partition S-conditionnelle de $S^+ - S$ et soit K l'ensemble des éléments de S qui n'ont aucun parent dans $S^+ - S$. Alors*

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

où

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

Avant de démontrer ce théorème, énonçons le corollaire suivant, où Q_s est la partition de S définie à la proposition 7 et dont la démonstration est évidente.

Corollaire 5 *Sous les hypothèses du théorème 6, la partition* Q_s $(= \{\{k\}; k \in K\} \cup \{T(C); C \in \mathcal{P}_{\bar{S}}\})$ *munit* S *de la structure de réseau bayésien de niveau deux, où la structure de graphe orienté,* G'*, est définie par :*

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

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

On remarque que le théorème 6 fournit en fait une structure de réseau bayésien de niveau deux 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 ci dessus, voici la représentation du réseau bayésien de niveau deux sur S

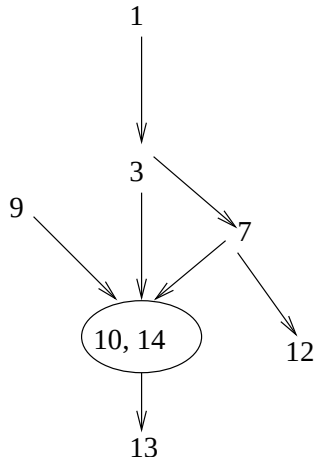


FIG. 2.8 – *Réseau bayésien de niveau deux représentant* P_S*.*

Démonstration du théorème 6 :

Comme S^+ est un réseau bayésien contenant S, on peut ici, sans perte de généralité, se limiter au cas où $S^+ = I$.

Soit $\mathcal{R}_I$ l'ensemble des parties de I constitué des singletons $\{k\}$, où $k \in K$, et des parties de la forme $C \cup T(C)$ $(= M(C) - R(C))$, où $C \in \mathcal{P}_{\bar{S}}$; il est clair que R_I est une partition de I puisqu'on peut la déduire des partitions Q_S de S et $\mathcal{P}_{\bar{S}}$ de $\bar{S}$ en conservant les $\{k\}$ (atomes de Q_S) et en faisant, pour tout C (atome de $\mathcal{P}_{\bar{S}}$) son union avec $T(C)$ (atome de Q_S).

On a donc, à partir de l'expression $P_I = \prod_{i\in I} P_{i/p(i)}$, en regroupant les i selon la partition $\mathcal{R}_I$,

$$P_I = \prod_{k\in K} P_{k/p(k)} \prod_{C\in\mathcal{P}_{\bar{S}}} \prod_{i\in C\cup T(C)} P_{i/p(i)}.$$

Notre but est de calculer $P_S = \sum_{\bar{S}} \prod_{i\in I} P_{i/p(i)}$. Or, par définition de K, on a, pour tout $k \in K$, $(k \cup p(k)) \cap \bar{S} = \emptyset$; de plus par définition de $T(C)$, un sommet j, appartenant à un atome C de $\mathcal{P}_{\bar{S}}$, ne peut intervenir comme variable de $P_{i/p(i)}$ (autrement dit appartenir à $i \cup p(i)$) que si $i \in C \cup T(C)$ (voir proposition 6 (3)). Il en résulte que le calcul de $P_S = \sum_{\bar{S}} P_I$ s'effectue sous la forme

$$P_S = \prod_{k\in K} P_{k/p(k)} \prod_{C\in\mathcal{P}_{\bar{S}}} \sum_{C} \prod_{i\in C\cup T(C)} P_{i/p(i)}.$$

Il reste à démontrer que, étant fixé $C \in \mathcal{P}_{\bar{S}}$,

$$\sum_{C} \prod_{i\in C\cup T(C)} P_{i/p(i)} = P_{T(C)/R(C)} \ ;$$

On rappelle que $T(C) \cup R(C) = F(C)$, frontière de Markov de C ; le calcul de $P_{F(C)}$ ayant été effectué dans le cadre de la démonstration du théorème 5, il reste à calculer $P_{R(C)}$ puis $P_{T(C)/R(C)} = \frac{P_{F(C)}}{P_{R(C)}}$.

On sait que

$$P_{F(C)} = [\sum_{C} \prod_{i\in C\cup T(C)} P_{i/p(i)}][\sum_{(T(C))^+ - M(C)} \prod_{i\in R(C)\cup((T(C))^+ - M(C))} P_{i/p(i)}] \qquad (2.3)$$

et que, dans ce produit, le deuxième facteur ne dépend que de $R(C)$.

Comme $F(C) = R(C) \cup T(C)$, on va effectuer le calcul

$$P_{R(C)} = \sum_{T(C)} P_{F(C)}.$$

Utilisant l'expression (2.3) de $P_{F(C)}$ et la remarque qui la suit, il vient

$$P_{R(C)} = [\sum_{T(C)} \sum_{C} \prod_{i\in C\cup T(C)} P_{i/p(i)}][\sum_{(T(C))^+ - M(C)} \prod_{i\in R(C)\cup((T(C))^+ - M(C))} P_{i/p(i)}],$$

c'est-à-dire

$$P_{R(C)} = [\sum_{C\cup T(C)} \prod_{i\in C\cup T(C)} P_{i/p(i)}][\sum_{(T(C))^+ - M(C)} \prod_{i\in R(C)\cup((T(C))^+ - M(C))} P_{i/p(i)}].$$

Dans cette dernière expression, le premier facteur vaut 1 ; de manière standard, il suffit pour s'en assurer d'ordonner les éléments de $C \cup T(C)$ selon un ordre hiérarchique et d'effectuer les sommations selon les éléments de $C \cup T(C)$ en remontant cet ordre.

Donc

$$P_{R(C)} = \sum_{(T(C))^+ - M(C)} \prod_{i \in R(C) \cup ((T(C))^+ - M(C))} P_{i/p(i)}.$$

Il résulte des expressions ci-dessus de $P_{F(C)}$ et $P_{R(C)}$ que

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

■

Chapitre 3

Techniques de Construction d'Arbres de Jonction. Applications aux calculs dans les Réseaux Bayésiens

3.1 Introduction

Soit donné, sur un ensemble $\Omega_I = \prod_{i\in I} \Omega_i$, la loi P_I d'une famille, indexée par l'ensemble fini I, de variables aléatoires discrètes, notée $X_I = (X_i)_{i\in I}$, et soit $\mathcal{J}$ un recouvrement de I, tel que $P_I = \frac{1}{\alpha}\prod_{J\in\mathcal{J}} f_J$, où pour tout J, f_J est une application connue de $\Omega_J = \prod_{i\in J}\Omega_i$ dans $\mathbb{R}_+$. C'est le cas si on dispose d'un réseau bayésien sur I ; alors $\mathcal{J}$ est l'ensemble des parties $i \cup p(i)$ où $p(i)$ est l'ensemble des parents de i et $f_{i\cup p(i)}$ est la loi de X_i conditionnellement à $X_{p(i)} = (X_j)_{j\in p(i)}$; dans ce cas α est connu et vaut 1.

Nous nous intéressons aux techniques de calcul des probabilités P_A (loi de $X_A = (X_i)_{i\in A}$ où $A \subseteq I$) et $P_{A/B}$ (loi de X_A conditionnellement à X_B, où A et B sont deux parties disjointes de I) ; nous dirons que A est une cible de calcul.

Plusieurs classes d'algorithmes ont été développées pour ces calculs. Nous nous intéressons ici à celle, classique, initiée par Lauritzen et Spiegelhalter (voir [39], [53]), qui fournit une méthode algorithmique de calcul dans le cas particulier où $\mathcal{J}$ possède la structure d'arbre de jonction de parties de I (voir section 3.2.2). Cet algorithme calcule alors les P_J (ou les $P_{J-B/B}$). Dans le cas général, son usage, pour calculer des lois P_K (où K appartient à une famille donnée de cibles $\mathcal{K}$) exige la construction préalable d'un arbre de jonction $\mathcal{C}$ tel que, pour tout $J \in \mathcal{J}$ (respectivement tout $K \in \mathcal{K}$) il existe une partie C de $\mathcal{C}$ telle que $J \subseteq C$ (respectivement $K \subseteq C$) (nous dirons que $\mathcal{C}$ est un sur-recouvrement de $\mathcal{J} \cup \mathcal{K}$). Cette problématique est présentée de manière détaillée en section 3.1.1.

En section 3.2, nous fournissons des compléments sur la notion d'arbre de jonction et sur

celle, qui lui est associée, de suite recouvrante propre possédant la propriété d'intersection courante.

En section 3.3 nous étudions deux techniques de construction d'un arbre de jonction qui soit un sur-recouvrement d'une famille donnée de parties de I ; en sous-section 3.3.1, nous présentons la technique, développée dans [39], [14] et qui est implémentée dans de nombreux logiciels (Netica, Hugin, Genie), dite de "moralisation et triangularisation", qui consiste à obtenir $\mathcal{C}$ comme l'ensemble des cliques pour une structure de graphe non orienté déduite de $\mathcal{J}$ et $\mathcal{K}$ (nous ne reprendrons pas ici intégralement ces algorithmes classiques de construction) ; en sous-section 3.3.2, reprenant une idée de Adnan Darwiche (voir [5]) nous présentons deux versions, dont l'une nouvelle, de construction récurrente d'un arbre de jonction ; nous présentons plusieurs exemples d'une telle construction.

En section 3.4 nous présentons et justifions l'algorithme LS (Lauritzen et Spiegelhalter) pour le calcul des P_C et des $P_{C-B/B}$, où $C \in \mathcal{C}$ (arbre de jonction) ; son déroulement comporte deux phases qualifiées par les informaticiens de "suites de transmissions de messages" ; nous mettons son déroulement en évidence sur un exemple.

3.1.1 Contexte et annonce du résultat fondamental

Soit $X_I = (X_i)_{i\in I}$ une famille de variables aléatoires discrètes : pour tout i, X_i est à valeurs dans l'ensemble fini Ω_i. Pour tout $J \subseteq I$, on note X_J la famille $(X_i)_{i\in J}$, à valeurs dans $\Omega_J = \prod_{i\in J} \Omega_i$ et P_J la loi de X_J ; pour tout $x_I = (x_i)_{i\in I} \in \Omega_I$, on note x_J sa restriction à (autrement dit sa projection sur) Ω_J.

Soit $\mathcal{C}$ un recouvrement de I. Rappelons que $\mathcal{C}$ et la loi P_I de la famille $(X_i)_{i\in I}$ sont dits **compatibles** si P_I peut s'écrire sous la forme

$$\forall\ x_I \in \Omega_I \quad P_I(x_I) \ = \ \frac{1}{\alpha} \prod_{C\in\mathcal{C}} f_C(x_C) \tag{3.1}$$

où, pour tout C, f_C est une application de Ω_C dans $\mathbb{R}_+$ connue (i.e. mise en mémoire d'ordinateur) et souvent appelée **potentiel** de C ; on pourra écrire (3.1) sous la forme abrégée

$$P_I \ = \ \frac{1}{\alpha} \prod_{C\in\mathcal{C}} f_C \ .$$

α est la constante de normalisation, non mise en mémoire ; elle s'exprime bien sûr formellement comme $\alpha = \sum_{x\in\Omega_I} \prod_{C\in\mathcal{C}} f_C(x_C)$, mais on désire éviter ce calcul en général très lourd. Le but des calculs présentés ici est justement de ne faire effectuer de sommations que sur des ensembles Ω_J où les J sont souhaités "bien plus petits" que I et convenablement choisis en fonction de ce que sont les lois, ou lois conditionnelles, de sous-familles de X_I que l'on veut effectivement calculer.

Or, si P_I est compatible avec $\mathcal{C}$, un algorithme fondamental, dû à Lauritzen [39], fournit une méthode de calcul pour obtenir les lois P_C de familles $X_C(= (X_i)_{i\in C})$ sous condition que le recouvrement $\mathcal{C}$ soit un **arbre de jonction** (voir définition en 3.2.2 ci-dessous)

3.1.2 Principe du calcul de familles de lois X_K (où $K \subseteq I$)

L'emploi du résultat fondamental ci-dessus est d'une grande flexibilité car, si on sait que la loi P_I est compatible avec un recouvrement $\mathcal{J}$, elle l'est aussi avec tout sur-recouvrement $\mathcal{C}$ de $\mathcal{J}$, c'est-à-dire tout recouvrement vérifiant

$$\forall\ J \in \mathcal{J}\ \ \exists\ C \in \mathcal{C}\ \ J \subseteq C\ ;$$

il suffit en effet de choisir une application ϕ de $\mathcal{J}$ dans $\mathcal{C}$ telle que

$$\forall\ J \in \mathcal{J}\ \ J \subseteq \phi(J)$$

et, partant de l'hypothèse $P_I = \frac{1}{\alpha} \prod_{J\in\mathcal{J}} g_J(x_J)$, de poser, pour tout $C \in \mathcal{C}$, $f_C(x_C) = \prod_{J\in\phi^{-1}(C)} g_J(x_J)$, ce qui a bien un sens puisque, pour tout $J \in \phi^{-1}(C)$, $J \subseteq C$ et x_J est la projection sur Ω_J de x_C $(\in \Omega_C)$; si $\phi^{-1}(C) = \emptyset$ on pose, comme il est usuel, $f_C(x_C) = 1$.

Soit donné par ailleurs un ensemble $\mathcal{K}$ de parties de I (parfois $\mathcal{J}$ lui même) et assignons nous pour but de calculer les lois P_K pour tout $K \in \mathcal{K}$. Les éléments de $\mathcal{K}$ seront appelés **cibles**.

Remarquons que, si $\mathcal{C}$ est un ensemble de parties de I telles que

$$\forall\ K \in \mathcal{K}\ \ \exists\ C \in \mathcal{C}\ \ K \subseteq C\ ,$$

et qu'on ait été capable d'obtenir les lois P_C, on peut en déduire chaque P_K en choisissant un C tel que $K \subseteq C$ et en effectuant une sommation sur les variables aléatoires d'indices dans $C - K$; autrement dit, si on décompose x_C en le couple (x_K, x_{C-K}), on a

$$P_K(x_K) = \sum_{x_{C-K}\in\Omega_{C-K}} P_C(x_C).$$

P_I étant donné compatible avec le recouvrement $\mathcal{J}$, une méthode de calcul des lois cibles P_K $(K \in \mathcal{K})$ est alors la suivante :

A. Fabriquer un sur-recouvrement $\mathcal{C}$ de $\mathcal{J} \cup \mathcal{K}$ qui soit un arbre de jonction.

B. Utiliser la compatibilité de P_I avec $\mathcal{C}$ pour calculer les lois P_C.

C. Calculer les lois P_K à partir des lois P_C.

3.1.3 Calcul de lois conditionnelles

Il est essentiel de remarquer que la méthode dont nous venons de présenter le principe s'adapte sans difficulté au calcul de lois conditionnelles ; soit en effet fixé A, partie stricte non vide de I et une valeur x_A de la variable aléatoire X_A (autrement dit, on connaît pour tout $a \in A$, une valeur x_a de X_a). Alors la loi conditionnelle de X_{I-A}, que nous noterons $P^{x_A}_{I-A/A}$ est, par application immédiate de la définition élémentaire des probabilités conditionnelles, compatible avec le recouvrement de $I - A$ constitué des parties $J - A$ telles que

$J \in \mathcal{J}$ et $J - A \neq \emptyset$; de manière précise , si par hypothèse on a

$$P_I(x_I) \;=\; \frac{1}{\alpha} \prod_{J \in \mathcal{J}} g_J(x_J)$$

il vient, avec une constante de normalisation $\alpha(x_A)$,

$$P^{x_A}_{I-A}(x_{I-A}) \;=\; \frac{1}{\alpha(x_A)} \prod_{J \in \mathcal{J}_A} g^{x_A}_{J-A}(x_{J-A})$$

où $\mathcal{J}_A = \{J \in \mathcal{J} \;;\; J - A \neq \emptyset\}$ et $g^{x_A}_{J-A}$ est l'application de Ω_{J-A} dans $\mathbb{R}_+$ déduite de g_J en fixant chaque variable d'indice a dans $A \cap J$ (s'il y en a) à la valeur conditionnante x_a. On pourra alors calculer les lois, conditionnellement à l'événement $[X_A = x_a]$, de n'importe quelle famille de cibles contenues dans $I - A$.

3.1.4 Application aux réseaux bayésiens et réseaux bayésiens de niveau 2

Le contexte que nous venons de présenter est en particulier celui des calculs de lois ou de lois conditionnelles dans les réseaux bayésiens ou les réseaux bayésiens de niveau 2 (voir [30]).

Pour un réseau bayésien, on dispose, pour tout $i \in I$, de l'ensemble de ses "parents" (éventuellement vide) et des lois conditionnelles $P^{x_{p(i)}}_{i/p(i)}$, autrement dit des tableaux de valeurs $g(x_{i \cup p(i)}) = P^{x_{p(i)}}_{i/p(i)}(x_i)$ (en particulier, si $p(i) = \emptyset$, $g(x_i) = P_i(x_i)$) ; on est bien dans le cadre de la compatibilité de P_I avec le recouvrement $\mathcal{J} = \{i \cup p(i) \;;i \in I\}$ car $P_I(x_I) = \prod_{i \in I} P^{x_{p(i)}}_{i/p(i)}(x_i)$ (ici la constante de normalisation vaut 1 ; par contre elle n'est pas donnée de manière immédiate pour les lois conditionnelles $P^{x_A}_{I-A}$).

Dans le cas des réseaux bayésiens de niveau deux, on dispose d'une partition $\mathcal{B}$ de I et, pour tout $B \in \mathcal{B}$, de l'ensemble de ses parents $p(B)$ qui est une partie, éventuellement vide, de $\mathcal{B}$, ainsi que des lois conditionnelles $P^{x_{p(B)}}_{B/p(B)}$, où $x_{p(B)}$ est un élément de $\Omega_{p(B)}$ (cette notation repose sur l'identification de $p(B)$, partie de $\mathcal{B}$, avec la partie de I, $(p(B))^\star$, qui est l'union des parties de I constituant $p(B)$; donc la notation $\Omega_{p(B)}$ signifie $\Omega_{(p(B))^\star}$) ; alors

$$P_I(x_I) \;=\; \prod_{B \in \mathcal{B}} P^{x_{p(B)}}_{B/p(B)}(x_B) \;,$$

ce qui exprime la compatibilité de P_I avec le recouvrement $\mathcal{J} = \{B \cup p(B) \;;B \in \mathcal{B}\}$.

3.1.5 Analyse critique de la méthode

Avant de la détailler, relevons trois difficultés inhérentes à la démarche que nous avons présentée en 3.1.2 et 3.1.3.

A. L'opération de "redescente" de P_C (resp. $P^{x_A}_{C/A}$) à P_K (resp. $P^{x_A}_{K/A}$), où $K \subseteq C$ (resp. $K \subseteq C \subseteq I - A$), peut être lourde si Ω_{C-K} est "gros", ceci pouvait être dû à la fois au cardinal des Ω_i, où $i \in C - K$ (taille intrinsèque au problème posé) et au cardinal de $C - K$ (taille liée à la méthode de calcul) ; il peut être intéressant, si on vise plusieurs cibles K ($\in \mathcal{K}$) de ne pas les traiter globalement (ce qui grossit les parties appartenant à $\mathcal{C}$) mais individuellement, ou "par petits paquets".

B. L'étude de $P^{x_A}_{I-A/A}$ (i.e. le calcul des $P^{x_A}_{C/A}$, où $\mathcal{C}$ est le recouvrement construit sur $I-A$) est totalement liée au choix de x_A dont dépend en particulier la constante de normalisation $\alpha(x_A)$; si on s'intéresse à la transition de conditionnement $P_{I-A/A}$ (i.e. l'application de Ω_A dans l'ensemble des probabilités sur $I - A$, $x_A \rightsquigarrow P^{x_A}_{I-A/A}$) cette démarche impose de faire autant de calculs distincts qu'il y a d'éléments dans Ω_A.

C. L'étape **A**, purement de théorie des graphes, qui consiste à construire un arbre de jonction qui est un sur-recouvrement de $\mathcal{J} \cup \mathcal{K}$, peut être lourde ; deux techniques se trouvent essentiellement dans la littérature (et dans les logiciels) :

- technique par "moralisation et triangularisation", dont nous donnerons en 3.3.1 le principe, mais ne détaillerons pas les algorithmes associés (voir [39], [16], [14]),
- technique par construction récurrente d'une suite recouvrante propre P.I.C. (voir définition en 3.2.4 ci-dessous), introduite dans [14], [5] et que nous détaillerons car nous proposons une variante qui, pensons nous, l'accélère sensiblement.

3.1.6 Plan

Nous avons introduit ci-dessus, en fin de section 3.1.2, les trois stades du calcul des lois cibles (calcul adapté en 3.1.3 au calcul des lois conditionnelles). Nous allons les détailler dans la suite. Voici donc le plan des sections suivantes de ce document :

3.2. Compléments sur la notion d'arbre de jonction et celle, qui lui est étroitement associée, de suite recouvrante propre possédant la propriété d'intersection courante (P.I.C.).

3.3. Techniques de construction des suites recouvrantes propres possédant la propriété d'intersection courante.

3.4. Algorithme de calcul des lois P_C, où $C \in \mathcal{C}$ (arbre de jonction sur I compatible avec P_I).

3.2 Arbres de jonction et suites recouvrantes propres P.I.C.

3.2.1 Rappels : Arbre, arborescence, numérotation topologique d'un arbre

Une structure de graphe non-orienté $\mathcal{H}$ sur un ensemble $\mathcal{J}$ est appelée **arbre** si, pour tout couple (J, J') de sommets distincts, il existe une chaîne unique les reliant (suite $(J_0, \ldots, J_k)$ telle que $J_0 = J, J_k = J'$ et $\forall\ \ell \in \{1, \ldots, k\}\ \{J_{\ell-1}, J_\ell\} \in \mathcal{H}$).

Une structure de graphe orienté $\mathcal{G}$ sur un ensemble $\mathcal{J}$ est appelée **arborescence** s'il existe

un sommet sans prédécesseurs, dit racine, soit R, relié à chaque autre sommet J par un **chemin** unique (suite $(J_0, \ldots, J_k)$ telle que $J_0 = R, J_k = J$ et $\forall\ \ell \in \{1, \ldots, k\}\ (J_{\ell-1}, J_\ell) \in \mathcal{G}$). Il existe plusieurs définitions équivalentes de l'arborescence ; une qui nous sera utile est que tout sommet a un unique prédécesseur, sauf un (la racine) qui n'en a aucun.

Etant donné un graphe orienté $(\mathcal{J}, \mathcal{G})$, le graphe non orienté **sous-jacent** est $(\mathcal{J}, \mathcal{H})$ tel que $\{J, J'\} \in \mathcal{H}$ si et seulement si $(J, J') \in \mathcal{G}$ ou $(J', J) \in \mathcal{G}$ et on appelle chaîne pour $\mathcal{G}$ toute chaîne pour le graphe orienté sous-jacent $\mathcal{H}$.

Il est clair que le graphe non orienté sous-jacent à une arborescence est un arbre. Inversement, étant donné un arbre $(\mathcal{J}, \mathcal{H})$, tout choix d'un de ses sommets comme racine implique une unique structure d'arbre à laquelle $\mathcal{H}$ est sous-jacent.

Toute chaîne de J à J' dans une arborescence est soit sous-jacente à un chemin de J à J', soit sous-jacente à un chemin de J' à J, soit telle qu'il existe J'' vérifiant : la chaîne de J à J' est la concaténée de deux chaînes sous-jacentes l'une à un chemin de J'' à J, l'autre à un chemin de J'' à J'.

Soit $(\mathcal{J}, \mathcal{H})$ un arbre tel que $\text{card}(\mathcal{J}) = n$; soit un numérotage (de 1 à n) des sommets de cet arbre. La numérotation est dite **topologique** si l'arborescence construite sur cet arbre en prenant pour racine le sommet numéroté 1 posséde la propriété suivante : pour tout $i \neq 1$, sur le chemin joignant la racine au sommet numéroté i, les sommets intermédiaires ont des numéros inférieurs à i ; il en résulte, en raison de l'unicité des chemins dans une arborescence, que tous les sommets joignant le sommet numéroté i à des feuilles de l'arborescence (sommets sans successeurs) ont des numéros supérieurs à i.

Les algorithmes permettant de procéder à une numérotation topologique d'un arbre $(\mathcal{J}, \mathcal{H})$ sont aisés à décrire : on initialise en choisissant librement le sommet affecté du numéro 1 puis, une fois affectés les numéros $1, \ldots, i$, on affecte le numéro $i + 1$ à l'un des sommets vérifiant la propriété suivante : tous les sommets situés sur la chaîne le reliant à la racine ont déja été numérotés (autrement dit à un sommet dont, dans l'arborescence définie en prenant pour racine le sommet numéroté 1, le prédécesseur a déjà été numéroté)

Les numérotage "en profondeur d'abord" ou "en largeur d'abord", classiques en théorie des graphes, sont des cas particuliers de numérotations topologiques.

3.2.2 Arbres de jonction - Séparateurs

Définition 13 *On appelle* **recouvrement propre** *d'un ensemble I tout ensemble de parties de I, soit $\mathcal{J}$ tel que*

$$\left|\begin{array}{l} \cup_{J\in\mathcal{J}} J = I, \\ \textit{pour tout couple } (J, J') \textit{ d'éléments distincts de } \mathcal{J},\ J - J' \neq \emptyset\ ; \end{array}\right.$$

en d'autres termes un recouvrement est propre s'il n y a aucune inclusion (large ou stricte) entre ses éléments.

Il est clair qu'à tout recouvrement non propre peut être associé un recouvrement propre

en en retirant toute partie qui est incluse strictement dans une autre et, en cas d'égalités entre parties, en en conservant une seule.

Définition 14 *Soit I un ensemble fini. On appele* **arbre de jonction** *sur I tout arbre $(\mathcal{C}, \mathcal{H})$, où $\mathcal{C}$ est un recouvrement propre de I vérifiant la propriété suivante :*
pour tout couple (C, C') d'éléments de $\mathcal{C}$ non-adjacents, tout sommet intermédiaire C'' sur la chaîne dans $\mathcal{H}$ joignant C à C' vérifie $C \cap C' \subset C''$.

Remarque : Si card $(\mathcal{C}) \leq 2$, la propriété intervenant dans la définition des arbres de jonction est vide. Sont donc trivialement des arbres de jonction :
-pour $k = 1$, $\mathcal{C} = \{I\}$ (arbre à un seul élément, et donc zéro arête),
-pour $k = 2$, toute paire $\{C, C'\}$ telle que $C \cup C' = I$, $C - C' \neq \emptyset$, $C' - C \neq \emptyset$.

Définition 15 *Soit $(\mathcal{C}, \mathcal{H})$ un arbre de jonction sur un ensemble I ; les séparateurs associés à cet arbre sont les intersections $C \cap C'$, où C et C' sont deux éléments de $\mathcal{C}$ qui sont adjacents pour la structure d'arbre $\mathcal{H}$.*

On remarque que, si le cardinal de $\mathcal{C}$ est k, il y a $k-1$ séparateurs (nombre d'arêtes dans un arbre à k sommets).

Notons que les séparateurs ne sont pas nécessairement non vides. Par exemple, une partition en parties non vides de I, munie de n'importe quelle structure d'arbre, est un cas particulier d'arbre de jonction pour lequel les séparateurs sont tous vides.

Une présentation équivalente des arbres de jonction est donnée par le théorème suivant.

Théorème 7 *Soit $\mathcal{C}$ un recouvrement propre de I et $\mathcal{H}$ une structure d'arbre sur $\mathcal{C}$.*
Pour que $(\mathcal{C}, \mathcal{H})$ soit un arbre de jonction, il faut et il suffit que, pour tout $i \in I$, l'ensemble des éléments de $\mathcal{C}$ auxquels appartient i, soit $\mathcal{C}_i$, constitue un sous-arbre de $(\mathcal{C}, \mathcal{H})$ (autrement dit soit une partie de $\mathcal{C}$ connexe pour $\mathcal{H}$).

Démonstration

1. Soit $(\mathcal{C}, \mathcal{H})$ un arbre de jonction ; supposons qu'il existe $i \in I$ tel que $\mathcal{C}_i$ ne soit pas un sous-arbre, c'est-à-dire qu'il existe $C \in \mathcal{C}_i$ et $C' \in \mathcal{C}_i$ tels que la chaîne reliant C à C' comporte au moins un sommet $C'' \notin \mathcal{C}_i$; autrement dit $i \in C \cap C'$, mais $i \notin C''$; il est alors faux que $C \cap C' \subset C''$, ce qui est contradictoire avec la définition des arbres de jonction.

2. Réciproquement, supposons que l'arbre $(\mathcal{C}, \mathcal{H})$ ne soit pas un arbre de jonction ; alors il existe 3 éléments de $\mathcal{C}$, notés C, C' et C'' tels que C'' soit sur la chaîne reliant C à C' mais que $C \cap C'$ ne soit pas un sous-ensemble de C'' ; il existe alors i tel que $i \in C \cap C'$ mais $i \notin C''$; C et C' ne sont pas donc reliés par une chaîne dans la restriction de $\mathcal{H}$ à $\mathcal{C}_i$.

■

3.2.3 Suite recouvrante propre P.I.C.

Définition 16 *Etant donné un ensemble I, on appele* **suite recouvrante propre** *de I toute suite $(C_1, \ldots, C_k)$ de parties de I telle que*

$$\left| \begin{array}{l} \cup_{i=1}^{k} C_i = I \\ \text{pour tout couple } (j, j') \text{ d'indices distincts, } C_j - C_{j'} \neq \emptyset \end{array} \right.$$

Définition 17 *Une suite de parties d'un ensemble I, $(C_1, \ldots, C_k)$, est dite posséder la propriété d'intersection courante (P.I.C.) si*

$$\forall\ j \in \{2, \ldots, k\}\ \ \exists\ \ell \in \{1, \ldots, j-1\}\ \ C_j \cap (\cup_{h=1}^{j-1} C_h) \subseteq C_\ell$$

La propriété d'intersection courante s'écrit de manière équivalente :

$$\forall\ j \in \{2, \ldots, k\}\ \ \exists\ \ell \in \{1, \ldots, j-1\}\ \ \forall h < j\ \ C_j \cap C_h \subseteq C_\ell$$

On remarque que cette propriété est toujours satisfaite pour $j = 2$, avec $\ell = 1$.

Nous nous intéressons particulièrement aux suites recouvrantes propres possédant la P.I.C. ; on remarque que sont dans ce cas :
- la suite à un élément (I)
- toute suite recouvrante propre à 2 éléments : couple (C_1, C_2) tel que $C_1 \cup C_2 = I$, $C_1 - C_2 \neq \emptyset$, $C_2 - C_1 \neq \emptyset$.

3.2.4 Lien entre les arbres de jonction et la P.I.C.

La propriété essentielle des arbres de jonction, en ce qui concerne nos besoins, est donnée par le théorème suivant :

Théorème 8 *Tout arbre de jonction sur I donne naissance à autant de suites recouvrantes propres possédant la P.I.C. qu'il existe de numérotations topologiques sur cet arbre. Inversement toute suite recouvrante propre possédant la P.I.C. peut être munie d'une structure d'arbre de jonction.*

Démonstration : Le théorème est évident si $k = 1$ ou $k = 2$. Soit désormais $k \geq 3$.

1. Soit donné un arbre de jonction sur I, $(\mathcal{C}, \mathcal{H})$ et soit $(C_1, \ldots, C_k)$ une numérotation topologique des éléments de $\mathcal{C}$, associé à une arborescence $\mathcal{G}$ dont $\mathcal{H}$ est l'arbre sous-jacent. Nous allons montrer que $(C_1, \ldots, C_k)$ posséde la P.I.C.

Définissons, pour tout $i \in \{2, \ldots, k\}$, $\sigma(i)$ de sorte que $C_{\sigma(i)}$ soit le prédécesseur de C_i pour l'arborescence $\mathcal{G}$ (en particulier $\sigma(2) = 1$). Il résulte du caractère topologique de la numérotation que $\sigma(i) < i$; de plus, pour tout $j < i$ et différent de $\sigma(i)$, l'unique chaîne reliant C_j à C_i admet $C_{\sigma(i)}$ pour sommet intermédiaire ; en effet les seuls éléments de $\mathcal{C}$ tels que les chaînes les reliant à C_i ne passent pas par $C_{\sigma(i)}$ sont ceux situés sur des chaînes reliant C_i à des feuilles de l'arborescence, et leurs numéros sont, dans la numérotation

topologique, supérieurs à i. Il résulte alors du fait que $(\mathcal{C}, \mathcal{H})$ est un arbre de jonction que $C_i \cap C_j \subseteq C_{\sigma(i)}$.

2. Réciproquement, soit $(C_1, \ldots, C_k)$ une suite recouvrante propre de I possédant la P.I.C. On définit sur $\mathcal{C} = \{C_1, \ldots, C_k\}$ une structure d'arborescence en affectant à C_2, comme prédécesseur, C_1 et, à tout C_i tel que $i > 2$, un prédécesseur unique, noté $C_{\sigma(i)}$ vérifiant, pour tout $j < i$, $C_j \cap C_i \subseteq C_{\sigma(i)}$ (un tel $\sigma(i)$ existe par définition de la P.I.C. ; s'il y en a plusieurs, on en fixe arbitrairement un). Cette numérotation de l'arborescence est évidemment topologique.

Nous allons vérifier que l'arbre sous-jacent à cette arborescence est un arbre de jonction. Soit un couple d'indices (i, j), avec $i < j$, et considérons la chaîne reliant C_i à C_j dans l'arbre que nous venons de construire. Deux cas sont à envisager selon que cette chaîne est ou non un chemin dans l'arborescence (voir exemples ci-dessous).

a. **Cas où la chaîne est un chemin ;** soit $(C_{x_0}, \ldots, C_{x_r})$ ce chemin (avec $i = x_0 < x_1 < \ldots < x_r = j$) ; par construction, pour tout $s \in \{1, \ldots, r\}$, $x_{s-1} = \sigma(x_s)$ donc, par P.I.C., $C_{x_0} \cap C_{x_s} \subseteq C_{x_0} \cap C_{x_{s-1}}$; on a donc la chaîne d'inclusions :

$$C_i \cap C_j = C_{x_0} \cap C_{x_r} \subseteq C_{x_0} \cap C_{x_{r-1}} \subseteq \ldots \subseteq C_{x_0} \cap C_{x_1},$$

d'où en particulier $C_i \cap C_j \subseteq C_{x_s}$ pour tout $s \in \{1, \ldots, r-1\}$.

b. **Cas où la chaîne n'est pas un chemin ;** on a vu en 3.2.1 qu'elle se compose alors, dans l'arborescence, d'une "partie montante" et d'une "partie descendante" ; autrement dit il existe un sommet C_ℓ (où $\ell < i$ et $\ell < j$) et deux chemins, l'un de C_ℓ à C_i, noté $(C_{x_0}, \ldots, C_{x_v})$ (avec $\ell = x_0 < x_1 < \ldots < x_v = i$), l'autre de C_ℓ à C_j, noté $(C_{y_0}, \ldots, C_{y_w})$ (avec $\ell = y_0 < y_1 < \ldots < y_w = j$) ; par construction, pour tout $s \in \{1, \ldots, v\}$, $x_{s-1} = \sigma(x_s)$ et, pour tout $t \in \{1, \ldots, w\}$, $y_{t-1} = \sigma(x_t)$.

