



**HAL**  
open science

## Factorisation des calculs dans un réseau bayésien : interprétation des facteurs

Zineb Azouz

► **To cite this version:**

Zineb Azouz. Factorisation des calculs dans un réseau bayésien : interprétation des facteurs. Journées Francophone sur les Réseaux Bayésiens, May 2008, Lyon, France. hal-00259627

**HAL Id: hal-00259627**

**<https://hal.science/hal-00259627>**

Submitted on 14 Apr 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Factorisation des calculs dans un réseau bayésien : interprétation des facteurs

**Zineb Azouz**

*Université de Constantine  
route d'Aïn el bey, Constantine Algérie  
zineb.azouz@gmail.com*

---

*RÉSUMÉ. Etant donné une famille de variables aléatoires munie d'une structure de réseau bayésien, la restriction de la loi à une sous famille de v.a. se présente sous forme d'un produit de facteurs; nous donnons une condition nécessaire et suffisante sur le graphe du réseau pour que ces facteurs s'interprètent en termes de probabilités conditionnelles; ces graphes seront dits "non pathologiques".*

*ABSTRACT. Given a family of random variables provided with a structure of a bayesian network, the restriction of the distribution to a sub family of r.v. is presented in the form of a product of factors; we give a necessary and sufficient condition on the graph of the network so that these factors are interpreted as conditional probabilities; these graphs will be called "nonpathological".*

*MOTS-CLÉS : Réseaux bayésiens, D-séparation, Réseaux bayésien de niveau 2.*

*KEYWORDS: Bayesian networks, D-separation, Bayesian Network of level 2.*

---

## 1. Introduction

### 1.1. Calcul des restrictions dans un Réseau Bayésien

Rappelons que (voir (JENSEN., 2001) et (NAIM *et al.*, 2004)), étant donné un ensemble  $I$  et un graphe orienté sans circuits  $\mathcal{G}$  sur  $I$ , un réseau bayésien (RB) est une famille de v.a.  $X_I = (X_i)_{i \in I}$  (où  $X_i$  est à valeurs dans  $\Omega_i$  et donc  $X_I$  à valeurs dans  $\Omega_I = \prod_{i \in I} \Omega_i$ ), telle que, pour tout  $i$ , la loi conditionnelle de  $X_i$  relativement à l'ensemble des v.a. autres qu'elle-même et ses descendants (dont l'ensemble est noté  $d(i)$ ), ne dépende en fait que de la valeur  $x_{p(i)}$  prise par l'ensemble de ses parents :

$$\forall i \quad \forall x_I \quad P_{i|I - (\{i\} \cup d(i))}(x_i | x_{I - (\{i\} \cup d(i))}) = P_{i|p(i)}(x_i | x_{p(i)})$$

avec la convention que si,  $i$  est une racine pour le graphe  $\mathcal{G}$  (i.e  $p(i) = \emptyset$ ),  $P_{i|p(i)}$  désigne la loi de la variable  $X_i$ . Il en résulte l'expression de la loi de la famille  $X_I$  :

$$P_I(x_I) = \prod_{i \in I} P_{i|p(i)}(x_i | x_{p(i)}).$$

Nous nous intéressons ici au calcul des restrictions de la loi  $P_I$  du RB, c'est à dire, étant donné une partie quelconque  $S$  de  $I$ , au calcul de  $P_S$ , loi de la famille  $X_S = (X_i)_{i \in S}$ . Autrement dit, étant donné  $x_S \in \Omega_S$  et en décomposant  $x_I$  sous la forme  $(x_S, x_{I-S})$ , on considère le calcul :  $P_S(x_S) = \sum_{x_{I-S} \in \Omega_{I-S}} P_I(x_S, x_{I-S})$ .

Il résulte de la définition que la restriction d'un réseau bayésien à une partie commençante  $J$  de  $I$  (c'est-à-dire une partie telle que  $\forall j \in J \quad p(j) \subset J$ ), hérite de la structure de RB ; donc, pour une partie quelconque  $S$ , le calcul de  $P_S$  peut être effectué en se restreignant à la partie commençante engendrée par  $S$ , noté  $S^+$  :

$$P_S(x_S) = \sum_{x_{S^+ - S} \in \Omega_{S^+ - S}} P_I(x_S, x_{S^+ - S}).$$

Nous ne perdons donc pas en généralité en supposant dans toute la suite de cet article que  $S^+ = I$ , autrement dit que toutes les feuilles de  $I$  appartiennent à  $S$ .

