

Chapitre 4

Modèles graphiques et Indépendances

Dans le chapitre d'initiation, nous avons présenté successivement les deux champs théoriques à la base des réseaux bayésiens, la théorie des graphes et la théorie des probabilités. Nous avons présenté une méthode intuitive de transposition d'un graphe causal vers un *espace probabilisé*. Enfin, nous avons mentionné l'existence d'un résultat important reliant une *propriété graphique* (la d-séparation) et une *propriété probabiliste* (l'indépendance conditionnelle).

Ce type de résultat est loin d'être évident au premier abord. En effet, on peut considérer intuitivement que les graphes permettent la représentation de relations binaires entre éléments d'un même ensemble alors que les probabilités induisent une relation qualitative *ternaire* – l'*indépendance conditionnelle* – qui ne semble pas autoriser le même genre de représentation.

Dans ce chapitre, nous allons présenter les outils théoriques permettant de concilier effectivement la théorie des graphes et la théorie des probabilités.

Cet exposé est plus général que ce qui serait strictement indispensable pour les réseaux bayésiens, puisque nous allons étudier dans le même cadre théorique les modèles non orientés, appelés aussi réseaux de Markov, et

les modèles orientés que sont les réseaux bayésiens.

En conclusion, nous présentons quelques arguments qui nous font préférer les *arcs* aux *arêtes*, ou autrement dit les réseaux bayésiens aux modèles de Markov.

4.1 Graphoïdes

La relation d'indépendance conditionnelle sur un ensemble de variables V est une relation ternaire sur l'ensemble des parties de V et peut donc se décrire, par extension, comme la liste des triplets de sous-ensembles disjoints de V vérifiant cette relation. Plus généralement, une telle liste de triplets détermine par extension une relation ternaire, quelle que soit la sémantique de la relation. Cette partie étudie les propriétés formelles et les structures intéressantes d'une telle relation.

4.1.1 Modèles d'indépendance

► DÉFINITION 4.1 (MODÈLE D'INDÉPENDANCE)

Soit V un ensemble fini et non vide de variables, on note $T(V)$ l'ensemble des triplets $\ll A \diamond B \mid C \gg$ de sous-parties disjointes A, B, C de V où A et B sont non vides. Toute sous-partie de $T(V)$ est un modèle d'indépendance.

$\ll A \diamond B \mid C \gg_M$ est la proposition logique qui indique que le triplet $\ll A \diamond B \mid C \gg$ appartient au modèle d'indépendance M .

$$\ll A \diamond B \mid C \gg_M \iff \ll A \diamond B \mid C \gg \in M$$

NOTE 4.1 [Pea88a] parle de *dependency model*. Cependant, comme chaque triplet sera interprété, dans un contexte probabiliste, comme une indépendance conditionnelle, il semble opportun de profiter de la traduction pour rectifier cette appellation.

4.1.2 Semi-graphoïde et graphoïde

Les modèles d'indépendance tels que définis ci-dessus sont des ensembles très peu contraints. Pour être utiles, ils doivent être structurés. Cette structure est donnée par de nouvelles notions, présentées ici telles qu'introduites par [Pea88a] : les semi-graphoïdes et les graphoïdes.

► DÉFINITION 4.2 (SEMI-GRAPHOÏDE)

Un modèle d'indépendance M est un semi-graphoïde s'il satisfait pour tout A, B, S, P sous-ensembles disjoints de V :

$$\begin{array}{l}
 \text{(Indépendance triviale)} \quad \ll A \diamond \emptyset \mid S \gg_M \\
 \\
 \begin{array}{l}
 \text{(Symétrie)} \quad \ll A \diamond B \mid S \gg_M \Rightarrow \ll B \diamond A \mid S \gg_M \\
 \text{(Décomposition)} \quad \ll A \diamond (B \cup P) \mid S \gg_M \Rightarrow \ll A \diamond B \mid S \gg_M \\
 \text{(Union faible)} \quad \ll A \diamond (B \cup P) \mid S \gg_M \Rightarrow \ll A \diamond B \mid (S \cup P) \gg_M \\
 \text{(Contraction)} \quad \text{et } \left\{ \begin{array}{l} \ll A \diamond B \mid (S \cup P) \gg_M \\ \ll A \diamond P \mid S \gg_M \end{array} \right\} \Rightarrow \ll A \diamond (B \cup P) \mid S \gg_M
 \end{array}
 \end{array}$$

NOTE 4.2 L'axiome d'indépendance triviale n'est pas explicitement donné dans [Pea88a]. Cependant, comme le remarque [Wil94], il semble être nécessaire et implicitement accepté¹.