Comme $i < j$, autrement dit $x_v < y_w$, on a

$$C_i \cap C_j = C_{x_v} \cap C_{y_w} \subseteq C_{x_v} \cap C_{\sigma(y_w)} = C_{x_v} \cap C_{y_{w-1}} \subseteq C_{y_{w-1}} \ ;$$

admettons, comme hypothèse de récurrence, avoir établi, pour un couple (s, t) autre que $(1, 1)$, qu'il est vrai que $C_i \cap C_j \subseteq C_{x_{s'}}$ pour tout s' tel que $s \leq s' \leq v$, et que $C_i \cap C_j \subseteq C_{x_{y'}}$ pour tout t' tel que $t \leq t' \leq w$.

Alors, si $x_s > y_t$, il vient

$$C_i \cap C_j \subseteq C_{x_s} \cap C_{y_t} \subseteq C_{\sigma(x_s)} \cap C_{y_t} \subseteq C_{x_{s-1}}$$

et, si $x_s < y_t$, il vient

$$C_i \cap C_j \subseteq C_{x_s} \cap C_{y_t} \subseteq C_{x_s} \cap C_{\sigma(y_t)} \subseteq C_{y_{t-1}} \ ;$$

on établit ainsi par récurrence que $C_i \cap C_j \subseteq C_{x_s}$ et $C_i \cap C_j \subseteq C_{y_t}$ pour tout s $(1 \leq s \leq v)$ et tout t $(1 \leq t \leq w)$, autrement dit $C_i \cap C_j \subseteq C_m$ pour tout sommet C_m sur la chaîne reliant C_i à C_j.

■

Exemples

Cas a.

$$i = 3, \quad j = 10$$

$$\sigma(5) = 3, \quad \sigma(6) = 5, \quad \sigma(10) = 6$$

$$C_3 \cap C_{10} \subseteq C_3 \cap C_6 \subseteq C_3 \cap C_5$$

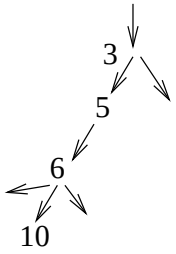


FIG. 3.1 – *Cas a.*

Cas b.

$$i = 10, \quad j = 13$$

$$\sigma(13) = 11, \quad \sigma(11) = 3, \quad \sigma(3) = 2$$

$$\sigma(10) = 8, \quad \sigma(8) = 7, \quad \sigma(7) = 2$$

$$C_{10} \cap C_{13} \subseteq C_{10} \cap C_{11} \subseteq C_8 \cap C_3 \subseteq C_3 \cap C_7 \subseteq C_3 \cap C_2$$

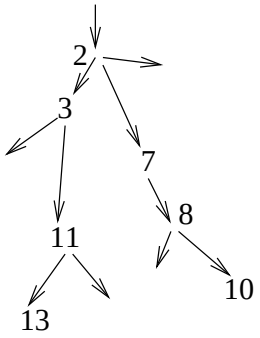


FIG. 3.2 – *Cas b.*

3.3 Deux méthodes de construction d'un arbre de jonction

3.3.1 Méthode par moralisation et triangularisation

Principe de la méthode :

Nous rappelons ici le principe de cette méthode, qu'on trouvera développée dans [39], [14], [2], [54], [22] et qui est en particulier implémentée dans de nombreux logiciels (Netica, Hugin, Genie, BNT, BKD)

Elle est fondée sur la notion de **clique** d'un graphe non orienté (I, H) ; on rappelle qu'une clique C est une partie de I complète et maximale, c'est-à-dire telle que :

- pour tout couple (c', c'') d'éléments distincs de C, $\{c', c''\} \in H$,
- si $C \neq I$, il existe un couple (c, d), où $c \in C$ et $d \notin C$, tel que $\{c, d\} \notin H$.

Rappelons qu'un **cycle**, dans un graphe non orienté, est une chaîne de sommets tous distincts, sauf l'origine et l'extrémité, qui coïncident ; on utilise alors la définition et le théorème suivants (théorème sur la preuve duquel nous reviendrons ci-dessous)

Définition 18 *Un graphe non-orienté (I, H) est dit **triangulé** (en anglais chordal) si et seulement si, pour tout cycle de longueur supérieure ou égale à 4, soit $(i_0, i_1, \ldots, i_{k-1}, i_k)$, (où $k \geq 4$, $i_0 = i_k$ et, pour tout $\ell \in \{0, \ldots, k-1\}$, $\{i_\ell, i_{\ell+1}\} \in H$), il existe au moins*

*une **corde**, c'est-à-dire un couple d'indices* (ℓ, ℓ') *tels que* $0 \leq \ell < \ell' \leq k$, $\ell' - \ell \geq 2$ *et* $\{i_\ell, i_{\ell'}\} \in H$.

Théorème 9 *L'ensemble des cliques d'un graphe triangulé peut être muni de la structure d'arbre de jonction.*

Décrivons, pour cette méthode, le calcul des P_K (où $K \in \mathcal{K}$) pour une probabilité P_I compatible avec un recouvrement $\mathcal{J}$; il se déroule en cinq étapes :

1. Moralisation : Construire sur I le graphe non orienté H dont les arêtes sont les paires $\{i, j\}$ telles que i et j appartiennent à un même élément J de $\mathcal{J}$ ou à un même élément $K \in \mathcal{K}$; tout $J \in \mathcal{J}$ et tout $K \in \mathcal{K}$ est alors une partie de I complète pour H ; H est appelé le graphe moral associé à $\mathcal{J} \cup \mathcal{K}$.
2. Triangularisation : Fabriquer sur I un graphe triangulé H^t contenant H ; chaque $J \in \mathcal{J}$ et chaque $K \in \mathcal{K}$, étant complet pour H, l'est aussi pour H^t et est donc contenu dans une clique de H^t.
3. Jonction : Ordonner l'ensemble $\mathcal{C}$ des cliques de H^t en une suite possédant la P.I.C. soit $(C_1, \ldots, C_k)$ et, s'il y a lieu ($k \geq 3$), choisir pour tout $i \in \{2, \ldots, k\}, \sigma(i)$ tel que $C_i \cap (\cup_{j<i} C_j) \subseteq C_{\sigma(i)}$.
4. Mise en forme de P_I : déduire, de l'écriture initiale de P_I (compatibilité avec $\mathcal{J}$) une écriture mettant en évidence sa compatibilité avec $\mathcal{C}$ (voir 3.1.2).
5. Calcul : calculer les P_C pour $C \in \mathcal{C}$ (voir 3.4) et en déduire les P_K pour tout $K \in \mathcal{K}$.

Remarques :

1. Le nom de "moralisation" pour la première étape provient de ce que, dans le cas des réseaux bayésiens, où $\mathcal{J} = \{i \cup p(i) \ ; \ i \in I\}$, et si on se limite à $\mathcal{K} = \mathcal{J}$, cette étape consiste à rajouter, aux liens $\{i, j\}$ tels que $j \in p(i)$, les liens $\{i', i''\}$ dans le cas où il existe j tel que $i' \in p(j)$ et $i'' \in p(j)$ ("on marie les parents d'un enfant commun")
2. L'étape 2 ("triangularisation") est évidemment toujours réalisable puisque le graphe complet construit sur I est triangulé ; mais celui-ci est sans intérêt pour les calculs, car conduisant à des sommations sur son unique clique, I lui-même. Le problème délicat de théorie des graphes est ici l'optimisation de cette étape (trouver un sur-graphe triangulé H^t de H ayant "le moins possible" d'arêtes et donc des cliques "aussi petites que possible") ; des algorithmes ont été fournis par Olmested et autres (voir [14]) à cet effet, ainsi que pour l'étape 3 (jonction) ; ils font usage de la notion de numérotage parfait (voir ci-dessous) ; pour leur étude, nous renvoyons le lecteur à [14], [16], [48].

Complément sur les graphes triangulés :

Le théorème (9) est en fait une part d'un résultat plus complet, qu'en peut trouver exposé par exemple dans [14] [41] et qui fait l'objet du théorème 10 ci-dessous. Ce théorème fait intervenir les notions suivantes, dans l'énoncé desquelles on considère un graphe non orienté (I, H).

Reprenons tout d'abord la définitions 5 du chapitre 2.

Définition 19 *Etant donné* 3 *parties disjointes A, B et S de I, où A et B sont non vides, on dit que S* ***sépare*** *A et B (où que S est un* ***séparateur*** *de A et B) si toute chaîne joignant un sommet a dans A à un sommet b dans B a au moins un sommet intermédiaire dans S*

On remarque que, si A et B sont séparés par S, il n'existe aucune arête entre un sommet de A et un sommet de B et, plus généralement, il n'existe aucune arête entre la fermeture connexe de A dans $I - S$ (plus petite partie connexe de $I - S$ contenant A) et la fermeture connexe de B dans $I - S$. Le cas particulier où S est vide exprime que A et B sont dans des composantes connexes distinctes de I.

Définition 20 *Une partition (A, B, S) de I est dite constituer une* ***décomposition propre*** *de I si $A \neq \emptyset$, $B \neq \emptyset$, S est complet et sépare A et B.*

Une décomposition propre de séparateur vide, $(A, B, \emptyset)$, signifie que I n'est pas connexe, A et B étant des unions de composantes connexes.

Définition 21 *(I, H) est dit* ***décomposable*** *s'il est complet ou bien s'il existe une décomposition propre de I, soit (A, B, S), telle que les graphes restreints à $A \cup S$ et $B \cup S$ soient décomposables.*

Cette définition récursive est valide car A et B sont non vides, de sorte de $A \cup S$ et $B \cup S$ ont strictement moins d'arêtes que I.

Définition 22 *Un numérotage des sommets de I est dit* ***parfait*** *si, pour tout sommet i, l'ensemble de ceux de ses voisins qui ont des numéros inférieurs à lui est complet.*

On peut alors énoncer :

Théorème 10 *Les* 5 *propriétés suivantes d'un graphe non orienté (I, H) sont équivalentes :*

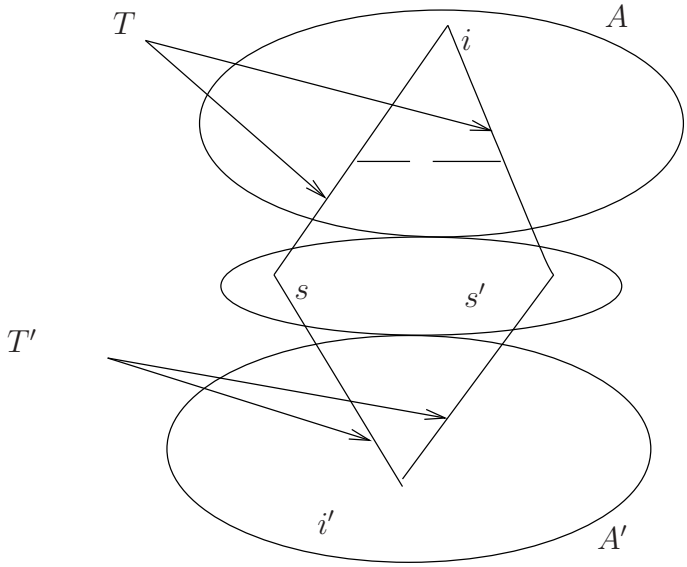
1. *(I, H) est triangulé,*
2. *pour tout couple (i, i') de sommets distincts situés dans une même composante connexe de I, tout séparateur non vide minimal (pour l'inclusion) de $\{i\}$ et $\{i'\}$ est complet,*
3. *(I, H) est décomposable,*
4. *l'ensemble des cliques peut être muni de la structure d'arbre de jonction,*
5. *il existe un numérotage parfait des sommets de I.*

Démonstration :

Nous donnons ici uniquement la démonstration de l'implication qui nous est nécessaire, et qui constitue le théorème 9 ci dessus ; à savoir (1) $\Longrightarrow$ (4) ; elle se décompose en les implications sucessives (1) $\Longrightarrow$ (2), (2) $\Longrightarrow$ (3) et (3) $\Longrightarrow$ (4). Pour les autres démonstrations nécessaires pour établir l'équivalence des 5 propriétés, nous renvoyons à [14].

(1) $\Longrightarrow$ **(2)**. Soit (I, H) triangulé, i et i' deux sommets distincts situés dans une même composante connexe et S un séparateur minimal (nécessairement non vide) de $\{i\}$ et $\{i'\}$.

Si S n'a qu'un sommet, il est complet.



Si non, soient s et s' deux éléments distincts de S ; notre but est de montrer que $\{s, s'\} \in H$. Comme S est minimal pour l'inclusion, $S - \{s\}$ n'est pas un séparateur de $\{i\}$ et $\{i'\}$, de sorte qu'il existe une chaîne reliant i et i' passant par s mais ne passant par aucun autre sommet de S ; de même il existe une chaîne reliant i' à i en passant par s' mais ne passant par aucun autre sommet de S ; concaténant ces deux chaînes, on obtient un cycle $(i, \ldots, s, \ldots, i', \ldots, s', \ldots, i)$ de longueur supérieure ou égale à 4 (car i, i', s et s' sont distincts) et que nous allons récrire $(s, \ldots, i', \ldots, s', \ldots, i, \ldots, s)$; seuls s et s' appartiennent à S, et les deux chaînes délimitées par s et s', soit T et T', sont contenues respectivement dans la composante connexe de i dans $I - S$ (notons la A) et dans la composante connexe de i' dans $I - S$ (notons la A').

Le graphe étant triangulé, ce cycle peut être raccourci à l'aide de cordes, qui ne peuvent pas relier un sommet de A à un sommet de A' (voir la remarque suivant la définition de la séparation).

Tout raccourcissement effectué en reliant deux sommets de T (voir par exemple la corde marquée – – sur la figure), ou deux sommets de T', ou encore s (ou s') à un sommet de A (ou A'), crée un cycle strictement plus court que le précédent, qui reste de la forme $(s,$ chaîne dans $A, s',$ chaîne dans $A', s)$; après avoir procédé à tous les raccourcissements de ce type possible, on arrive nécessairement à un cycle de longueur 4 (s, j, s', j', s) où $j \in A$ et $j' \in A'$; ce cycle possède encore une corde, qui ne peut pas être $\{j, j'\}$ et qui donc est $\{s, s'\}$.

(2) $\Longrightarrow$ **(3)**. Supposons que, pour tout couple (i, i') de sommets distincts de I situés

dans une même composante connexe, tout séparateur minimal de $\{i\}$ et $\{i'\}$ est complet. Nous allons mettre en évidence, si I n'est pas complet, une décomposition propre (A, B, S) telle que les graphes restreints à $A \cup S$ et $B \cup S$ soient décomposables.

La démonstration se fait par récurrence sur le cardinal de I.

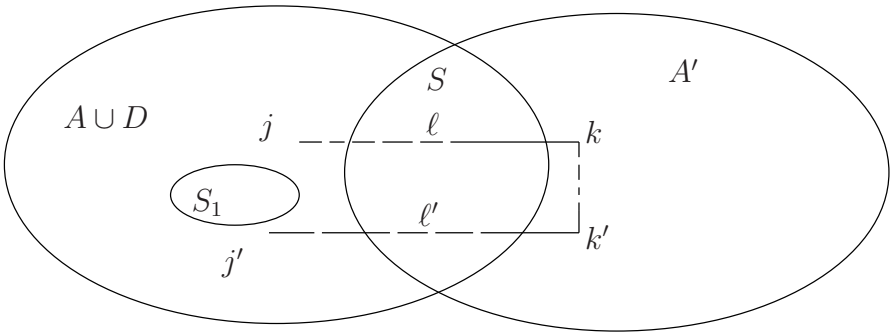
Le plus petit cardinal de I pour lequel l'hypothèse ait un sens est 3, avec le graphe $i - s - i'$, où le séparateur de $\{i\}$ et $\{i'\}$ est $\{s\}$; ce graphe est décomposable, car il admet la décomposition $(\{i\}, \{i'\}, \{s\})$ où $\{i, s\}$ et $\{i', s\}$ sont eux-mêmes décomposables car complets.

Supposons donc satisfaite la propriété à démontrer pour $\text{card}(I) \leq n$; soit I de cardinal $n + 1$, vérifiant la propriété (2), et non complet.

Soit i et i' deux sommets non adjacents ; soit S un séparateur minimal de $\{i\}$ et $\{i'\}$ et soit respectivement A et A' les fermetures connexes de i et i' dans $I - S$; soit enfin $D = I - (A \cup A' \cup S)$ (éventuellement vide). S étant complet, $(A \cup D, A', S)$ forme une décomposition propre de I.

Il reste à montrer que les restrictions de H aux parties, toutes deux de cardinal inférieur ou égal à n, $A \cup D \cup S$ et $A' \cup S$ sont décomposables, ce qui, en vertu de l'hypothèse de récurrence, est assuré si la propriété (2) est assurée pour chacune de ces parties.

Etablissons cette propriété (2) pour $A \cup D \cup S$ (la démonstration est identique pour $A' \cup S$). Soit donc j et j' deux sommets distincts de $A \cup D \cup S$, et S_1 un séparateur minimal de $\{j\}$ et $\{j'\}$, pour le graphe restreint à $A \cup D \cup S$; nous allons montrer que S_1 est aussi un séparateur de $\{j\}$ et $\{j'\}$ pour le graphe initial sur I, et alors il sera nécessairement minimal et donc complet par hypothèse. En effet, supposons que S_1 n'est pas un séparateur de $\{j\}$ et $\{j'\}$ dans $I = (A \cup D \cup S) \cup A'$; il existerait alors une chaîne de j à j', évitant S_1 et ayant nécessairement au moins un sommet intermédiaire dans A' ; plus précisément, il existerait deux couples de sommets successifs sur cette chaîne, soit (ℓ, k) et (k', ℓ'), tels que $\ell \neq \ell'$, $\{\ell, k\} \in H$, $\{k', \ell'\} \in H$, $\ell \in A \cup D \cup S$, $k \in A'$, $k' \in A'$, $\ell' \in A \cup D \cup S$. S séparant $A \cup D$ de A', il est impossible que $\ell \in A \cup D$ et $\ell' \in A \cup D$; donc $\ell \in S$ et $\ell' \in S$, et comme S est complet, on pourrait raccourcir cette chaîne en utilisant l'arête $\{\ell, \ell'\}$; on aurait alors construit une chaîne reliant j à j' dans $A \cup D \cup S$ en évitant S_1, ce qui est contraire au fait que S_1 sépare $\{j\}$ et $\{j'\}$ pour la restriction à $A \cup D \cup S$.



(3) $\Longrightarrow$ **(4)**. Supposons que H est décomposable ; nous allons construire une structure d'arbre de jonction sur l'ensemble de ses cliques.

La démonstration se fait par récurrence sur le cardinal de I.

La propriété est évidente si card$(I)= 1$, où H est décomposable car complet et l'unique clique est I lui-même.

Supposons la propriété démontrée pour card$(I) \leq n$.

Soit I de cardinal $n+1$, décomposable. S'il est complet, la propriété est évidente puisque l'unique clique est I lui-même. Si non, soit (A_1, A_2, S) une décomposition propre de I, les restrictions de H à $A_1 \cup S$ et $A_2 \cup S$ étant décomposables.

Il résulte de la décomposabilité qu'une clique ne peut contenir à la fois des éléments de A_1 et des éléments de A_2, puisqu'entre deux tels éléments ne peut pas exister d'arête. Donc toute clique C vérifie l'une au moins des deux propriétés $C \cap A_1 = \emptyset$ et $C \cap A_2 = \emptyset$, autrement dit l'ensemble $\mathcal{C}$ des cliques vérifie $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$ où

$$\mathcal{C}_1 = \{C \in \mathcal{C} \ ; \ C \subseteq A_1 \cup S\} = \{C \in \mathcal{C} \ ; \ C \cap A_2 = \emptyset\}$$

et

$$\mathcal{C}_2 = \{C \in \mathcal{C} \ ; \ C \subseteq A_2 \cup S\} = \{C \in \mathcal{C} \ ; \ C \cap A_1 = \emptyset\}.$$

On distingue deux cas :

(a) S est lui même une clique ; alors $\mathcal{C}_1 \cap \mathcal{C}_2 = \{S\}$ et on construit un arbre sur $\mathcal{C}$ en faisant l'union des arêtes des arbres de jonction déjà connus sur $\mathcal{C}_1$ et $\mathcal{C}_2$ en vertu de l'hypothèse de récurrence.

(b) S n'est pas une clique, car non maximal parmi les parties complètes ; alors $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ et il existe $C_1 \in \mathcal{C}_1$ et $C_2 \in \mathcal{C}_2$, distincts et tels que $S \subseteq C_1$ et $S \subseteq C_2$; on construit un arbre sur $\mathcal{C}$ en faisant l'union des arêtes des arbres de jonction sur $\mathcal{C}_1$ et $\mathcal{C}_2$ et d'une arête "nouvelle", $\{C_1, C_2\}$.

Il reste à vérifier que l'arbre ainsi construit est bien un arbre de jonction, c'est-à-dire (voir théorème 8) que, pour tout $i \in I$, l'ensemble $\mathcal{C}_i$ des cliques auxquelles appartient i est connexe pour cet arbre.

Si $i \in A_1$ (resp A_2) toutes les cliques auxquelles il appartient sont dans $\mathcal{C}_1$ (resp $\mathcal{C}_2$) et donc leur ensemble est connexe car il l'est dans l'arbre de jonction connu sur $\mathcal{C}_1$ (resp $\mathcal{C}_2$).

Si $i \in S$, on doit distinguer selon les cas (a) et (b) :
(a) S est une clique : alors toutes les cliques auxquelles appartient i sont S lui-même et éventuellement des cliques de $\mathcal{C}_1$ et $\mathcal{C}_2$, constituant des sous-arbres qui se raccordent en S en un sous-arbre de $\mathcal{C}$.
(b) S n'est pas une clique : alors les cliques auxquelles appartient i sont des cliques de $\mathcal{C}_1$, dont C_1, et des cliques de $\mathcal{C}_2$, dont C_2 ; on a ainsi deux sous-arbres qui se raccordent, par l'arête créée entre C_1 et C_2, en un sous-arbre de $\mathcal{C}$.

■

3.3.2 Méthode par construction récurrente d'une suite recouvrante propre P.I.C.

Contexte :

Soit donné un recouvrement fini de I, $\mathcal{F}$. Notre but est de construire pas à pas une suite recouvrante propre, $(C_1, \ldots, C_p)$, qui possède la propriété d'intersection courante (P.I.C.) et telle que tout $F \in \mathcal{F}$ soit contenu dans au moins l'un des C_j $(1 \leq j \leq p)$.

L'algorithme exige que soit adopté sur I un numérotage, de sorte qu'on va noter $I = \{1, \ldots, n\}$; on considère pour chaque F ($F \in \mathcal{F}$) son plus petit élément pour ce numérotage et on fait l'union des parties F qui ont même plus petit élément ; ordonnant ces unions selon leurs plus petits éléments, on obtient une suite recouvrante $(Q_1, \ldots, Q_r)$ (où $r \leq n$) telle que, si on note α_k (où $1 \leq k \leq r$) le plus petit élément de Q_k, la suite $(\alpha_1, \ldots, \alpha_r)$ est strictement croissante ; la condition, imposée par l'énoncé de notre problème, que tout F soit inclus dans l'un des C_j équivaut évidemment à l'inclusion de chaque Q_k dans l'un des C_j.

La suite $(Q_1, \ldots, Q_r)$ n'est pas nécessairement propre et on ne change rien au problème posé en la "nettoyant" c'est-à-dire en en retirant tout terme qui soit contenu strictement dans un autre ; remarquons à ce sujet que, par construction, l'inclusion $Q_j \subseteq Q_{j'}$ ne peut pas avoir lieu avec $j < j'$, car alors on aurait $\alpha_j \in Q_j$ mais $\alpha_j \notin Q_{j'}$, puisque tous les éléments de $Q_{j'}$ sont supérieurs ou égaux à $\alpha_{j'}$ qui est strictement supérieurs à α_j ; le nettoyage consiste donc à éliminer tous les Q_j vérifiant $\exists\ j' < j \quad Q_j \subset Q_{j'}$. Mais les algorithmes que nous proposons n'exigent pas que cette suite $(Q_1, \ldots, Q_r)$ soit propre.

Le problème que doit résoudre l'algorithme s'énonce alors :

Soit donné une suite recouvrante $(Q_j)_{1 \leq j \leq r}$ de parties non vides de $I = \{1, \ldots, n\}$ telle que, si on note $\alpha_j = \inf(Q_j)$, la suite $(\alpha_1, \ldots, \alpha_r)$ est strictement croissante. On doit construire une suite recouvrante propre $(C_1, \ldots, C_p)$ qui posséde la P.I.C. et telle que, pour tout j $(1 \leq j \leq r)$ il existe k $(1 \leq k \leq p)$ vérifiant $Q_j \subseteq C_k$.

On remarque que, nécessairement , $\alpha_1 = 1$.

Dans le cas des réseaux bayésiens, où $\mathcal{F}$ contient $\mathcal{J} = \{i \cup p(i)\ ; i \in I\}$, on réalise natu-

rellement la construction de la suite $(Q_1, \ldots, Q_r)$ de la manière suivante :
- on adopte sur I un ordre antihiérarchique relativement à la structure de graphe orienté sans circuits : pour tout i, ses parents sont supérieurs à lui (autrement dit, I étant $\{1, \ldots, n\}$, l'ordre $(n, n-1, \ldots, 1)$ est hiérarchique) ;
- on définit, pour tout i, $\tilde{Q}_i$ comme l'union de i, de $p(i)$ et, s'il y a lieu, de celles des parties appartenant à $\mathcal{F}$, non de la forme $i \cup p(i)$, dont i est le plus petit élément ; remarquons que ceci pourrait s'énoncer aussi de la manière suivante : $\tilde{Q}_i = i \cup p'(i)$ où p' est le réseau bayésien déduit du réseau bayésien donné en rajoutant les liens (j, i), chaque fois que j, différent de i, appartient à un F, non de la forme $i \cup p(i)$, et dont i est le plus petit élément ;
- facultativement, on nettoie la suite $(\tilde{Q}_1, \ldots, \tilde{Q}_n)$.

Nous allons proposer deux versions de l'algorithme.

La première version (A) s'effectue en deux étapes : on construit d'abord une suite $(C_1^\star, \ldots, C_r^\star)$ de parties non vides de I, de longueur égale à celle de la suite donnée $(Q_1, \ldots, Q_r)$, qui posséde la P.I.C. et telle que, pour tout j, il existe k vérifiant $Q_j \subseteq C_k^\star$; celle-ci n'est pas nécessairement propre mais n'admet d'inclusions $C_{j'}^\star \subseteq C_j^\star$ que si $j' < j$ dans un deuxième temps, s'il y a lieu, on la "nettoie" en en retirant des éléments inclus dans d'autres, d'où une suite $(C_1, \ldots, C_p)$ qui évidemment vérifie encore $\forall\ j\ \ \exists\ k\ Q_j \subseteq C_k$; nous verrons que ce "nettoyage" peut bien être effectué en respectant la propriété d'intersection courante.

La seconde version de l'algorithme (B), qui est à notre connaissance originale, construit d'emblée la suite propre $(C_1, \ldots, C_p)$, en éliminant au fur et à mesure de leur construction les parties de I qui seraient incluses dans d'autres déjà construites.

L'une comme l'autre des deux versions construit en fait successivement des parties, que nous noterons $D_1, \ldots, D_r$ (dans la version A) ou $D_1^\#, \ldots, D_p^\#$ (dans la version B) telles que la suite inversée $(D_r, \ldots, D_1)$ ou $(D_p^\#, \ldots, D_1^\#)$ (selon la version) soit la suite $(C_1^\star, \ldots, C_r^\star)$ ou $(C_1, \ldots, C_p)$ cherchée ; en d'autres termes la suite construite vérifie la propriété d'intersection courante inversée (P.I.C.I.), c'est à dire, selon la version :

$$\forall\ s \in \{1, \ldots, r-1\}\ \exists\ t \in \{s+1, \ldots, r\}\ D_s \cap (\cup_{u>s} D_u)\ \subseteq D_t$$

ou

$$\forall\ s \in \{1, \ldots, p-1\}\ \exists\ t \in \{s+1, \ldots, p\}\ D_s^\# \cap (\cup_{u>s} D_u^\#)\ \subseteq D_t^\#$$

Version A de l'algorithme :

Phase 1 : Construction d'une suite non nécessairement propre

L'algorithme (adapté de [14], [5], [1]) a exactement r pas ; à chaque pas j, il construit une partie D_j, contenant Q_j et contenue dans $\cup_{\ell \geq j} Q_\ell$, d'où il résulte que $\cup_{\ell > j} D_\ell = \cup_{\ell > j} Q_\ell$.

Cette suite $(D_1, \ldots, D_r)$ vérifie la P.I.C.I. ; elle n'est pas nécessairement propre mais, par construction, ses éléments sont tous non vides et distincts et les inclusions $D_{j'} \subseteq D_j$ ne peuvent avoir lieu que si $j' > j$.

Notons, pour tout j, $K_j = \{\alpha_j, \ldots, \alpha_{j+1} - 1\}$ (dans le cas particulier où $r = n$ et donc, pour tout $j \in \{1, \ldots, n\}$, $\alpha_j = j$, on a $K_j = \{j\}$).

Rappelons que la P.I.C.I. est automatiquement assurée pour l'avant-dernier terme de toute suite de parties ; afin que la suite $(D_j)_{1 \leq j \leq r}$ vérifie la P.I.C.I., l'algorithme met en mémoire, pour tout j tel que $1 \leq j \leq r-2$, $E_j = D_j \cap (\cup_{\ell > j} Q_\ell)$. Une condition suffisante pour qu'une partie D_k (avec $k > j$) assure la P.I.C.I. pour D_j (c'est à dire vérifie $D_j \cap (\cup_{\ell > j} D_\ell) \subseteq D_k$) est que $E_j \subseteq D_k$. Si $E_j = \emptyset$, cela signifie que $D_j \cap (\cup_{\ell > j} D_\ell) = \emptyset$ et donc la P.I.C.I. est trivialement assurée pour D_j ; si non, la P.I.C.I. sera naturellement assurée pour D_j en incluant dans D_k, outre Q_k, E_j dès que $K_k \cap E_j \neq \emptyset$ (autrement dit quand $\alpha_k \leq \inf(E_j) < \alpha_{k+1}$, avec la convention $\alpha_{r+1} = r+1$), ce qui n'introduit bien dans D_k que des éléments supérieurs ou égaux à α_k, donc contenus dans $\cup_{\ell \geq k} Q_\ell$.

Voici donc la description de l'algorithme.

Données : suite recouvrante $(Q_1, \ldots, Q_r)$ de parties non vides de $\{1, \ldots, n\}$;
$\alpha_i = \inf(Q_i)$; $\alpha_{r+1} = r+1$;
la suite $(\alpha_1, \ldots, \alpha_r)$ est strictement croissante
Objets construits : suites de parties de I : $(D_1, \ldots, D_r)$, $(E_1, \ldots, E_{r-2})$; pour tout j tel que $E_j \neq \emptyset$, $e_j = \inf(E_j)$
Initialisation : $D_1 = Q_1$; $E_1 = Q_1 - \{1\}$
Pas courant : $j \geq 2$:

$$D_j = Q_j \cup [\cup_{k<j,\ \alpha_j \leq e_k < \alpha_{j+1}} E_k]$$

$$\text{si } \ D_j = \emptyset, E_j = \emptyset \ ; \quad \text{si } \ D_j \neq \emptyset, \ E_j = D_j \cap (\cup_{k>j} Q_k)$$

Fin de l'algorithme : $j = n$.

Exemple

Soit le réseau bayésien ci-dessous, muni d'un ordre antihiérarchique.

On veut calculer les lois $P_{i \cup p(i)}$ (pour tout $i \in I$).

On définit donc les parties :

$$Q_1 = \{1,2,3,5\},\ Q_2 = \{2,3\},\ Q_3 = \{3,7\},\ Q_4 = \{4,5,6\}$$

$$Q_5 = \{5,7\},\ Q_6 = \{6,8\},\ Q_7 = \{7,9\},\ Q_8 = \{8,9\},\ Q_9 = \{9\}.$$

Nous reprendrons cet exemple en 3.4.8 ci-dessous pour montrer comment les calculs de probabilités se déroulent après avoir obtenu une suite recouvrante propre possédant la P.I.C..

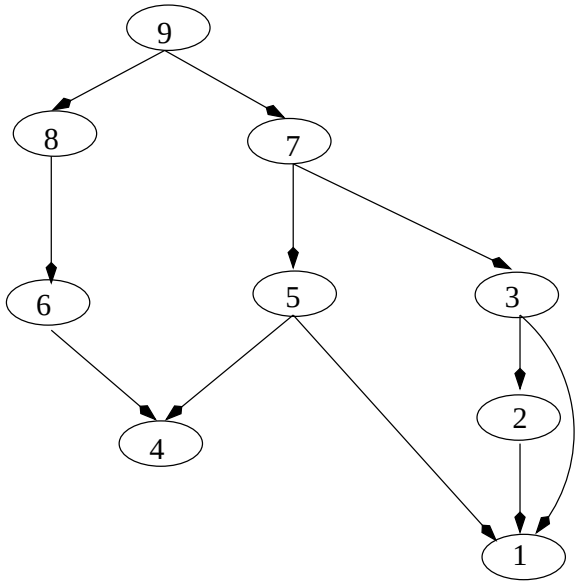


FIG. 3.3 – *Exemple* 1.

Si on ne procéde pas à un nettoyage de la suite, on a le déroulement suivant de l'algorithme

Pas i	D_i	E_i
1	$D_1 = Q_1 = \{1,2,3,5\}$	$E_1 = \{2,3,5\}$
2	$D_2 = Q_2 \cup E_1 = \{2,3,5\}$	$E_2 = \{3,5\}$
3	$D_3 = Q_3 \cup E_2 = \{3,5,7\}$	$E_3 = \{5,7\}$
4	$D_4 = Q_4 = \{4,5,6\}$	$E_4 = \{5,6\}$
5	$D_5 = Q_5 \cup E_3 \cup E_4 = \{5,6,7\}$	$E_5 = \{6,7\}$
6	$D_6 = Q_6 \cup E_5 = \{6,7,8\}$	$E_6 = \{7,8\}$
7	$D_7 = Q_7 \cup E_6 = \{7,8,9\}$	$E_7 = \{8,9\}$
8	$D_8 = Q_8 \cup E_7 = \{8,9\}$	$E_8 = \{9\}$
9	$D_9 = Q_9 \cup E_8 = \{9\}$	$E_9 = \emptyset$

TAB. 3.1 – Déroulement de l'algorithme version A sans nettoyage préalable.