Les applications effectives des RB faisant souvent intervenir des ensembles  $I$  de très grande taille, relativement à laquelle celle de  $S$  est faible, il est essentiel de mettre en place des techniques de simplification de la sommation sur les  $x_{I-S} \in \Omega_{I-S}$ . Une large fraction de la littérature relative aux calculs dans les RB est consacrée à de telles techniques et à leurs implémentations informatiques.

Nous prolongeons ici le travail effectué par Linda Smail dans sa thèse (voir (SMAIL, 2004)). A l'occasion de la mise en place d'un nouvel algorithme de calcul

(dit "des restrictions successives"), elle avait rappelé les factorisations possibles de ces calculs et posé le problème de l'interprétation des facteurs en termes de probabilités conditionnelles, type de préoccupation qu'on trouve aussi dans ((STUDENY, 2005)).

## 1.2. Cas particulier : les chaînes de Markov

La simplification du calcul de  $P_S$  est classique dans le cas d'une chaîne de Markov finie, avec  $I = \{1, \dots, n\}$  et  $X_I = (X_1, X_2, \dots, X_n)$ . Le graphe  $\mathcal{G}$  est dans ce cas composé des couples élémentaires  $(i-1, i)$  ( $2 \leq i \leq n$ ); alors, pour tout  $i \in \{1, 2, \dots, n\}$ , son unique parent est  $i-1$  ( $p(i) = \{i-1\}$ ) et pour tout  $i \in \{1, 2, \dots, n-1\}$ , sa descendance est  $d(i) = \{i+1, i+2, \dots, n\}$ .

Ainsi, si  $S = \{s_1, s_2, \dots, s_m\}$  où  $(s_1 < s_2 < \dots < s_m)$  et si  $s_m = n$  (afin que  $S^+ = \{1, \dots, n\}$ ),  $X_S = (X_{s_1}, \dots, X_{s_m})$  est aussi une chaîne de Markov et  $P_S$  s'écrit comme :

$$P_S(x_{s_1}, \dots, x_{s_m}) = \prod_{k=1}^m P_{s_k|s_{k-1}}(x_{s_k} | x_{s_{k-1}}) \quad (1)$$

avec la convention, si  $k = 1$ ,  $P_{s_1|s_0}(x_{s_1} | x_{s_0}) = P_{s_1}(x_{s_1})$ .

Soit  $L$  l'ensemble des indices  $\ell$  tels que  $s_\ell$  n'ait pas de parents dans  $I - S$  (soit que  $\ell = 1$  avec  $s_1 = 1$ , soit que  $s_{\ell-1} \in S$ ). Autrement dit, si  $\ell \in L$ ,  $s_\ell$  est soit 1, soit l'enfant de  $s_{\ell-1}$  et dans ce cas  $p_{s_\ell|s_{\ell-1}}$  est l'une des données de la chaîne de Markov. Si  $k \notin L$ , le calcul de  $P_{s_k|s_{k-1}}$  est aisé et ne fait intervenir que les éléments de  $C_k = \{s_{k-1} + 1, \dots, s_k - 1\}$  (avec, si  $s_1 > 1$ , le cas particulier  $C_1 = \{1, \dots, s_1 - 1\}$ ); les parties  $C_k$  (où  $k \notin L$ ) constituent une partition de  $I - S$  et on a alors :

$$P_{s_k|s_{k-1}}(x_{s_k} | x_{s_{k-1}}) = \sum_{x_{C_k} \in \Omega_{C_k}} \prod_{i=s_{k-1}}^{s_k} P_{i|i-1}(x_i | x_{i-1}). \quad (2)$$

En récapitulant, il résulte de (1) que le calcul de  $P_S(x_{s_1}, \dots, x_{s_m})$  s'effectue sous la forme suivante :

$$P_S(x_{s_1}, \dots, x_{s_m}) = \left[ \prod_{k \in L} P_{s_k|s_{k-1}}(x_{s_k} | x_{s_{k-1}}) \right] \cdot \left[ \prod_{k \notin L} \sum_{x_{C_k} \in \Omega_{C_k}} \prod_{i=s_{k-1}+1}^{s_k} P_{i|i-1}(x_i | x_{i-1}) \right]. \quad (3)$$