On peut définir pour chaque modèle d'indépendance $M \subset T(V)$ un semi-graphoïde $SG(M)$ qui est le semi-graphoïde minimal (au sens de l'inclusion) contenant M . $SG(M)$ est la *fermeture* de semi-graphoïde de M .

► DÉFINITION 4.3 (GRAPHOÏDE)

Un modèle d'indépendance M est un graphoïde s'il est un semi-graphoïde et s'il satisfait :

$$\text{(Intersection)} \quad \text{et } \left\{ \begin{array}{l} \ll A \diamond B \mid (S \cup P) \gg_M \\ \ll A \diamond P \mid (S \cup B) \gg_M \end{array} \right\} \Rightarrow \ll A \diamond (B \cup P) \mid S \gg_M$$

De même que plus haut, $G(M)$ est la *fermeture* de graphoïde de M .

[Pea88a] donne de ces structures une représentation visuelle qui fixe assez bien les idées sur les intuitions qui les sous-tendent (voir figure 4.1).

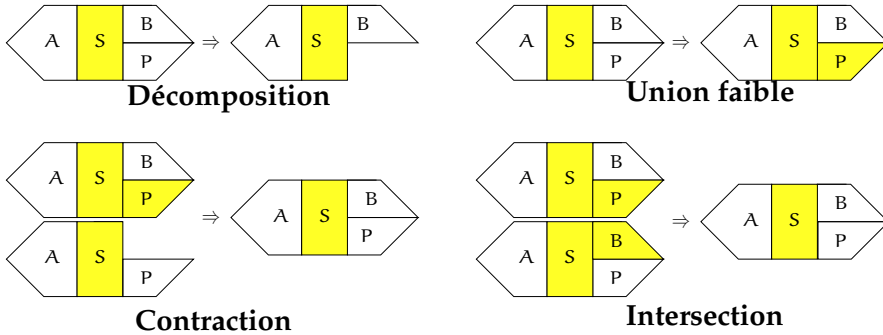


FIG. 4.1 Représentations graphiques [Pea88a] des différents axiomes caractérisant les graphoïdes et les semi-graphoïdes.

Les semi-graphoïdes et graphoïdes sont des structures très abstraites qui possèdent bien plus d'un champ d'application. [Daw98], [Stu97] ainsi

¹ Voir, par exemple, [Pea88a], paragraphe 3.2.1, page 97 :
 (...) Note that $B_1(\alpha)$ is nonempty because $I(X, S, \emptyset)$ guarantees that (...).

que [CDLS99] en citent un certain nombre telles que : l'indépendance conditionnelle probabiliste, l'indépendance conditionnelle pour les fonctions de croyance, la dépendance multivaluée, les fonctions conditionnelles naturelles, la théorie des bases de données relationnelles, la séparation dans les graphes, l'orthogonalité d'espaces vectoriels, etc. Pour ce qui nous intéresse ici, deux champs d'application sont particulièrement pertinents : l'indépendance conditionnelle et la séparation dans les graphes.

4.2 Modèle d'indépendance et loi de probabilité

Le lien entre modèle d'indépendance et loi de probabilité est assez simple. Toute loi de probabilité \mathcal{P} sur un ensemble V de variables définit un modèle d'indépendance. Il suffit en effet de lister l'ensemble des triplets (A, B, C) de sous-ensembles disjoints de V qui vérifient la propriété d'indépendance conditionnelle $A \perp\!\!\!\perp B \mid C^2$. En notant $M_{\mathcal{P}}$ le modèle d'indépendance ainsi créé, on a :

THÉORÈME 4.1

- $M_{\mathcal{P}}$ possède une structure de semi-graphoïde.
- Si \mathcal{P} est positive alors $M_{\mathcal{P}}$ possède une structure de graphoïde.