On remarque que la suite obtenue n'est pas propre (on a $D_2 \subset D_1$, $D_8 \subset D_7$, $D_9 \subset D_8$).
Si on procéde en préalable à un nettoyage de la suite donnée, on conserve

$$Q_1 = \{1,2,3,5\}, \quad Q_3 = \{3,7\}, \; Q_4 = \{4,5,6\}, \; Q_5 = \{5,7\}$$

$$Q_6 = \{6,8\}, \; Q_7 = \{7,9\}, \; Q_8 = \{8,9\}$$

d'où le déroulement de l'algorithme :

Pas j	K_j	D_j	E_j
1	$1,2$	$D_1 = Q_1 = \{1,2,3,5\}$	$E_1 = \{2,3,5\}$
2	3	$D_2 = Q_3 \cup E_1 = \{2,3,5,7\}$	$E_2 = \{5,7\}$
3	4	$D_3 = Q_4 = \{4,5,6\}$	$E_2 = \{5,6\}$
4	5	$D_4 = Q_5 \cup E_2 = \{5,6,7\}$	$E_4 = \{6,7\}$
5	6	$D_5 = Q_6 \cup E_4 = \{6,7,8\}$	$E_5 = \{7,8\}$
6	7	$D_6 = Q_7 \cup E_5 = \{7,8,9\}$	
7	$8,9$	$D_7 = Q_8 = \{8,9\}$	

TAB. 3.2 – Déroulement de l'algorithme version A avec nettoyage préalable.

On remarque que, ici encore, la suite obtenue n'est pas propre (on a $D_7 \subset D_6$)

Phase 2 : Nettoyage de la suite obtenue à la phase 1 :

En inversant la suite $(D_1, \ldots, D_n)$ obtenue par l'algorithme précédent, on obtient une suite recouvrante $(C_1^\star, \ldots, C_n^\star)$ à éléments tous distincts, vérifiant la P.I.C., et telle que, pour tout couple (i,j) tel que $1 \leq i < j \leq n$, il est faux que $C_i^\star \supset C_j^\star$. Si cette suite n'est pas propre, la possibilité de la "nettoyer", c'est-à-dire d'en retirer tout C_i pour lequel il existe j $(> i)$ vérifiant $C_i^\star \subset C_j^\star$ est donnée par le théorème suivant

Théorème 11 *Soit $\mathcal{C} = (C_1, \ldots, C_k)$ une suite de parties non vides distinctes de I possédant les propriétés suivantes :*

1. *elle est recouvrante*
2. *elle vérifie la P.I.C.*
3. *les seuls couples (j,j') pour lesquels on puisse avoir l'inclusion $C_j \subset C_{j'}$ vérifient $j < j'$,*
4. *il existe au moins un couple (j,j') tel que $j < j'$ et $C_j \subset C_{j'}$.*

Soit p le plus petit élément de $\{1, \ldots, k\}$ tel que l'ensemble des j strictement inférieurs à p et vérifiant $C_j \subset C_p$ soit non vide ; notons $(v_1, \ldots, v_s)$ (où $s \geq 1$) la suite croissante de ces indices j $(< p)$ tels que $C_j \subset C_p$; alors la suite $\mathcal{C}' = (C'_1, \ldots, C'_{k-s})$ déduite de $\mathcal{C}$ en substituant C_p à C_{v_1} et en en retirant le terme d'indice p ainsi que, si $s > 1$, les termes d'indices $v_2, \ldots, v_s$, vérifie les propriétés (1), (2) *et* (3).

Démonstration

$\mathcal{C}'$ est recouvrante car, dans le passage de $\mathcal{C}$ à $\mathcal{C}'$, les seules suppressions d'éléments sont celles de parties contenues dans C_p qui, elle, subsiste (avec déplacement à la place v_1).

Le fait que $\mathcal{C}'$ posséde la propriété (3) résulte immédiatement de ce que, dans le passage de $\mathcal{C}$ à $\mathcal{C}'$, le seul déplacement d'un élément de la suite initiale est celui de C_p, qui vient occuper une place d'indice inférieur à celui de sa place initiale.

Il reste donc à établir la propriété d'intersection courante pour
$\mathcal{C}' = (C_1, \ldots, C_{v_1-1}, C_p, C_{v_1+1}, \ldots, C_{v_2-1}, C_{v_2+1}, \ldots, C_{v_s-1}, C_{v_s+1}, \ldots, C_{p-1}, C_{p+1}, \ldots, C_k)$.

La suite des indices (tous distincts) des termes de $\mathcal{C}'$ est :
$\mathcal{S} = (1, \ldots, v_1 - 1,\ p,\ v_1 + 1, \ldots, v_2 - 1, v_2 + 1, \ldots, v_s - 1, v_s + 1, \ldots, p - 1, p + 1, \ldots, k)$;
pour tout $j \in \mathcal{S}$ notons $\mathcal{S}_j^-$ (resp. $\mathcal{S}_j^+$) l'ensemble des indices classés strictement avant (resp. après) j dans cette suite.

Prouver que cette suite vérifie la propriété d'intersection courante revient à montrer que, pour tout j autre que 1 ou 2 dans cette suite, il existe $i \in \mathcal{S}_j^-$ tel que $C_j \cap (\cup_{\ell \in \mathcal{S}_j^-} C_\ell) \subset C_i$.

Nous distinguons les quatre cas suivants (dont certains, sauf le numéro 3, peuvent être vides)

1. $j \in \mathcal{S}_{v_1}^-$; dans ce cas $\mathcal{S}_{v_1}^- = (1, \ldots, j-1)$ et la P.I.C. est évidente puisque aucune modification n'a été opérée sur $(C_1, \ldots, C_j)$.
2. $j \in \mathcal{S}_{p+1}^+$; dans ce cas $\cup_{\ell \in \mathcal{S}_j^-} C_\ell = \cup_{m<j} C_m$ (car les seules modifications, dans l'ensemble initial $\{C_1, \ldots, C_{j-1}\}$, ont consisté en la suppression de parties contenues dans C_p, qui appartient à cet ensemble ; alors, par hypothèse, $C_j \cap (\cup_{\ell \in \mathcal{S}_j^-} C_\ell) \subseteq C_h$, où $1 \leq h \leq j-1$; si h n'est pas l'un des v_u $(1 \leq u \leq s)$, il appartient à $\mathcal{S}_j^-$ et la propriété d'intersection courante est établie ; si $h = v_u$, on a

   $$C_j \cap (\cup_{\ell \in \mathcal{S}_j^-} C_\ell) \subseteq C_{v_u} \subset C_p$$

   et la P.I.C. est encore valable, car $p \in \mathcal{S}_j^-$.
3. $j = p$; dans ce cas $\mathcal{S}_j^- = \{1, \ldots, v_1 - 1\}$ et on doit montrer l'existence de $h (\in \{1, \ldots, v_1 - 1\})$ tel que $S'_p = C_p \cap (C_1 \cup \ldots \cup C_{v_1 - 1}) \subseteq C_h$.
   Notons $S_p = C_p \cap (C_1 \cup \ldots \cup C_{p-1})$; nous allons tout d'abord établir que $S_p = C_{v_1}$. En effet d'une part $S_p \supseteq C_p \cap C_{v_1} = C_{v_1}$; d'autre part, en vertu de la P.I.C. de $\mathcal{C}$, il existe $k < p$ tel que $S_p \subseteq C_k$; on a donc $C_{v_1} \subseteq C_k$. Or $C_{v_1} \subset C_p$. Il en résulte que $k = v_1$; en effet on ne peut avoir ni $k < v_1$, car ceci serait contraire à l'hypothèse (3), ni $v_1 < k < p$, car alors on aurait l'inclusion stricte $C_{v_1} \subset C_k$ (car $C_{v_1} \neq C_k$ par hypothèse) et p est par définition le plus petit élément tel qu'il existe ℓ vérifiant $C_\ell \subset C_p$.
   De l'égalité $k = v_1$ résulte que $C_{v_1} \subseteq S_p \subseteq C_{v_1}$, donc $S_p = C_{v_1}$.
   Alors

   $$S'_p = C_p \cap (C_1 \cup \ldots \cup C_{v_1 - 1}) \subseteq C_p \cap (C_1 \cup \ldots \cup C_{p-1}) = S_p = C_{v_1},$$

   et donc, en utilisant à nouveau $C_{v_1} \subset C_p$,

   $$S'_p = S'_p \cap C_{v_1} = [C_p \cap (C_1 \cup \ldots \cup C_{v_1 - 1})] \cap C_{v_1} = C_{v_1} \cap (C_1 \cup \ldots \cup C_{v_1 - 1}) = S_{v_1},$$

   d'où l'existence de $h, \in \{1, \ldots, v_1 - 1\}$, tel que $S'_p \subseteq C_h$.
4. $j \in \mathcal{S}_{v_1}^+ \cap \mathcal{S}_p^-$, autrement dit $v_1 < j < p$ et $j \notin \{v_2, \ldots, v_s\}$; dans ce cas on considère

   $$S'_j = C_j \cap (\cup_{l \in \mathcal{S}_j^-} C_\ell)\ ;$$

or pour tout $v_i \leq j-1,\ C_{v_i} \subset C_p$; $p > j$ mais $p \in \mathcal{S}_j^-$; alors

$$\cup_{l \in \mathcal{S}_j^-} C_\ell = \cup_{h<j} C_h \cup (C_p - C_{v_1})$$

d'où,

$$\begin{aligned} S'_j &= [C_j \cap (\cup_{l \in \mathcal{S}_j^-} C_\ell)] \cup [C_j \cap (C_p - C_{v_1})] \\ &= S_j \cup [C_j \cap (C_p - C_{v_1})] \ , \end{aligned}$$

et, comme $C_j \cap C_p \subseteq S_p = C_{v_1}$ (égalité établie dans le cas (3) ci-dessus), on a $C_j \cap (C_p - C_{v_1}) = \emptyset$, et donc $S'_j = S_j$; or, en vertu de la P.I.C. de la suite $\mathcal{C}$, il existe $h < j$ tel que $S_j \subseteq C_h$, ce qui établit la validation de la P.I.C. dans la suite $\mathcal{C}'$ (le cas où h est l'un des v_u se résolvant comme dans le cas (2) ci-dessus).

■

Par application réitérée de ce théorème on arrive à une suite recouvrante propre contenue dans $\mathcal{C}$, qui possède la P.I.C. (la propriété 3 dans l'énoncé du théorème était nécessaire pour cela et est conservée dans la suite propre).

Récapitulons cet algorithme :

Donnée : suite recouvrante $(C_1^\star, \ldots, C_n^\star)$ possédant la P.I.C. et telle que $C_i^\star \subset C_j^\star$ implique $i < j$
Objet courant : suite recouvrante $(C_1, \ldots, C_k)$ possédant les mêmes propriétés que la suite donnée
Objet construit : suite recouvrante propre $(C_1, \ldots, C_k)$ à l'arrêt de l'algorithme
Initialisation : $k = n$, $(C_1, \ldots, C_k) = (C_1^\star, \ldots, C_n^\star)$
Pas courant : Déterminer s'il existe, $p = \min\{j \ ; \ \exists\ i < j \ C_i \subset C_j\}$
Si p n'existe pas, fin de l'algorithme
Si p existe, soit $K = \{i < p \ ; \ C_i \subset C_p\}$ et $k = \min K$;
- retirer de la suite, s'ils existent, les C_i tels que $i \in K$ et $i \neq k$
- placer C_p à la place de C_k

L'algorithme s'arrête bien car à chaque pas la longueur de la suite diminue strictement.

Remarque : Tout autre algorithme de nettoyage, appliqué à la suite $(C_1^\star, \ldots, C_n^\star)$ conduirait à un ensemble de parties identique à $\{C_1, \ldots, C_k\}$ tel qu'il est constitué à l'arrêt de l'algorithme ci-dessus, et que donc aurait la structure d'arbre de jonction. Mais l'algorithme que nous proposons a l'avantage de l'obtenir muni d'un ordre qui assure la P.I.C.

Exemple : la suite $(C_1^\star, \ldots, C_9^\star)$, trouvée dans l'exemple précédent est la suivante :

$$C_1^\star = \{9\},\ C_2^\star = \{8,9\},\ C_3^\star = \{7,8,9\},\ C_4^\star = \{6,7,8\},\ C_5^\star = \{5,6,7\},$$

$$C_6^\star = \{4,5,6\},\ C_7^\star = \{3,5,7\},\ C_8^\star = \{2,3,5\},\ C_9^\star = \{1,2,3,5\}\ ,$$

nous remarquons que $C_1^\star \subset C_2^\star \subset C_3^\star$ et $C_8^\star \subset C_9^\star$; par application de l'algorithme ci-dessus, nous trouvons que la suite propre vérifiant la P.I.C., $(C_1, \ldots, C_6)$, correspondant à la suite $(C_1^\star, \ldots, C_9^\star)$, est la suivante :

$$C_1 = \{7,8,9\},\ C_2 = \{6,7,8\},\ C_3 = \{5,6,7\},$$

$$C_4 = \{4,5,6\},\ C_5 = \{3,5,7\},\ C_6 = \{1,2,3,5\}\ ,$$

On reconnaitra (l'indexage de sommets étant différent) la suite utilisée ci-dessous dans les calculs en 3.4.8.

Version B de l'algorithme :

Principe et justification de l'algorithme :

L'algorithme construit successivement r parties $D_1, \ldots, D_r$ de telle sorte que, après éventuellement élimination, au fur et à mesure de leur construction, de certaines d'entre elles, on obtienne une suite propre $(D_1^\#, \ldots, D_p^\#)(= (D_{s_1}, \ldots, D_{s_p}))$, possédant la P.I.C.I. et telle que tout Q_j $(1 \leq j \leq r)$ soit contenu dans au moins un $D_k^\#$ $(1 \leq k \leq p)$.

Les termes de la suite $(D_1, \ldots, D_r)$ non éliminés seront dits "validés" ; un D_j validé pourra être noté $\tilde{D}_j$; ainsi la notation, qui nous sera souvent utile, $\cup_{\ell > j} \tilde{D}_\ell$ désigne l'union de ceux des D_ℓ d'indices strictement supérieurs à j qui sont validés, et eux-seul ; par exemple, si $r = 5,\ p = 3$ et $(s_1, s_2, s_3) = (1, 2, 5)$, on a :

$$\cup_{\ell > 2} \tilde{D}_\ell = \tilde{D}_5\ (= D_5).$$

L'algorithme a exactement r pas.

Il est initialisé en construisant et validant $\tilde{D}_1 = Q_1$.

Au pas j, > 1, on a en mémoire $j' < j$, qui est le dernier pas, avant j, où a été construite une partie validée $(\tilde{D}_{j'})$; on construit une partie D_j de I (le détail de cette construction sera donné ultérieurement) qui contient Q_j et qui est contenue dans $\cup_{\ell \geq j'+1} Q_\ell$.

On valide D_j si et seulement si elle n'est pas vide et n'est pas contenue dans une partie déjà construite et validée ; cette opération suffit à assurer que l'on obtient une suite recouvrante propre ; remarquons en effet qu'il n'est pas possible qu'une partie $\tilde{D}_j$ $(j > 1)$ contienne une partie $\tilde{D}_k$ antérieurement construite et validée $(1 \leq k \leq j' < j)$ car $\alpha_k \in \tilde{D}_k$ alors que $\tilde{D}_j \subseteq \cup_{\ell \geq j'+1} Q_\ell$, et donc $\alpha_k \notin \tilde{D}_j$

Deux conséquences des conditions posées sur les D_j sont les suivantes :

- pour tout j, il existe bien un $\tilde{D}_k$ tel que $Q_j \subseteq \tilde{D}_k$ (c'est $\tilde{D}_j$ lui-même si celui-ci est validé, ou un $\tilde{D}_k$ antérieurement construit et contenant D_j si celui-ci n'est pas validé) ;

- pour tout j, si D_j est validée, $\cup_{k \geq j} \tilde{D}_k = \cup_{\ell \geq j'+1} Q_k$.

A l'entrée de chaque pas $j \geq 2$ on a en mémoire l'ensemble $\mathcal{M}_j$, éventuellement vide, des $\ell < j$ tels que D_ℓ ait été validé, mais que la propriété P.I.C.I. n'ait pas encore été assurée

pour $\tilde{D}_\ell$, ce qui signifie que ce n'est pas avec un indice h tel que $\ell < h < j$ que l'on peut déjà affirmer que $\tilde{D}_\ell \cap (\cup_{k>\ell} \tilde{D}_k) \subseteq \tilde{D}_h$.

Si $\mathcal{M}_j \neq \emptyset$, on a aussi en mémoire, à l'entrée du pas j, pour tout $\ell \in \mathcal{M}_j$, une partie E_ℓ, dépendant aussi de j, telle que, compte tenu des $\tilde{D}_h$ (où $\ell < h < j$) déjà construits et validés, on puisse affirmer que

$$\tilde{D}_\ell \cap (\cup_{h>\ell} \tilde{D}_h) \subseteq E_\ell$$

et ce quels que soient les $\tilde{D}_h$ (avec $h \geq j$) non encore construits et validés.

Cette partie E_ℓ est créée pour la première fois au pas ℓ (où $\tilde{D}_\ell$ avait été construite et validée) et c'est alors

$$E_\ell = \tilde{D}_\ell \cap (\cup_{h>\ell} Q_h).$$

Tant que $\ell \in \mathcal{M}_j$, E_j subsiste en mémoire mais peut éventuellement, à certains pas, être privé d'éléments qu'il possédait à l'entrée du pas, si on acquiert une information supplémentaire sur ce que sera $\tilde{D}_\ell \cap (\cup_{h>\ell} \tilde{D}_h)$ (cette opération sera précisée ci-dessous). Pour assurer la P.I.C.I. pour $\tilde{D}_\ell$, il suffit que, à un certain pas $j > \ell$, on construise D_j en y incluant E_ℓ et qu'on le valide, ou bien que l'on constate que $E_\ell = \emptyset$ ou encore que l'on constate que $\ell = r - 1$ (car alors la P.I.C.I. est automatiquement assurée pour $\tilde{D}_\ell$). Une fois effectuée la validation d'un $\tilde{D}_\ell$ qui contient E_ℓ, on retire ℓ de $\mathcal{M}_k$ pour $k > j$.

Le choix fait pour l'inclusion de $E_\ell (\neq \emptyset)$ dans D_j est de tenter cette opération à partir du premier pas j tel que $\{\alpha_j, \ldots, \alpha_{j+1} - 1\} \cap E_\ell \neq \emptyset$ autrement dit, en notant $e_\ell = \inf(E_\ell)$, au premier pas j tel que $\alpha_j \leq e_\ell < \alpha_{j+1}$; elle sera effective si D_j est validé. En d'autres termes, au pas j, on construit

$$\begin{aligned} D_j &= Q_j \cup \big(\cup_{\alpha_{j'+1} \leq e_\ell < \alpha_{j'+2}} E_\ell \big) \cup \ldots \cup \big(\cup_{\alpha_j \leq e_\ell < \alpha_{j+1}} E_\ell \big) \\ &= Q_j \cup \big(\cup_{\alpha_{j'+1} \leq e_\ell < \alpha_{j+1}} E_\ell \big) \ ; \end{aligned}$$

en effet, par définition de j', il n'y a eu de validation à aucun des pas $j'+1, \ldots, j-1$, s'ils existent (c'est-à-dire si $j' < j - 1$) et donc il faut tenter à nouveau, au pas j, l'inclusion dans D_j de tous les E_ℓ ; en revanche, il n'est pas nécessaire d'inclure $Q_{j'+1}, \ldots, Q_{j-1}$ dans D_j car la non validation à chacun de ces pas implique que ces parties étaient contenues dans un $\tilde{D}_k$ avec $k \leq j'$.

Il nous reste à détailler la gestion des E_ℓ et à vérifier que celle-ci assure bien la propriété suivante : à tout j ($> \ell$), tel que $\ell \in \mathcal{M}_j$, compte tenu, s'il y a lieu, des $\tilde{D}_h$ déjà construits et validés postérieurement à l'introduction de ℓ dans les mémoires $\mathcal{M}_h$ (c'est-à-dire les $\tilde{D}_h$ tels que $\ell + 1 \leq h \leq j - 1$) et quels que soient éventuellement les $\tilde{D}_h$, avec $h \geq j$, non encore construits et validés, on a bien, comme annoncé, $\tilde{D}_\ell \cap (\cup_{h>\ell} \tilde{D}_h) \subseteq E_\ell$.

Cette inclusion est bien assurée, on l'a vu plus haut, à la création de E_ℓ au pas ℓ en prenant $E_\ell = \tilde{D}_\ell \cap (\cup_{h>\ell} Q_h)$.

Soit un pas ultérieur j ($j > \ell$) à l'entrée duquel $\ell \in \mathcal{M}_j$ ce qui signifie que, si E_ℓ est non vide, il vérifie $e_\ell \geq \alpha_{j'+1}$, où j' est le dernier pas, strictement avant j, où a été effectuée une validation.

Détaillons, de manière arborescente, les circonstances qui peuvent se produire alors :

(**1**) $\{\alpha_{j'+1}, \ldots, \alpha_{j+1} - 1\} \cap E_\ell = \emptyset$. Alors E_ℓ ne figure pas parmi les E_h dont en tente l'inclusion dans D_j.

(**1.1**) D_j est validé (i.e. D_j n'est pas contenu dans un D_h déjà construit et validé)

(**1.1.1**) $E_\ell \subseteq \tilde{D}_j$ (ceci peut se produire car, quoique E_ℓ ne figure pas parmi les E_h dont en impose l'inclusion dans $\tilde{D}_j$, $\tilde{D}_j$ contient également Q_j et éventuellement certains des E_h ($h \neq j$)). Alors la P.I.C.I. est assurée pour D_ℓ et ℓ n'appartient plus aux mémoires $\mathcal{M}_h$ pour $h > j$.

(**1.1.2**) $E_\ell \not\subseteq \tilde{D}_j$. Alors ℓ est maintenu dans $\mathcal{M}_{j+1}$, sans modification de E_ℓ.

(**1.2**) D_j n'est pas validé. Alors ℓ est maintenu dans $\mathcal{M}_{j+1}$, sans modification de E_ℓ, sauf si $\ell = j'$ et $j = r - 1$; en effet, dans ce cas particulier, seul D_r peut être validé avec un indice strictement supérieur à ℓ, autrement dit $\tilde{D}_\ell$ sera l'avant dernier ou le dernier élément de la suite construite et validée et donc pour ce $\tilde{D}_\ell$ la P.I.C.I. est vide ; on ne maintient donc pas ℓ dans $\mathcal{M}_{j+1}$ ($= \mathcal{M}_r$).

(**2**) $\{\alpha_{j'+1}, \ldots, \alpha_{j+1} - 1\} \cap E_\ell \neq \emptyset$. Alors D_j est construit de telle sorte que $D_j \supseteq E_\ell$.

(**2.1**) D_j est validé. La P.I.C.I. est alors assurée pour $\tilde{D}_\ell$ et ℓ n'appartient plus aux mémoires $\mathcal{M}_h$ pour $h > j$.

(**2.2**) D_j n'est pas validé ; ℓ appartiendra encore à $\mathcal{M}_{j+1}$; en ce que concerne E_ℓ deux cas peuvent se produire :

(**2.2.1**) Pour tout h tel que $\ell + 1 \leq h \leq j'$ et que D_h ait été validé, $i \notin \tilde{D}_h$ (cette condition concerne tous les $\tilde{D}_h$ construits et validés depuis l'introduction de ℓ dans les mémoires $\mathcal{M}$; elle est en particulier satisfaite s'il n'existe pas de tel h, c'est-à-dire si $j' = \ell$).

Alors, si i appartient à E_ℓ, on l'en retire. Nous allons expliquer pourquoi ce retrait (qui allège les parties $\tilde{D}_h$ à construire) ne compromet pas la condition

$$\tilde{D}_\ell \cap (\cup_{h>\ell} \tilde{D}_h) \subseteq E_\ell.$$

Notons, s'il existe, j'' le premier pas après j où aura lieu une validation, et remarquons que i ne peut figurer ni dans $\cup_{\ell<h<j} \tilde{D}_h$ ($= \cup_{l<h\leq j'} \tilde{D}_h$), en vertu de l'hypothèse introductive de ce cas (2.2.1), ni dans $\cup_{h\geq j''} \tilde{D}_h$ puisque, par construction des $\tilde{D}_h$, tous les éléments y seront $\geq \alpha_{j''}$ (on prend $\cup_{h>j''} \tilde{D}_h = \emptyset$ si j'' n'existe pas, c'est-à-dire qu'il n'existe plus aucune validation après j') ; autrement dit le seul $\tilde{D}_h$ (avec $h > \ell$) auquel puisse ici appartenir i est $\tilde{D}_{j''}$ (et en particulier aucun $\tilde{D}_h$ si j'' n'existe pas).

On distingue deux cas, selon que j'' existe ou non.

(**2.2.1.1**) j'' existe. Alors, lors du pas j'', $\ell \in \mathcal{M}_{j''}$, car $\ell \in \mathcal{M}_j$ et qu'il ne peut être retiré d'une mémoire $\mathcal{M}_h$ qu'en sortie de pas où il y a validation ; E_ℓ (éventuellement vide) doit alors par construction être inclus dans $\tilde{D}_{j''}$, en vue d'assurer $\tilde{D}_\ell \cap (\cup_{h>\ell} \tilde{D}_h) \subseteq \tilde{D}_{j''}$; cette inclusion sera assurée aussi bien que i appartienne à $\tilde{D}_{j''}$ ou ne lui appartienne pas, car alors il n'appartient pas à $\cup_{h>\ell} \tilde{D}_h$.

(**2.2.1.2**) j'' n'existe pas. Alors, une fois achevé l'algorithme, ayant comporté le retrait

de i de E_ℓ et aucune nouvelle validation de D_h avec $h \geq j$, i n'appartiendra pas à $\tilde{D}_\ell \cap (\cup_{h>\ell}\tilde{D}_h)$ ($= \tilde{D}_\ell \cap (\cup_{\ell<h\leq j'}\tilde{D}_h)$) ; le retrait de i de E_ℓ ne faisait donc pas obstacle à la validation de la P.I.C.I. pour $\tilde{D}_\ell$.

(**2.2.2**) Il existe h tel que $\ell + 1 \leq h \leq j'$, D_h a été validé et $i \in \tilde{D}_h$.
Alors i n'est pas retiré de E_ℓ (en effet $i \in \cup_{h>\ell}\tilde{D}_h$ et donc i doit appartenir au $\tilde{D}_m$ qui assurera la P.I.C.I. pour D_ℓ ; il ne peut donc être retiré de E_ℓ).

Nous pouvons maintenant décrire le déroulement de cet algorithme.

Description de l'algorithme :

Données :
suite recourante $(Q_1, \ldots, Q_r)$ de parties non vides de $\{1, \ldots, n\}$;
$\alpha_i = \inf(Q_i)$; $\alpha_{r+1} = r + 1$;

Objets construits :
$(D_1, \ldots, D_r)$ parties de I
$(v_1, \ldots, v_r)$ indicateurs de validation ($v_1 = 1$ si D_1 validé, $v_1 = 0$ si non)
$\mathcal{M}_j$ ($j \leq r-2$) partie de $\{j+1, \ldots, r\}$, créée au pas ℓ et évoluant aux pas suivants
pour tout $\ell \in \mathcal{M}_j$, E_ℓ, partie de I créée au pas j et évoluant aux pas suivants
pour tout $j \geq 2$, j', plus grand indice $< j$ en lequel l'indicateur de validation vaut 1.

Initialisation : $D_1 = Q_1$, $v_1 = 1$, $\mathcal{M}_1 = \{1\}$, $E_1 = Q_1 \cap (\cup_{h>1} Q_h)$, $j' = 1$.

Pas courant : j (≥ 2) $D_j = Q_j \cup (\cup_{\alpha_{j'+1} \leq e_\ell < \alpha_{j+1}} E_\ell)$
On distingue les deux cas suivants :

1. Validation (il existe pas de D_m tel que $m < j$, $v_m = 1$, $D_j \subseteq D_m$)
   - $v_j = 1$, $(j+1)' = j$
   - $\mathcal{M}_{j+1}$ se déduit de $\mathcal{M}_j$ en y opérant les retraits et ajouts indiqués ci-dessous,
   - pour tout $\ell \leq j'$ tel que $\ell \in \mathcal{M}_j$ et $\{\alpha_{j'+1}, \ldots, \alpha_{j+1} - 1\} \cap E_\ell = \emptyset$
     - si $E_\ell \subseteq D_j$, ne plus faire figurer ℓ dans $\mathcal{M}_{j+1}$
     - si $E_\ell \not\subseteq D_j$, conserver ℓ dans $\mathcal{M}_{j+1}$ et ne pas modifier E_ℓ
   - pour tout $\ell \leq j'$ tel que $\ell \in \{\alpha_{j'+1}, \ldots, \alpha_{j+1} - 1\} \cap E_\ell \neq \emptyset$, retirer ℓ de $\mathcal{M}_{j+1}$
   - introduire j dans $\mathcal{M}_{j+1}$ et créer $E_j = D_j \cap (\cup_{h>j} Q_h)$.
2. Non validation (il existe D_m tel que $m < j$, $v_m = 1$, $D_j \subseteq D_m$)
   - $v_j = 0$, $(j+1)' = j'$
   - $\mathcal{M}_{j+1} = \mathcal{M}_j$ (et on ne crée pas E_j) sauf dans le cas $j = r-1$ (voir ci-dessous)
   - pour tout ℓ tel que $\ell \in \mathcal{M}_j$ et $\{\alpha_{j'+1}, \ldots, \alpha_{j+1} - 1\} \cap E_\ell = \emptyset$, E_ℓ n'est pas modifié, sauf si $\ell = j'$ et $j = r-1$, auquel cas ℓ est retiré de $\mathcal{M}_j$
   - pour tout ℓ tel que $\ell \in \mathcal{M}_j$ et $\{\alpha_{j+1}, \ldots, \alpha_{j+1} - 1\} \cap E_\ell \neq \emptyset$
     - s'il existe $h \in \{\alpha_{j+1}, \ldots, \alpha_{j'} - 1\}$ tel que $v_h = 1$ et $j \in D_h$, E_ℓ n'est pas modifié
     - sinon, E_ℓ est modifié par le retrait de j

Fin de l'algorithme : $j = r$. Alors la suite des D_j tels que $v_j = 1$ est la suite recouvrante propre possédant la P.I.C. cherchée.

Exemples :

Exemple 1 **:** Reprenons l'exemple traité à l'occasion de la version A de l'algorithme

$$Q_1 = \{1,2,3,5\},\ Q_2 = \{2,3\},\ Q_3 = \{3,7\},\ Q_4 = \{4,5,6\}$$

$$Q_5 = \{5,7\},\ Q_6 = \{6,8\},\ Q_7 = \{7,9\},\ Q_8 = \{8,9\},\ Q_9 = \{9\}.$$

Voici alors le déroulement de l'algorithme :

Pas j	j'	D_j	Validation	$E_\ell(\ell \in \mathcal{M}_j)$
1		$D_1 = Q_1 = \{1,2,3,5\}$	oui	$E_1 = \{2,3,5\}$
2	1	$D_2 = Q_2 \cup E_1 = \{2,3,5\}$	non	$E_1 = \{3,5\}^{(\star)}$
3	1	$D_3 = Q_3 \cup E_1 = \{3,5,7\}$	oui	$E_3 = \{5,7\}$
4	3	$D_4 = Q_4 = \{4,5,6\}$	oui	$E_3 = \{5,7\}$, $E_4 = \{5,6\}$
5	4	$D_5 = Q_5 \cup E_3 \cup E_4 = \{5,6,7\}$	oui	$E_5 = \{6,7\}$
6	5	$D_6 = Q_6 \cup E_5 = \{6,7,8\}$	oui	$E_6 = \{7,8\}$
7	6	$D_7 = Q_7 \cup E_6 = \{7,8,9\}$	oui	$E_7 = \{8,9\}$
8	7	$D_8 = Q_8 \cup E_7 = \{8,9\}$	non	$E_7 = \{9\}$
9	8	$D_9 = Q_9 \cup E_7 = \{9\}$	non	$E_7 = \emptyset$

TAB. 3.3 – Déroulement de l'algorithme version B.

($\star$) Le sommet 2 a pu être retiré à E_1 car il n'existe pas de D_i validé postérieurement à D_1 et avant le pas 2 auquel appartiendrait 2 (en fait il n'existe pas du tout de pas strictement entre 1 et 2)

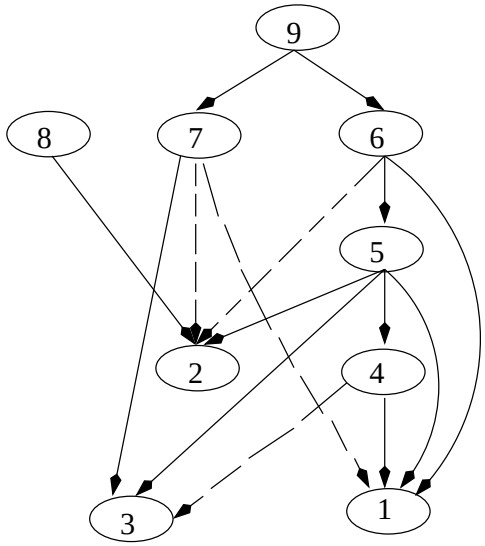


FIG. 3.4 – *Exemple* 2.

Exemple 2 **:** Soit le réseau bayésien dont le graphe orienté sans circuits est représenté par le schéma ci-dessus, muni d'un ordre antihiérarchique, et donnons nous comme but de calculer les lois des triplets (X_2, X_6, X_7), (X_1, X_7) et (X_3, X_4).
Ceci conduit à rajouter les liens $(7,2)$, $(6,2)$, $(7,1)$ et $(4,3)$ figurés en pointillé.

Si l'on ne se préoccupe pas de faire un nettoyage préalable, on doit appliquer l'algorithme à la suite

$$Q_1 = \{1,4,5,6,7\},\ Q_2 = \{2,5,6,7,8\},\ Q_3 = \{3,4,5,7\},\ Q_4 = \{4,5\},$$

$$Q_5 = \{5,6\},\ Q_6 = \{6,9\},\ Q_7 = \{7,9\},\ Q_8 = \{8\},\ Q_9 = \{9\}.$$

Voici alors le déroulement de l'algorithme :

j	j'	D_j	v_j	$E_\ell(\ell \in \mathcal{M}_j)$
1		$D_1 = Q_1 = \{1,4,5,6,7\}$	1	$E_1 = \{4,5,6,7\}$
2	1	$D_2 = Q_2 = \{2,5,6,7,8\}$	1	$E_1 = \{4,5,6,7\},\ E_2 = \{5,6,7,8\}$
3	2	$D_3 = Q_3 = \{3,4,5,7\}$	1	$E_1 = \{4,5,6,7\}, E_2 = \{5,6,7,8\}, E_3 = \{5,7\}$
4	3	$D_4 = Q_4 \cup E_1 = \{4,5,6,7\}$	0	$E_1 = \{4,5,6,7\}^{(\star)}, E_2 = \{5,6,7,8\}\ E_3 = \{5,7\}$
5	3	$D_5 = Q_5 \cup E_1 \cup E_2 \cup E_3 = \{4,5,6,7,8\}$	1	$E_5 = \{6,7,8\}$
6	5	$D_6 = Q_6 \cup E_5 = \{6,7,8,9\}$	1	$E_6 = \{7,8,9\}$
7	6	$D_7 = Q_7 \cup E_6 = \{7,8,9\}$	0	$E_7 = \{8,9\}$
8	7	$D_8 = Q_8 \cup E_7 = \{8,9\}$	0	$\mathcal{M}_8$ est vide car 8 est l'avant dernier indice
9	8	$D_9 = Q_9 = \{9\}$	0	

TAB. 3.4 – Déroulement de l'algorithme version B.