## 1.3. Retour au cas général

La généralisation, pour les réseaux bayésiens, de la décomposition valide pour la chaîne de Markov va faire intervenir la remarque élémentaire suivante : dans une

chaîne de Markov, si  $j'$  et  $j''$  (où  $j' < j''$ ) sont deux indices non dans  $S$ , appartenant respectivement à  $C_{k'}$  et  $C_{k''}$  telles que  $k' \neq k''$ , l'unique chemin qui relie  $j'$  à  $j''$  contient au moins un élément de  $S$ . Dans le cas général des RB, les parties  $C_k$  sont alors introduites comme des classes d'équivalence associées à une relation d'équivalence sur  $I - S$ ; pour cette relation  $i'$  et  $i''$  sont équivalents si et seulement s'ils ne sont pas D-séparés par  $I$  (notion classique qui remonte à (PEARL *et al.*, 1998)), c'est à dire si et seulement s'il existe, dans le graphe moral associé à  $\mathcal{G}$ , une chaîne reliant  $i'$  à  $i''$ ; on rappelle que le graphe moral associé au graphe orienté sans circuits  $\mathcal{G}$  est le graphe non orienté  $\mathcal{H}$  dont les arêtes sont les paires  $\{i', i\}$  telles que l'un des deux éléments soit parent de l'autre, ou sinon qu'il ait un enfant en commun.

Remarquons que dans une chaîne de Markov, le graphe moral  $\mathcal{H}$  a pour arêtes les paires  $\{i - 1, i\}$  ( $2 \leq i \leq n$ ).

La partition S-conditionnelle, notée  $\mathcal{C}$ , est alors définie comme la partition de  $I - S$  en classes d'équivalence pour cette relation. Notons par ailleurs  $L$  l'ensemble des éléments de  $S$  qui n'ont aucun parent dans  $I - S$  (soit que ce soient des racines, soit que leurs parents soient tous dans  $S$ ).

Pour toute classe  $C \in \mathcal{C}$ , soit  $M(C)$  la couverture de Markov de  $C$ , c'est à dire l'ensemble des éléments de  $I$  qui soit sont dans  $C$ , soit sont voisins d'au moins un élément de  $C$  dans le graphe moral; autrement dit les éléments de  $M(C)$  soit sont dans  $C$ , soit sont parent ou enfant d'au moins un élément de  $C$ , soit encore ont un enfant en commun avec au moins un élément de  $C$ .

On appelle frontière de Markov de  $C$  l'ensemble  $F(C)$  des éléments de  $M(C)$  non dans  $C$ .

On remarque que  $F(C)$  est contenu dans  $S$ ; soit en effet  $i \in F(C)$ ; si  $i$  n'appartient pas à  $S$ , il appartiendrait à une classe  $C'$  autre que  $C$ , ce qui est absurde puisqu'il est relié à au moins un élément de  $C$  dans le graphe moral.

Nous introduisons également la partition  $(T(C), R(C))$  de  $F(C)$ , que nous appelons partition canonique, où  $T(C)$  est l'ensemble des éléments de  $F(C)$  qui sont enfants d'au moins un élément de  $C$ .

Nous avons déjà dit qu'on peut supposer que  $I = S^+$ , il en résulte immédiatement que, pour tout  $C$  non vide de  $S^+ - S$  (ce qui est le cas des éléments de la partition S-conditionnelle),  $T(C)$  est non vide et  $C$  est contenu dans  $T(C)^+$ .

Dans le cas de la chaîne de Markov de longueur  $n$ , si  $C = \{s_{k-1} + 1, \dots, s_k - 1\}$  (où  $k \geq 2$  et  $s_k > s_{k-1}$ ), on a  $M(C) = \{s_{k-1}, \dots, s_k\}$ ,  $F(C) = \{s_{k-1}, s_k\}$ ,  $T(C) = \{s_k\}$  et  $R(C) = \{s_{k-1}\}$  (et d'autre part, si  $s_1 > 1$  et  $C = \{1, \dots, s_{2-1}\}$ ,  $M(C) = \{1, \dots, s_1\}$ ,  $F(C) = T(C) = \{s_1\}$  et  $R(C) = \emptyset$ ).

Pour tout réseau bayésien, on démontre (voir (SMAIL, 2004))

$$P_S(x_S) = \prod_{\ell \in L} P_{\ell|p(\ell)}(x_\ell | p(x_\ell)) \cdot \prod_{C \in \mathcal{C}} \sum_{x_C \in \Omega_C} \prod_{i \in C \cup T(C)} P_{i|p(i)}(x_i | x_{p(i)}) \quad (4)$$

ce qui généralise la formule (3) connue pour les chaînes de Markov. Considérons, dans la formule (4) le facteur  $\sum_{x_C \in \Omega_C} \prod_{i \in C \cup T(C)} P_{i|p(i)}(x_i | x_{p(i)})$ . On vérifie qu'il ne dépend que de  $x_{F(C)}$ ; en effet pour tout  $i$  appartenant à  $C \cup T(C)$ ,  $P_{i|p(i)}$  ne fait intervenir que des variables dans le couverture de Markov  $M(C)$ , par définition même de celle ci; après sommation sur les variables à indices dans  $C$ , il subsiste une fonction de  $x_{M(C)-C}$ , c'est à dire une fonction de  $x_{F(C)}$ .