Démonstration

Soit \mathcal{P} une loi de probabilité et $M_{\mathcal{P}}$ le modèle d'indépendance vérifiant :

$$\ll A \diamond B \mid C \gg_{M_{\mathcal{P}}} \iff A \perp\!\!\!\perp B \mid C$$

$M_{\mathcal{P}}$ doit alors vérifier :

① Symétrie

$$\begin{aligned} \ll A \diamond B \mid S \gg_{M_{\mathcal{P}}} &\iff A \perp\!\!\!\perp B \mid S \\ &\Rightarrow B \perp\!\!\!\perp A \mid S \\ &\iff \ll B \diamond A \mid S \gg_{M_{\mathcal{P}}} \end{aligned}$$

② Décomposition

$$\begin{aligned} \ll A \diamond B \cup P \mid S \gg_{M_{\mathcal{P}}} &\iff A \perp\!\!\!\perp B \cup P \mid S \\ &\Rightarrow A \perp\!\!\!\perp B \mid S \\ &\iff \ll A \diamond B \mid S \gg_{M_{\mathcal{P}}} \end{aligned}$$

² Rappelons que la notion d'indépendance conditionnelle est dépendante de la loi \mathcal{P} . En toute rigueur, elle devrait donc s'écrire : $A \perp\!\!\!\perp B \mid C[\mathcal{P}]$ (voir page 355).

③ *Union faible*

$$\begin{aligned} \ll A \diamond B \cup P \mid S \gg_{M_P} &\iff A \perp\!\!\!\perp B \cup P \mid S \\ &\text{or } P = F(S \cup P) \text{ (projection)} \\ &\Rightarrow A \perp\!\!\!\perp B \cup P \mid S \cup P \\ &\Rightarrow A \perp\!\!\!\perp B \mid S \cup P \\ &\iff \ll A \diamond B \mid (S \cup P) \gg_{M_P} \end{aligned}$$

④ *Contraction et Intersection*

Ces deux démonstrations sont des utilisations directes des propriétés (P4) et (P5) de la sous-section B.2.2 page 357.

⑤ *Indépendance triviale*

$$\ll A \diamond \emptyset \mid S \gg_{M_P} \text{ car } A \perp\!\!\!\perp \emptyset \mid S.$$

□

Ce théorème énonce une implication : le modèle d'indépendance issu d'une loi de probabilité a une structure de semi-graphoïde. Il s'agit maintenant de se poser la question réciproque : qu'en est-il d'un semi-graphoïde ? Représente-t-il nécessairement une loi de probabilité ?

CONJECTURE 4.2 (PEARL AND PAZ, 1985)

Soit M un modèle d'indépendance. Si M est un semi-graphoïde alors il existe une probabilité \mathcal{P} telle que

$$\mathcal{P}(X \mid Y, Z) = \mathcal{P}(X \mid Z) \iff \ll X \diamond Y \mid Z \gg_M$$

De plus, si M est un graphoïde alors il existe une probabilité \mathcal{P} positive vérifiant cette relation.

Malheureusement, [Stu92] montre que cette conjecture est fausse.

THÉORÈME 4.3 ([STU92])

Soit une famille de propositions de type :

$$\left[\ll A_1 \diamond B_1 \mid C_1 \gg \& \dots \& \ll A_r \diamond B_r \mid C_r \gg \right] \Rightarrow \ll A_{r+1} \diamond B_{r+1} \mid C_{r+1} \gg \ (\diamond)$$

Si $|V| > 4$ alors aucune famille finie de propositions de type (\diamond) ne peut caractériser complètement les relations d'indépendance conditionnelle sur V .

NOTE 4.3 Une telle famille finie est une tentative d'axiomatisation de la relation d'indépendance conditionnelle (voir, par exemple, [GPP91] ou [Mal91]).

La définition d'un graphoïde prend effectivement la forme d'une famille de type (\diamond) page précédente. Donc, d'après ce théorème, un modèle d'indépendance muni d'une structure de graphoïde ne peut caractériser complètement toute relation d'indépendance conditionnelle. La conjecture 4.2 page précédente est donc fautive dans le cas général. Cependant, pour des classes particulières de modèles d'indépendance, cette réciproque est vérifiée, par exemple pour des modèles d'indépendance où toutes les variables de V apparaissent dans au moins un triplet ([GP90], [Mal91]), ou encore pour des modèles dits marginaux où les variables de conditionnement sont fixées ([GPP91]). De tels résultats négatifs ou restrictifs remettent en cause l'utilisation des modèles d'indépendance pour manipuler les indépendances conditionnelles probabilistes. Heureusement, [Stu97] propose un théorème (assez technique) qui permet d'établir que l'intuition de Pearl qui avait conduit à l'introduction de la notion de semi-graphoïde et à la conjecture 4.2 page précédente était bien fondée.

THÉORÈME 4.4

La fermeture F de sous-graphoïde de tout couple d'éléments de $T(V)$ est un modèle d'indépendance conditionnelle probabiliste.