($\star$) A ce pas numéro 4, où il n'y a pas de validation, le sommet 4 ne peut pas être retiré de E_1 car il appartient à D_3, validé.

Exemple 3 **:** Soit le réseau bayésien dont le graphe orienté est représenté par le schéma ci-dessous.

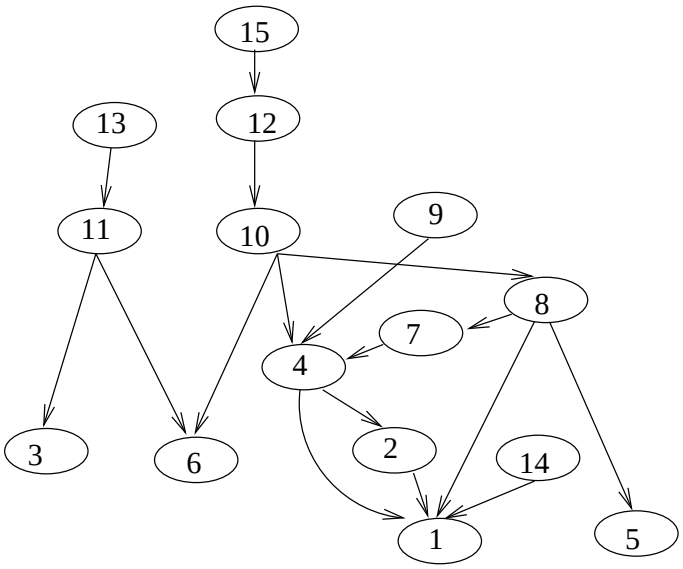


FIG. 3.5 – *Exemple* 3.

Si le but est de calculer les lois $P_{i\cup p(i)}$, on est conduit à appliquer l'algorithme à la suite

(ici nettoyée au préalable)

$$Q_1 = \{1,2,4,8,14\},\ Q_2 = \{3,11\},\ Q_3 = \{4,7,9,10\},\ Q_4 = \{5,8\},\ Q_5 = \{6,10,11\},$$

$$Q_6 = \{7,8\},\ Q_7 = \{8,10\},\ Q_8 = \{10,12\},\ Q_9 = \{11,13\},\ Q_{10} = \{12,15\},$$

Voici alors le déroulement de l'algorithme :

j	j'	D_j	v_j	$E_\ell(\ell \in \mathcal{M}_j)$
1		$D_1 = Q_1 = \{1,2,4,8,14\}$	1	$E_1 = \{4,8\}$
2	1	$D_2 = Q_2 = \{3,11\}$	1	$E_1 = \{4,8\}$, $E_2 = \{11\}$
3	2	$D_3 = Q_3 \cup E_1 = \{4,7,8,9,10\}$	1	$E_2 = \{11\}$, $E_3 = \{7,8,9,10\}$
4	3	$D_4 = Q_4$	1	$E_2 = \{11\}$ $E_3 = \{7,8,9,10\}$, $E_4 = \{8\}$
5	4	$D_5 = Q_5 = \{6,10,11\}$	1	$E_2 = \{11\}$ $E_3 = \{7,8,9,10\}$, $E_4 = \{8\}$, $E_5 = \{10,11\}$
6	5	$D_6 = Q_6 \cup E_3 = \{7,8,9,10\}$	0	$E_2 = \{11\}$ $E_3 = \{8,9,10\}^{(\star)}$, $E_4 = \{8\}$, $E_5 = \{10,11\}$
7	5	$D_7 = Q_7 \cup E_3 \cup E_4 = \{8,9,10\}$	0	$E_2 = \{11\}$ $E_3 = \{8,10\}^{(\star\star)}$, $E_4 = \emptyset^{(\star\star\star)}$, $E_5 = \{10,11\}$
8	5	$D_8 = Q_8 \cup E_3 \cup E_5 = \{8,10,11,12\}$	1	$E_2 = \{11\}$, $E_8 = \{11,12\}$
9	8	$D_9 = Q_9 \cup E_2 \cup E_8 = \{11,12,13\}$	1	
10	1	$D_{10} = Q_{10} = \{12,15\}$	1	

TAB. 3.5 – Déroulement de l'algorithme version B.

($\star$) A ce pas 7 est retiré de E_3 car $7 \notin D_4$ (validé), $7 \notin D_5$ (validé).

($\star\star$) A ce pas 8 n'est pas retiré de E_3 car $8 \in D_4$ (validé), 9 est retiré car $9 \notin D_4$, $9 \notin D_5$.

($\star\star\star$) A ce pas 8 est retiré de E_4 car $8 \notin D_5$.

3.4 Calcul de lois de sous-familles dans le cadre de la loi d'une famille de variables aléatoires compatible avec un arbre de jonction (algorithme LS)

3.4.1 Présentation de la méthode

Soit $\mathcal{C}$ un arbre de jonction sur un ensemble I ; soit choisi un numérotage des éléments de $\mathcal{C}$ tel que $(C_1, \ldots, C_k)$ soit une suite recouvrante propre possédant la propriété d'intersection courante (théorème 8) et soit fixé, pour tout j $(2 \leq j \leq k)$, $\sigma(j)$ tel que $C_j \cap (\cup_{h=1}^{j-1} C_h) \subseteq C_{\sigma(j)}$.

Soit $P = P_I$ compatible avec $\mathcal{C}$; il est donc mis en mémoire, pour tout j $(1 \leq j \leq k)$, une application f_j de Ω_{C_j} dans $\mathbb{R}^+$ de sorte que $P = \frac{1}{\alpha} \prod_{j=1}^{k} f_j$ (où le facteur de normalisation α n'est pas mis en mémoire).

Notons, pour tout j,

$$\begin{cases} R_j = C_j - (\cup_{h=1}^{j-1} C_h) \ ; \\ S_j = C_j \cap (\cup_{h=1}^{j-1} C_h) \ ; \end{cases}$$

en particulier $S_1 = \emptyset$ et $R_1 = C_1$. On a donc par hypothèse, pour tout $j \geq 2$, $S_j \subseteq C_{\sigma(j)}$, d'où $S_j = C_j \cap C_{\sigma(j)}$.

Le calcul des P_{C_j} s'effectue en deux phases :

Phase 1 : Algorithme qui a pour effet d'écrire P sous la forme d'un produit de transitions de conditionnement

$$P = \prod_{j=1}^{k} P_{R_j/S_j} \quad ;$$

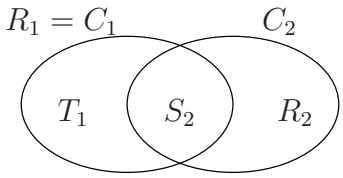
en particulier $P_{R_1/S_1} = P_{C_1}$; il peut éventuellement exister des valeurs de j autres que 1 pour lesquelles le séparateur S_j soit vide, et alors $P_{R_j/S_j} = P_{R_j}$.

Phase 2 : Algorithme qui a pour but, partant de $P = \prod_{j=1}^{k} P_{R_j/S_j}$, de calculer, pour tout j $(1 \leq j \leq k)$, P_{C_j} et P_{S_j}.

3.4.2 Première phase, dans le cas particulier élémentaire où $k = 2$

On connaît les fonctions f_1 et f_2 de sorte que

$$P(x) = \frac{1}{\alpha} f_1(x_{C_1}) f_2(x_{C_2}) \ .$$



La P.I.C. est ici vide ; la suite recouvrante (C_1, C_2) étant propre, on a : $T_1 = C_1 - C_2 \neq \emptyset$ et $R_2 \stackrel{\Delta}{=} C_2 - C_1 \neq \emptyset$; on a donc une partition de I sous la forme $I = T_1 \cup S_2 \cup R_2$ (où $S_2 = C_1 \cap C_2$). Les calculs qui suivent, qui visent à écrire P sous la forme $P_{R_2/S_1} P_{R_1} (= P_{R_2/S_2} P_{C_1})$, sont classiques ; on distingue selon que $S_2 \neq \emptyset$ ou $S_2 = \emptyset$.

Si $S_1 = C_1 \cap C_2 \neq \emptyset$, on a

$$P(x) = P(x_{T_1}, x_{S_2}, x_{R_2}) = \frac{1}{\alpha} f_1(x_{T_1}, x_{S_2}) f_2(x_{S_2}, x_{R_2}),$$

forme qui, rappelons le, exprime que X_{T_1} et X_{R_2} sont indépendants conditionnellement à X_{S_2} ; il vient :

$$P_{C_2}(x_{C_2}) = P_{S_2,R_2}(x_{S_2}, x_{R_2}) = \frac{1}{\alpha} \left[\sum_{x_{T_1}} f_1(x_{T_1}, x_{S_2}) f_2(x_{S_2}, x_{R_2}) \right] \stackrel{\Delta}{=} \frac{1}{\alpha} \tilde{f}_1(x_{S_2}) f_2(x_{S_2}, x_{R_2}),$$

$$P_{S_2}(x_{S_2}) = \sum_{x_{R_2}} P_{S_2,R_2}(x_{S_2}, x_{R_2}) = \frac{1}{\alpha} \left[\sum_{x_{T_1}} f_1(x_{T_1}, x_{S_2}) \right] \left[\sum_{x_{R_2}} f_2(x_{S_2}, x_{R_2}) \right] \stackrel{\Delta}{=} \frac{1}{\alpha} \tilde{f}_1(x_{S_2}) h_2(x_{S_2}),$$

$$P_{R_2/S_2}(x_{R_2/S_2}) = \frac{P_{S_2,R_2}(x_{S_2},x_{R_2})}{P_{S_2}(x_{S_2})} = \frac{f_2(x_{S_2},x_{R_2})}{h_2(x_{S_2})},$$

$$P_{C_1}(x_{C_1}) = P_{T_1,S_2}(x_{T_1},x_{S_2}) = \frac{1}{\alpha} f_1(x_{T_1},x_{S_2}) \sum_{x_{R_2}} f_2(x_{S_2},x_{R_2}) = \frac{1}{\alpha} f_1(x_{T_1},x_{S_2}) h_2(x_{S_2}),$$

où α peut se calculer en appliquant $\sum_{x_{C_1}} P_{C_1}(x_{C_1}) = 1$, d'où

$$\alpha = \sum_{x_{C_1}} f_1(x_{T_1},x_{S_2}) h_2(x_{S_2});$$

Relevons que les calculs à effectuer pour obtenir P_{R_2/S_2} et P_{C_1} se sont décomposés en 2 sommations, l'une sur Ω_{R_2} (calcul de $h_2(x_{S_2})$), puis l'autre sur Ω_{C_1} (calcul de α).

Si $S_1 = C_1 \cap C_2 = \emptyset$, on a $T_1 = C_1$ et $R_2 = C_2$; X_{C_1} et X_{C_2} sont indépendantes et les calculs se simplifient en

$$P(x) = \frac{1}{\alpha} f_1(x_{C_1}) f_2(x_{C_2})$$

$$P_{C_2}(x_{C_2}) = \frac{1}{\alpha} \big[\sum_{x_{C_1}} f_1(x_{C_1}) \big] f_2(x_{C_2}) = \frac{1}{\alpha} \tilde{f}_1 f_2(x_{C_2}),$$

$$P_{C_1}(x_{C_1}) = \frac{1}{\alpha} f_1(x_{C_1}) \big[\sum_{x_{C_2}} f_2(x_{C_2}) \big] = \frac{1}{\alpha} f_1(x_{C_1}) h_2,$$

et α peut se calculer par : $\alpha = \sum_{x_{C_1}} f_1(x_{C_1}) h_2$

Pour $k \geq 3$, l'algorithme que nous allons présenter maintenant généralise, en s'appuyant sur la P.I.C., la succession des calculs élémentaires ci-dessus, où en particulier la constante de normalisation α ne se calcule qu'en fin de course.

3.4.3 Présentation de la première phase du calcul, de l'algorithme dans le cas général

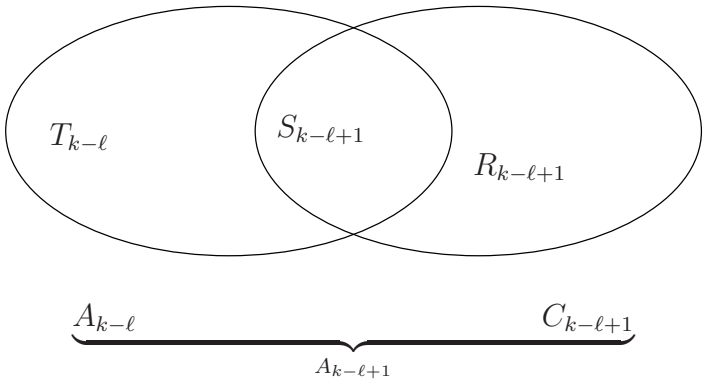
L'algorithme, à chaque pas, récrit P sous la forme

$$P = \frac{1}{\alpha} \prod_{j=1}^{k} g_j$$

où, pour tout j, g_j est une application de Ω_{C_j} dans $\mathbb{R}^+$, de telle sorte que, à l'issue du pas ℓ, tout g_j tel que $j \geq k - \ell + 1$ soit égal à P_{R_j/S_j}.

A l'initialisation on a, pour tout j, $g_j = f_j$.

Au pas ℓ ($1 \leq \ell \leq k$), ne sont pas modifiés les g_j tels que $j > k - \ell + 1$ (ils auront été définitivement calculés aux pas précédents) et les g_j tels que $j \leq k - \ell$ et $j \neq \sigma(k - \ell + 1)$. Les seules modifications sont opérées sur $g_{k-\ell+1}$ et $g_{\sigma(k-\ell+1)}$ (si $\ell \leq k-1$) ; pour les décrire on va distinguer selon que $S_{k-\ell+1}$ est vide ou non vide.



Si $S_{k-\ell+1} \neq \emptyset$, le "$\ell$-ième facteur de normalisation" est la fonction $h_{k-\ell+1}$ définie sur $\Omega_{S_{k-\ell+1}}$ par la formule suivante, qui utilise le fait que $(S_{k-\ell+1}, R_{k-\ell+1})$ est une partition de $C_{k-\ell+1}$ et que donc on peut écrire $x_{C_{k-\ell+1}}$ comme le couple $(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}})$:

$$h_{k-\ell+1}(x_{S_{k-\ell+1}}) \stackrel{\Delta}{=} \sum_{x_{R_{k-\ell+1}}} g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}}) \ ;$$

on modifie ensuite $g_{k-\ell+1}$ par l'instruction

$$g_{k-\ell+1}(x_{C_{k-\ell+1}}) \stackrel{\Delta}{=} \frac{g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}})}{h_{k-\ell+1}(x_{S_{k-\ell+1}})}$$

(et alors, **comme on le démontrera ci-dessous en** 3.4, $g_{k-\ell+1} = P_{R_{k-\ell+1}/S_{k-\ell+1}}$) ; puis on modifie $g_{\sigma(k-\ell+1)}$ par l'instruction

$$g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) := g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) h_{k-\ell+1}(x_{S_{k-\ell+1}})$$

(formule qui a un sens car, par la P.I.C., $S_{k-\ell+1} \subseteq C_{\sigma(k-\ell+1)}$).

Si $S_{k-\ell+1} = \emptyset$, autrement dit $R_{k-\ell+1} = C_{k-\ell+1}$, le "$\ell$-ième facteur de normalisation" est la constante $h_{k-\ell+1}$ définie par

$$h_{k-\ell+1} = \sum_{x_{C_{k-\ell+1}}} g_{k-\ell+1}(x_{C_{k-\ell+1}}) \ \ ;$$

on modifie $g_{k-\ell+1}$ par l'instruction

$$g_{k-\ell+1}(x_{C_{k-\ell+1}}) := \frac{g_{k-\ell+1}(x_{C_{k-\ell+1}})}{h_{k-\ell+1}} \ \ ;$$

(et alors, **comme on le démontrera ci-dessous en** 3.4, $g_{k-\ell+1} = P_{C_{k-\ell+1}}$) ; on modifie ensuite $g_{\sigma(k-\ell+1)}$ par l'instruction :

$$g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) := g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) \cdot h_{k-\ell+1}$$

Remarquons que, au dernier pas de l'algorithme (pas numéro k), on manipule $S_1 = \emptyset$ et $R_1 = C_1$; donc on calcule la constante $h_1 = \sum_{x_{C_1}} g_1(x_{C_1})$, g_1 étant tel qu'il figure à l'entrée de ce pas numéro k. Or, à l'entrée de ce pas, on a obtenu

$$\begin{aligned} P(x) &= \frac{1}{\alpha}\, g_1(x_{C_1}) P_{R_2/S_2}(x_{C_2}) \dots P_{R_k/S_k}(x_{C_k}) \\ &= \frac{1}{\alpha}\, g_1(x_{C_1}) P_{R_2/S_2}(x_{S_2}, x_{R_2}) \dots P_{R_k/S_k}(x_{S_k}, x_{R_k}); \end{aligned}$$

donc la constante de normalisation de cette formule, α, vérifie

$$\alpha = \sum_x g_1(x_{C_1}) P_{R_2/S_2}(x_{S_2}, x_{R_2}) \dots P_{R_k/S_k}(x_{S_k}, x_{R_k}) \ ;$$

or $(C_1, R_2, \dots, R_k)$ est une partition de I, donc par sommations successives, sur $x_{R_k}, x_{R_{k-1}}, \dots, x_{R_2}$, on vérife que :

$$\alpha = \sum_{x_{C_1}} g_1(x_{C_1}) = h_1 \ ;$$

la constante α est donc obtenue en fin d'algorithme, par une sommation sur Ω_{C_1}, et il aurait été plus lourd de la calculer initialement par une sommation sur Ω_I.

3.4.4 Justification de la première phase du calcul de l'algorithme

Notons, pour tout j, $A_j = \cup_{s \le j} C_s$ (donc $A_k = I$).

Il s'agit de démontrer la proposition suivante :

Proposition 8 *A l'entrée du pas ℓ $(1 \le \ell \le k)$*

$$P_{A_{k-\ell+1}} = \frac{1}{\alpha} \prod_{s=1}^{k-\ell+1} g_s$$

et, dans la réalisation de ce pas, il vient

$$P_{R_{k-\ell+1}/S_{k-\ell+1}} = \frac{g_{k-\ell+1}}{h_{k-\ell+1}}$$

(c'est à dire $P_{R_{k-\ell+1}}$ si $S_{k-\ell+1} = \emptyset$).

Démonstration

A l'entrée du pas 1, on a bien, puisque les g_j sont initialisés aux f_j, et que $A_k = I$, $\frac{1}{\alpha} \prod_{s=1}^{k} g_s = P_{A_k}$.

Soit donc, par hypothèse de récurrence, $P_{A_{k-\ell+1}} = \frac{1}{\alpha} \prod_{s=1}^{k-\ell+1} g_s$, autrement écrit, pour tout $x_{A_{k-\ell+1}}$,

$$P_{A_{k-\ell+1}}(x_{A_{k-\ell+1}}) = \frac{1}{\alpha} \big[\prod_{s=1}^{k-\ell} g_s(x_{C_s}) \big] \cdot g_{k-\ell+1}(x_{C_{k-\ell+1}}) \ ;$$

comme $A_{k-\ell} = \cup_{s \leq k-\ell} C_s$, on peut définir $\tilde{g}_{k-\ell}(x_{A_{k-\ell}}) = \prod_{s=1}^{k-\ell} g_s(x_{C_s})$, d'où :

$$P_{A_{k-\ell+1}}(x_{A_{k-\ell+1}}) = \frac{1}{\alpha}\ \tilde{g}_{k-\ell}(x_{A_{k-\ell}}) \cdot g_{k-\ell+1}(x_{C_{k-\ell+1}})$$

ou encore, en effectuant la partition de $A_{k-\ell+1}$ en $R_{k-\ell+1}\ \cup\ S_{k-\ell+1}\ \cup\ T_{k-\ell}$, où
$R_{k-\ell+1} = C_{k-\ell+1} - A_{k-\ell} \neq \emptyset$,
$S_{k-\ell+1} = C_{k-\ell+1} \cap A_{k-\ell}$ (vide ou non vide selon les cas),
$T_{k-\ell} = A_{k-\ell} - S_{k-\ell+1} \neq \emptyset$,

$$P_{A_{k-\ell+1}}(x_{T_{k-\ell}}, x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}}) = \frac{1}{\alpha}\ \tilde{g}_{k-\ell}(x_{T_{k-\ell}}, x_{S_{k-\ell+1}}) \cdot g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}})\ ;$$

ceci exprime que $X_{T_{k-\ell}}$ et $X_{R_{k-\ell+1}}$ sont indépendants conditionnellement à $X_{S_{k-\ell+1}}$ (qui dégénère en l'indépendance de $X_{T_{k-\ell}}$ et $X_{R_{k-\ell+1}}$ si $S_{k-\ell+1} = \emptyset$).

Dans la suite de la démonstration, nous ne détaillons pas les simplifications évidentes si $S_{k-\ell+1} = \emptyset$.

On a

$$\begin{aligned} P_{C_{k-\ell+1}}(x_{C_{k-\ell+1}}) &= P_{S_{k-\ell+1} \cup R_{k-\ell+1}}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}}) \\ &= \frac{1}{\alpha}\Big[\sum_{x_{T_{k-\ell}}} \tilde{g}_{k-\ell}(x_{T_{k-\ell}}, x_{S_{k-\ell+1}})\Big]\ g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}}) \\ &\triangleq \frac{1}{\alpha}\ \tilde{g}_{k-\ell+1}(x_{S_{k-\ell+1}})\ g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}}) \end{aligned}$$

et

$$\begin{aligned} P_{S_{k-\ell+1}}(x_{S_{k-\ell+1}}) &= \sum_{x_{R_{k-\ell+1}}} P_{S_{k-\ell+1} \cup R_{k-\ell+1}}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}}) \\ &= \frac{1}{\alpha}\ \tilde{g}_{k-\ell+1}(x_{S_{k-\ell+1}})\Big[\sum_{x_{R_{k-\ell+1}}} g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}})\Big] \\ &= \frac{1}{\alpha}\ \tilde{g}_{k-\ell+1}(x_{S_{k-\ell+1}})\ h_{k-\ell+1}(x_{S_{k-\ell+1}})\ \ ; \end{aligned}$$

or

$$P_{R_{k-\ell+1}/S_{k-\ell+1}}(x_{R_{k-\ell+1}}/x_{S_{k-\ell+1}}) = \frac{P_{S_{k-\ell+1}, R_{k-\ell+1}}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}})}{P_{S_{k-\ell+1}}(x_{S_{k-\ell+1}})}$$

d'où, comme annoncé,

$$P_{R_{k-\ell+1}/S_{k-\ell+1}}(x_{R_{k-\ell+1}}/x_{S_{k-\ell+1}}) = \frac{g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}})}{h_{k-\ell+1}(x_{S_{k-\ell+1}})}.$$

Vérifions par ailleurs la conservation de l'hypothèse de récurrence, c'est à dire que, après initialisation des g_j au cours du pas ℓ, il vient, à l'entrée du pas $\ell+1$, $\frac{1}{\alpha}\ \prod_{s=1}^{k-\ell} g_s = P_{A_{k-\ell}}$.

Les g_j étant encore tels qu'à l'entrée du pas $k-\ell+1$, il vient, en utilisant l'égalité $A_{k-\ell+1} = A_{k-\ell} \cup R_{k-\ell+1}$ et l'hypothèse de récurrence :

$$\begin{aligned} P_{A_{k-\ell}}(x_{A_{k-\ell}}) &= \sum_{x_{R_{k-\ell+1}}} P_{A_{k-\ell+1}}(x_{A_{k-\ell}}, x_{R_{k-\ell+1}}) \\ &= \frac{1}{\alpha}\, \tilde{g}_{k-\ell}(x_{A_{k-\ell}}) \sum_{x_{R_{k-\ell+1}}} g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}}) \\ &= \frac{1}{\alpha}\, \big[\prod_{s=1}^{k-\ell} g_s(x_{C_s})\big] h_{k-\ell+1}(x_{S_{k-\ell+1}})\ ; \end{aligned}$$

par l'hypothèse d'intersection courante de la suite $(C_h)_{1\leq h\leq k}$, on sait que $S_{k-\ell+1} \subseteq C_{\sigma(k-\ell+1)}$; donc en actualisant, pour $\ell \leq k-1$, $g_{\sigma(k-\ell+1)}$ sous la forme $g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) \cdot h_{k-\ell+1}(x_{S_{k-\ell+1}})$, et en conservant tous les autres g_j (avec $1 \leq j \leq k-\ell$), il vient bien

$$P_{A_{k-\ell}}(x_{A_{k-\ell}}) = \frac{1}{\alpha} \prod_{s=1}^{k-\ell} g_s(x_{C_s}).$$

■

On peut alors donner le déroulement de l'algorithme comme suit

3.4.5 Déroulement de la première phase du calcul de l'algorithme

Données :
Suite recouvrante propre de I, P.I.C., $(C_1, \ldots, C_k)$
Pour tout j $(1 \leq j \leq k)$, fonctions f_j de Ω_{C_j} dans $\mathbb{R}^+$ de sorte que P_I soit proportionnel à $\prod_{j=1}^k f_j$

Objets manipulés :
Fonctions g_j $(1 \leq j \leq k)$ de Ω_{C_j} dans $\mathbb{R}^+$ et h_j $(1 \leq j \leq k-1)$ de S_j $(= C_j \cap (\cup_{h=1}^{j-1}))$ dans $\mathbb{R}^+$
Probabilités conditionnelles P_{R_j/S_j} $(1 \leq j \leq k)$ $(= P_{R_j}$ si $S_j = \emptyset)$

Initialisation : pour tout j $g_j = f_j$

Pas courant : ℓ $(1 \leq \ell \leq k)$. Pour tout $x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}}$ et $x_{C_{\sigma(k-\ell+1)}}$

si $S_{k-\ell+1} \neq \emptyset$:
$h_{k-\ell+1}(x_{S_{k-\ell+1}}) := \sum_{x_{R_{k-\ell+1}}} g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}})$
$g_{k-\ell+1}(x_{C_{k-\ell+1}}) = P_{R_{k-\ell+1}/S_{k-\ell+1}}(x_{R_{k-\ell+1}}/x_{S_{k-\ell+1}}) := \frac{g_{k-\ell+1}(x_{S_{k-\ell+1}}, x_{R_{k-\ell+1}})}{h_{k-\ell+1}(x_{S_{k-\ell+1}})}$.
$g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) := g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) \cdot h_{k-\ell+1}$

si $S_{k-\ell+1} = \emptyset$ (i.e. $C_{k-\ell+1} = R_{k-\ell+1}$) :
$h_{k-\ell+1}(x_{k-\ell+1}) := \sum_{x_{C_{k-\ell+1}}} g_{k-\ell+1}(x_{C_{k-\ell+1}})$
$g_{k-\ell+1}(x_{C_{k-\ell+1}}) := P_{R_{k-\ell+1}}(x_{R_{k-\ell+1}}) := \frac{g_{k-\ell+1}(x_{R_{k-\ell+1}})}{h_{k-\ell+1}}$
$g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) := g_{\sigma(k-\ell+1)}(x_{C_{\sigma(k-\ell+1)}}) \cdot h_{k-\ell+1}$

Fin de l'algorithme : $\ell = k$

3.4.6 Deuxième phase du calcul

A l'entrée de cette phase on dispose d'une loi de probabilité vérifiant

$$P = \prod_{j=1}^{k} P_{R_j/S_j},$$

où les P_{R_j/S_j} sont mis en mémoire.

On obtient en sortie toutes les lois P_{C_j} $(1 \leq j \leq k)$ et P_{S_j} $(2 \leq j \leq k)$. L'algorithme au pas j calcule P_{C_j} $(j \geq 2)$.

Au premier pas de calcul on lit P_{R_1/S_1}, qui est en fait P_{C_1}.

Au pas j ($j \geq 2$) :

1. on utilise le fait que $S_j \subseteq C_{\sigma(j)}$, et que $P_{C_{\sigma(j)}}$ a déjà été calculé lors d'un pas précédent car $\sigma(j) < j$; on a donc, posant $B_j = C_{\sigma(j)} - S_j$,

$$P_{S_j}(x_{S_j}) = \sum_{x_{B_j} \in \Omega_{B_j}} P_{C_{\sigma(j)}}(x_{B_j}, x_{S_j}),$$

2. on utilise le fait que $C_j = S_j \cup R_j$, on calcule P_{C_j} par la formule

$$P_{C_j} = P_{S_j} \cdot P_{R_j/S_j} \ .$$

3.4.7 Rôle de l'hypothèse de propreté de la suite P.I.C. $(C_j)_{1 \leq j \leq k}$

L'hypothèse de propreté de la suite P.I.C. $(C_j)_{1 \leq j \leq k}$ est intervenue, dans le déroulement de l'algorithme LS, dans chacune des deux phases :

- dans la phase 1, on effectue des sommations du type $h_j(x_{S_j}) = \sum_{x_{R_j}} g_j(x_{S_j}, x_{R_j})$, ce qui suppose implicitement que $R_j \neq \emptyset$, propriété mise en défaut si $C_j \subseteq C_{\sigma(j)}$, car alors $S_j = C_j \cap C_{\sigma(j)} = C_j$, d'où $R_j = C_j - S_j = \emptyset$ et $h_j(x_{S_j}) = g_j(x_{S_j})$,

- dans la phase 2, on effectue des sommations du type $P_{S_j}(x_{S_j}) = \sum_{x_{B_j} \in \Omega_{B_j}} P_{C_{\sigma(j)}}(x_{B_j}, x_{S_j})$, ce qui suppose implicitement que $B_j \neq \emptyset$, propriété mise en défaut si $C_{\sigma(j)} \subseteq C_j$ car alors $S_j = C_{\sigma(j)}$ d'où $B_j = C_{\sigma(j)} - S_j = \emptyset$ et $P_{S_j}(x_{S_j}) = P_{C_{\sigma(j)}}(x_{C_{\sigma(j)}})$.

De telles circonstances de "non propreté" peuvent se produire en particulier dans le calcul de probabilités conditionnelles. En fait étant donné une suite P.I.C. $(C_j)_{1 \leq j \leq k}$, on utilise la suite $(C_j^A)_{1 \leq j \leq k}$ où pour tout j $C_j^A = C_j - A$; il résulte de la définition de la P.I.C. que $(C_j^A)_{1 \leq j \leq k}$ est P.I.C. ; en revanche, elle n'est pas nécessairement propre. Dans ce cas deux solutions sont possibles :

1) travailler avec la suite P.I.C. $(C_j^A)_{1 \leq j \leq k}$ avec les conventions ci dessus,

2) nettoyer la suite $(C_j^A)_{1 \leq j \leq k}$ afin d'obtenir une suite P.I.C. propre.

3.4.8 Terminologie informatique

Dans l'algorithme qui constitue la première phase de la méthode, au pas ℓ est utilisée seulement la fonction $g_{k-\ell+1}$, pour modifier elle-même et $g_{\sigma(k-\ell+1)}$; en terminologie informatique, on dit que ce pas comporte la transmission d'un "message" de $C_{k-\ell+1}$ (ensemble de définition de $g_{k-\ell+1}$) à $C_{\sigma(k-\ell+1)}$.

Dans l'algorithme qui constitue la seconde phase de la méthode, au pas j sont utilisés seulement $P_{\sigma(j)}$ (défini sur $C_{\sigma(j)}$) et P_{R_j/S_j} (défini sur $R_j \cup S_j = C_j$), pour calculer P_{C_j} ; en terminologie informatique, on dit que ce pas comporte la transmission d'un "message" de $C_{\sigma(j)}$ à C_j.

3.4.9 Etude d'un exemple

Nous allons illuster l'algorithme de calcul presenté ci-dessus sur un réseau bayésien dont le graphe est représenté ci dessous, sur $I = \{a, b, c, d, e, f, g, h, i\}$.

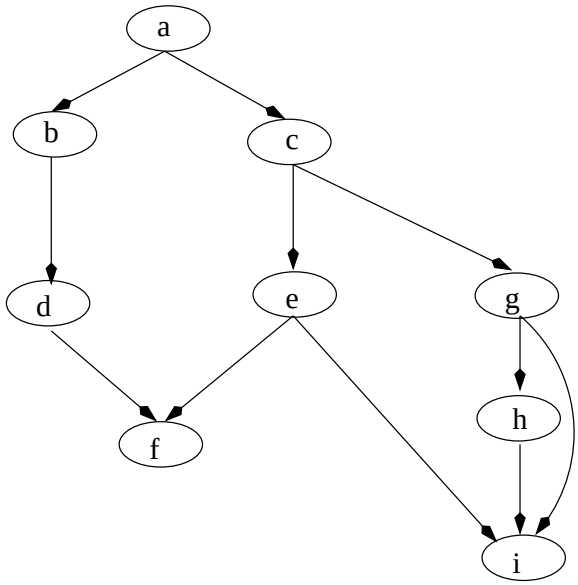


FIG. 3.6 – *Exemple* 1.

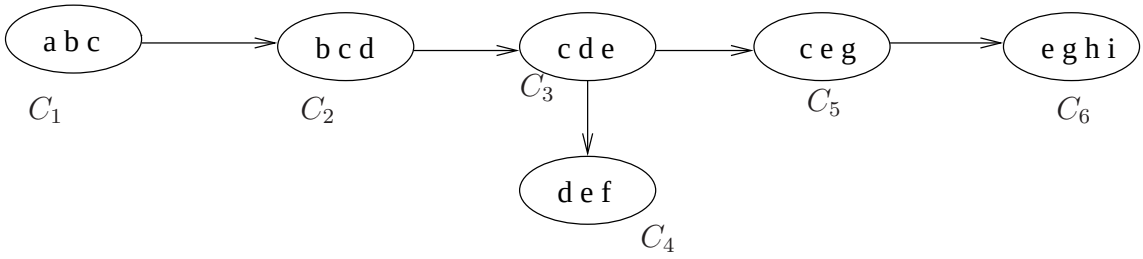
L'ensemble $\mathcal{J}$, des parties de I composées chacune d'un élément et de ses parents est le suivant :

$$\mathcal{J} = \{\{a\}, \{a, b\}, \{a, c\}, \{d, b\}, \{e, c\}, \{g, c\}, \{h, g\}, \{f, d, e\}, \{i, h, g, e\}\}$$

La suite obtenue $(C_1, \ldots, C_6)$, ainsi que les séparateurs S_j $(2 \leq j \leq 6)$, les $R_j = C_j - S_j$ et les $\sigma(j)$ caractérisant la P.I.C. sont donnés dans le tableau ci-dessous.

j	C_j	S_j	R_j	$\sigma(j)$
1	abc		abc	
2	bcd	bc	d	1
3	cde	cd	e	2
4	def	de	f	3
5	ceg	ce	g	3
6	$eghi$	eg	h, i	5

Voici une représentation graphique de cette arborescence.