Or, dans le cas d'une chaîne de Markov,  $C = \{s_{k-1} + 1, \dots, s_k - 1\}$ ,  $T(C) = \{s_k\}$  et il est vrai que :

$$\sum_{x_C \in \Omega_C} \prod_{i \in C \cup T(C)} P_{i|p(i)}(x_i | x_{p(i)}) = P_{s_k|s_{k-1}}(x_{s_k} | x_{s_{k-1}}).$$

Nous sommes donc amenés à nous demander si, dans le cas général, ce résultat subsiste en s'exprimant alors :

$$\sum_{x_C \in \Omega_C} \prod_{i \in C \cup T(C)} P_{i|p(i)}(x_i | x_{p(i)}) = P_{T(C)|R(C)}(x_{T(C)} | x_{R(C)}).$$

Le but de cet article est d'établir que cette égalité n'est pas vérifiée en toute généralité et de caractériser les RB pour lesquels elle est satisfaite.

Les RB pour lesquels elle ne l'est pas seront dits ici "pathologiques"; cette terminologie se justifie car il s'agit de situations, relatives au graphe  $\mathcal{G}$ , nous semble-t-il, que les praticiens effectuant des modélisations en termes de RB cherchent à éviter (voir 4.1 ci-dessous).

## 2. Les réseaux bayésiens pathologiques

Soit fixée une classe  $C$  de la partition S-Conditionnelle; notons  $M$  sa couverture de Markov et  $(T, R)$  la partition canonique de sa frontière de Markov  $F$ . Rappelons que  $T$  est défini comme l'ensemble des nœuds qui sont enfants d'au moins un élément de  $C$ ; il paraît donc naturel de les voir comme situés "globalement" après les éléments de  $C$  et de  $R$ , autrement dit, il paraîtrait "anormal" qu'on puisse trouver un parent appartenant à  $T$  pour un élément de  $T^+$  (partie commençante engendrée par  $T$ ) qui ne serait lui même ni dans  $T$  ni dans  $C$  (i.e. qui appartiendrait à  $R \cup J$ , où  $J = T^+ - M$ ).

Cette situation peut cependant se produire, comme c'est le cas dans les exemples suivants qui sont les cas les plus simples de situations où il existe  $i \in R \cup J$  dont le parent soit dans  $T$  (dans l'exemple 1,  $3 \in J$  et dans l'exemple 2,  $3 \in R$ ) :



C'est cette circonstance, nécessaire et suffisante pour l'apparition de la pathologie, qui fera l'objet du théorème qui est présentée en section 3.

Dans ce qui va suivre intervient la partition  $(C, T, R, J)$  de  $T^+$ , avec les regroupements suivants :

$$F = T \cup R, M = C \cup F = C \cup T \cup R, T^+ = M \cup J = C \cup T \cup R \cup J$$

### Définition

Etant donné une partie  $C$  de  $I$ , le graphe  $\mathcal{G}$  présente une pathologie en  $C$  si et seulement s'il existe  $i \in R \cup J$  tel que  $p(i) \cap T \neq \emptyset$ .

Les deux lemmes suivants fournissent des conditions nécessaires et suffisantes de pathologie.

### Lemme 1

Pour que le Graphe  $\mathcal{G}$  présente une pathologie en  $C$ , il faut et il suffit que  $R \cup J$  ne soit pas une partie commençante.

**Lemme 2**

Pour que  $\mathcal{G}$  présente une pathologie en  $C$ , il faut et il suffit qu'il existe un chemin dans  $\mathcal{G}$ , soit  $(j_0, \dots, j_n)$  (où  $n \geq 1$ ), tel que :

- $j_0 \in T$ ,
- $j_n \in R$ ,
- si  $n \geq 2$ , alors, pour tout  $k$  tel que  $1 \leq k \leq n - 1$ , on a  $j_k \in J$ .

**3. théorème**

Le théorème qui suit vaut pour toute partie  $C \subset I$  satisfaisant des hypothèses dont on sait qu'elles sont satisfaites pour les classes de la partition  $S$ -conditionnelle. A la section 4, nous commenterons son usage pour le calcul de la loi  $P_s$ .