C'est-à-dire : il existe une loi de probabilité \mathcal{P} sur l'ensemble des variables V telle que $\mathcal{P}(X | Y, Z) = \mathcal{P}(X | Z) \iff \langle\langle X \diamond Y | Z \rangle\rangle_F$.

4.3 Modèles d'indépendance et séparation dans les graphes

De la même façon que pour les probabilités dans la sous-section précédente, les modèles d'indépendance permettent aussi de décrire certaines propriétés en théorie des graphes. Réciproquement, représenter un modèle par un graphe permettrait de visualiser beaucoup plus facilement la relation représentée par ce modèle.

Comme il a été dit plus haut, l'écueil principal est qu'un graphe est une relation binaire entre les éléments d'un ensemble alors qu'un modèle d'indépendance est une relation ternaire entre sous-parties de cet ensemble. Le lien entre ces deux types de relation est apporté par la notion de *séparation*. La séparation établit, quel que soit le type de graphe, s'il est possible de séparer (dans un certain sens) deux sous-ensembles de nœuds par un troisième. Le « certain sens » dépend du type de graphe qui est utilisé (voir les sections suivantes qui décrivent précisément ces séparations).

► DÉFINITION 4.4 (SÉPARATION)

Soit un graphe $G = (V, E)$, pour tout triplet (X, Y, S) de sous-parties de V , disjointes deux à deux, on note $\langle X | S | Y \rangle_G$ la propriété « X et Y sont séparés par S

dans le graphe G ».

NOTE 4.4 La propriété contraposée de la séparation est appelée la *connexion*. X et Y , sous-ensembles de V , sont donc soit connectés, soit séparés par Z .

La séparation permet d'introduire une relation ternaire sur les sous-ensembles de nœuds d'un graphe. Tout comme dans la sous-section précédente, il s'agit maintenant de préciser la formalisation de cette relation comme modèle d'indépendance.

► DÉFINITION 4.5 (I-MAP, D-MAP, P-MAP, GRAPHE-ISOMORPHISME)

soit $G = (V, E)$ un graphe et $M \subset T(V)$ un modèle d'indépendance,

- G est une D-map de M ssi $\ll X \diamond Y | Z \gg_M \Rightarrow \langle X | Z | Y \rangle_G$.
- G est une I-map de M ssi $\ll X \diamond Y | Z \gg_M \Leftarrow \langle X | Z | Y \rangle_G$.
- G est une P-map de M ssi $\ll X \diamond Y | Z \gg_M \iff \langle X | Z | Y \rangle_G$.

Un modèle d'indépendance est dit *graphe-isomorphe* si et seulement s'il existe un graphe G qui soit une P-map de M .

NOTE 4.5 Les termes de D-map (*dependency map*), I-map (*independency map*) et P-map (*perfect map*) ont été gardés comme définis par [Pea88a].

Si un graphe G est une D-map d'un modèle M , toute connexion de sous-ensembles de nœuds indique une dépendance dans M (contraposée de la définition ci-dessus). Réciproquement, si le graphe est une I-map, toute séparation est alors l'indication d'une indépendance dans le modèle. Enfin, une P-map est à la fois une D-map et une I-map.

En notant M_G le modèle d'indépendance induit par la séparation dans le graphe G (c'est-à-dire $\langle X | Z | Y \rangle_G \iff \ll X \diamond Y | Z \gg_{M_G}$), on peut aussi dire que pour un modèle d'indépendance M :

- G est une D-map de M si et seulement si $M \subset M_G$ (certaines indépendances lues par séparation dans le graphe G ne sont pas dans le modèle M).
- G est une I-map de M si et seulement si $M \supset M_G$ (toute indépendance lue par séparation dans le graphe G est présente dans le modèle M ; cependant, certaines indépendances du modèle M ne sont pas représentées).
- G est une P-map de M si et seulement si $M = M_G$.

EXEMPLE 4.6 Pour un ensemble de variables V , le graphe (V, \emptyset) est une I-map de tout modèle d'indépendance. De même, le graphe non orienté complet $(V, V \times V)$ est une D-map de tout modèle d'indépendance.

De fait, l'expressivité de chaque type de graphe va dépendre de l'exacte définition de la séparation puisque cette dernière aura des répercussions

fortes sur la classe des modèles qui peut y être représentée. Ainsi il s'avère que certains modèles n'ont de P-map dans aucun type de graphe.