On a $P_I = \prod_{j=1}^{6} f_{C_j}$, les potentiels f_{C_j} étant choisis comme suit :

1. $f_1 = P(a)P(b/a)P(c/a)$
2. $f_2 = P(d/b)$
3. $f_3 = P(e/c)$
4. $f_4 = P(f/d, e)$
5. $f_5 = P(g/c)$
6. $f_6 = P(h/g)P(i/e, g, h)$

Le déroulement de la première phase de l'algorithme est alors le suivant (nous indiquons à chaque fois les modifications effectuées)

Initialisation : $g_1 = f_1$, $g_2 = f_2$, $g_3 = f_3$, $g_4 = f_4$, $g_5 = f_5$, $g_6 = f_6$

Pas 1 :

$$\left|\begin{array}{l} h_6(x_e, x_g) = \sum_{h,i} g_6(x_e, x_g, x_h, x_i) \\ g_6(x_e, x_g, x_h, x_i) := \frac{g_6(x_e, x_g, x_h, x_i)}{h_6(x_e, x_g)} \\ g_5(x_c, x_e, x_g) := g_5(x_c, x_e, x_g) \cdot h_6(x_e, x_g) \quad (\text{car } \sigma(6) = 5) \end{array}\right.$$

qui, du fait que $g_6 = f_6 = P(h/g)P(i/e, g, h)$, donne

$$h_6 = 1,\ g_6 := g_6,\ g_5 := g_5 (= f_5)\ ;$$

On a obtenu $P_{R_6/S_6} = P_{h,i/e,g} = g_6 = f_6$ (qui en fait se déduit par produit à partir des données du réseau bayésien, à savoir $P(h/g)P(i/e, g, h)$).

Pas 2 :

$$h_5(x_c, x_e) = \sum_g g_5(x_c, x_e, x_g)$$

qui, du fait que $g_5 = f_5 = P(g/c)$, donne $h_5 = 1$ d'où $g_5 := g_5$ et, comme $\sigma(5) = 3$ $g_3 := g_3 (= f_3)$

On a obtenu $P_{R_5/S_5} = P_{g/c,e} = g_5 = f_5$ (qui en fait est la donnée $P_{g/c}$; l'algorithme a mis en évidence que, conditionnellement à X_c, X_g est indépendant de X_e).

Pas 3 :

$$h_4(x_d, x_e) = \sum_f g_4(x_d, x_e, x_f)$$

qui, du fait que $g_4 = f_4 = P(f/d, e)$, donne $h_4 = 1$ d'où $g_4 := g_4$ et, comme $\sigma(4) = 3$, $g_3 := g_3 (= f_3)$

On a obtenu $P_{R_4/S_4} = P_{f/d,e} = g_4 = f_4$ (qui est en fait l'une des données du réseau bayésien).

Pas 4 :

$$h_3(x_c, x_d) = \sum_e g_3(x_c, x_e, x_d)$$

qui, du fait que $g_3 = f_3 = P(e/c)$, donne $h_3 = 1$ d'où $g_3 := g_3$ et, comme $\sigma(3) = 2$, $g_2 := g_2 (= f_2)$

On a obtenu $P_{R_3/S_3} = P_{e/c} = g_3 = f_3$ (qui en fait est la donnée $P_{e/c}$; l'algorithme a mis en évidence que, conditionnellement à X_c, X_e est indépendant de X_d).

Pas 5 :

$$h_2(x_b, x_c) = \sum_d g_2(x_b, x_c, x_d)$$

qui, du fait que $g_2 = f_2 = P(d/b)$, donne $h_2 = 1$ d'où $g_2 := g_2$ et, comme $\sigma(2) = 1$, $g_1 := g_1 (= f_1)$

On a obtenu $P_{R_2/S_1} = P_{d/b,c} = g_2 = f_2$ (qui en fait la donnée $P_{d/b}$; l'algorithme a mis en évidence que, conditionnellement à X_b, X_c est indépendant de X_d).

Pas 6 :

Comme $S_1 = \emptyset$, on a $h_1 = 1$ d'où, $P_{R_1/S_1} = P_{R_1} = P_{a,b,c} = g_1 = f_1$.

Maintenant après avoir obtenu toutes les transitions de conditionnement P_{R_i/S_i} pour tout $1 \leq i \leq 6$, nous pouvons appliquer la deuxième phase d'algorithme afin de calculer toutes les distributions de probabilités P_{C_i} $(1 \leq i \leq 6)$.

Pas 1 : Nous commençons par la racine, C_1 ; sa distribution de probabilité est donnée par

$$P_{a,b,c}(x_a, x_b, x_c) = g_1.$$

Pas 2 : Comme $S_2 \subset C_1$, il vient que

$$P_{S_2}(x_{S_2}) = P_{b,c}(x_b, x_c) = \sum_a P_{C_1}(x_a, x_b, x_c)$$

d'où $P_{b,c,d}(x_b, x_c, x_d) = P_{b,c}(x_b, x_c) P_{d/b,c}(x_d/x_b, x_d)$.

Pas 3 : Comme $S_3 \subset C_2$, il vient que

$$P_{S_3}(x_{S_3}) = P_{c,d}(x_c, x_d) = \sum_b P_{C_2}(x_d, x_b, x_c)$$

d'où $P_{c,d,e}(x_c, x_d, x_e) = P_{c,d}(x_c, x_d) P_{e/c,d}(x_e/x_c, x_d)$.

Pas 4 : Comme $S_4 \subset C_3$, il vient que

$$P_{S_4}(x_{S_4}) = P_{d,e}(x_d, x_e) = \sum_c P_{C_3}(x_c, x_d, x_e)$$

d'où $P_{d,e,f}(x_d, x_e, x_f) = P_{d,e}(x_d, x_e)P_{f/d,e}(x_f/x_d, x_e)$.

Pas 5 : Comme $S_5 \subset C_3$, il vient que

$$P_{S_5}(x_{S_5}) = P_{c,e}(x_c, x_e) = \sum_d P_{C_3}(x_c, x_d, x_e)$$

d'où $P_{c,e,g}(x_c, x_e, x_g) = P_{c,e}(x_c, x_e)P_{g/c}(x_g/x_c)$.

Pas 6 : Comme $S_6 \subset C_5$, il vient que

$$P_{S_6}(x_{S_6}) = P_{e,g}(x_e, x_g) = \sum_c P_{C_5}(x_g, x_c, x_e)$$

d'où $P_{e,g,h,i}(x_e, x_g, x_h, x_i) = P_{e,g}(x_e, x_g)P_{h/g}(x_h/x_i)P_{i/e,g,h}(x_i/x_e, x_g, x_h)$.

Chapitre 4

Algorithme des Restrictions Successives

4.1 Introduction

Nous avons introduit, au chapitre précédent, l'algorithme LS (Lauritzen et Spiegelhalter) destiné à calculer, pour des sous-ensembles de I que nous applelons "cibles", la loi marginale (ou des lois conditionnelles) et fondé sur les notions d'arbre de jonction et de suite recouvrante propre possédant la propriétés d'intersection courante (P.I.C.). Nous avions remarqué que cet algorithme se restreint dans sa version originale à des cibles assez particulières qui sont contenues dans les cliques du graphe non orienté déduit du graphe orienté définissant le réseau bayésien par moralisation et triangularisation (voir chapitre 3 section 3.2) ; cet algorithme nécessite, si on s'intéresse à des cibles plus générales, une adaptation [16] que nous avons présentée pour notre part, au chapitre 3 section 2.3, par création de liens nouveaux par rapport à ceux du réseau bayésien ; **ceci peut conduire à des cliques trop grosses pour que l'algorithme reste efficace**.

L'objectif, dans ce chapitre, est d'introduire un algorithme de calcul sur les réseaux bayésiens dit "algorithme des restrictions successives" et qui fait usage de la notion de réseau bayésien de niveau deux. Une première version de cet algorithme permet de calculer la loi marginale, hors conditionnement, d'un sous-ensemble quelconque des variables aléatoires d'un réseau bayésien ; ceci fait l'objet de la première partie de ce chapitre. Dans la deuxième partie nous présentons une extension de l'algorithme des restrictions successives, qui permet de calculer la loi d'un sous-ensemble de variables aléatoires conditionnée par un autre sous-ensemble de variables du réseau bayésien initial. Dans la dernière section nous comparons (sur quelques exemples) les résultats obtenus par l'algorithme des restrictions successives avec ceux obtenus par l'algorithme d'arbre de jonction.

4.2 Calcul des lois marginales

Etant donné un réseau bayésien, associé à une famille $(X_i)_{i\in I}$ de variables aléatoires, nous nous intéressons au calcul de la loi d'une sous-famille $(X_i)_{i\in A}$ où A est une partie quelconque de I, dite "cible". Nous ne formulons sur la cible aucune hypothèse particulière en connexion avec la structure du réseau bayésien.

L'algorithme que nous proposons est fondé sur les notions de réseau bayésien de niveau deux (chapitre 1, section 2), de descendance proche, et de transformation du réseau par restrictions successives. On rappelle que, dans un RB2, sur un ensemble I, les nœuds sont les éléments d'une partition J de I, qu'on appellera aussi "**atomes**". Nous commençons par donner la définition de la descendance proche, ensuite un exemple de calcul avant d'expliquer le principe de calcul des lois marginales, et nous terminons par une démonstration du résultat qui justifie cet algorithme.

Comme dans les chapitres précédents, nous présentons cet algorithme dans le cas de variables aléatoires discrètes, mais cette technique reste valable dans le cas général : les opérations de **marginalisation** (vocabulaire usuel des praticiens des réseaux bayésiens), présentées ici comme des sommations, sont alors des intégrations relativement à des mesures dominantes σ-finies.

4.2.1 Descendance proche

Nous allons introduire ici une nouvelle notion relative aux graphes orientés. Cette notion nous sera utile dans la suite en l'appliquant à des graphes sous-jacents à des réseaux bayésiens ou des réseaux bayésiens de niveau 2, pour introduire l'algorithme des restrictions successives.

Définition 23 *Soit (I, G) un graphe orienté sans circuits et soit $i \in I$. $dp_G(i)$ (ou $dp(i)$) est l'ensemble (éventuellement vide) appelé "descendance proche" de i, qui se compose des enfants de i et, s'il y a lieu, des sommets situés sur un chemin entre i et l'un de ses enfants.*

Exemple de descendance proche (sur un RB2)

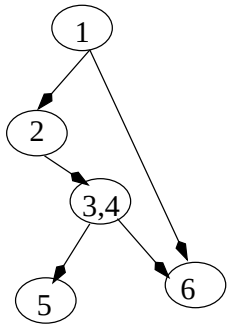


FIG. 4.1 – *réseau bayésien de niveau deux*

$dp(1) = \{2, \{3, 4\}, 6\}$

qu'on pourra écrire aussi $dp(1) = \{2, 3, 4, 6\}$.

On remarque l'abus de langage (déjà indiqué en 1.2), consistant à identifier un ensemble d'atomes de la partition avec leur union.

4.2.2 Exemple introductif à l'algorithme

Soit le graphe orienté sans circuits (I, G) donné en figure (4.2) (a).

La loi jointe associée au réseau bayésien de niveau 1 ainsi défini s'écrit sous la forme

$$\begin{aligned} P_I(x_I) \;&=\; P_1(x_1)\; P_2(x_2)\; P_3(x_3)\; P_{4/1}(x_4/x_1)\; P_{5/2,4}(x_5/x_2, x_4)\; P_{6/5}(x_6/x_5) \\ &\quad P_{7/3,5}(x_7/x_3, x_5)\; P_{8/1,6}(x_8/x_1, x_6)\; P_{9/2,6}(x_9/x_2, x_6)\; P_{10/2,7}(x_{10}/x_2, x_7) \end{aligned}$$

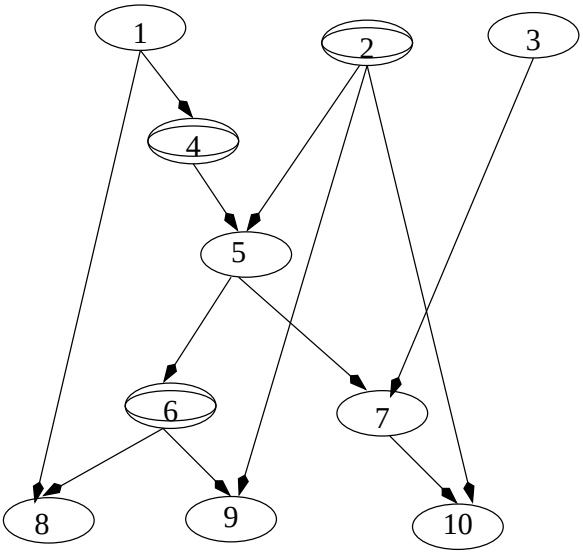
Le calcul de P_A, loi jointe des variables aléatoires dans la cible $A = \{1, 3, 5, 7, 8, 9, 10\}$, revient à marginaliser la loi jointe totale du réseau relativement aux variables indexées dans $\overline{A} = \{2, 4, 6\}$, c'est-à-dire sommer sur les valeurs prises par X_2, X_4 et X_6.

Nous allons choisir les variables de $\overline{A}$ l'une après l'autre pour la marginalisation. En premier lieu remarquons que seule la variable X_6 n'a pas de descendants dans $\overline{A}$; nous commençons par marginaliser par rapport à cette variable, et, en regroupant les facteurs, dans l'expression de P_I, où figure X_6, nous obtenons (résultat vérifiable directement, mais qui résulte aussi du lemme ci-dessous)

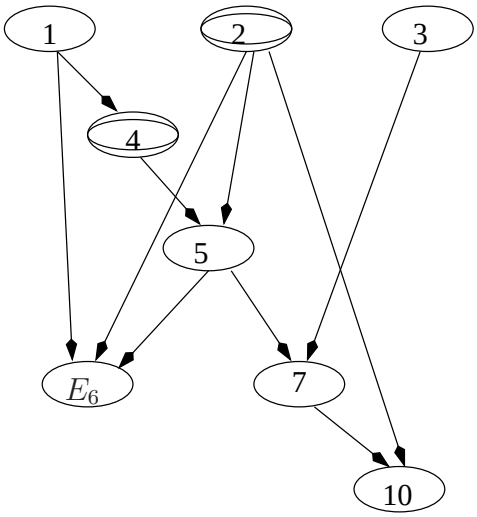
$$\sum_{x_6} P_{8/1,6}(x_8/x_1, x_6)\; P_{9/2,6}(x_9/x_2, x_6)\; P_{6/5}(x_6/x_5) \;=\; P_{8,9/1,2,5}(x_8, x_9/x_1, x_2, x_5)\;;$$

ceci explique que le graphe qui résulte de cette sommation doit être muni d'une structure de réseau bayésien de niveau 2 (figure 4.1 (b)) par suppression de 6 et création d'un sommet $E_6 = \{8, 9\}$ dont les parents sont 1, 2 et 5, c'est-à-dire à la fois les parents initiaux de 6 et ceux de la descendance proche de 6 ; on a donc

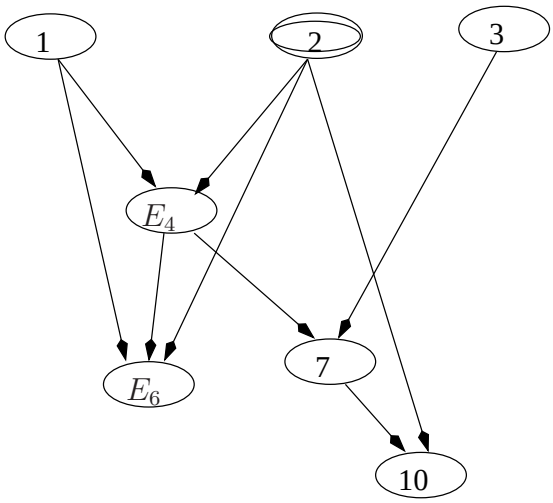
$$\begin{aligned} P_{I-6}(x_{I-6}) \;&=\; P_1(x_1)P_2(x_2)P_3(x_3)P_{4/1}(x_4/x_1)P_{5/2,4}(x_5/x_2, x_4)P_{E_6/1,2,5}(x_{E_6}/x_1, x_2, x_5) \\ &\quad P_{7/3,5}(x_7/x_3, x_5)P_{10/2,7}(x_{10}/x_2, x_7) \end{aligned}$$



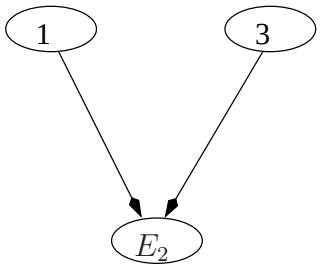
(a)



(b) $E_6 = \{8, 9\}$



(c) $E_4 = \{5\}$



(d) $E_2 = \{E_4, E_6, 7, 10\}$
$= \{5, 7, 8, 9, 10\}$

FIG. 4.2 – *(a) : Réseau bayésien initial (les nœuds dans le complémentaire de la cible sont entourés). (b) : Réseau bayésien de niveau* 2 *obtenu après marginalisation sur* 6*. (c) : Réseau bayésien de niveau* 2 *obtenu après marginalisation sur* 4*. (d) : Réseau bayésien de niveau* 2 *obtenu après marginalisation sur* 2*.*

Si nous marginalisons maintenant par rapport à X_4, nous effectuons le calcul suivant

$$\sum_{x_4} P_{4/1}(x_4/x_1)\ P_{5/2,4}(x_5/x_2,x_4) \ = \ P_{E_4/1,2}(x_{E_4}/x_1,x_2)\ ;$$

dans ce cas $E_4 = \{5\}$, et le réseau bayésien de niveau deux obtenu est celui de la figure 4.2 (c), sur l'ensemble $I - \{4,6\}$.

Maintenant il ne reste à faire que la marginalisation par rapport à X_2 qui s'effectue par la sommation suivante

$\sum_{x_2} P_{E_6/1,2,E_4}(x_{E_6}/x_1,x_2,x_{E_4})P_{E_4/1,2}(x_{E_4}/x_1,x_2)P_{10/2,7}(x_{10}/x_2,x_7)P_{7/3,5}(x_7/x_3,x_5)P_2(x_2)$
$= P_{E_6,E_4,7,10/1,3}(x_{E_6},x_{E_4},x_7,x_{10}/x_1,x_3),$

d'où la nécessité de créer un sommet $E_2 = \{E_6, E_4, 7, 10\}$, qui nous donne la structure de réseau bayésien de niveau 2 représenté sur la figure 4.1 (d) .

La loi de ce dernier réseau bayésien de niveau deux coïncide avec la loi de $(X_i)_{i\in A}$:

$$P_A(x_A) \ = \ P_1(x_1)\ P_3(x_3)\ P_{E_2/1,3}(x_{E_2}/x_1,x_3).$$

4.2.3 Algorithme des restrictions successives version 1

Soit $(I, G, (P_{i/p(i)})_{i\in I})$ un réseau bayésien de niveau 1. Etant donné une partie A de I, dite cible, considérons le problème de calcul de la loi des variables aléatoires $(X_i)_{i\in A}$. Pour résoudre ce problème nous proposons un algorithme de calcul dit *algorithme des restrictions successives*; l'idée générale est de gérer la succession des intégrations (ici où les variables aléatoires sont discrètes il s'agit de sommations) à effectuer relativement aux variables aléatoires X_ℓ pour $\ell \in \overline{A}$ (complémentaire de A dans I). C'est cette façon de gérer les sommations qui permet de garder une représentation sous forme de réseau bayésien de niveau deux de la loi cible.

On suppose que toutes les feuilles du réseau donné sont contenues dans la cible ; s'il n'en est pas ainsi, on commence par supprimer du graphe tous les nœuds qui n'ont pas de descendants dans A car ils n'interviennent pas dans le calcul de P_A ; autrement dit, comme en l'avait déjà indiqué au chapitre 1 (1.1.3), on se restreint à la partie commençante engendrée par A, notée A^+.

4.2.4 Principe de l'algorithme

Etant donné un RB1 sur l'ensemble I, l'idée générale de l'algorithme des restrictions successives, afin de calculer P_A, où $A \subset I$, est de construire une suite de parties $(I_0, \ldots, I_\ell)$ avec $\ell = card(\overline{A})$, et, pour tout $0 \leq s \leq \ell$, un réseau bayésien de niveau deux sur I_s, noté $R_s \ = \ (\mathcal{I}_s, G_s, (P_{J/p(J)})_{J\in\mathcal{I}_s})$, caractérisant $P_{\mathcal{I}_s}$, loi de $X_{I_s} = (X_i)_{i\in I_s}$, de telle sorte que :

1. $\mathcal{I}_0$ est le réseau initial (et donc $I_0 = I$).
2. La partie I_{s+1} contient un élément de moins que I_s, pris dans $\overline{A}$.

3. Tout atome de $\mathcal{I}_s$ est contenu soit dans A, soit dans $I_s - A$; s'il est contenu dans $I_s - A$, c'est un singleton.
4. Toute feuille de R_s est contenue dans A.
5. A la fin de l'algorithme, $I_\ell = A$ et la loi de $(X_i)_{i\in A}$ s'obtient simplement par produit des lois conditionnelles dans le réseau bayésien de niveau deux obtenu sur I_ℓ .

Précisons le pas général de cet algorithme : nous disposons en entrée de ce pas d'un RB2, sur $L \subset I$, $R = (\mathcal{I}, G, (P_{J/p(J)})_{J\in\mathcal{I}})$, caractérisant P_L, où $L = \cup_{J\in\mathcal{I}} J$, nous choisissons $i \in L - A$, tel que aucun de ses descendants ne soit dans $L - A$, nous obtenons en sortie le RB2, sur $L' = L - \{i\}$, $R' = (\mathcal{I}', G', (P_{J/p(J)})_{J\in\mathcal{I}'})$ caractérisant $P_{L'}$.

Le déroulement de ce pas peut être visualisé sur l'exemple de la figure 4.3 où $L = \{1, \ldots, 14\}$ et où on a choisi $i = 6$ (ce qui implique que $7, 8, 9, 10, 11, 12$ et 13 sont tous dans la cible A). Dans R comme dans R', les éléments de la partition ($\mathcal{I}$ en (a) ou $\mathcal{I}'$ en (b)) sont entourés par un trait plein ; les sous-ensembles qui seront à considérer par les besoins de la description du pas général, et du lemme qui la suit, sont entourés d'un trait tireté.

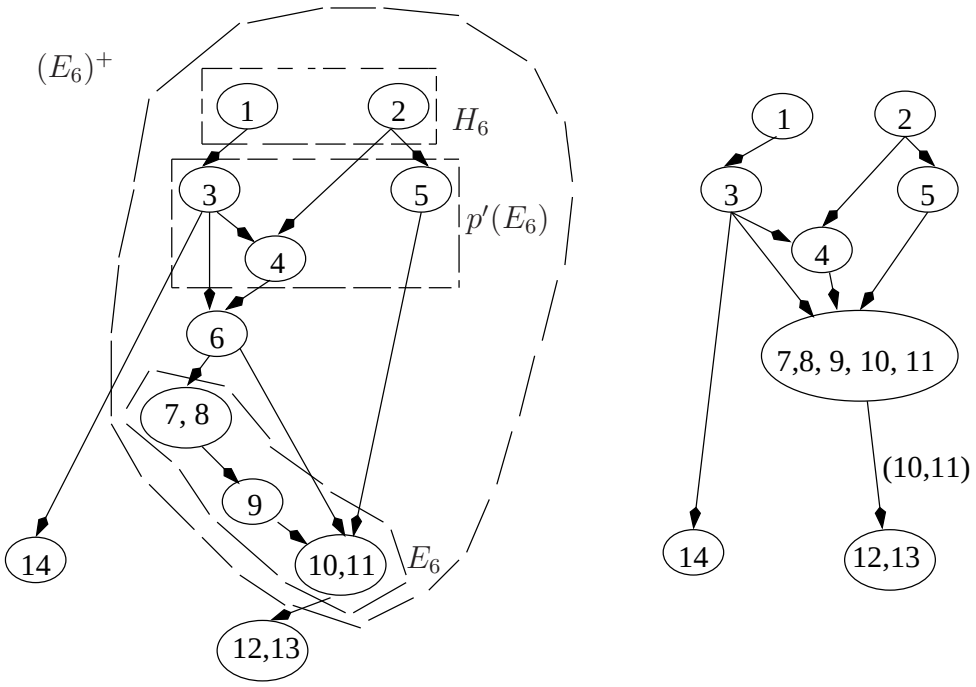


FIG. 4.3 – *(a) : Différents sous-ensembles définis à partir du sommet* 6. *(b) : RB*2 *obtenu après marginalisation sur* 6 *:* $E_6 = \{7, 8, 9, 10, 11\}$.

Choisissons un élément i de $\overline{A} \cap L$; d'après l'hypothèse de récurrence, le singleton $\{i\}$ appartient à $\mathcal{I}$; rappelons que ce choix est effectué de manière que i n'ait aucun descendant dans $\overline{A}$. D'après la propriété 4 (déjà vraie pour R_0), $\{i\}$ n'est pas une feuille de R et donc

la descendance proche de i est non vide ; nous la noterons E_i.

Alors le RB2 R' a pour sommets les atomes de la partition $\mathcal{I}'$ qui sont :

1. la descendance proche de i dans R, E_i, qui est, en vertu du choix de i, contenue dans A ;
2. tous les atomes de $\mathcal{I}$ autres que i et ceux dans E_i.

Il en résulte que la propriété 3 passe du RB2 R au RB2 R' : les atomes de $\mathcal{I}'$ sont contenus soit dans A, soit dans $L' - A$ (et alors ce sont des singletons).

Le graphe G' se déduit du graphe G de la manière suivante :

1. on supprime tous les liens concernant i et les éléments de E_i dans R,
2. on conserve tous les autres liens de G,
3. on affecte à E_i un ensemble de parents, $p'(E_i)$, qui comprend :
   (a) tous les parents de i dans G ;
   (b) tous les parents des sommets appartenant à la descendance proche de i, autres que i et ceux dans E_i lui-même.
4. on affecte à E_i comme enfants tous les enfants des sommets de R regroupés dans E_i, autres que ceux dans E_i lui-même.

Les données de probabilité associées à R' se déduisent de celles associées à R de la manière suivante :

1. on conserve la loi, conditionnellement à ses parents, pour tout sommet dont le passage de R à R' n'a modifié ni lui-même ni ses parents (c'est-à-dire tout sommet autre que i, ceux dans E_i et les enfants de ceux dans E_i).
2. pour tout enfant J de E_i (dans G), sa loi, conditionnellement à ses parents, se conserve en substituant E_i à l'ensemble de ceux des parents de J (dans G) qui appartenaient à E_i (on conserve l'information que seuls ces derniers interviennent effectivement dans p'_{J/E_i}) ; en effet, conditionnellement à ses parents dans G, J est indépendant de ceux des sommets de E_i qui ne sont pas des parents de J ; ceci résulte du fait qu'aucun de ces sommets n'est un descendant de J.
3. on crée la loi de E_i conditionnellement à $p'(E_i)$, selon la formule :

$$P_{E_i/p'(E_i)}(x_{E_i}/x_{p'(E_i)}) \;=\; \sum_{x_i \in \Omega_i} \Big[\Big(\prod_{J \in E_i} P_{J/p(J)}(x_J/x_{p(J)}) \Big) \; P_{i/p(i)}(x_i/x_{p(i)}) \Big].$$

Remarque

Nous pouvons écrire autrement la formule qui permet de calculer la loi de E_i conditionnellement à $p'(E_i)$:

$$P_{E_i/p'(E_i)}(x_{E_i}/x_{p'(E_i)}) \;=$$
$$\textstyle\prod_{J \in E_i - e(i)} P_{J/p(J)}(x_J/x_{p(J)}) \sum_{x_i \in \Omega_i} \Big[\Big(\prod_{J \in e(i)} P_{J/p(J)}(x_J/x_{p(J)}) \Big) \; P_{i/p(i)}(x_i/x_{p(i)}) \Big]$$

cela provient du fait que $\prod_{J \in E_i - e(i)} P_{J/p(J)}(x_J/x_{p(J)})$ ne dépend pas de X_i, et donc nous pouvons le sortir de la somme, et faire le produit avec le résultat de la sommation.

4.2.5 Justification de l'algorithme

Toutes les relations d'indépendance conditionnelle qui assurent le passage du RB2 R au RB2 R' (voir ci-dessus la description du pas général de l'algorithme) sont évidentes par construction. La seule démonstration à effectuer est celle qui justifie le calcul de la loi du "nouveau nœud" E_i conditionnellement à ses parents.

Lemme 1 *La loi de X_{E_i} conditionnellement à $X_{p'}(E_i)$ vérifie :*

$$P_{E_i/p'(E_i)}(x_{E_i}/x_{p'(E_i)}) \;=\; \sum_{x_i\in\Omega_i} \Big[\Big(\prod_{J\in E_i} P_{J/p(J)}(x_J/x_{p(J)}) \Big)\; P_{i/p(i)}(x_i/x_{p(i)}) \Big].$$

Démonstration

Pour alléger les notations, notons $\sum_i$ pour $\sum_{x_i\in\Omega_i}$ et, pour toute partie $B=\{b_1,\ldots,b_m\}$ de L, $\sum_B$ pour $\sum_{x_{b_1}\in\Omega_{b_1}} \ldots \sum_{x_{b_m}\in\Omega_{b_m}}$; nous omettons d'écrire les variables (par exemple on écrit $P_{J/p(J)}$ pour $P_{J/p(J)}(x_J/x_{p(J)})$).

Le lecteur peut visualiser les différents objets intervenant dans cette démonstration sur l'exemple donné ci-dessus (figure 4.3(a)) dans la description du pas général de l'algorithme : $i=6$; les J appartenant à E_i (autrement dit, avec l'abus de langage que nous commettons entre famille de parties disjointes de I et union des éléments de cette famille, les J contenus dans E_i) sont $\{7,8\}$, $\{9\}$ et $\{10,11\}$; $p'(E_i)$ se compose de 3 et 4 (car parents de i) et 5 (car parent de $\{10,11\}$ qui est dans E_i).

Le calcul de $P_{E_i/p'(E_i)}$ peut s'effectuer en se limitant à $(E_i)^+$ partie commençante définie par E_i , qui contient $p'(E_i)$ par construction ; décomposons $(E_i)^+$ selon la partition $(E_i,\{i\},p'(E_i),H_i)$, où $\Big(p'(E_i)\Big)^+ = p'(E_i)\cup H_i$. Alors

$$P_{E_i\cup p'(E_i)} \;=\; \sum_i \sum_{H_i} \Big[\Big(\prod_{J\in E_i} P_{J/p(J)} \Big)\; P_{i/p(i)} \Big(\prod_{k\in p'(E_i)} P_{k/p(k)} \Big) \Big(\prod_{h\in H_i} P_{h/p(h)} \Big) \Big].$$

L'indice i n'intervient que dans $\Big(\prod_{J\in E_i} P_{J/p(J)}\Big)\; P_{i/p(i)}$ alors que les ℓ appartenant à H_i ne peuvent intervenir que dans $\Big(\prod_{k\in p'(E_i)} P_{k/p(k)}\Big) \Big(\prod_{h\in H_i} P_{h/p(h)}\Big)$; donc

$$P_{E_i\cup p'(E_i)} \;=\; \Big\{ \sum_i \Big[\Big(\prod_{J\in E_i} P_{J/p(J)} \Big)\; P_{i/p(i)} \Big] \Big\} \Big\{ \sum_{H_i} \Big[\Big(\prod_{k\in p'(E_i)} P_{k/p(k)} \Big) \Big(\prod_{h\in H_i} P_{h/p(h)} \Big) \Big] \Big\}.$$

Or $P_{p'(E_i)} \;=\; \sum_{H_i} \Big[\Big(\prod_{k\in p'(E_i)} P_{k/p(k)} \Big) \Big(\prod_{h\in H_i} P_{h/p(h)} \Big) \Big]$. Donc

$$P_{E_i/p'(E_i)} \;=\; \frac{P_{E_i\cup p'(E_i)}}{P_{p'(E_i)}} \;=\; \sum_i \Big[\Big(\prod_{J\in E_i} P_{J/p(J)} \Big)\; P_{i/p(i)} \Big],$$

■

4.2.6 Evaluation de nombre d'opérations élémentaires dans le déroulement de l'algorithme

Dans la présentation du pas général de l'algorithme des restrictions successives, nous avons donné le choix suivant de la variable aléatoire, ne faisant pas partie de la cible, relativement à laquelle s'effectue la marginalisation : celle-ci ne doit pas avoir de descendants dans le complémentaire de la cible. Dans le cas où nous disposons de plusieurs choix entre les variables du complémentaire de la cible, nous proposons de choisir la variable qui minimise le nombre d'opérations arithmétiques que nous avons à faire lors du calcul de la distribution de probabilité donné en lemme 1.

Déterminons tout d'abord le nombre d'opérations arithmétiques effectuées à chaque pas du déroulement de l'algorithme ; rappelons la formule de calcul de la probabilité conditionnelle constituée lors de la marginalisation sur le nœud i :

$$P_{E_i/p'(E_i)}(x_{E_i}/x_{p'(E_i)}) \;=\; \sum_{x_i\in\Omega_i} \Big[\Big(\prod_{J\in E_i} P_{J/p(J)}(x_J/x_{p(J)}) \Big) \; P_{i/p(i)}(x_i/x_{p(i)}) \Big]. \quad (4.1)$$

Le nombre exact d'opérations arithmétiques effectuées à chaque pas i de l'algorithme est obtenu à partir de cette formule ; pour le donner introduisons les notations suivantes :

N_i : le nombre d'opérations arithmétiques effectuées à chaque pas i.

$|E_i| = \text{card}(E_i)$: le nombre de variables aléatoires regroupées dans E_i.

$|\Omega_i| = \text{card}(\Omega_i)$: cardinal de Ω_i, en d'autres termes le nombre de valeurs de la variable aléatoire X_i.

$|\Omega_J|$ où $J \subset I$: nombre de valeurs de la variables X_J, $|\Omega_J| = \prod_{j\in J} |\Omega_j|$.

L'équation (4.1) montre que le calcul de $P_{E_i|p'(E_i)}$ nécessite $|\Omega_{E_i\cup p'(E_i)}|$ termes. Chaque terme nécessite $|\Omega_i|$ sommations et $|E_i+1|$ multiplications ; le nombre total des opérations arithmétiques est alors donné par :

$$N_i = |\Omega_{E_i\cup p'(E_i)}| \; |\Omega_i| \; |E_i+1| \;=\; \Big(\prod_{j\in\{i\}\cap E_i\cup p'(E_i)} |\Omega_j| \Big) \; |E_i+1|$$

Soit $d =$ Card $(\bar{C})$ et $(i_1, \ldots, i_d)$ la suite des éléments de $\bar{C}$ dans l'ordre dans lequel l'algorithme procède aux marginalisations ; alors le nombre total d'opérations arithmétiques est :

$$\sum_{\ell=1}^{d} N_{i_\ell} = \sum_{s=1}^{d} \Big(\prod_{j\in\{i_\ell\}\cap E_{i_\ell}\cup p'(E_{i_\ell})} |\Omega_j| \Big) \; |E_{i_\ell}+1|.$$

4.2.7 Généralisation facile et importante

Nous avons présenté cette première version de l'algorithme des restrictions successives en prenant comme point de départ un réseau bayésien de niveau 1, et de ce fait, quoique, dès le second pas, l'algorithme manipule des RB2, il se trouve que les regroupements de nœuds

effectués ne portent que sur des nœuds qui sont dans la cible et donc les sommations ne portent, à chaque pas, que sur des variables aléatoires encore isolées, non dans la cible.

IL n'y aurait évidemment aucune différence fondamentale si le point de départ était un RB2 R_0 et la cible une union d'atomes de la partition $\mathcal{I}_0$ entrant dans la définition de ce RB2.