**Théorème**

Soit  $C$  une partie non vide de  $I$ . Pour laquelle on suppose que  $T$ , ensemble des enfants, non dans  $C$ , des éléments de  $C$  est non vide et que  $C \subset T^+$ . Pour qu'il existe un RB  $(X_i)_{i \in I}$  tel que :

$$\sum_{x_C \in \Omega_C} \prod_{i \in C \cup T(C)} P_{i|p(i)}(x_i | x_{p(i)}) \neq P_{T|R}(x_T | x_R)$$

il faut et il suffit que le graphe  $\mathcal{G}$  présente une pathologie en  $C$ .

**3.1. Préliminaires de la démonstration**

L'expression  $P_{T|R}(x_T | x_R)$  qui intervient dans le théorème et qui vaut  $\frac{P_F(x_F)}{P_R(x_R)}$  sera calculée dans le cas général en passant par le calcul de  $P_M(x_M)$ , puis de  $P_F(x_F) = \sum_{x_C \in \Omega_C} P_M(x_C, x_F)$  et enfin de  $P_R(x_R) = \sum_{x_T \in \Omega_T} P_F(x_T, x_R)$ .

Nous verrons alors à quel moment exactement la condition du théorème, à savoir la présence de pathologie en  $C$ , fera en sorte que puisse être mise en défaut l'égalité entre  $\sum_{x_C \in \Omega_C} \prod_{i \in C \cup T(C)} P_{i|p(i)}(x_i | x_{p(i)})$  et  $P_{T|R}(x_T | x_R)$ .

**3.1.1. Calcul de  $P_M(x_M)$** 

$M$  et  $J$  sont contenus dans  $T^+$  qui est, par définition, une partie commençante, donc conservant la structure de RB ; nous sommes donc autorisés, pour l'obtention de  $P_M(x_M)$ , à nous limiter à  $T^+$ .



$$P_M(x_M) = \sum_{x_J \in \Omega_J} \prod_{i \in T^+} P_{i|p(i)}(x_i | x_{p(i)})$$

Il est utile d'étudier, selon l'appartenance de  $i$  à chacune des parties  $C, T, R$  et  $J$ , quelle composante de  $x_{T^+}$  intervient dans  $x_i$  et  $x_{p(i)}$ ; en effet, si ni  $i$  ni ses parents n'appartiennent à  $J$ ,  $P_{i|p(i)}(x_i | x_{p(i)})$  pourra être mis en facteur lors de la sommation  $\sum_{x_J \in \Omega_J}$ . C'est le cas si  $i \in C$ , car alors ses parents sont dans  $M$  par définition même de  $M$ . De même si  $i \in T$ , l'un au moins de ses parents (appelons le  $i'$ ) appartient à  $C$ , et alors, s'il y a lieu les autres parents de  $i$  appartiennent aussi à  $C$  et donc à  $M$  (car liés à  $i'$  dans le graphe moral) et donc pas à  $J$ . En revanche, si  $i \in R$ , il peut se produire que des parents de  $i$  soient dans  $J$ .

On peut donc écrire :

$$P_M(x_M) = \sum_{x_J \in \Omega_J} \left( \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}) \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \right)$$

c'est à dire :

$$P_M(x_M) = \left( \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}) \right) \left( \sum_{x_J \in \Omega_J} \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \right). \quad (5)$$

### 3.1.2. Calcul de $P_F(x_F)$

$$P_F(x_F) = \sum_{x_C \in \Omega_C} P_M(x_C, x_F).$$

Dans l'expression (5), le facteur  $\sum_{x_J \in \Omega_J} \left( \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \right)$  ne dépend que de  $x_F$ , car, si  $i \in R \cup J$ , ses parents sont dans  $F \cup J$ , autrement dit ne peuvent appartenir à  $C$  (en effet, par définition de  $T$ , tous les enfants des éléments de  $C$  sont dans  $C \cup T$  et donc pas dans  $R \cup J$ ). Il résulte donc de (5) que :

$$P_F(x_F) = \left( \sum_{x_C \in \Omega_C} \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}) \right) \left( \sum_{x_J \in \Omega_J} \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \right).$$

### 3.2. Démonstration de la condition nécessaire du théorème

En utilisant la contraposée, ceci revient à montrer que si l'on est en présence d'un Graphe  $\mathcal{G}$  qui ne présente pas de pathologie en la partie  $C$ , alors on a :

$$\sum_{x_C \in \Omega_C} \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}) = P_{T|R}(x_T | x_R).$$

Supposons donc (voir Lemme1) que  $R \cup J$  soit une partie commençante et établissons l'égalité ci-dessus.