Si un modèle n'a pas de P-map, aucune représentation graphique ne sera complète. Utiliser une représentation graphique d'un modèle d'indépendance n'est pas intéressant si cette représentation est capable de « mentir » sur le modèle. Un moindre mal est que le mensonge soit par omission et le plus faible possible. C'est la raison pour laquelle la notion la plus importante est celle de *I-map minimale*.

Les trois sections suivantes s'attachent à décrire plus exactement la séparation dans les différents types de graphe.

4.4 Modèles non orientés : réseaux de Markov

4.4.1 Définition

► DÉFINITION 4.6 (SÉPARATION NON ORIENTÉE)

Soit $G = (V, E)$ un graphe non orienté ; pour tout triplet (X, Y, Z) de sous-ensembles disjoints de G , X est séparé de Y par Z dans G (noté $\langle X | Z | Y \rangle_G$) si et seulement si toute chaîne d'un nœud de X vers un nœud de Y passe par un nœud de Z .

$$\langle X | Z | Y \rangle_G \iff \left[\begin{array}{l} \forall (x_i)_{i \in \{1 \dots p\}} \subset \mathbb{N} \text{ chaîne de } G, \\ \left. \begin{array}{l} x_1 \in X \\ x_p \in Y \end{array} \right\} \Rightarrow \exists i \in \{1 \dots p\}, x_i \in Z \end{array} \right]$$

EXEMPLE 4.7 Dans la figure 4.2, toutes les chaînes de $\{1, 4, 7\}$ vers $\{3, 5, 6\}$ passent nécessairement par 2 : $\langle \{1, 4, 7\} | \{2\} | \{3, 5, 6\} \rangle$. De même, toutes les chaînes de 2 vers 5 passent nécessairement par 6 ou 3 : $\langle \{2\} | \{6, 3\} | \{5\} \rangle$.

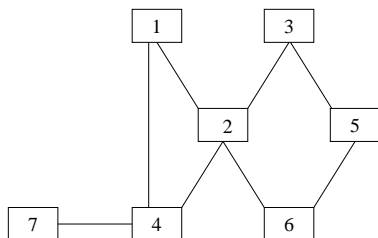


FIG. 4.2 Séparation dans un graphe non orienté

4.4.2 Séparation et indépendances : propriétés de Markov

Soit un graphe $G = (V, E)$ et une probabilité \mathcal{P} sur l'ensemble des variables V . La question qui se pose ici est de trouver le rapport entre séparation et probabilité (ou plus exactement *indépendance conditionnelle*), qui permet de mesurer la précision des informations que le graphe G apporte sur \mathcal{P} . Cette mesure est donnée par une série de propriétés dites de Markov qui sont plus ou moins fortes : plus le graphe vérifie de propriétés fortes, mieux il représente la loi \mathcal{P} .

► DÉFINITION 4.7 (PROPRIÉTÉS DE MARKOV)

Le graphe G et la loi \mathcal{P} peuvent vérifier :

(P) **la propriété de Markov par paire**

si et seulement si $\forall x, y \in V$, x et y non adjacents dans G ,

$$x \perp\!\!\!\perp y \mid (V \setminus \{x, y\})$$

(L) **la propriété de Markov locale**

si et seulement si $\forall x \in V$,

$$x \perp\!\!\!\perp (V \setminus \bar{\vartheta}_x) \mid \vartheta_x$$

où ϑ_x est le voisinage du nœud x et $\bar{\vartheta}_x$ est la fermeture de voisinage de x , c'est-à-dire $\bar{\vartheta}_x = \vartheta_x \cup \{x\}$ (voir section A.3 page 341).

(G) **la propriété de Markov globale**

si et seulement si $\forall A, B, S \subset V$ disjoints,

$$\langle A \mid S \mid B \rangle_G \Rightarrow A \perp\!\!\!\perp B \mid S$$

► DÉFINITION 4.8 (RÉSEAU MARKOVIEN)

Un graphe vérifiant (G) pour \mathcal{P} est un réseau markovien de \mathcal{P} .

La propriété (G) de Markov globale correspond à un graphe G , I-map du modèle d'indépendance conditionnelle probabiliste engendré par \mathcal{P} . Les propriétés sont ici données dans l'ordre de force croissante. Plus précisément, ces trois propriétés sont en relation comme suit :

THÉORÈME 4.5

$$(G) \Rightarrow (L) \Rightarrow (P)$$

Démonstration

- (G) \Rightarrow (L) : on a toujours $\langle x \mid \vartheta_x \mid (V \setminus \bar{\vartheta}_x) \rangle_G$.