Il suffit de modifier la rédaction des principes 2 et 3 de l'algorithme, énoncés en 4.2.4, de la manière suivante :

2. I_{s+1} se déduit de I_s en en retirant un atome de I_s, contenu dans $\bar{A}$.

3. Tout atome de $\mathcal{I}_s$ est contenu soit dans A soit dans $I_s - A$; s'il est contenu dans $I_s - A$, c'est l'un des atomes de la partition initiale $\mathcal{I}_0$.

Dans le pas général de l'algorithme, le passage du RB2 R au RB2 R' se fait par sommation sur un atome de $\mathcal{I}_s$ contenu dans $\bar{A}$, choisi de manière qu'il n'ait aucun descendant contenu dans $\bar{A}$.

La formule de calcul démontrée au lemme 1 aurait alors la forme suivante où, pour en rappeler la définition, nous écrirons $dp(J)$ (descendance proche de J dans R) pour noter le nouveau nœud crée lors du pas général de l'algorithme :

$$P_{E_J/p'(E_J)}(x_{E_J}/x_{p'(E_i)}) \;=\; \sum_{x_J} \Big[\Big(\prod_{K \in dp(J)} P_{K/p(K)}(x_K/x_{p(K)}) \Big)\; P_{J/p(J)}(x_J/x_{p(J)}) \Big].$$

4.3 Calcul des lois conditionnelles

4.3.1 Introduction

L'objectif, dans cette section, est d'introduire un algorithme de calcul des lois conditionnelles dans les réseaux bayésiens. Soit $(I, G, (P_{i/p(i)})_{i \in I})$ un réseau bayésien de niveau un. Etant donné deux parties disjointes A et B de I , on pose le problème de calcul de la loi des variables aléatoires X_a (où $a \in A$) conditionnées par les variables aléatoires X_b (où $b \in B$).

Rappelons que la technique liée à l'algorithme LS (voir chapitre 3, section 4) utilise un produit de fonctions "potentiels", proportionnel à la loi de $X_{A \cup B}$ quand X_B (famille des variables aléatoires X_b, où $b \in B$) est fixée à la valeur donnée x_B, et donc aussi proportionnel à la loi de X_A conditionnellement à $[X_B = x_B]$; le calcul de la constante de normalisation nécessaire pour obtenir cette loi conditionnelle est intégré à l'algorithme.

Par contre ici, nous recourons à l'emploi du théorème de Bayes :

$$P_{A/B}(x_A/x_B) \;=\; \frac{P_{A \cup B}(x_{A \cup B})}{P_B(x_B)},$$

donc le calcul de cette loi conditionnelle revient à calculer en premier lieu la loi de la variable aléatoire $X_{A \cup B}$, et ensuite à calculer la constante de normalisation sous la forme de la loi de la variables aléatoire X_B. L'idée générale est de gérer la succession des intégrations

(sommations si les variables aléatoires sont discrètes) relativement aux variables aléatoires X_i pour $i \in \overline{A \cup B}$, puis relativement aux variables aléatoires X_i pour $i \in A$.

Dans la section précédente nous avons introduit l'algorithme des restrictions successives, qui permet de calculer des lois marginales d'une sous famille $(X_c)_{c \in C}$, où C est un sous-ensemble du réseau bayésien initial. Il serait donc possible pour calculer la loi conditionnelle, d'appliquer cet algorithme séparément à $A \cup B$ et à B, mais ceci conduirait en général à des calculs redondants.

Nous proposons donc ici une nouvelle version de l'algorithme des restrictions successives qui permet de calculer X_B en utilisant le calcul préalable de $X_{A \cup B}$; la différence avec la version précédente de l'algorithme tient à ce que ce dernier était initialisé sur un réseau bayésien ordinaire (RB1) alors que celui présenté ici est initialisé sur le RB2 représentant de $X_{A \cup B}$.

Il n'est pas possible d'appliquer ici directement la généralisation de la version 1 de l'algorithme présentée en 4.4.7 car ceci supposerait que B soit union d'atomes du RB2 issu du calcul de la loi de $X_{A \cup B}$ et en général il n'en sera pas ainsi.

Plus généralement soit un RB2 sur un ensemble 1 $(\mathcal{I}, G_{\mathcal{I}}, (P_{J/p(J)})_{J \in \mathcal{I}})$; on pose le problème de calcul de la loi jointe des variables aléatoires $(X_i)_{i \in C}$, pour $C \subset I$ (et non nécessairement compatible avec la partition $\mathcal{I}$) ; ce calcul revient à marginaliser par rapport aux variables aléatoires $X_{\overline{C}} = (X_\ell)_{\ell \in \overline{C}}$.

On va voir par la suite que, par notre nouvelle version de l'algorithme des restrictions successives, le passage d'un réseau bayésien de niveau deux à la loi jointe d'un sous-ensemble de variables aléatoires se passe de manière analogue au passage de la loi d'un réseau bayésien de niveau un à la loi d'un sous-ensemble de variables aléatoires de ce réseau.

Nous commençons par donner un exemple de calcul, nous expliquons par la suite le principe de calcul des lois conditionnelles dans ces différentes phases, et nous terminons par une démonstration du résultat.

4.3.2 Exemple introductif à l'algorithme

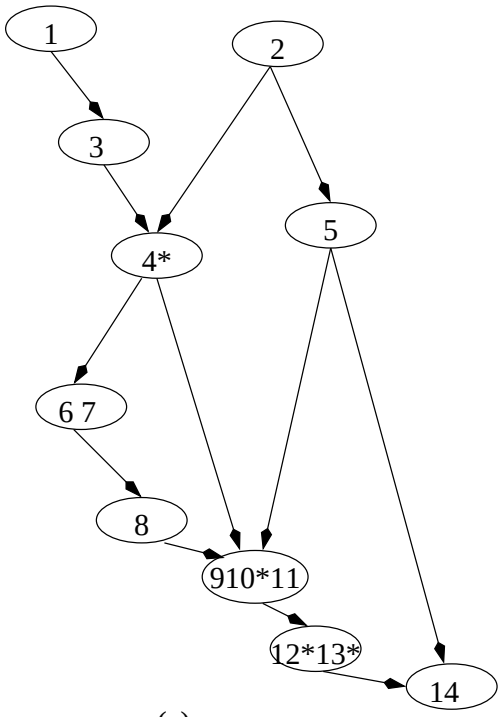
Soit le graphe (I, G), orienté sans circuits donné en figure 4.4 (a).

La loi jointe associée au réseau bayésien de niveau 2 ainsi défini s'écrit sous la forme

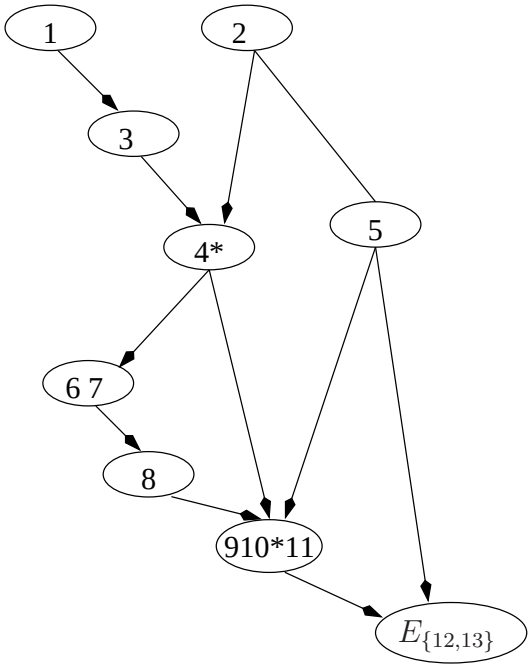
$$\begin{aligned} P_I(x_I) \;=\; & P_1(x_1)P_2(x_2)P_{3/1}(x_3/x_1)P_{4/2,3}(x_4/x_2,x_3)P_{5/2}(x_5/x_2)P_{6,7/4}(x_6,x_7/x_4) \\ & P_{8/6,7}(x_8/x_6,x_7)P_{9,10,11/4,5,8}(x_9,x_{10},x_{11}/x_4,x_5,x_8)P_{15/3}(x_{15}/x_3) \\ & P_{14/5,12,13}(x_{14}/x_5,x_{12},x_{13})P_{12,13/9,10,11}(x_{12},x_{13}/x_9,x_{10},x_{11}) \end{aligned}$$

On considère la cible $C = \{1,2,3,5,6,7,8,9,11,14,15\}$. Le calcul de P_C revient à marginaliser la loi jointe totale du réseau relativement aux variables indexées dans $\overline{C} = \{4,10,12,13\}$, c'est-à-dire sommer sur les valeurs prises par X_4, X_{10}, X_{12} et X_{13}.

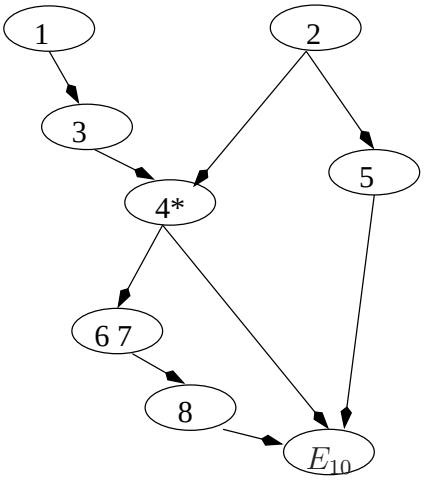
Nous allons choisir les variables de $\overline{C}$ l'une après l'autre pour la marginalisation.



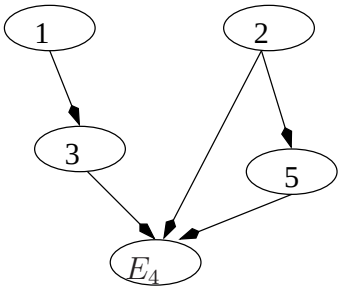
(a)



(b) $E_{\{12,13\}} = \{14\}$



(c) $E_{10} = \{9, 11, E_{\{12,13\}}\}$
$= \{9, 11, 14\}$



(d) $E_4 = \ 6, 7, 8, E_{10}\}$
$= \{6, 7, 8, 9, 11, 14\}$

FIG. 4.4 – *(a) : Réseau bayésien de niveau deux initial (les sommets non dans la cible sont signalés par une étoile). (b) : Réseau bayésien de niveau* 2 *obtenu après marginalisation sur* $12, 13$. *(c) : Réseau bayésien de niveau* 2 *obtenu après marginalisation sur* 10. *(d) : Réseau bayésien de niveau* 2 *obtenu après marginalisation sur* 4.

En premier lieu nous remarquons que parmi les sommets du RB2 contenant des éléments de $\bar{C}$, seul $\{12, 13\}$ n'a pas de descendants contenant aussi des éléments de $\bar{C}$. Il se trouve que 12 et 13 sont tous deux dans $\bar{C}$; nous commençons par marginaliser par rapport aux deux variables X_{12} et X_{13} et obtenons

$$\begin{aligned}&\textstyle\sum_{x_{12},x_{13}} P_{12,13/9,10,11}(x_{12}, x_{13}/x_9, x_{10}, x_{11}) P_{14/5,12,13}(x_{14}/x_5, x_{12}, x_{13})\\ &= P_{14/5,9,10,11}(x_{14}/x_5, x_9, x_{10}, x_{11});\end{aligned}$$

ceci explique que le graphe qui résulte de cette intégration est muni d'une structure de réseau bayésien de niveau 2 (figure 4.4 (b)) par suppression du sommet $\{12, 13\}$ et création d'un sommet $E_{\{12,13\}} = \{14\}$ dont les parents sont 5, 9, 10 et 11, c'est-à-dire à la fois les parents initiaux du sommet $\{12, 13\}$ et ceux de la descendance proche de $\{12, 13\}$.

Nous considérons ensuite l'élément de $\bar{C}$ 10 qui, dans le RB2, appartient au sommet $\{9, 10, 11\}$, alors que les autres éléments de ce sommet, 9 et 11, sont dans C ; pour marginaliser par rapport à X_{10}, nous effectuons le calcul suivant (résultat vérifiable directement, mais qui résulte aussi du lemme ci-dessous)

$$\begin{aligned}&\textstyle\sum_{x_{10}} P_{9,10,11/4,5,8}(x_9, x_{10}, x_{11}/x_4, x_5, x_8) P_{14/5,9,10,11}(x_{14}/x_5, x_9, x_{10}, x_{11})\\ &= P_{E_{10}/4,5,8}(x_{E_{10}}/x_4, x_5, x_8);\end{aligned}$$

dans ce cas on crée $E_{10} = \{9, 11, E_{\{12,13\}}\}$, et le réseau bayésien de niveau deux obtenu est celui sur la figure 4.4 (c).

Maintenant il ne reste à effectuer que la marginalisation par rapport à X_4 qui s'effectue par la sommation suivante

$$\begin{aligned}&\textstyle\sum_{x_4} P_{E_{10}/4,5,8}(x_{E_{10}}/x_4, x_5, x_8)\ P_{8/6,7}(x_8/x_6, x_7)\ P_{6,7/4}(x_6, x_7/x_4)\ P_{4/2,3}(x_4/x_2, x_3)\\ &= P_{E_{10},6,7,8/2,3}(x_{E_{10}}, x_6, x_7, x_8/x_2, x_3),\end{aligned}$$

d'où la nécessité de créer un sommet $E_4 = \{E_{10}, 6, 7, 8\}$, qui nous donne la structure de réseau bayésien de niveau 2 représenté sur la figure 4.4 (d) .

La loi de ce dernier réseau bayésien de niveau deux coïncide avec la loi de $(X_i)_{i\in C}$:

$$P_C(x_C) \;=\; P_1(x_1)P_2(x_2)P_{3/1}(x_3/x_1)P_{5/2}(x_5/x_2)P_{E_4/2,3}(x_{E_4}/x_2, x_3)P_{15/3}(x_{15}/x_3).$$

4.3.3 Algorithme des restrictions successives version 2

Soit $(\mathcal{I}, G, (P_{J/p(J)})_{i\in I})$ un réseau bayésien de niveau 2 sur un ensemble I, caractérisant la loi de X_I (c'est-à-dire que $\mathcal{I}$ est une partition de I) ; soit C une partie de I, non nécessairement union d'atomes de la partition ; notre but est de construire un réseau bayésien de niveau 2 sur C, caractérisant la loi de X_C.

$\bar{C}$ désignant le complémentaire de C dans I, on distingue deux types de sommets dans le réseau bayésien de niveau 2 initial : ceux qui ne contiennent aucun élément de $\bar{C}$ (dits

"$\bar{C}$-vides") et ceux qui en contiennent (dits "$\bar{C}$-non vides") ; soit ℓ le nombre de sommets $\bar{C}$-non vides. L'idée générale de l'algorithme est de construire une suite de parties de I, soit $(I_0, \ldots, I_\ell)$, et pour tout $0 \leq s \leq \ell$ un réseau bayésien de niveau deux sur I_s, noté R_s, de telle sorte que :

1. $\mathcal{I}_0$ est le réseau bayésien de niveau 2 initial (et donc $I_0 = I$).
2. La partie I_{s+1} se déduit de I_s par considération d'un sommet $\bar{C}$-non vide, soit K ; c'est la marginalisation par rapport à ceux des éléments de K qui appartiennent à $\bar{C}$ qui permet de passer de R_s à R_{s+1}, et I_{s+1} est donc égal à I_s privé de ces éléments.
3. A la fin de cet algorithme, $I_\ell = C$ et la loi de X_C s'obtient simplement par produit des lois conditionnelles dans le réseau bayésien de niveau deux obtenu sur I_ℓ .

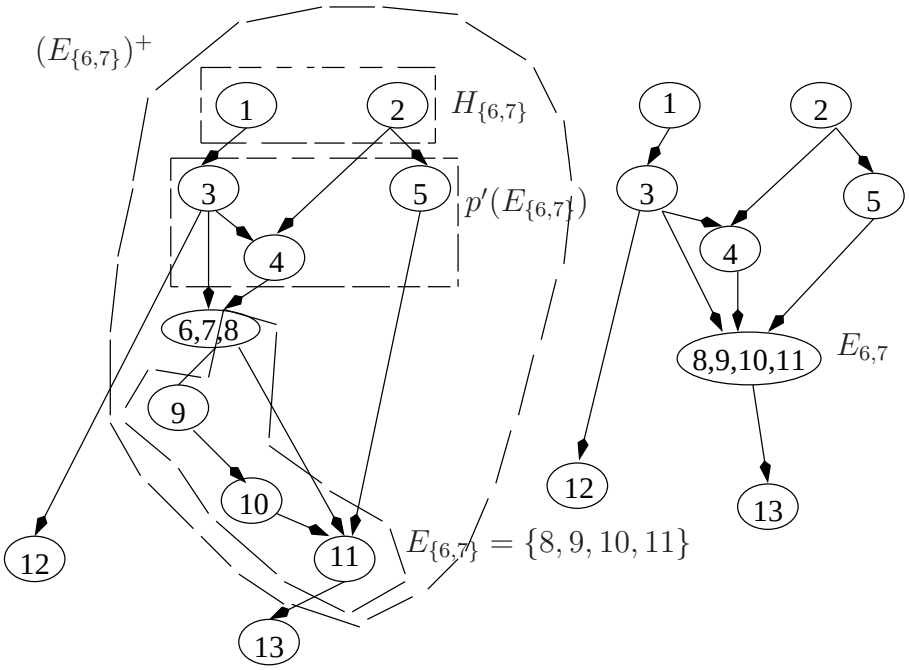


FIG. 4.5 – *(a) : Différents sous-ensembles définis à partir de* $\{6, 7\}$*. (b) : RB*2 *obtenu après marginalisation sur* $\{6, 7\}$ *:* $E_{6,7} = \{8, 9, 10, 11\}$.

Précisons le pas général de cet algorithme : nous disposons en entrée de ce pas d'un RB2, sur $L \subset I$, $R = (\mathcal{I}, G, (P_{J/p(J)})_{J \in \mathcal{I}})$ et nous obtiendrons en sortie $R' = (\mathcal{I}', G', (P_{J/p(J)})_{J \in \mathcal{I}'})$ qui est un RB2 sur une partie L' de L dont la construction est détaillée ci-dessous.

Choisissons un sommet J de $\mathcal{I}$, $\bar{C}$-non vide ; ce choix est effectué de manière que J n'ait aucun descendant lui même $\bar{C}$-non vide. Posons $J = J_1 \cup J_2$, où $J_1 \subset \bar{C}$ et $J_2 \subset C$.

Le déroulement de cette étape peut être visualisé sur l'exemple de la figure (4.5) où $J = \{6, 7\}$: dans R comme dans R', les éléments de la partition ($\mathcal{I}$ en (a) ou $\mathcal{I}'$ en (b)),

et du lemme qui la suit, sont entourés par un trait plein ; les sous-ensembles qui seront à considérer par les besoins de la description du pas général de cette phase sont entourés d'un trait pointillé.

Le réseau bayésien de niveau 2 R' a pour sommets les atomes de la partition $\mathcal{I}'$ de $L' = L - J_1$ qui sont :

1. l'union de J_2 et des atomes de $\mathcal{I}$ qui sont dans la descendance proche de J dans R, qu'on va noter E_{J_1},
2. tous les atomes de $\mathcal{I}$ autres que J et ceux dans E_{J_1} .

Le graphe G' se déduit du graphe G de la manière suivante :

1. on supprime tous les liens concernant J et les éléments de E_{J_1} dans R.
2. on conserve tous les autres liens de G.
3. on affecte à E_{J_1} un ensemble de parents, $p'(E_{J_1})$, qui comprend :
   (a) tous les parents de J pour G ;
   (b) tous les parents des sommets appartenant à la descendance proche de J, autres que J et ceux dans E_{J_1} lui-même.
4. on affecte à E_{J_1} comme enfants tous les enfants des sommets de R regroupés dans E_{J_1}, autres que ceux dans E_{J_1} lui-même.

Les données de probabilité associées à R' se déduisent de celles associées à R de la manière suivante :

1. on conserve la loi, conditionnellement à ses parents, pour tout sommet dont le passage de R à R' n'a modifié ni lui-même ni ses parents (c'est-à-dire tout sommet autre que J, ceux dans E_{J_1} et les enfants de ceux dans E_{J_1}).
2. pour tout enfant (pour G) de E_{J_1}, soit K, sa loi, conditionnellement à ses parents, se conserve en substituant E_{J_1} à l'ensemble de ceux des parents (pour G) de K qui appartenaient à E_{J_1} ; on conserve l'information que seuls ces derniers interviennent effectivement dans $p'_{K/E_{J_1}}$; en effet, conditionnellement à ses parents pour G, K est indépendant de ceux des sommets de E_{J_1} qui ne sont pas des parents de K ; ceci résulte du fait qu'aucun de ces sommets n'est un descendant de K.
3. on crée la loi de E_{J_1} conditionnellement à $p'(E_{J_1})$, par la formule :

$$P_{E_{J_1}/p'(E_{J_1})}(x_{E_{J_1}}/x_{p'(E_{J_1})})$$
$$= \sum_{x_{J_1} \in \Omega_{J_1}} \left[\left(\prod_{K \in dp(J)} P_{K/p(K)}(x_K/x_{p(K)}) \right) P_{J/p(J)}(x_{J_1}, x_{J_2}/x_{p(J)}) \right].$$

4.3.4 Justification de l'algorithme

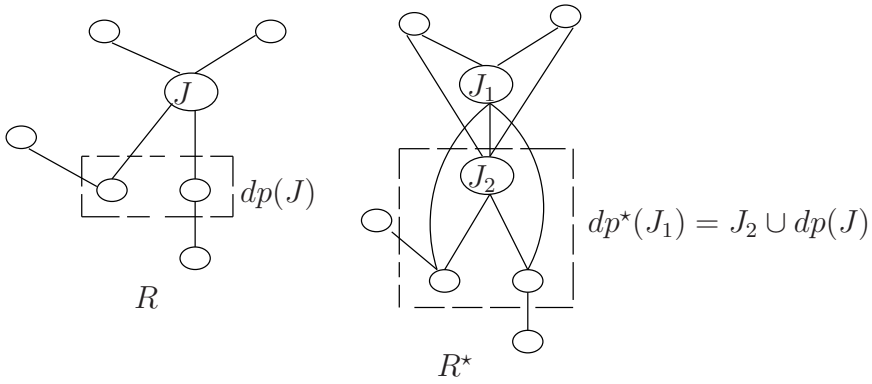
Comme pour la première version de l'algorithme, la seule justification à fournir est celle relative au calcul de $P_{E_{J_1}/p'(E_{J_1})}$

Lemme 2 *La loi de $X_{E_{J_1}}$ conditionnellement à $X_{p'(E_{J_1})}$ se calcule comme*

$$P_{E_{J_1}/p'(E_{J_1})}(x_{E_{J_1}}/x_{p'(E_{J_1})})$$
$$= \sum_{x_{J_1}\in\Omega_{J_1}} \left[\left(\prod_{K\in dp(J)} P_{K/p(K)}(x_K/x_{p(K)}) \right) P_{J/p(J)}(x_{J_1}, x_{J_2}/x_{p(J)}) \right].$$

Démonstration

Considérons les mêmes notations que pour le lemme 1. Nous cherchons à calculer $\sum_{J_1} P_{d_p(J)/p(d_p(J)} P_{J/p(J)}$; pour cela, il suffit de considérer un réseau bayésien de niveau deux équivalent à celui d'entrée du pas général de l'algorithme et distinguant ente J_1 ($\subset \bar{C}$) et J_2 ($\subset C$) comme le montre la figure suivante :



Autrement dit, étant fixé J ($\in \mathcal{I}$), qui contient strictement J_1 sur lequel nous voulons effectuer la sommation, nous introduisons un RB2, équivalent à R (c'est à dire qu'il est construit sur le même ensemble L et détermine la même loi pour les variables aléatoires indexées par L), noté $R^\star = (\mathcal{I}^\star, \mathcal{G}^\star, (P_{J/p(J)})_{J\in\mathcal{I}^\star})$ et qui se déduit de R par les opérations suivantes :

- toutes les parties de L constituant $\mathcal{I}^\star$ sont les mêmes constituant $\mathcal{I}$, à l'exception de J qui est remplacé par J_1 et J_2 (où $J_2 = J - J_1$)

- les parties dans $R^\star$ des éléments de $\mathcal{I}$ autres que J sont inchangés, sauf si un tel parent est J, auquel cas il est remplacé par les deux parents J_1 et J_2

- les parents de J_1 dans $R^\star$ sont les parents de J dans R

- les parents de J_2 dans $R^\star$ sont J_1 et les parents de J dans R.

Les seules probabilités conditionnelles à recalculer sont donc

$$\left| \begin{array}{l} P_{J_1/p^\star(J_1)} = P_{J_1/p(J_1)} = \sum_{J_2} P_{J_1\cup J_2/p(J)} \\ P_{J_2/p^\star(J_2)} = P_{J_2/J_1\cup p(J_1)} = \frac{P_{J_1\cup J_2/p(J)}}{P_{J_1/p^\star(J)}} \end{array} \right.$$

Dans le RB2 $R^\star$ la descendance proche de J_1 est clairement $dp^\star(J_1) = J_2 \cup dp(J)$.

Il résulte de la construction de R' que celui-ci se déduit de R par sommation sur J_1 et donc le lemme 2 n'est autre que l'application à $R^\star$ de la remarque faite en 4.2.7. ■

4.4 Application

Dans cette section, nous allons comparer les résultats obtenus par l'algorithme LS (Lauritzen et Spiegelhalter) et l'algorithme des restrictions successives ; la comparaison portera sur le nombre d'opérations arithmétiques effectuées par chaque algorithme pour calculer la loi de probabilité d'une même cible qu'on précisera.

Ces opérations arithmétiques élémentaires sont des additions, des multiplications et des divisions. Notons que les variables que nous allons considérer sont binaires ; pour une telle variable X, le calcul de la loi P_X peut être effectué soit comme les calculs séparés de $P(X = 0)$ et $P(X = 1)$ (qui, nous le verrons, font intervenir des multiplications) soit comme le calcul de (par exemple) $P(X = 0)$ puis $P(X = 1) = 1 - P(X = 0)$; la même remarque s'applique aux lois conditionnelles ; nous avons choisi ici de faire le dénombrement d'opérations en nous basant sur la première option ; le choix de la seconde ne modifierait pas l'équilibre entre les nombres globaux d'opérations par les deux méthodes.

Nous allons effectuer cette comparaison sur l'exemple "Visite en Asie" déjà présenté au chapitre 1.

Rappelons le graphe orienté associé à cette application, avec le codage suivant :

V : Visite en Asie

F : Fumer

B : Bronchite

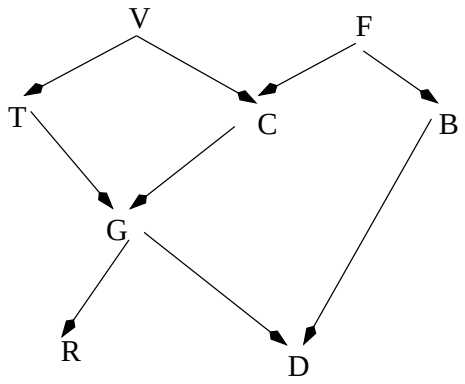
C : Cancer

T : Tuberculose

G : Grave maladie (Tuberculose ou Cancer)

D : Dyspnée

R : Radiologie



Supposons, dans cette application, que nous disposions d'informations complémentaires sur le réseau bayésien : le malade a visité l'Asie et il a des troubles de respiration (Dyspnée).

Nous allons chercher la loi de la variable "Fumer" conditionnellement à cette information.

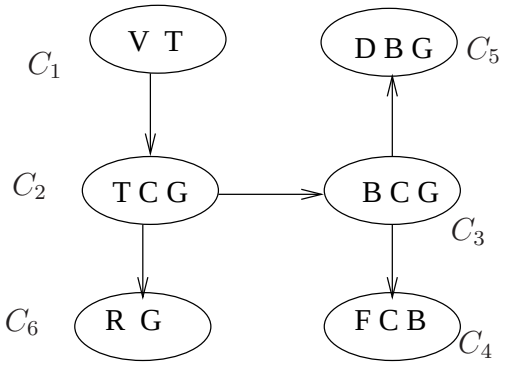
4.4.1 L'algorithme LS

L'application de l'algorithme LS, comme nous l'avons déjà présenté au chapitre 3, commence par la construction d'un arbre de jonction associé au réseau bayésien. Il s'agit d'un arbre construit sur un recouvrement $\mathcal{I}$ de I, qui doit être tel que la cible est contenue dans l'un des éléments de $\mathcal{I}$; ici, comme la cible est restreinte à un seul nœud (F), cette condition est automatiquement satisfaite.

Les éléments de $\mathcal{I}$ sont ordonnés de manière à constituer une suite possédant la propriété d'intersection courante (P.I.C.) (voir chapitre 3 section 3.3) ; l'application de l'un des algorithmes de construction d'une suite P.I.C. conduit à la suite $(C_1, \ldots, C_6)$ qui est donnée dans le tableau ci dessous, ainsi que les séparateurs $S_j = C_j \cap \cup_{k<j} C_k$ (pour $1 \leq j \leq 6$), les $R_j = C_j - S_j$ et les $\sigma(j)$ caractérisant la P.I.C. ($S_j \subseteq C_{\sigma(j)}$) ; on rappelle que dans l'arborescence définie par la suite P.I.C. tout arc est de la forme $(C_{\sigma(j)}, C_j)$ (où $2 \leq j \leq 6$)

j	C_j	S_j	R_j	$\sigma(j)$
1	VT		VT	
2	TCG	T	CG	1
3	BCG	CG	B	2
4	FCB	CB	F	3
5	DBG	BG	D	3
6	RG	G	R	2

Voici une représentation graphique de cette arborescence.

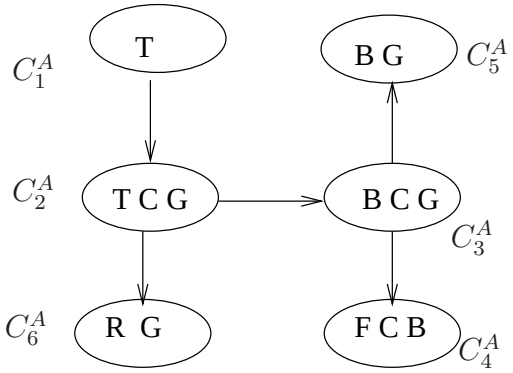


Notons que nous ne prenons pas en compte le déroulement de l'algorithme de construction de cette arborescence dans l'évaluation de la complexité du calcul de $P(F/V = 1, D = 1)$ par la méthode LS.

Afin de pouvoir appliquer l'algorithme LS, nous mettons en évidence la compatibilité de P_I avec le recouvrement $\mathcal{I}$: $P_I = \prod_{j=1}^{6} \tilde{f}_j$ où, pour tout j, $\tilde{f}_j$ est une fonction de variables constituant C_j, c'est à dire :

$\tilde{f}_1(V,T) = P(V)P(T/V)$
$\tilde{f}_2(T,C,G) = P(G/T,C)$
$\tilde{f}_3(B,C,G) = 1$
$\tilde{f}_4(F,C,B) = P(F)P(B/F)P(C/F)$
$\tilde{f}_5(D,B,G) = P(D/B,G)$
$\tilde{f}_6(R,G) = P(R/G)$.

Comme notre but est le calcul d'une loi conditionnellement à la fixation (à la valeur 1) des variables V et D, l'algorithme va utiliser en fait l'arborescence définie sur $A = I - \{V,D\} = \{F,T,C,B,G,R\}$



On constate que cette suite n'est pas propre puisque $C_1^A \subset C_2^A$ et $C_5^A \subset C_3^A$.

Comme nous l'avons indiqué en section 3.4.7, nous pouvons ici soit "nettoyer l'arborescence" (c'est à dire supprimer C_1^A et C_5^A) soit faire tourner l'algorithme sur cette arborescence non propre, en sachant qu'il y a deux pas qui ne conduiront à aucun calcul effectif : un pas dans la première phase car $C_5^A \subset C_{\sigma(5)}^A$ et un dans la seconde phase car $C_{\sigma(2)}^A \subset C_2^A$; nous allons ici rédiger le déroulement de l'algorithme en conservant l'arborescence sans la nettoyer mais préciserons les modifications qu'aurait entraîné le nettoyage, c'est à dire la suppression de C_1^A et C_5^A.

Commençons par factoriser $P_{I-\{V,D\}}$: $P_{I-\{V,D\}/V=1,D=1} = \alpha \prod_{j=1}^{6} f_j$ où, pour tout j, f_j se déduit de $\tilde{f}_j$ en fixant à la valeur 1 celles des variables V et D qui appartiennent à C_j ; de manière précise :

$f_1(T) = \tilde{f}_1(V=1,T) = P(V=1)P(T/V=1)$
$f_2(T,C,G) = \tilde{f}_2(T,C,G) = P(G/T,C)$
$f_3(B,C,G) = \tilde{f}_3(B,C,G) = 1$
$f_4(F,C,B) = \tilde{f}_4(F,C,B) = P(F)P(B/F)P(C/F)$
$f_5(B,G) = \tilde{f}_5(D=1,B,G) = P(D=1/B,G)$
$f_6(R,G) = \tilde{f}_6(R,G) = P(R/G)$

Une "phase zéro" de l'algorithme consiste à dresser les tableaux des potentiels f_j ($1 \leq$

$j \leq 6$) ; dénombrons les multiplications nécessaires :

Calcul de f_1 : 2 multiplications.

Calcul de f_4 : 16 ($= 8 \times 2$) multiplications,

soit en tout 18 multiplications.

Le déroulement de la première phase de l'algorithme est alors le suivant (nous indiquons à chaque fois le nombre d'opérations effectuées)

Initialisation : $g_1(T) = f_1(T)$, $g_2(T,C,G) = f_2(T,C,G)$, $g_3(B,C,G) = f_3(B,C,G)$, $g_4(F,C,B) = f_4(F,C,B)$, $g_5(B,G) = f_5(B,G)$, $g_6(R,G) = f_6(R,G)$

Pas 1 : Modification de g_6 et $g_{\sigma(6)} = g_2$

$$\left|\begin{array}{l} h_6(G) = \sum_R g_6(R,G) \quad \text{2 additions} \\ g_6(R,G) := \frac{g_6(R,G)}{h_6(G)} \quad \text{4 divisions} \\ g_2(T,C,G) := g_2(T,C,G) \cdot h_6(G) \quad \text{8 multiplications} \end{array}\right.$$

Pas 2 : Modification de g_5 et $g_{\sigma(5)} = g_3$

$$\left|\begin{array}{l} h_5(B,G) = g_5(B,G) \quad \text{0 addition} \\ g_5(B,G) := \frac{g_5(B,G)}{h_5(B,G)} = 1 \quad \text{0 division} \\ g_3(B,C,G) := g_3(B,C,G) \cdot h_5(B,G) = g_3(B,C,G) \quad \text{0 multiplication} \end{array}\right.$$

(*Ce pas aurait être absent en cas de nettoyage, mais il ne comporte aucune opération.*)

Pas 3 : Modification de g_4 et $g_{\sigma(4)} = g_3$

$$\left|\begin{array}{l} h_4(B,C) = \sum_F g_4(B,C,F) \quad \text{4 additions} \\ g_4(B,C,F) := \frac{g_4(B,C,F)}{h_4(B,C)} \quad \text{8 divisions} \\ g_3(B,C,G) := g_3(B,C,G) \cdot h_4(B,C) \quad \text{8 multiplications} \end{array}\right.$$

Pas 4 : Modification de g_3 et $g_{\sigma(3)} = g_2$

$$\left|\begin{array}{l} h_3(C,G) = \sum_B g_3(B,C,G) \quad \text{4 additions} \\ g_3(B,C,G) := \frac{g_3(B,C,G)}{h_3(C,G)} \quad \text{8 divisions} \\ g_2(T,C,G) := g_2(T,C,G) \cdot h_3(C,G) \quad \text{8 multiplications} \end{array}\right.$$

Pas 5 : Modification de g_2 et $g_{\sigma(2)} = g_1$

$$\left|\begin{array}{l} h_2(T) = \sum_{C,G} g_2(T,C,G) \quad \text{6 additions} \\ g_2(T,C,G) := \frac{g_2(T,C,G)}{h_2(T)} \quad \text{8 divisions} \\ g_1(T) := g_1(T) \cdot h_2(T) \quad \text{2 multiplications} \end{array}\right.$$

Pas 6 :

Comme $S_1 = \emptyset$, on a $h_1 = 1$.