Comme  $P_{T|R}(x_T | x_R) = \frac{P_F(x_T)}{P_R(x_R)}$ , il nous reste à calculer

$$\begin{aligned} P_R(x_R) &= \sum_{x_T \in \Omega_T} P_F(x_T, x_R). \\ &= \sum_{x_T \in \Omega_T} \left( \sum_{x_C \in \Omega_C} \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}) \right) \left( \sum_{x_J \in \Omega_J} \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \right). \end{aligned}$$

Or  $R \cup J$  est une partie commençante, donc le facteur  $\prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)})$  ne dépend que de  $x_{R \cup J}$ , et, après sommations sur les  $x_J \in \Omega_J$ , l'expression  $\sum_{x_J \in \Omega_J} \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)})$  ne dépend plus que de  $x_R$  et certainement pas de  $x_T$ , d'où :

$$\begin{aligned} P_R(x_R) &= \left( \sum_{x_{C \cup T} \in \Omega_{C \cup T}} \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}) \right) \left( \sum_{x_J \in \Omega_J} \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \right) \\ &= \sum_{x_J \in \Omega_J} \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \end{aligned}$$

car il est élémentaire que, pour toute partie  $A$  de  $I$ ,  $\sum_{x_A \in \Omega_A} \prod_{i \in A} P_{i|p(i)}(x_i | x_{p(i)}) = 1$ .

Il vient alors :

$$P_{T|R}(x_T | x_R) = \frac{\left( \sum_{x_C \in \Omega_C} \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}) \right) \left( \sum_{x_J \in \Omega_J} \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \right)}{\left( \sum_{x_J \in \Omega_J} \prod_{i \in R \cup J} P_{i|p(i)}(x_i | x_{p(i)}) \right)}$$

$$= \sum_{x_C \in \Omega_C} \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}).$$

### 3.3. Démonstration de la condition suffisante du théorème

Soit  $(j_0, \dots, j_n)$  une suite telle que celle fabriquée dans le lemme 2. On va construire un réseau bayésien, à variables booléennes  $X_i$  (autrement dit, pour tout  $i$ ,  $\Omega_i = \{0, 1\}$ ), tel que :

$$\exists (x_T, x_R) \in \{0, 1\}^{\text{card}(F)} \quad P_{T|R}(x_T | x_R) \neq \sum_{x_C \in \{0, 1\}^C} \prod_{i \in C \cup T} P_{i|p(i)}(x_i | x_{p(i)}).$$

Pour cela, si  $i \notin \{j_1, \dots, j_n\}$ , on prend pour  $P_{i|p(i)}$  l'équiprobabilité quelles que soient les valeurs des variables conditionnantes ; autrement dit :

$$\forall x_{p(i) \in \{0, 1\}^{\text{card}(p(i))}} \quad P_{i|p(i)}(0 | x_{p(i)}) = P_{i|p(i)}(1 | x_{p(i)}) = \frac{1}{2}$$

En revanche, pour tout  $j_k$  (avec  $1 \leq k \leq n$ ),  $P_{x_{j_k}|p(x_{j_k})}$  va pouvoir dépendre effectivement de la variable conditionnante  $x_{j_{k-1}}$  (et d'elle seule si  $j_k$  a d'autres parents que  $j_{k-1}$ ) ; on notera :

$$P_{j_k|p(j_k)}(x_{j_k} | x_{p(j_k)}) = a_{k, x_{j_k}}^{x_{j_{k-1}}}$$

avec  $a_{k,0}^{x_{j_{k-1}}} + a_{k,1}^{x_{j_{k-1}}} = 1$ .

$P_{j_k|p(j_k)}$  est donc caractérisé par deux nombres,  $a_{k,0}^0$  et  $a_{k,0}^1$ .

En décomposant  $x_F$  sous la forme  $(x_R, x_T)$ , il vient

$$P_F(x_R, x_T) = \left(\frac{1}{2}\right)^{\text{card}(T) + \text{card}(R) - 1} \sum_{(y_1, \dots, y_{n-1}) \in \{0, 1\}^{n-1}} \prod_{1 \leq k \leq n} a_{k, y_k}^{y_{k-1}}$$

La recherche de contre-exemple revient alors à trouver une famille de couples  $(a_{k,0}^0, a_{k,1}^0)_{1 \leq k \leq n}$  telle que l'expression, ne dépendant en fait que de  $y_0$  et  $y_n$ ,

$$\sum_{(y_1, \dots, y_{n-1}) \in \{0,1\}^{n-1}} \prod_{1 \leq k \leq n} a_{k, y_k}^{y_{k-1}}$$

prenne, pour  $y_n$  fixé (par exemple  $y_n = 0$ ), des valeurs différentes pour  $y_0 = 0$  et  $y_0 = 1$ .