- (L) \Rightarrow (P) : soit un graphe G et une loi \mathcal{P} pour lesquels (L) est vérifiée.

$$x \perp\!\!\!\perp (V \setminus \bar{\vartheta}_x) \mid \vartheta_x$$

Soit x et y non adjacents dans G. D'après la propriété (P3) de la sous-section B.2.2 page 357,

$$x \perp\!\!\!\perp (V \setminus \bar{\vartheta}_x) \mid [\vartheta_x \cup ((V \setminus \bar{\vartheta}_x) \setminus \{y\})]$$

ou encore $x \perp\!\!\!\perp (V \setminus \bar{\vartheta}_x) \mid V \setminus \{x, y\}$

Or $y \in (V \setminus \bar{\vartheta}_x)$ (en tant que non adjacent de x) et donc, d'après la propriété (P2) de la sous-section B.2.2 page 357,

$$x \perp\!\!\!\perp y \mid V \setminus \{x, y\}$$

□

Les réciproques (P) \Rightarrow (L) \Rightarrow (G) seraient bien plus intéressantes que le théorème lui-même. En effet, tester (P) ne demande que de tester chaque paire de variables non adjacentes alors que tester (G) demande un calcul sur un grand nombre de triplets de sous-ensembles de V . Avec les réciproques, il serait aisé d'obtenir des résultats sur la qualité de représentation du modèle d'indépendance de \mathcal{P} par le graphe G. Malheureusement, elles ne sont pas toujours vraies. Le résultat³ exact est le suivant :

THÉORÈME 4.6

Si la loi \mathcal{P} vérifie la propriété suivante :

$$\forall A, B, C, D \text{ sous-ensembles disjoints de } V,$$

$$\text{Si } A \perp\!\!\!\perp B \mid C \cup D \text{ et } A \perp\!\!\!\perp C \mid B \cup D \text{ alors } A \perp\!\!\!\perp B \cup C \mid D$$

Alors, l'équivalence (G) \iff (L) \iff (P) est vérifiée.

NOTE 4.8 On note que la propriété que doit vérifier \mathcal{P} est exactement la propriété P5 page 357.

4.4.3 Réseaux de Markov et factorisation

Dans le chapitre précédent, on a montré la relation entre indépendance conditionnelle et factorisation de la probabilité (voir théorème B.2 page 356). Nécessairement, une relation existe entre propriétés de Markov et factorisation.

³ Pearl and Paz

► DÉFINITION 4.9 (FACTORISATION)

Soit une loi de probabilité jointe $\mathcal{P}(V)$ et un graphe non orienté $G = (V, E)$, on dit que \mathcal{P} possède une factorisation selon G si et seulement si pour tout sous-graphe complet S de G , il existe une fonction $\Psi_S(V)$ ne dépendant que des nœuds de S telle que :

$$\mathcal{P}(V) = \prod_{S \text{ sous-graphe complet de } G} \Psi_S(V)$$

Ψ_S est appelée un *potentiel*. Cette factorisation de \mathcal{P} en potentiels n'est pas unique. De fait, par multiplication des potentiels, on peut restreindre la factorisation à l'ensemble des cliques de G .

PROPRIÉTÉ 4.10

Soit \mathcal{C} l'ensemble des cliques de G . \mathcal{P} se factorise selon G si et seulement si pour toute clique C de \mathcal{C} , il existe un potentiel Ψ_C tel que :

$$\mathcal{P}(V) = \prod_{C \in \mathcal{C}} \Psi_C(V)$$

Toutes les probabilités ne se factorisent pas ainsi. On note souvent $\mathcal{M}_F(G)$ l'ensemble des probabilités pouvant se factoriser ainsi selon G . Pour une probabilité \mathcal{P} , on note (F) la propriété « $\mathcal{P} \in \mathcal{M}_F(G)$ ». La relation entre factorisation et propriétés de Markov peut alors s'énoncer comme suit :

THÉORÈME 4.7

Une probabilité \mathcal{P} pouvant se factoriser selon G vérifie alors la propriété globale de Markov (G) (qui, elle-même, implique les deux autres propriétés de Markov : locale (L) puis par paire (P)).

$$(F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P).$$

De plus, si \mathcal{P} est positive,

$$(F) \iff (G) \iff (L) \iff (P).$$

4.4.4 Limites

Grâce à ces résultats, la représentation graphique d'une probabilité sur plusieurs variables par un graphe non orienté sur ces variables semble bien cernée. Il est possible de discerner assez aisément les I-map d'une loi de probabilité et même d'essayer de les améliorer pour obtenir une I-map minimale. Pourquoi chercher plus loin et essayer de trouver d'autres repré-

sentations certainement plus complexes, en tout cas moins intuitives ? La raison principale est la limitation du modèle non orienté.