(*Ce pas aurait être absent en cas de nettoyage, mais il ne comporte aucune opération.*)

Le nombre total d'opérations effectuées dans cette première phase est 16 additions, 26 multiplications, et 28 divisions.

Maintenant, après avoir obtenu les transitions de conditionnement $g_j = P_{R_j/S_j}$ pour tout $1 \leq j \leq 6$, nous pouvons appliquer la deuxième phase d'algorithme afin de calculer toutes les distributions de probabilités P_{C_i} $(1 \leq i \leq 6)$.

Pas 1 : nous commençons par la racine, C_1 ;

$$P(T/V = 1, D = 1) = g_1(T).$$

donc nous effectuons 0 addition.

(*Ce pas aurait être absent en cas de nettoyage, mais il ne comporte aucune opération.*)

Pas 2 : comme $S_2 \subset C_1$, il vient que

$$P(S_2/V = 1, D = 1) = P(T/V = 1, D = 1) = g_1(T)$$

donc nous effectuons 0 addition.

$P(C, T, G/V = 1, D = 1) = P(C, G/T, V = 1, D = 1)P(T/V = 1, D = 1)$, ce qui donne 8 multiplications.

Pas 3 : comme $S_3 \subset C_2$, il vient que

$$P(S_3/V = 1, D = 1) = P(C, G/V = 1, D = 1) = \sum_T P(T, C, G/V = 1, D = 1)$$

donc nous effectuons 4 additions.

$P(B, C, G/V = 1, D = 1) = P(B/C, G, V = 1, D = 1)P(C, G/V = 1, D = 1)$, ce qui donne 8 multiplications.

Pas 4 : comme $S_4 \subset C_3$, il vient que

$$P(S_4/V = 1, D = 1) = P(B, C/V = 1, D = 1) = \sum_C P(B, C, F/V = 1, D = 1)$$

donc nous effectuons 4 additions.

$P(B, C, F/V = 1, D = 1) = P(F/C, B/V = 1, D = 1)P(B, C/V = 1, D = 1)$, ce qui donne 8 multiplications.

Pas 5 : comme $S_5 \subset C_3$, il vient que

$$P(S_5/V = 1, D = 1) = P(B, G/V = 1, D = 1) = \sum_C P(B, C, G/V = 1, D = 1)$$

donc nous effectuons 4 additions.

$P(B, G) = P(B, G/V = 1, D = 1)$, ce qui donne 0 multiplication.

(*Ce pas aurait être absent en cas de nettoyage, les 4 additions qu'il comporte sont effectivement inutiles comme surcroît d'information.*)

Pas 6 : comme $S_6 \subset C_2$, il vient que

$$P(S_6/V=1,D=1) = P(G/V=1,D=1) = \sum_{T,C} P(T,C,G/V=1,D=1)$$

donc nous effectuons 6 additions.

$P(R,G/V=1,D=1) = P(R,G/V=1,D=1)P(G/V=1,D=1)$, ce qui donne 4 multiplications.

Le nombre total d'opérations effectuées dans cette deuxième phase est 18 additions et 28 multiplications

Maintenant pour le calcul de $P(F/D=1,V=1)$, nous effectuons encore une marginalisation à partir de la clique FCB de la façon suivante :

$$P(F/D=1,V=1) = \sum_{B,C} P(F,C,B/V=1,D=1)$$

donc on effectue 6 additions.

Le nombre total d'opérations effectuées pour le calcul de $P(F/D=1,V=1)$ est alors :

$18+26+28 = 72$ multiplications

$16+18+6 = 40$ additions (*ou* 36 *en cas de nettoyage préalable*)

28 divisions.

4.4.2 L'algorithme des restrictions successives

Détails du calcul de la loi de F, conditionnée par $D=1$, $V=1$, par la méthode des restrictions successives.

1. Calcul de $P(D=1,V=1,F)$ pour les deux valeurs de F,
   En premier lieu l'algorithme supprime la variable aléatoire R (Résultats de radiologie) puisque elle n'intervient pas dans ce calcul.
   (a) Calcul de $P(D=1/F,G) = \sum_B P(D=1/B,G)P(B/F)$.
   Nous avons 4 couples de valeurs de (F,G) ; pour chaque valeur de B nous effectuons 2 multiplications et 1 addition, ce qui donne au total pour ce calcul 8 multiplications et 4 additions.
   (b) Calcul de $P(D=1/F,T,C) = \sum_G P(D=1/F,G)P(G/T,C)$.
   Nous avons 8 triplets de valeurs de (F,T,C) ; pour chaque valeur de G nous effectuons 2 multiplications et 1 addition, ce qui donne au total pour ce calcul 16 multiplications et 8 additions.
   (c) Calcul de $P(D=1/V=1,F,C) = \sum_T P(D=1/F,T,C)P(T/V=1)$.
   Pour les 4 couples de valeurs de (F,C) nous effectuons 8 multiplications et 4 additions.

(d) Calcul de $P(D = 1/V = 1, F) = \sum_C P(D = 1/V = 1, F, C)P(C/F)$.
Pour les 2 valeurs de F nous effectuons 4 multiplications et 2 additions.

2. Calcul de $P(D = 1, V = 1, F) = P(D = 1/V = 1, F)P(V = 1)P(F)$.
Pour les deux valeurs de F nous effectuons 4 multiplications.
3. Calcul de $P(D = 1, V = 1) = \sum_F P(D = 1, V = 1, F)$.
Nous effectuons 1 addition.
4. Calcul de $P(F/D = 1, V = 1)$.
Pour les deux valeurs de F, nous effectuons 2 divisions.

Le nombre total d'opérations effectuées pour le calcul de $P(F/D = 1, V = 1)$ est alors :

$8 + 16 + 8 + 4 + 4 = 40$ multiplications

$4 + 8 + 4 + 2 + 1 = 19$ additions

2 divisions.

Pour chaque type d'opération, son nombre d'occurrences dans l'algorithme est nettement inférieur à ce qu'il était pour l'algorithme LS.

4.5 Conclusion

Notre but initial, en introduisant l'algorithme des restrictions successives, étant de présenter un outil de calcul de loi marginale pour un sous-ensemble de variables aléatoires quelconque d'un réseau bayésien ou d'un réseau bayésien de niveau deux. Cet outil permet d'effectuer le calcul de façon symbolique afin de trouver une représentation sous forme d'un réseau bayésien de niveau deux dont la loi conjointe coïncide avec la loi de la cible.

Cet algorithme permet aussi de calculer des lois conditionnelles et il a l'avantage de s'adapter à n'importe quel sous-ensemble de sommets du graphe et aussi de présenter à chaque étape de calcul des résultats interprétables en termes de probabilités conditionnelles qu'on peut utiliser pour répondre à différentes questions dans différents domaines de diagnostic.

Un gain essentiel par rapport aux méthodes classiques est que celles ci ne peuvent, quelle que soit la procédure de calcul qu'elles mettent en œuvre, qu'aboutir à former la loi d'une cible sous forme extensive (et de ce fait se contentent en général, dans leur mise en œuvre, de cibles à un élément, en passant par l'extraction de l'information relative à des cibles à plusieurs éléments particulières, qui sont en général les cliques de l'arbre de jonction associé au réseau bayésien) ; dans le cas de variables aléatoires discrètes, exprimer la loi sous forme extensive signifie fournir la liste des valeurs de probabilités d'atteinte de chaque famille de valeurs x_i où pour tout i appartenant à la cible A, x_i est l'une des valeurs distinctes que peut prendre la variable aléatoire X_i, comme nous l'avons déjà expliqué dans l'introduction des réseaux bayésiens de niveaux deux (le nombre de telles familles est le produit des n_i où, pour tout i appartenant à A, n_i est le nombre de valeurs distinctes que peut prendre la variable aléatoire X_i). Avec l'algorithme des restrictions successives une information peut

rester disponible sur des causalités intérieures dans la cible (information qui est tuée par les résultats extensifs).

Chapitre 5

Conclusion

L'objectif que nous avons poursuivi tout au long de cette thèse est l'amélioration et le développement de méthodes de calcul de lois et de lois conditionelles dans les réseaux bayésiens.

Dans un premier temps, nous avons présenté les réseaux bayésiens, ainsi que les réseaux bayésiens de niveau deux, leurs propriétés ainsi que leur utilité dans le calcul de lois de probabilités. Ce calcul, comme nous l'avons déjà vu, se fait grâce à plusieurs algorithmes d'inférence exacte dans les réseaux bayésiens, le plus célébre étant celui de Lauritzen et Spiegelhalter, implémenté dans plusieurs logiciels de calcul, par exemple : Netica, Hugin, Genie, BNT, Analytica, Ideal, Java Bayes, etc. Des dérivés de cet algorithme sont également implémentés dans différents logiciels commerciaux ou académiques. La plupart de ces logiciels sont à télécharger gratuitement sur internet, où on trouve des versions d'évaluation, ou des versions limitées dans le nombre de variables aléatoires.

Dans la seconde partie de la thèse nous avons présenté un algorithme de calcul de lois et de lois conditionnelles dans les réseaux bayésiens (algorithme des restrictions successives) ; cet algorithme a l'avantage de s'adapter à n'importe quel sous-ensemble de variables aléatoires du réseau bayésien, et aussi de fournir à chaque étape de calcul des résultats interprétables en termes de probabilités conditionnelles qu'on peut utiliser pour répondre à différentes questions dans différents domaines de diagnostic.

L'application de l'algorithme des restrictions successives à différents problèmes a montré son efficacité. Malheureusement, pour des applications compliquées, le calcul exact est rarement possible sous les contraintes de temps et de mémoire. C'est particulièrement le cas des problèmes probabilistes impliquant un grand nombre de variables aléatoires, ou de dépendances, ou encore qui prennent un grand nombre de valeurs. Dans ce cas l'inférence exacte devient insurmontable et des méthodes d'approximation doivent être employées ; ceci est dit "l'inférence approchée". Beaucoup de méthodes approchées sont en voie de développement pour des calcul approchés dans les réseaux bayésiens de grandes tailles ; nous citons à titre d'exemple les travaux de [32], [29], [42], [46], [28] et [20]. La confrontation de nos méthodes et ces techniques d'approximation doit être un prolongement indispen-

sable à notre travail.

L'application de l'algorithme des restrictions successives pour le calcul de la loi marginale d'un sous-ensemble de variables aléatoires (cible) d'un réseau bayésien nous donne, comme nous l'avons déjà présenté dans le chapitre 4 de cette thèse, une structure de réseau bayésien de niveau deux représentant la loi de ce sous-ensemble de variables aléatoires. L'algorithme des restrictions successives dans ses étapes choisit un ordre sur les variables du complémentaire de la cible afin d'arriver à une structure d'un réseau bayésien de niveau deux défini sur cette cible ; on se pose la question de savoir si ce réseau bayésien de niveau deux dépend de l'ordre choisi sur les variables et s'il est interprétable dans le cadre des séparations étudiées au chapitre 2 de cette thèse ; nous comptons prochainement effectuer une mise au point sur cette question.

L'algorithme des restrictions successives permet de structurer le calcul de la loi marginale d'un sous-ensemble de variables aléatoires d'un réseau bayésien, c'est à dire le calcul dans le cas d'une seule cible (ou de deux cibles emboîtées pour les calculs de lois conditionnelles). Une autre question qu'on peut se poser est : pouvons nous utiliser l'algorithme des restrictions successives pour calculer les lois marginales de plusieurs cibles à la fois en évitant d'éventuels calculs répétitifs ? Sur ce sujet, l'utilisation des propriétés des réseaux bayésiens de niveau deux caractérisant une loi semble être une voie prometteuse ; nous avons commencé à l'explorer et comptons en proposer l'implémentation à nos partenaires informaticiens.

Annexe A

Graphe orienté sans circuits

Dans cette section nous rappelons quelques concepts élémentaires de la théorie des graphes nécessaires pour la compréhension de la thèse. L'utilisation des graphes et de leurs algorithmes s'est généralisée pour devenir un des outils mathématiques de base pour la modélisation et la résolution de très nombreux problèmes scientifiques et techniques. En effet les graphes sont un outil essentiel pour représenter les modèles probabilistes ou autres utilisations de l'intelligence artificielle et des systèmes experts. Beaucoup de résultats théoriques et algorithmiques de la théorie des graphes peuvent être utilisés pour analyser les différents aspects de ces domaines. On se réfèrera par exemple à [3], [4] et [21].

Nous nous sommes efforcés de respecter le vocabulaire français de théorie des graphes en évitant les transpositions à partir du vocabulaire anglais (par exemple nous parlons de "circuit" là où le terme anglais est "cycle"). En revanche, nous avons maintenu le vocabulaire usuel en théorie des réseaux bayésiens, qui fait parfois double emploi avec celui de la théorie des graphes et qui fait largement usage des analogies avec la généalogie : par exemple nous parlons de "parent" et non de "prédécesseur" ainsi que "d'enfant" et non de "successeur" ; de même nous désignons par le terme "racine" un nœud sans prédécesseurs, et par le terme "feuille" un nœud sans successeurs, même si le graphe orienté considéré n'est pas une arborescence.

A.1 Préliminaire sur les graphes orientés sans circuits

A.1.1 Graphe sans circuits

Un graphe orienté est un couple (I, G) où I est un ensemble fini, et G est une partie de l'ensemble des couples (i, j) d'éléments de I.

Si cela ne prête pas à confusion, nous pourons dire "le graphe G" (en précisant éventuellement "sur I") pour désigner "le graphe (I, G)".

Les éléments de I sont appelés **nœuds**, ou **sommets**.

Les éléments de G sont appelés **arcs**.

Un **chemin**, relativement à G, est une suite d'éléments, de I, $(i_0, \ldots, i_k)$ (où $k \geq 1$) vérifiant : $\forall\ s \in \{1, \ldots, k\}\ \ (i_{s-1}, i_s) \in G$; i_0 est l'origine du chemin et i_k son extrémité.

Un **circuit** est un chemin dont l'origine et l'extrémité sont identiques.

Nous nous intéressons essentiellement dans cette thèse aux graphes orientés sans circuits, et donc en particulier sans boucles : pour tout i, $(i, i) \notin G$.

Voici un exemple sur lequel on présentera toutes les notions et notations dans la suite de cette annexe :

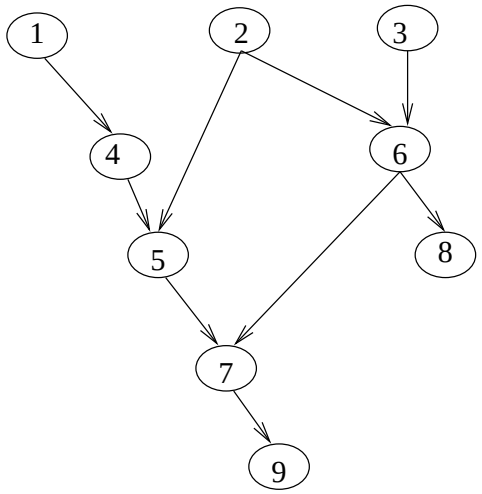


FIG. A.1 – Exemple d'un graphe orienté sans circuits (I, G).

$I = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$.
$G = \{(1, 4), (2, 5), (2, 6), (3, 6), (4, 5), (5, 7), (6, 7), (6, 8), (7, 9)\}$.

Remarque : On distinguera bien le vocabulaire des **graphes orientés sans circuits** de celui des **graphes non-orientés** : un graphe non-orienté est un couple (I, H) où H est un sous ensemble de l'ensemble des singletons $\{i\}$ et des paires $\{i, j\}$ d'éléments de I distincts (une paire est une partie à deux éléments).

A tout graphe orienté sans circuits (I, G) on peut associer le graphe non orienté **sous-jacent** (I, H) où H est l'ensemble des paires $\{i, j\}$ telles que l'un des couples (i, j) ou (j, i) appartienne à G (un seul de ces deux couples peut appartenir à G en raison de l'absence de circuits).

Les éléments de H sont appelés **arêtes**.

Dans un graphe non-orienté on appele **chaîne** toute suite $(i_0, \ldots, i_k)$ (où $k > 1$) vérifiant $\forall\ s \in \{1, \ldots, k\}\ \ \{i_{s-1}, i_s\} \in H$, et **cycle** une chaîne dont l'origine et l'extrémité sont identiques. Il peut se produire que dans le graphe non-orienté sous-jacent au graphe orienté sans circuits existent des cycles (dans l'exemple ci-dessus, on a le cycle $(2, 5, 7, 6, 2)$).

On doit signaler ici un risque de confusion entre le vocabulaire français et le vocabulaire

anglais, où circuit se dit “cycle” et un graphe orienté sans circuits se dit “Directed Acyclic Graph” (DAG).

A.1.2 Parties caractéristiques associées à un nœud.

Etant donné un graphe orienté sans circuits (I, G) on associe à tout élément i de I :

1. l'ensemble $p(i)$ de ses **parents** (aussi appelés prédécesseurs ou antécédents immédiats, ensemble noté également, en théorie des graphes, $\Gamma^-(i)$ ou $N^-(i)$) :
$$p(i) \;=\; \{j \;;\; (j,i) \in G\} \;;$$
2. l'ensemble $e(i)$ de ses **enfants** (aussi appelés successeurs, ou successeurs immédiats, ensemble noté également, en théorie des graphes, $\Gamma^+(i)$ ou $N^+(i)$) :
$$e(i) \;=\; \{j \;;\; (i,j) \in G\} \;;$$
3. l'ensemble $\ell(i)$ dit **lignée ascendante** de i, aussi appelé ensemble de tous ses antécédents, qui s'obtient par itération de la relation “être parent” :
$$\ell(i) \;=\; \{j \;;\; \exists (l_0, \ldots, l_k)\; l_0 = i \;, l_k = j \;,\; \forall\; s \in \{1, \ldots, k\} \;\; (l_s, l_{s-1}) \in G\} \;;$$
4. l'ensemble $d(i)$ de ses **descendants**, aussi dit descendance, ou lignée descendante, de i qui s'obtient par itération de la relation “être enfant” :
$$d(i) \;=\; \{j \;;\; \exists (l_0, \ldots, l_k)\; l_0 = i \;,\; l_k = j \;,\; \forall\; s \in \{1, \ldots, k\} \;\; (l_{s-1}, l_s) \in G\} \;;$$
5. l'ensemble $c(i)$ des **collatéraux** à i, qui est l'ensemble des éléments, autres que i lui même, qui ne sont ni dans sa lignée ni dans sa descendance :
$$c(i) \;=\; I \;-\; (\{i\} \cup \ell(i) \cup d(i)) \;;$$
6. l'ensemble $a(i)$ des **aïeux** ou ancêtres de i, qui est l'ensemble des éléments de la lignée ascendante de i autre que ses parents :
$$a(i) = \ell(i) - p(i).$$

Une **racine** d'un graphe est un nœud sans parent (les sommets 1, 2 et 3 du graphe de l'exemple sont ses racines) alors qu'une **feuille** est un nœud sans enfant (les nœuds 8 et 9 du graphe de l'exemple sont ses feuilles).

En théorie des graphes, les termes “racine” et “feuille” sont plutôt réservés au cas des arborescences ; dans le cas des graphes orientés les plus généraux, les termes correspondants seraient alors “source” et “puits”.

Remarque : Certains de ces ensembles peuvent bien sûr être vides.

Exemple :

Dans l'exemple on a pour l'élément 5 :

1. $p(5) = \{2, 4\}$.
2. $\ell(5) = \{1, 2, 4\}$.
3. $a(5) = \{1\}$.
4. $e(5) = \{7\}$.
5. $d(5) = \{7, 9\}$.
6. $c(5) = \{3, 6, 8\}$.

A.2 Numérotages des nœuds.

Soit I un ensemble fini tel que $\text{Card}(I) = n$.
Un numérotage des nœuds de I est une suite $(i_1, \ldots, i_n)$ telle que $\{i_1, \ldots, i_n\} = I$

A.2.1 Numérotation hiérarchique (aussi appelée ordre hiérarchique)

Définition
Une numérotation $(i_1, \ldots, i_n)$ *est dite* **hiérarchique** *pour* G *si :*

$$\forall\ s \in \{1, \ldots, n\}\ \ p(i_s) \subset \{i_1, \ldots, i_{s-1}\}.$$

Autrement dit tout élément est classé après ses parents.

Remarques :
Il résulte immédiatement de la définition que :

1. $$\forall\ s \in \{1, \ldots, n\}\ \ \ell(i_s) \subset \{i_1, \ldots, i_{s-1}\},$$
   c'est à dire que tout élément est classé après tous ceux de sa lignée.
2. $$\forall\ s \in \{1, \ldots, n\}\ \ d(i_s) \cap \{i_1, \ldots, i_{s-1}\} = \emptyset,$$
   c'est à dire qu'avant un élément donné n'est classé aucun de ses descendants.
3. En particulier : $p(i_1) = \emptyset$.
4. L'absence de circuits (en anglais : acyclicity) est une condition suffisante d'existence de numérotation hiérarchique, autrement dit, si dans un graphe il n'existe pas de numérotation hiérarchique, il existe des circuits ; c'est le cas, par exemple, du graphe suivant :

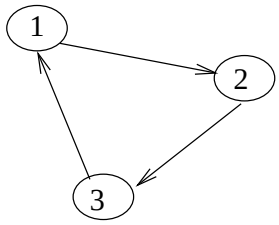


FIG. A.2 – Exemple d'un graphe avec circuit.

Définition

*Une partie J de I est appellée "***partie commençante***" si elle vérifie la propriété suivante :*

$$\forall j \in J \quad \ell(j) \subset J.$$

Il suffit évidemment, pour que J soit commençante, que pour tout $j \in J$ autre qu'une racine, ses parents soient aussi dans J. En effet le fait que tous les aïeux de j soient encore dans J s'en déduit par récurrence.

Il est clair que, si $(i_1, \ldots, i_n)$ est une numérotation hiérarchique, toute partie de la forme $\{i_1, \ldots, i_s\}$ (c'est à dire ensemble des sommets de numéro inférieur ou égal à s) est commençante, la réciproque étant fausse. Mais, si on se donne une partie commençante J de cardinal s, il est toujours possible de fabriquer une numérotation hiérarchique telle que $J = \{i_1, \ldots, i_s\}$.

A.2.2 Décomposition de I en classes selon G, Numérotation topologique (aussi appelée ordre topologique)

Définition

Etant donné un graphe orienté sans circuits sur I, il existe une partition unique de I dont les éléments seront appelés **générations** *ou* **classes***, notées $(C_0, \ldots, C_m)$ et définies de la manière suivante :*

- C_0 est la classe des sommets sans antécédents, aussi appelés racines.

- C_s $(1 \leq s \leq m)$ est l'ensemble des éléments i de I tels que le plus long chemin d'une racine à i soit de longueur égale à s.

Il est clair que les générations autres que C_0 peuvent également se définir par récurrence : $i \in C_s$ si et seulement s'il existe un chemin $(i_0, \ldots, i_s)$, $i_s = i$, où pour tout ℓ $(0 \leq \ell \leq s-1)$ $i_\ell \in C_\ell$.

Exemple :

Dans l'exemple précédent on constate la décomposition de I en cinq générations : C_0, C_1, C_2, C_3, C_4 telles que :

$C_0 = \{1, 2, 3\}$, $C_1 = \{4, 6\}$, $C_2 = \{5, 8\}$, $C_3 = \{7\}$, $C_4 = \{9\}$.

La décomposition en générations permet de définir un cas particulier des numérotations hiérarchiques : les numérotations topologiques.

Définition

Une numérotation $(i_1, \ldots, i_n)$ des éléments de I est dite **topologique** *si, étant donné deux générations C_j et $C_{j'}$, telles que $j < j'$, tous les éléments de C_j sont numérotés avant ceux de $C_{j'}$.*

En d'autres termes un tel numérotage décrit successivement $C_0, \ldots, C_m$

Un graphe orienté sans circuits possède toujours au moins un ordre topologique et contient au moins un nœud X_i sans parent.

Deux ordres topologiques distincts se distinguent uniquement par des permutation internes aux générations ; il existe en général plusieurs ordres topologiques sur un graphe orienté sans circuits sauf dans le cas particulier où celui ci est "linéaire" c'est à dire qu'il existe un numérotage $(i_1, \ldots, i_n)$ tel que les seuls arcs soient les couples (i_s, i_{s+1}) où $1 \leq s \leq n - 1$.

Ces ordres résultent de l'algorithme élémentaire obtenu en itérant à partir du graphe donné l'opération qui consiste à retirer un sommet sans parents ainsi que les arcs dont il est l'origine et à numéroter les sommets dans l'ordre de leurs suppressions.

A.3 Cas particuliers : arborescence, arbre, polyarbre

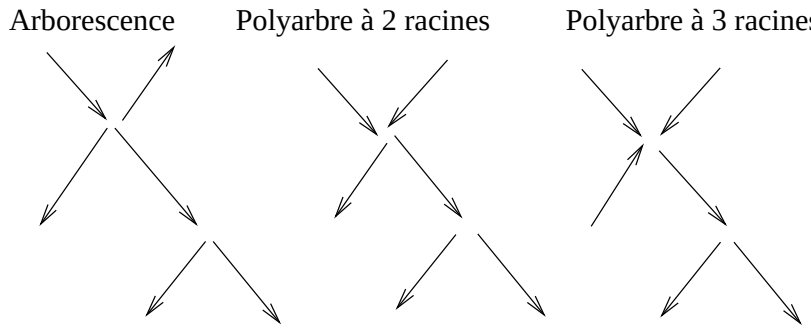
Une **arborescence** est un graphe orienté sans circuits dans lequel chaque nœud possède au plus un parent ; le graphe non-orienté sous-jacent à une arborescence est appelé **arbre** ; il est caractérisé par le fait qu'il ne contient pas de cycle (ou par le fait que son nombre d'arêtes est égal à celui des nœuds diminué d'une unité).

La terminologie anglaise pour arborescence est "tree", d'où une certaine confusion entre les acceptions du mot "arbre" dans des textes mathématiques ou informatiques en français. Ces notions sont utilisées dans cette thèse au chapitre 3 ; les définitions essentielles y sont rappelées, et les autres propriétés qui nous sont utiles y sont données en section 3.2.1.

Plus généralement, un graphe orienté tel que, s'il existe un chemin entre deux nœuds, celui-ci est unique, est appelé **polyarbre** ; le mot **polyarborescence** serait plus conforme aux régles de la terminologie française en théorie des graphes, mais à notre connaissance il n'est jamais employé.

Le graphe orienté sous-jacent à un polyarbre est un arbre.

Exemple : Arborescence et deux polyarbres admettant le même arbre sous-jacent.



Annexe B

Conditionnement et indépendance conditionnelle

Nous rappelons dans cette section les fondements du calcul de probabilités et probabilités conditionnelles nécessaires à la compréhension de la thèse. Nous présentons la notion de conditionnement et celle d'indépendance conditionnelle, et nous précisons également les conventions particulières ainsi que des notions adoptées pour ce travail.

B.1 Notions et notations de calcul des probabilités

Dans tout ce texte, nous manipulons des variables aléatoires X à valeurs dans des espaces mesurables $(\Omega, \mathcal{F})$; chaque $(\Omega, \mathcal{F})$ introduit est muni d'une mesure σ-finie μ dite dominante, relativement à laquelle toutes les lois de probabilités sont supposées absolument continues ; les densités de ces lois sont les outils de base de nos calculs ; si cela ne prête pas à une confusion, nous parlerons de l'espace mesurable Ω, en omettant la mention de la tribu $\mathcal{F}$.

On rappelle que si Ω est fini, on prend systématiquement pour tribu l'ensemble de toutes les parties de Ω et pour mesure μ la mesure de dénombrement ; nous ne le préciserons donc pas ; alors la densité de la loi d'une variable aléatoire X à valeurs dans Ω est l'application :

$$x \quad \rightsquigarrow \quad P(X = x).$$

Si on considère une famille d'espaces mesurables $(\Omega_i)_{i\in I}$, muni chacun de la mesure μ_i, l'espace produit $\prod_{i\in I} \Omega_i$ est systématiquement muni de la mesure produit $\prod_{i\in I} \mu_i$:

$$(\prod_{i\in I} \mu_i)(\prod_{i\in I} A_i) \;=\; \prod_{i\in I} \mu_i(A_i).$$

B.2 Conditionnement

Soient deux variables aléatoires Y_1 et Y_2, à valeurs respectivement dans Ω_1 (de mesure dominante μ_1) et Ω_2 (de mesure dominante μ_2).

Soit P_1 la loi de Y_1, P_2 la loi de Y_2 et $P_{1,2}$ la loi du couple (Y_1, Y_2). On rappelle qu'on appelle loi de Y_2 conditionnellement à Y_1, notée $P_{2/1}$, toute transition de probabilité :

$$y_1 \quad \rightsquigarrow \quad P_{2/1}(./y_1).$$

vérifiant :

- pour tout y_1, $P_{2/1}(./y_1)$ est une probabilité sur Ω_2.
- pour tout événement A_2 dans Ω_2, $P_{2/1}(A_2/.)$ est mesurable,
- pour tout événement A_1 dans Ω_1 et tout événement A_2 dans Ω_2, on a :

$$P_{1,2}(A_1 \times A_2) \; = \; \int_{A_1} P_{2/1}(A_2/y_1) P_1(dy_1).$$

Si $f_{1,2}$ est une densité de $P_{1,2}$ par rapport à $\mu_1 \times \mu_2$, P_1 admet pour densité par rapport à μ_1 l'application :

$$f_1 : y_1 \quad \rightarrow \quad \int_{\Omega_1} f_{1,2}(y_1, y_2) \mu_2(dy_2).$$

et, pour tout y_1, $P_{2/1}(./y_1)$ admet pour densité par rapport à μ_2 l'application :

$$y_2 \quad \rightsquigarrow \quad f_{2/1}(y_2/y_1) \; = \; \frac{f_{1,2}(y_1, y_2)}{f_1(y_1)} \; ; \tag{B.1}$$

on sait que $f_{2/1}(y_2/y_1)$ ainsi définie n'a pas de sens pour $f_1(y_1) = 0$, mais ceci n'est pas gênant car $P_1(\{y_1; f_1(y_1) = 0\}) = 0$ et car la loi de Y_2 conditionnellement à Y_1 n'est définie que P_1 presque sûrement relativement aux valeurs prises par Y_1. En conséquence, dans toute la thèse, nous divisons par des densités sans que nous nous préoccupions de leur nullité éventuelle et égalons des densités (conditionnelles ou absolues) sans avoir besoin de rappeler à chaque fois qu'il s'agit d'égalités presque sûres.

La formule (B.1) a dans la pratique un double usage :
- soit, connaissant $f_{1,2}$ (densité du couple), en déduire une densité de la loi de Y_2 conditionnellement à Y_1,
- soit, connaissant une densité marginale de Y_1 (f_1) et une densité conditionnelle de Y_2 relativement à Y_1 ($f_{2/1}$), en déduire la densité du couple :

$$f_{1,2}(y_1, y_2) \; = \; f_{2/1}(y_2/y_1) f_1(y_1). \tag{B.2}$$

Dans le cas général d'une famille de variables aléatoires $(X_i)_{i \in I}$, ces variables aléatoires étant numérotées $(i_1, \ldots, i_n)$, la formule (B.2) se généralise en la **formule des conditionnements emboîtés** :

$$f_{i_1,\ldots,i_n}(y_{i_1}, \ldots, y_{i_n}) \; = \; \prod_{\ell=1}^{n} f_{i_\ell/i_1,\ldots,i_{\ell-1}}(y_{i_\ell}/y_{i_1}, \ldots, y_{i_{\ell-1}}), \tag{B.3}$$

le facteur pour $\ell = 1$ étant par convention la densité marginale $f_{i_1}(y_{i_1})$.

B.3 Couple de variables aléatoires conditionnellement indépendantes

Soit donné trois espaces mesurés $(\Omega_k, \mathcal{F}_k, \mu_k)$ où $k \in \{0,1,2\}$, et, pour tout k une variable aléatoire Y_k à valeurs dans Ω_k, de loi absolument continue par rapport à μ_k.

Le théorème ci-dessous exprime l'équivalence de six propriétés, qui expriment l'indépendance de Y_1 et Y_2 **conditionnellement** à Y_0, ce qu'on note $Y_1 \perp\!\!\!\perp Y_2 \ / \ Y_0$

Théorème 12 *Les six propriétés suivantes (où les égalités sont presque-sûres) sont équivalentes*

(1) $f_{1,2/0}(y_1, y_2/y_0) = f_{1/0}(y_1/y_0) f_{2/0}(y_2/y_0)$

(1′) *il existe deux fonctions g_1, définie sur $\Omega_0 \times \Omega_1$, et g_2, définie sur $\Omega_0 \times \Omega_2$, telles que* $f_{1,2/0}(y_1, y_2/y_0) = g_1(y_0, y_1) g_2(y_0, y_2)$

(2) $f_{1/2,0}(y_1/y_2, y_0) = f_{1/0}(y_1/y_0)$

(2′) *il existe une fonction h_1, définie sur $\Omega_0 \times \Omega_1$, telle que* $f_{1/2,0}(y_1/y_2, y_0) = h_1(y_0, y_1)$

(3) $f_{2/1,0}(y_2/y_1, y_0) = f_{2/0}(y_2/y_0)$

(3′) *il existe une fonction h_2, définie sur $\Omega_0 \times \Omega_2$, telle que* $f_{2/1,0}(y_2/y_1, y_0) = h_2(y_0, y_2)$

Démonstration :

(3) et (3′) sont les analogues de (2) et (2′) par échange de 1 et 2.

Les propriétés (1′), (2′) et (3′) sont des atténuations respectivement de (1), (2) et (3) (par exemple (1) $\Longrightarrow$ (1′). Il suffit donc, pour établir le théorème de démontrer les implications suivantes :

$$(1') \implies (2') \implies (2) \implies (1)$$

(1′) $\Longrightarrow$ (2′)

Par hypothèse, $f_{1,2/0}(y_1, y_2/y_0) = g_1(y_0, y_1) g_2(y_0, y_2)$.