Cette propriété s'interprète très aisément relativement à la suite  $(X_{j_0}, \dots, X_{j_n})$  qui est un chaîne de Markov de variables aléatoires booléennes, avec pour tout  $k$  (où  $1 \leq k \leq n$ ),

$$P_{k|k-1}(y_k|y_{k-1}) = a_{k, y_k}^{y_{k-1}};$$

alors

$$P_{n|0}(y_n|y_0) = \sum_{(y_1, \dots, y_{n-1}) \in \{0,1\}^{n-1}} \prod_{1 \leq k \leq n} a_{k, y_k}^{y_{k-1}}$$

et on dispose du contre-exemple cherché dès lors que  $P_{n|0}(0|0) \neq P_{n|0}(0|1)$ , autrement dit dès lors que les v.a.  $Y_0$  et  $Y_n$  ne sont pas indépendantes et une condition nécessaire et suffisante pour cela est que ; pour tout  $k$ ,  $a_{k,0}^0 \neq a_{k,0}^1$ , autrement dit :  $Y_k$  et  $Y_{k-1}$  ne sont pas indépendantes.

## 4. Commentaires

### 4.1. Situation à risque de pathologie

Il peut être utile de s'assurer qu'aucune pathologie ne risque de se produire, et ce, quel que soit  $S$ . une condition élémentaire mais apparemment assez forte pour qu'il en soit ainsi est qu'il n'existe aucun couple  $(i, i')$  vérifiant la condition suivante :  $i$  parent de  $i'$  et de plus il existe un chemin  $(i_0, \dots, i_n)$  ( $n \geq 3$ ) pour lequel aucun  $i_j$  ( $1 \leq j \leq n-2$ ) n'est parent de  $i'$ .

Dans la pratique, cette condition n'est peut-être pas très contraignante. Nous ne connaissons pas d'exemple de modélisation par un réseau bayésien au sein duquel on ait fait figurer une chaîne d'au moins quatre nœuds, dont chacun (sauf évidemment le dernier) influe sur le suivant et où de plus le premier influe aussi sur le dernier, mais sans qu'aucun des intermédiaires (sauf bien sûr l'avant-dernier) fasse de même. En revanche, si tous les éléments de la chaîne contribuent à influencer sur le dernier, ceci traduit simplement qu'il n'y a entre eux aucune forme d'indépendance (la loi de leur ensemble est "quelconque").

#### 4.2. Une situation pathologique peut elle s'interpréter en terme de réseau bayésien de niveau 2 ?

Soit  $Q_S$  l'ensemble des parties de  $S$  constituée des singletons  $\{\ell\}$  où,  $\ell$  parcourt l'ensemble  $L$  des éléments de  $S$  n'ayant aucun parent dans  $S^+ - S$ , et des parties  $T(C)$  où  $C$  parcourt la partition  $S$ -conditionnelle  $\mathcal{C}$ . Remarquons que l'un de ces deux types d'éléments de  $Q_S$  peut être absent.

Il est démontré (voir (SMAIL, 2004), sous section 2.3) que  $Q_S$  ainsi défini constitue une partition de  $S$ .

Il résulte de la formule (4) et du théorème que si le réseau ne présente de pathologie relativement à aucune partie  $C \in \mathcal{C}$ ,

$$P_S(x_S) = \left[ \prod_{\ell \in L} P_{\ell|p(\ell)}(x_\ell | p(x_\ell)) \right] \cdot \left[ \prod_{C \in \mathcal{C}} P_{T(C)|R(C)}(x_{T(C)} | x_{R(C)}) \right], \quad (5)$$

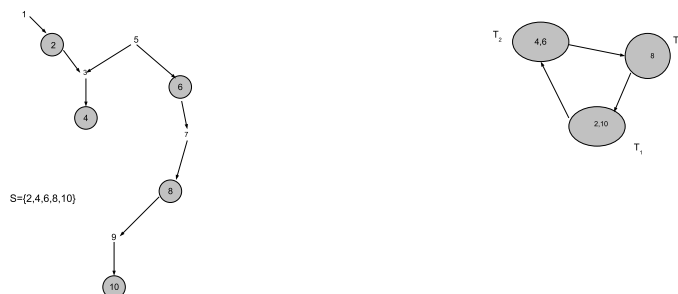
où les parties  $p(\ell)$  ( $\ell \in L$ ) et  $R(C)$  (où  $C \in \mathcal{C}$ ) sont contenues dans  $S$  (éventuellement vides).