Soit un système de trois variables $\{D_1, D_2, S\}$. D_1 et D_2 représentent toutes deux le tirage d'un dé (valeur entre 1 et 6); S représente la somme de ces deux tirages (valeur entre 2 et 12). Ce système vérifie le modèle d'indépendance représenté dans le tableau 4.1.

$D_1 \perp\!\!\!\perp D_2$	Les deux tirages sont indépendants
non $D_1 \perp\!\!\!\perp S$ non $D_2 \perp\!\!\!\perp S$	Chaque tirage et la somme sont dépendants
non $D_1 \perp\!\!\!\perp D_2 S$	La connaissance de la somme rend dépendants les deux tirages

TAB. 4.1 *Modèle d'indépendance de $\{D_1, D_2, S\}$*

Pour représenter un tel modèle, il faudrait pouvoir :

- ne pas relier D_1 et D_2 ;
- relier D_1 et S , relier D_2 et S ;
- trouver un moyen pour qu'il n'y ait pas $\langle D_1 | S | D_2 \rangle$.

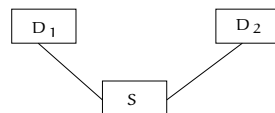


FIG. 4.3 *Représentation non orientée des relations dans $\{D_1, D_2, S\}$.*

Il n'est pas possible de représenter par une P-map un tel modèle dans un graphe non orienté. La figure 4.3 est une représentation possible. Ce graphe ne vérifie que les deux premiers points mais ne vérifie pas le troisième. En fait, il ne vérifie ni (G), ni (L), ni même (P). Une autre représentation serait un graphe complet entre les trois variables, mais le premier point ne serait pas vérifié et la représentation ne serait pas meilleure.

D'où l'intérêt d'aller chercher des modèles plus complexes qui permettent de mieux représenter (ou au moins différemment) ces lois de probabilité en utilisant des graphes orientés.

4.5 Modèles orientés : réseaux bayésiens

4.5.1 Définitions

La séparation dans les graphes orientés est plus complexe que dans les graphes non orientés. En effet, il ne suffit pas de savoir si au moins un nœud de tout chemin entre X et Y appartient à Z ; il faut aussi que ce nœud vérifie des conditions supplémentaires, apportées par les orientations des arcs. Cette notion provient également de [Pea87a]. Tout comme lui, on présen-

tera d’abord la notion de *chaîne active* qui facilite grandement celle de séparation orientée (*directed separation* ou *d-séparation*). Dans toute cette partie, on considérera que les graphes orientés dont on parle sont sans circuit.

Soit une chaîne $C = (x_i)_{i \in I}$ dans un graphe orienté \vec{G} . On dira que x_i est un *puits* de la chaîne C s’il est du type : $x_{i-1} \rightarrow x_i \leftarrow x_{i+1}$; c’est-à-dire s’il est un sommet à arcs convergents dans la chaîne.

► DÉFINITION 4.11 (CHAÎNE ACTIVE, BLOQUÉE)

Soit une chaîne $C = (x_i)_{i \in I}$ dans \vec{G} et Z un sous-ensemble de nœuds de \vec{G} . C est une chaîne active par rapport à Z si les deux conditions suivantes sont réunies :

- Tout puits de C a l’un de ses descendants dans \vec{G} ou lui-même qui appartient à Z .
- Aucun élément de C qui n’y est pas un puits n’appartient à Z .

Une chaîne non active par rapport à Z est dite *bloquée* par Z .

Le type de modèles d’indépendance que peuvent prendre en compte les graphes orientés (et que ne pouvaient pas prendre en compte les graphes non orientés) est caractérisé dans cette définition un peu complexe : comment représenter la situation où deux variables sont indépendantes mais où la connaissance d’une troisième les rendrait dépendantes⁴ ? Pour deux variables qui ne sont reliées que par une unique chaîne, cette situation se présente si l’unique chaîne est bloquée par la troisième variable en question.