Il en résulte que

$$\begin{aligned} f_{2/0}(y_2/y_0) &= \int_{\Omega_1} f_{1,2/0}(y_1, y_2/y_0) d\mu_1(y_1) \\ &= \int_{\Omega_1} g_1(y_0, y_1) g_2(y_0, y_2) d\mu_1(y_1) \\ &= A_1(y_0) g_2(y_0, y_2) \end{aligned}$$

où $A_1(y_0) = \int_{\Omega_1} g_1(y_0, y_1) d\mu_1(y_1)$.

Alors

$$\begin{aligned} f_{1,2/0}(y_1/y_2, y_0) &= \frac{f_{0,1,2}(y_0, y_1, y_2)}{f_{2,0}(y_2, y_0)} \\ &= \frac{f_{1,2/0}(y_1, y_2/y_0) f_0(y_0)}{f_{2/0}(y_2/y_0) f_0(y_0)} \\ &= \frac{g_1(y_0, y_1) g_2(y_0, y_2)}{A_1(y_0) g_2(y_0, y_2)} \\ &= \frac{g_1(y_0, y_1)}{A_1(y_0)} \end{aligned}$$

On a bien mis en évidence une fonction h_1 de (y_0, y_1) telle que

$$f_{1/0,2}(y_1/y_0, y_2) = h_1(y_0, y_1).$$

$(2') \implies (2)$

Par hypothèse, $f_{1/2,0}(y_1/y_2, y_0) = h_1(y_0, y_1)$.

Alors

$$\begin{aligned} f_{1/0}(y_1/y_0) &= \frac{f_{0,1}(y_0, y_1)}{f_0(y_0)} \\ &= \frac{\int_{\Omega_2} f_{0,1,2}(y_0, y_1, y_2) d\mu_2(y_2)}{f_0(y_0)} \\ &= \frac{\int_{\Omega_2} h_1(y_0, y_1) f_{0,2}(y_0, y_2) d\mu_2(y_2)}{f_0(y_0)} \\ &= \frac{h_1(y_0, y_1) \int_{\Omega_2} f_{0,2}(y_0, y_2) d\mu_2(y_2)}{f_0(y_0)} \\ &= \frac{h_1(y_0, y_1) f_0(y_0)}{f_0(y_0)} \\ &= h_1(y_0, y_1) \end{aligned}$$

On a bien établi que $f_{1/2,0}(y_1/y_2, y_0) = f_{1/0}(y_1/y_0)$.

$(2) \implies (1)$

Par hypothèse, $f_{1/2,0}(y_1/y_2, y_0) = f_{1/0}(y_1/y_0)$.

Alors

$$\begin{aligned} f_{1,2/0}(y_1, y_2/y_0) &= \frac{f_{0,1,2}(y_0, y_1, y_2)}{f_0(y_0)} \\ &= \frac{f_{1/0,2}(y_1/y_0, y_2) f_{0,2}(y_0, y_2)}{f_0(y_0)} \\ &= f_{1/0}(y_1/y_0) \frac{f_{0,2}(y_0, y_2)}{f_0(y_0)} \\ &= f_{1/0}(y_1/y_0) f_{2/0}(y_2/y_0). \end{aligned}$$

■

B.4 Famille de variables aléatoires conditionnellement indépendantes

Etant donné $n+1$ variables aléatoires $Y_0, Y_1, \ldots, Y_n$, on doit distinguer (comme pour l'indépendance usuelle) entre l'indépendance, conditionnellement à Y_0, de tous les couples (Y_i, Y_j) (où $1 \leq j < j \leq n$) et l'indépendance, conditionnellement à Y_0, des variables aléatoires $(Y_1, \ldots, Y_n)$ "dans leur ensemble".

Nous dirons que $(Y_1, \ldots, Y_n)$ sont indépendantes (la précision "dans leur ensemble" étant sous-entendue), conditionnellement à Y_0, si on a la relation entre densités

$$f_{1,2,\ldots,n/0}(y_1, y_2, \ldots, y_n/y_0) \;=\; \prod_{i=1}^{n} f_{i/0}(y_i/y_0).$$

Ceci implique, par intégration, que pour tout couple (J, K) de parties disjointes de $\{1, \ldots, n\}$, si on note $Y_J = (Y_i)_{i \in J}$ et $Y_K = (Y_i)_{i \in K}$, les variables aléatoires Y_J et Y_K sont indépendantes conditionnellement à Y_0.

Une condition nécessaire et suffisante d'indépendance de $Y_1, \ldots, Y_n$ conditionnellement à Y_0 est que, pour tout $i \in \{2, \ldots, n\}$, Y_i et $Y_{\{1,\ldots,i-1\}}$ soient indépendantes conditionnellement à Y_0 ; en effet, à partir des $n-1$ égalités

$$f_{1,\ldots,i/0}(y_1, \ldots, y_i/y_0) = f_{1,\ldots,i-1/0}(y_1, \ldots, y_{i-1}/y_0) f_{i/0}(y_i/y_0)$$

on reconstate immédiatement que $f_{1,\ldots,n/0}(y_1, \ldots, y_n/y_0) = \prod_{i=1}^{n} f_{i/0}(y_i/y_0)$.

Evidemment, cette condition peut s'énoncer relativement à n'importe quel numérotage $(s_1, \ldots, s_n)$ de $\{1, \ldots, n\}$: $Y_1, \ldots, Y_n$ sont indépendantes conditionnellement à Y_0 si et seulement si, pour tout $j \geq 2$, Y_{s_j} et $Y_{\{s_1,\ldots,s_{j-1}\}}$ sont indépendantes conditionnellement à Y_0.

Annexe C

Complément sur les réseaux bayésiens

C.1 Notations

Cette annexe est consacrée à l'énoncé d'un certain nombre de propriétés toutes équivalentes à la définition des réseaux bayésiens et à la démonstration de leur équivalence.

Elle se situe dans le cadre d'une famille $(X_i)_{i\in I}$ de variables aléatoires définies sur un même espace de probabilité, I étant muni de la structure (G) de graphe orienté sans circuits. Les notions et notations relatives aux graphes orientés sans circuits figurent en Annexe A ci-dessus.

Pour tout i, X_i est à valeurs dans un ensemble Ω_i ; Ω_i est muni d'une structure d'espace mesurable, c'est à dire d'une tribu (σ-algèbre) pour laquelle nous n'avons pas besoin de fournir de notation. Chaque Ω_i introduit est muni d'une mesure σ-finie μ_i relativement à laquelle la loi P_i de X_i est absolument continue ; f_i désigne une densité de X_i par rapport à μ_i

Pour simplifier l'expression, on applique aux variables aléatoires la terminologie des graphes orientés sans circuits ; on dit par exemple que X_j est parent de X_i pour exprimer que l'indice j est parent de l'indice i dans le graphe (I, G).

Inversement, étant donné 3 parties disjointes de I, J, K, et L, on dit que J est indépendant de K conditionnellement à L pour exprimer que $X_J \perp\!\!\!\perp X_K \,/\, X_L$.

Pour simplifier l'écriture, pour toute partie J de I, on note $X_J = (X_i)_{i\in J}$; c'est une variable aléatoire à valeurs dans $\Omega_J = \prod_{i\in J} \Omega_i$; sa loi est notée P_J et f_J est une densité de P_J par rapport à μ_J, mesure produit des μ_i $(i \in J)$. Etant donné deux parties disjointes J et K de I, on note $P_{K/J}$ la loi de X_K conditionnellement à X_J et, pour tout $x_J \in X_J$, $f_{K/J}(./x_J)$ désigne une densité de la loi de probabilité $P_{K/J}(./J)$.

C.2 Théorème fondamental

C.2.1 Théorème

Les cinq propriétés suivantes sont équivalentes :

1. $$\forall\ i\ \in I \quad X_i\ \perp\!\!\!\perp\ X_{c(i)\ \cup\ a(i)} / \ X_{p(i)}.$$

2. $$f_I(x_I)\ =\ \prod_{i\ \in\ I} f_{i/p(i)}(x_i/x_{p(i)}).$$

3. *Il existe une famille de fonctions à valeurs positives ou nulles* $(g_i)_{i\in I}$ *telles que :*
   - pour tout i, g_i *est définie sur* $\Omega_i \times \Omega_{p(i)}$.
   - pour tout i, *il existe* $A_i \in]0,+\infty[$ *tel que, pour tout* $x_{p(i)} \in \Omega_{p(i)}$,
   $$\int_{\Omega_i} g_i(x_i, x_{p(i)})\mu_i(dx_i)\ =\ A_i,$$
   - il existe $A \in]0,+\infty[$ *tel que* $f_I(x_I)\ =\ \frac{1}{A}\prod_{i\in I} g_i(x_i, x_{p(i)})$.

4. *Il existe une numérotation hiérarchique* $\{s_1, \ldots, s_n\}$ *telle que :*
   $$\forall\ j\ \in\ \{2, \ldots, n\} \quad X_{s_j}\ \perp\!\!\!\perp\ X_{\{s_1,\ldots,s_{j-1}\}-\ p(s_j)}\ /\ X_{p(s_j)}.$$

5. *Pour toute numérotation hiérarchique* $\{s_1, \ldots, s_n\}$ *on a :*
   $$\forall\ j\ \in\ \{2, \ldots, n\} \quad X_{s_j}\ \perp\!\!\!\perp\ X_{\{s_1,\ldots,s_{j-1}\}-\ p(s_j)}\ /\ X_{p(s_j)}.$$

Avant de donner le schéma de la démonstration du théorème, on va faire quelques commentaires, et introduire un lemme qui seront utiles pour la démonstration, ainsi que pour la suite du travail.

Commentaires.

La propriété 1 exprime que toute variable est, conditionnellement à ses parents, indépendante de ses aïeux et de ses collatéraux ; on remarque que ses descendants ne sont pas évoqués dans cette propriété.

Si la variable considérée est de génération 0, elle n'a ni aïeux ni parents, et la propriété 1 exprime qu'elle est indépendante des variables qui ne sont pas dans sa descendance (c'est-à-dire de ses collatéraux).

La propriété faisant intervenir une numérotation topologique, qui figure en 4 et 5, exprime que toute variable autre que celle d'indice 1 est, conditionnellement à ses parents, indépendante de celles classées avant elle dans ce numérotage, autres que ses parents ; rappelons que le fait que la numérotation soit topologique a pour effet que sont classés, avant cette variable, tous ses parents et ses aïeux si elle en a, et, éventuellement, certains de ses collatéraux.

C.2.2 Un résultat essentiel

Lemme

Si la propriété (3) est vérifiée, alors pour toute partie commençante J on a :

$$f_J(x_J) \;=\; \frac{\prod_{k\in I-J} A_k}{A} \prod_{j\,\in\, J} g_j(x_j, x_{p(j)}).$$

De plus, $A = \prod_{i\in I} A_i$.

Démonstration

Choisissons un ordre hiérarchique tel que les éléments de J soient classés avant ceux de $\overline{J}$.

Soit $\mathrm{Card}(J) = m$; on note sans perte de généralité : $J = \{1,\ldots,m\}$ $\overline{J} = \{m+1,\ldots,n\}$.

Notons pour $\ell \in \{m,\ldots,n\}$ $J_\ell = \{1,\ldots,l\}$; on remarque que $J = J_m$ et $I = J_n$.

La définition de numérotation hiérarchique implique que pour chaque $\ell \in \{m,\ldots,n\}$ la partie J_ℓ contient tous les parents et aïeux de ℓ, mais aucun des descendants de ℓ.

Il en résulte que tous les J_ℓ sont des parties commençantes et on va démontrer par récurrence descendante sur ℓ la propriété :

$$f_{J_\ell}(x_{J_\ell}) \;=\; \frac{\prod_{k=\ell+1}^{n} A_k}{A} \prod_{j=1}^{\ell} g_j(x_j, x_{p(j)}).$$

La propriété est vraie pour $\ell = n$ (par hypothèse).

Nous supposons qu'elle est vraie pour ℓ ; démontrons la pour $\ell - 1$.

Pour calculer la densité par rapport à $\prod_{j\,\in\, J_{\ell-1}} \mu_j$ de $X_{J_{\ell-1}}$, on doit intégrer $f_{J_\ell}(x_1,\ldots,x_\ell)$ par rapport à x_ℓ. On obtient, en faisant usage de l'hypothèse de récurrence :

$$\begin{aligned}
f_{J_{\ell-1}}(x_1,\ldots,x_{\ell-1}) &= \int_{\Omega_\ell} f_{J_\ell}(x_1,\ldots,x_\ell)\mu_\ell(dx_\ell) \\
&= \frac{\prod_{k=\ell+1}^{n} A_k}{A} \int_{\Omega_\ell} \prod_{j=1}^{\ell} g_j(x_j, x_{p(j)})\mu_\ell(dx_\ell) \\
&= \frac{\prod_{k=\ell+1}^{n} A_k}{A} \int_{\Omega_\ell} \Big[\prod_{j=1}^{\ell-1} g_j(x_j, x_{p(j)})\Big] g_\ell(x_\ell, x_{p(\ell)})\mu_\ell(dx_\ell).
\end{aligned}$$

La numérotation étant hiérarchique, ℓ n'appartient à aucun des ensembles $j \cup p(j)$, où

$1 \leq j \leq \ell - 1$; donc

$$\begin{aligned} f_{J_{\ell-1}}(x_1, \ldots, x_{\ell-1}) &= \frac{\prod_{k=\ell+1}^{n} A_k}{A} \prod_{j=1}^{\ell-1} g_j(x_j, x_{p(j)}) \int_{\Omega_\ell} g_\ell(x_\ell, x_{p(\ell)}) \mu_\ell(dx_\ell) \\ &= \frac{\prod_{k=\ell}^{n} A_k}{A} \prod_{j=1}^{\ell-1} g_j(x_j, x_{p(j)}). \end{aligned}$$

Pour établir que $A = \prod_{i \in I} A_i$, considérons le cas particulier de la partie commençante $J = \{1\}$. On a alors

$$f_1(x_1) = \frac{\prod_{k=1}^{n} A_k}{A} g_1(x_1)$$

d'où

$$1 = \int_{\Omega_1} f_1(x_1) d\mu_1(dx_1) = \frac{\prod_{k=1}^{n} A_k}{A} \int_{\Omega_1} g_1(x_1) d\mu_1(dx_1) = \frac{\prod_{i=1}^{n} A_i}{A}.$$

■

Corollaire

Si la propriété (3) est vérifiée, on a pour tout i : $f_{i/p(i)} = \frac{g_i}{A_i}$.

Démonstration

Calculons $f_{i/p(i)}$; pour cela nous allons considérer les deux parties commençantes $\ell(i)$ et $\{i\} \cup \ell(i)$.

Il résulte du lemme que :

$$f_{\ell(i)}(x_{\ell(i)}) = \frac{\prod_{k \notin \ell(i)} A_k}{A} \prod_{j \in \ell(i)} g_j(x_j, x_{p(j)}).$$

et

$$f_{\{i\} \cup \ell(i)}(x_{\{i\} \cup \ell(i)}) = \frac{\prod_{k \notin \{i\} \cup \ell(i)} A_k}{A} g_i(x_i, x_{p(i)}) \prod_{j \in \ell(i)} g_j(x_j, x_{p(j)})$$

d'où :

$$f_{i/\ell(i)}(x_i / x_{\ell(i)}) = \frac{1}{A_i} g_i(x_i, x_{p(i)}).$$

Comme $\ell(i) = p(i) \cup a(i)$, $a(i)$ et $p(i)$ étant disjointes, ceci implique que :

$$f_{i/p(i)}(x_i / x_{p(i)}) = \frac{1}{A_i} g_i(x_i, x_{p(i)}).$$

(on utilise l'implication $(2') \implies (2)$ dans le théorème de l'annexe B, en prenant $Y_0 = X_{p(i)}$, $Y_1 = X_i$, $Y_2 = X_{a(i)}$)

■

C.2.3 Démonstration du théorème

On va établir les implications suivantes :

$$2 \iff 3 \quad \text{et } 1 \implies 5 \implies 4 \implies 2 \implies 1.$$

(2) $\implies$ (3) : évident, car (3) est une atténuation de (2).

(3) $\implies$ (2) : il résulte du corollaire ci-dessus que

$$f_I(x_I) = \frac{1}{A} g_i(x_i, x_{p(i)}) = \frac{1}{A} \prod_{i \in I} A_i f_{i/p(i)}(x_i / x_{p(i)})$$

d'où, comme $A = \prod_{i \in I} A_i$ (voir lemme ci-dessus)

$$f_I(x_I) = \prod_{i \in I} f_{i/p(i)}(x_i / x_{p(i)}).$$

1 $\implies$ 5 . Résulte immédiatement de ce que, par définition de la numérotation hiérarchique, on a pour tout j $(2 \leq j \leq n)$ $p(s_j) \subset \{s_1, \ldots, s_{j-1}\}$ et $\{s_1, \ldots, s_{j-1}\} - p(j) \subset e(s_j) \cup a(s_j)$.

5 $\implies$ 4 . C'est une évidence.

4 $\implies$ 2 . D'après la propriété générale d'emboîtement des conditionnement, on a :

$$f_I(x_{s_1}, \ldots, x_{s_n}) = \prod_{i=1}^{n} f_{s_i/s_1,\ldots,s_{i-1}}(x_{s_i} / x_{s_1}, \ldots, x_{s_{i-1}}) \ ;$$

qui devient par hypothèse

$$f_I(x_{s_1}, \ldots, x_{s_n}) = \prod_{i=1}^{n} f_{s_i/p(s_i)}(x_{s_i} / x_{p(s_i)}),$$

et par commutativité du produit on obtient :

$$f_I(x_I) = \prod_{i \in I} f_{i/p(i)}(x_i / x_{p(i)}).$$

2 $\implies$ 1 . Pour simplifier la notation, supposons que $I = \{1, \ldots, n\}$ (la numérotation naturelle est hiérarchique) ; alors la propriété (2) s'écrit :

$$f_I(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{i/p(i)}(x_i / x_{p(i)}).$$

On va calculer pour tout i :$f_{i/\ell(i) \cup c(i)}(x_i / x_{\ell(i) \cup c(i)})$.

$\ell(i) \cup c(i)$ est une partie commençante, et d'après le lemme précédent (appliqué avec $g_j(x_j, x_{p(j)}) = f_{j/p(j)}(x_j/x_{p(j)})$ et donc avec $A = 1$ et pour tout i $A_i = 1$)

$$f_{\ell(i)\cup c(i)}(x_{\ell(i)\cup E(i)}) = \prod_{j\in\ \ell(i)\cup c(i)} f_{j/p(j)}(x_j/x_{p(j)}).$$

Il en est de même pour $\{i\} \cup \ell(i) \cup c(i)$ qui est une partie commençante ;

$$f_{\{i\}\cup\ell(i)\cup c(i)}(x_{i\cup\ell(i)\cup c(i)}) = \prod_{j\in i\cup\ell(i)\cup E(i)} f_{j/p(j)}(x_j/x_{p(j)}),$$

d'où :

$$\begin{aligned} f_{i/\ell(i)\cup c(i)}(x_i/x_{\ell(i)\cup c(i)}) &= \frac{f_{\{i\}\cup\ell(i)\cup c(i)}(x_{\{i\}}, x_{\cup\ell(i)\cup c(i)})}{f_{\ell(i)\cup c(i)}(x_{\ell(i)\cup c(i)})} \\ &= \frac{\prod_{j\in\{i\}\cup\ell(i)\cup c(i)} f_{j/p(j)}(x_j/x_{p(j)})}{\prod_{j\in\ell(i)\cup c(i)} f_{j/p(j)}(x_j/x_{p(j)})} \\ &= f_{i/p(i)}(x_i/x_{p(i)}). \end{aligned}$$

■

C.3 Usage de la décomposition en générations

A l'aide de la décomposition de I en générations (ou classes) (voir Annexe A.2.2) on peut établir une nouvelle caractérisation des réseaux bayésiens. Nous supposons que I se compose de $n+1$ générations $C_0, \ldots, C_n$; la génération C_0 se compose des racines de I (individus sans parents).

Proposition 9 *Pour que $(I, G, (X_i)_{i\in I})$ soit un réseau bayésien, il faut et il suffit que soient vérifiées les deux conditions suivantes :*

(C_1') pour tout nœud i non dans la génération 0, i est, conditionnellement à ses parents, indépendant des nœuds, autres que ses parents, appartenant aux générations précédentes ;

(C_2') tous les nœuds d'une même génération sont indépendants conditionnellement à ceux des générations précédentes (en particulier ceux de la génération 0 sont indépendants).

Démonstration

Sans perte de généralité, on considère que la numérotation naturelle $\{1, \ldots, n\}$ est topologique.

a) Condition nécessaire.

Soit $i \in I$, tel que i n'appartient pas à la génération 0 ; on cherche la loi de X_i conditionnellement aux $(X_k)_{k\in\{1,\ldots,n\}}$.

On suppose vérifiées les propriétés figurant dans l'énoncé du théorème fondamental.

(C'_1) en résulte immédiatement car, pour tout i, l'ensemble des nœuds appartenant aux générations qui précédent strictement la sienne est contenu dans $\ell(i) \cup a(i)$.

Pour assurer la condition (C'_2), nous munissons I d'une numérotation topologique ; nous considérons l'ordre que ceci induit sur une génération s $(0 \leq s \leq n)$ (la numérotation étant topologique, les nœuds de cette génération sont consécutifs dans ce numérotage) ; nous allons établir (voir Annexe B) que, pour tout $i \in C_s$ qui n'est pas classé premier dans cette génération, on a, J_i désignant les nœuds appartenant à C_s classés avant i, $X_i \perp\!\!\!\perp X_{J_i} \,/\, X_{D_s}$, où $D_s = C_0 \cup \ldots \cup C_{s-1}$ (union des générations strictement antérieures à C_s ; si $s = 0$, $D_s = \emptyset$ et on considère l'indépendance usuelle) ; on remarque que $J_i \subset c(i)$ (en fait, tous les éléments de la génération C_s autres que i sont collatéraux à i).

Calculons $f_{i,J_i/D_s}(x_i, x_{J_i}/x_{D_s})$

$$\begin{aligned} f_{i,J_i/D_s}(x_i, x_{J_i}/x_{D_s}) &= \frac{f_{i,J_i,D_s}(x_i, x_{J_i}, x_{D_s})}{f_{D_s}(x_{D_s})} \\ &= f_{i/J_i,D_s}(x_i/x_{J_i}, x_{D_s}) f_{J_i/D_s}(x_{J_i}/x_{D_s}) \end{aligned}$$

Or $J_i \cup D_s \subseteq \ell(i) \cup c(i)$ et donc, par définition du réseau bayésien, on a

$$X_i \perp\!\!\!\perp X_{J_i \cup D_s - p(i)} \,/\, X_{p(i)}$$

et

$$X_i \perp\!\!\!\perp X_{D_s - p(i)} \,/\, X_{p(i)}$$

d'où

$$f_{i/J_i,D_s}(x_i/x_{J_i}, x_{D_s}) = f_{i/p(i)}(x_i/x_{p(i)}) = f_{i/D_s}(x_i/x_{D_s}).$$

On a donc

$$f_{i,J_i/D_s}(x_i, x_{J_i}/x_{D_s}) = f_{i/D_s}(x_i/x_{D_s}) f_{J_i/D_s}(x_{J_i}/x_{D_s})$$

ce qui établit l'indépendance des X_i $(i \in C_s)$ conditionnellement à X_{D_s}.

b) Condition suffisante.

On suppose que le découpage en générations vérifie les deux conditions (C'_1) et (C'_2). Utilisant la propriété (4) dans le théorème fondamental et le fait qu'une numérotation topologique est un cas particulier de numérotation hiérarchique, on va montrer que quel que soit i $(2 \leq i \leq m)$, X_i est, conditionnellement à ses parents, indépendante de ceux de ses prédécesseurs, dans la suite considérée, autres que ses parents.

Soit i tel que $2 \leq i \leq n$, on distingue deux cas :

Premier cas : i est le premier de la génération C_s. Alors $\{x_1, \ldots, x_{i-1}\} = C_0 \cup \ldots \cup C_{s-1}$ $(= \emptyset$ si $s = 0)$ et d'après la condition (C'_1) :

$$f(x_i/x_1, \ldots, x_{i-1}) = f(x_i/x_{p(i)}).$$

Deuxième cas : i n'est pas le premier de sa génération. Alors, si on suppose que le premier de cette génération est $\ell + 1$, on peut écrire :

$$\begin{aligned} f(x_i/x_1, \ldots, x_{i-1}) &= f(x_i/x_1, \ldots, x_\ell, x_{\ell+1}, \ldots, x_{i-1}) \\ &= f(x_i/x_1, \ldots, x_\ell) \\ &= f(x_i/x_{p(i)}), \end{aligned}$$

car $X_{\ell+1}, \ldots, X_{j-1}$ sont dans la même génération que X_i, et, d'après la condition (C_2') on a l'indépendance, conditionnellement aux variables des générations précédentes $(X_1, \ldots, X_\ell)$, de $(X_{\ell+1}, \ldots, X_{i-1})$ et X_i.

■

Bibliographie

[1] S. M. Aji and R. J. McEliece. The generalized distributive law. *IEEE Trans. Inform. Theory*, 46 :325–343, 2000.

[2] Ann Becker and Patrick Naim. *Les Réseaux Bayésiens*. Springer, 1999.

[3] C. Berge. *Graphs and Hypergraphs*. North-Holland Publishing compagny, Amsterdam, 1973.

[4] E. Castillo, J.M. Gutièrrez, and A.S. Hadi. *Expert Systems and Probabilistic Network Models*. Monographs in Computer Science, 1997.

[5] Huang Cecil and Darwiche Adnan. Inference in belief networks : A procedural guide. *International Journal of Approximate Reasoning*, 15 :225–263, 1996.

[6] Eugene Charniak. Bayesian networks without tears : making bayesian networks more accessible to the probabilistically unsophisticated. *AI Mag.*, 12(4) :50–63, 1991.

[7] G. Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, 42 :393–405, 1990.

[8] G. Cooper and E. Herskovits. Determination of the entropy of a belief network is np-hard. *Knowledge Systems Lab*, 1991.

[9] G. F. Cooper. A diagnostic method that uses casual knowledge and linear programming in the application of bayes' formula. *Computer Methods and Programs in Biomedicine*, 1986.

[10] G. F. Cooper. Probabilistic inference using belief networks is np-hard. *Knowledge Systems Laboratory*, 42 :393–405, 1987.

[11] G. F. Cooper. Computer-based medical diagnosis using belief networks and bounded probabilities. *In : Miller P.L. (Ed.), Selected Topics in Medical Artificial Intelligence (Springer-Verlag)*, pages 85–97, 1988.

[12] G. F. Cooper. Expert systems based on belief networks. *Current Research Directions*, 1988.

[13] G. Roberto Cowell. Advanced inference in bayesian networks. *Statistics*, 19 :301–312, 2000.

[14] Roberto G. Cowell, A. Philip Dawid, Stefenl L. Lauritzen, and David.J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, 1999.

[15] R. Dechter. Bucket elimination : A unifying framework for probabilistic inference. *In E. Horvits and F. Jensen, editor, Proc. Twelthth Conf. on Uncertainty in Artificial Intelligence.*, pages 211–219, 1996.

[16] V. Jensen Finn. *An introduction to Bayesian Networks.* UCL Press, 1999.

[17] Corset Franck. *Aide à l'optimisation de maintenance à partir de réseaux bayésiens et fiabilité dans un contexte doublement censuré.* PhD thesis, Thèse soutenue au sein d'IS2, 2003.

[18] D. Geiger and D. Heckerman. Parameter priors for directed acyclic graphical models and the characterization of several probability distributions. *The Annals of Statistics*, 30 :1412–1440, 2002.

[19] D. Geiger, D. Heckerman, H. King, and C. Meek. Stratified exponential families : Graphical models and model selection. *The Annals of Statistics*, 29 :505–526, 2001.

[20] Zoubin Ghahramani and Matthew J. Beal. Graphical models and variational methods.

[21] M. Gondran and M. Minoux. *Graphes et algorithmes.* Collection de la direction des études et recherches d'électricité de France, 1974.

[22] Petr Hájek, Tomás Havránek, and Radim Jiroušek. *Information Processing in Expert Systems.* CRC Press, 1992.

[23] D. Heckerman. A tractable inference algorithm for diagnosing multiple diseases. *Elsevier Science Publishers B.V, North Holland*, 1989.

[24] D. Heckerman. A tutorial on learning with bayesian networks. *In Learning in Graphical Models, M. Jordan, ed.. MIT Press, Cambridge, MA,*, 1999.

[25] D. Heckerman and J. Breese. Causal independence for probability assessment and inference using bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 26 :826–831, 1996.

[26] David Heckerman and Michael P. Wellman. Bayesian networks. *Commun. ACM*, 38(3) :27–30, 1995.

[27] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The lumiere project : Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 256–265. Madison, WI, Morgan Kaufmann, 1998.

[28] T. Jaakkola. Variational methods for inference and estimation in graphical models, 1997.

[29] T. Jaakkola and M. Jordan. Variational probabilistic inference and the qmr-dt database. *Journal of Artificial Intelligence Research*, 10 :291–322, 1998.

[30] Raoult Jean-Pierre and Smail Linda. Algorithme des restrictions successives. *5ème Congrès International Pluridisciplinaire Qualité et Sûreté de Fonctionnement*, 2003.

[31] F. V. Jensen and S. L. Lauritzen. Probabilistic networks. In D. M. Gabbay and Ph. Smets, editors, *Defeasible Reasoning and Uncertainty management Systems*, pages 289–320. Kluwer, netherlands, 2000.

[32] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. In *Proceedings of the NATO Advanced Study Institute on Learning in graphical models*, pages 105–161. Kluwer Academic Publishers, 1998.

[33] Pearl Judea. Bayesian networks : a model of self-activated memory for evidential reasoning. *Cognitive Science Society, UC Irvine*, pages 329–334, 1985.

[34] Pearl Judea. A constraint-propagation approach to probabilistic reasoning. *In : L. N. Kanal and J. F. Lemmer (eds), Uncertainty in Artificial Intelligence, Amsterdam, NorthHolland*, pages 3718–1986, 1986.

[35] Pearl Judea. Fusion, propagation and structuring in belief networks. *UCLA Computer Science Department Technical Report 850022 (R-42) ; Artificial Intelligence,*, 29 :241–288, 1986.

[36] Pearl Judea and A. Paz. Graphoids : a graph-based logic for reasoning about relevance relations. *UCLA Computer Science Department Technical Report 850038 ; In B. Du Boulay, et.al. (Eds.), Advances in Artificial Intelligence-II, North-Holland Publishing Co.*, 1987.

[37] Pearl Judea and T. Verma. Influence diagrams and d-separation. *UCLA Cognitive Systems Laboratory, Technical Report 880052*, 1988.

[38] Daphne Koller and Avi Pfeffer. Object oriented bayesian networks. *In Proceeding of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence*, 97, 1997.

[39] Stefen L. Lauritzen and David.J. Spiegelhalter. Local computation with probabilities on graphical structures and their application to expert systems. *Proceedings of the Royal Statistical Society*, Series B 50 (2), 1988.

[40] Steffen Lauritzen. Propagation and probabilities, means and variances in mixed graphical association models. *JASA*, 87 :1098–1108, 1992.

[41] Steffen L.Lauritzen. *Graphical Models*. Oxford Science Publications, 1996.

[42] D. J. C. MacKay. Introduction to Monte Carlo methods. In M. I. Jordan, editor, *Learning in Graphical Models*, NATO Science Series, pages 175–204. Kluwer Academic Press, 1998.

[43] A. L. Madsen and F. V. Jensen. Lazy propagation in junction trees. *In Gregory F. Cooper and Serafin Moral, editors, Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers*, pages 362–369, 1998.

[44] M. Meila and D Heckerman. An experimental comparison of several clustering and initialization methods. *Machine Learning*, 42 :9–29, 2001.

[45] M. Middleton, B.and Shwe, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper. Probabilistic diagnosis using a reformulation of the internist-1/qmr knowledge base ii. evaluation of diagnostic performance. *Methods Inf Med*, 30 :256–267, 1991.

[46] R. M. Neal. Probabilistic inference using markov chain monte carlo methods. Technical Report CRG-TR-93-1, University of Toronto, 1993.

[47] Richard E. Neapolitan. *Probabilistic Reasoning in Expert Systems.* A Wiley-Interscience Publication, John Wiley and Sons, Inc., 1990.

[48] G. Cowell Roberto. Introduction to inference for bayesian networks. *Statistics*, 19 :301–312, 2000.

[49] P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In *In Shachter, R. et al., eds., Uncertainty in Artificial Intelligence.*, volume 4, pages 169–198. Univ. Califormia Press, 1990.

[50] Duncan Smith. The efficient propagation of arbitrary subsets of beliefs in discrete-valued bayesian belief networks,. *Statistics*, 19 :301–312, 2000.

[51] P. Smyth, D. Heckerman, and M. Jordan. Probabilistic independence networks for hidden markov probability models. *Neural Computation*, 9 :227–269, 1997.

[52] D. J. Spiegelhaletr, A. P. Dawid, S. L. Lauritzen, and R. G. Cowwel. Bayesian analysis in expert systems. *Statistical science*, 8 :219–1247, 1993.

[53] D. J. Spiegelhaletr and S. L. Lauritzen. Statistical reasoning and learning in knowledgebases represented as causal networks. *Lecture Notes in Medical Informatics*, 36 :105–112, 1988.

[54] D. J. Spiegelhaletr and S. L. Lauritzen. Sequential updating of conditional probabilities on directed graphical structures. *Networks*, 20 :579–605, 1990.

[55] J. Stillman. On heuristics for finding loop cutsets in multiply connected belief networks. In P. B. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 233–243. North-Holland, Amsterdam, 1991.

[56] H. J. Suermondt and G. F. Cooper. Probabilistic inference in multiply connected brief networks using loop cutsets. *Journal of Approximate Reasoning*, 4 :283–306, 1990.

[57] H. J. Suermondt and G. F. Cooper. A combination of exact algorithms for inference on bayesian belief networks. *int. J. Approximate Reasoning*, 5 :521–542, 1991.

[58] H. J. Suermondt and G. F. Cooper. Initialization for the method of conditioning in bayesian belief networks. *Artificial Intelligence*, 50 :83–94, 1991.

[59] H. J. Suermondt, G. F. Cooper, and D. E. Heckerman. A combination of cutset conditioning with clique-tree propagation in the pathfinder system. In P. B. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 245–253. North-Holland, Amsterdam, 1991.

[60] Robert E. Tarjan and Mihalis Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM J. Comput.*, 13(3) :566–579, 1984.

[61] Kjærulff Uffe. A note on triangulated graphs and junction trees. *Lecture Note, Department of Mathematics and Computer Science, Aalborg University, Denmark*, 1992.

[62] Kjærulff Uffe. Inference in bayesian networks using nested junction trees. *Statistics*, 19 :301–312, 2000.

[63] Yang Xiang and F. V. Jensen. Inference in multiply sectioned bayesian networks with extended shafer-shenoy and lazy propagation. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 680–687, 1999.

[64] Yanping Xiang and Finn Verner Jensen. Inference in multiply sectioned bayesian networks with extended shafer-shenoy and lazy propagation. *UAI*, 43 :680–687, 1999.

[65] N. L. Zhang and D. Poole. A simple approach to bayesian network computations. *In Proc. of the Tenth Canadian Conference on Artificial Intelligence*, 71 :171–178, 1994.

Index