On voit ici une expression de  $P_S$  qui est analogue à celle que l'on a pour la loi d'un réseau bayésien qui aurait pour nœud du graphe non pas les éléments de  $S$ , mais les éléments de la partition  $Q_S$ ; c'est ce qu'on appelle (voir (SMAIL *et al.*, 2005a)) un réseau bayésien de niveau 2 noté RB2; mais, pour qu'il en soit ainsi il faudrait étudier au préalable si  $Q_S$  est muni d'une structure de graphe orienté sans circuits, noté  $\mathcal{G}'_S$ , de telle sorte que pour tout  $\ell$ ,  $p(\ell)$  soit contenu dans l'ensemble des parents relativement à  $\mathcal{G}'_S$ , de  $\{\ell\}$  et pour tout  $C$ ,  $R(C)$  soit contenu dans l'ensemble des parents, relativement à  $\mathcal{G}'_S$  de  $T(C)$ ; ceci conduit à définir  $\mathcal{G}'_S$  comme suit :

- . si  $\ell \in L$  et  $\ell' \in L$ ,  $(\{\ell\}, \{\ell'\}) \in \mathcal{G}'_S$  si  $(\ell, \ell') \in \mathcal{G}$ ,
- . si  $\ell \in L$  et  $C \in \mathcal{C}$ ,  $(\{\ell\}, T(C)) \in \mathcal{G}'_S$  si  $\ell \in R(C)$ ,
- . si  $C \in \mathcal{C}$  et  $C' \in \mathcal{C}$ ,  $(T(C), T(C')) \in \mathcal{G}'_S$  si  $T(C) \cap R(C') \neq \emptyset$ .

Or, on constate que le graphe orienté ainsi défini sur  $Q_S$  n'est pas nécessairement sans circuit, comme le montre le contre exemple suivant construit à partir d'un graphe non pathologique  $\mathcal{G}$  et avec  $S = \{2, 4, 6, 8, 10\}$ ,  $L = \emptyset$ ,  $C_1 = \{1, 9\}$ ,  $C_2 = \{3, 5\}$ ,  $C_3 = \{7\}$ ,

$$T(C_1) = \{2, 10\}, T(C_2) = \{4, 6\}, T(C_3) = \{8\} \text{ et } R(C_1) = \{8\}, R(C_2) = \{2\}, R(C_3) = \{6\};$$



Quoique non-pathologique, ce graphe présente un phénomène "d'effet lointain" (arc de 1 vers 9) analogue à ceux qui caractérisent les graphes pathologiques ; nous nous proposons de chercher ultérieurement nous nous proposons de chercher ultérieurement une caractérisation des graphes  $\mathcal{G}$  et des parties  $S$  tels que le graphe associé  $\mathcal{G}'_S$  soit sans circuits.

## Remerciements

Ce travail a été réalisé sous la direction du Professeur Jean-Pierre Raoult, durant un séjour au Laboratoire d'Analyse et de Mathématiques Appliquées de l'Université Paris Est Marne La Vallée (CNRS, UMR 8050).

## 5. Bibliographie

- JENSEN. F. V., *Bayesian networks and Decision Graphs*, Springe, Springe, 2001.
- NAIM P., WUILLEMIN P.-H., LERAY P., POURRET O., BECKER A., *Réseaux bayésiens*, Eyrolles, 2004.
- PEARL J., VERMA. T., « Influence Diagrams and d-Separation », *Cognitive Systems Laboratory, Computer Science Dept. UCLA, Technical Report 880052*, 1998.
- SMAIL L., *Algorithmique pour les réseaux bayésiens et leurs extensions*, Thèse en Mathématiques Appliquées, Université Paris Est Marne La Vallée, 2004.
- SMAIL L., RAOULT J.-P., *Réseaux bayésiens de niveau 2 et D-séparation*, Colloque EGC 2005 (Extraction et Gestion de Connaissances ; section "Modèles graphiques probabilistes pour la modélisation des connaissances : inférence, apprentissage et applications, Paris, 2005a.
- SMAIL L., RAOULT J.-P., *Successive Restrictions Algorithm in Bayesian Networks*, IDA (6th international symposium on Intelligence Data Analysis, Madrid, 2005b.
- STUDENY M., *Probabilistic Conditional Independent Structures*, Springer-Verlag (Information Science and Statistics), 2005.