► DÉFINITION 4.12 (D-SÉPARATION)

Soit $\vec{G} = (V, E)$ un graphe orienté, pour tout triplet (X, Y, Z) de sous-ensembles disjoints de V , X est *d-séparé* de Y par Z dans \vec{G} (noté $\langle X | Z | Y \rangle_{\vec{G}}$) si et seulement si toute chaîne $(x_i)_{i \in \{1 \dots p\}}$ avec $x_1 \in X$ et $x_p \in Y$ est bloquée par Z .

La d-séparation est certainement moins lisible que la séparation. Cependant, il est facile d’automatiser une procédure de reconnaissance ou même, avec un peu d’expérience, de repérer directement les sous-ensembles d-séparés d’un graphe.

EXEMPLE 4.9 Dans la figure 4.4 ci-après,

- non $\langle \{1\} | \{2\} | \{3\} \rangle$ - il y a deux chaînes : $\{1, 2, 3\}$ et $\{1, 4, 2, 6, 5, 3\}$. La seconde est bloquée par 2 (qui n’est pas un puits dans cette chaîne) et par 6. En revanche, la première est active puisque 2 y est un puits.
- De même, si on considère la d-séparation de $\{1\}$ et $\{3\}$ par $\{7\}$, la chaîne $\{1, 2, 3\}$ est active puisque 7 est un descendant de 2, puits de la chaîne.
- $\langle \{3\} | \{4\} | \{7\} \rangle$ - toutes les chaînes de 3 à 7 passent par 4 et 4 n’y est jamais un puits. Donc toutes les chaînes de 3 à 7 sont bloquées par 4.

⁴ C’est exactement le cas dans la sous-section 4.4.4 page 83.

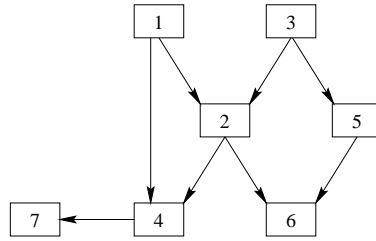


FIG. 4.4 Séparation dans un graphe orienté

4.5.2 Propriétés de Markov dans les graphes orientés

L'étude des propriétés de Markov dans les graphes orientés – et donc de la qualité de représentation des modèles orientés – est à la base du développement du domaine des réseaux bayésiens. On peut ainsi citer [KSC84], [Pea87a], [Smi89], etc.

La définition de la d-séparation montre que la symétrie entre tous les voisins d'un nœud dans un graphe est brisée par l'orientation : les puits et les descendants des puits jouent un rôle très particulier. Ce rôle se répercute dans les propriétés orientées de Markov par l'importance de la notion de *non-descendant* dans ces énoncés.

► DÉFINITION 4.13 (PROPRIÉTÉS ORIENTÉES DE MARKOV)

Le graphe \vec{G} et la loi \mathcal{P} peuvent vérifier :

(OP) **Propriété orientée de Markov par paire**

si et seulement si $\forall x, y \in V$, x et y non adjacents dans \vec{G} et $y \in \text{nd}(x)$,

$$x \perp\!\!\!\perp y \mid (\text{nd}(x) \setminus \{y\})$$

(OL) **Propriété orientée de Markov locale**

si et seulement si $\forall x \in V$,

$$\{x\} \perp\!\!\!\perp \text{nd}(x) \mid \Pi_x$$

(OG) **Propriété orientée de Markov globale**

si et seulement si $\forall A, B, S \subset V$ disjoints,

$$\langle A \mid S \mid B \rangle_{\vec{G}} \Rightarrow A \perp\!\!\!\perp B \mid S$$

► DÉFINITION 4.14 (RÉSEAU BAYÉSIEN)

Un graphe orienté vérifiant (OG) pour \mathcal{P} est un réseau bayésien de \mathcal{P} .

Comme pour les graphes non orientés, (OG) représente la plus forte propriété que peut vérifier un graphe orienté pour y lire directement une *indépendance conditionnelle*.

De même que pour les réseaux markoviens, on a :

THÉORÈME 4.8

De plus,

$$(OG) \iff (OL) \Rightarrow (OP)$$

Il est à remarquer qu’il y a toujours équivalence entre (OG) et (OL), contrairement au cas non orienté. Par contre, le théorème 4.6 page 82 reste vrai pour l’équivalence entre (OP) et (OL). En particulier, si la loi \mathcal{P} est positive, alors il y a équivalence entre les trois propriétés orientées de Markov.

4.5.3 Réseaux bayésiens et factorisation

De même que pour les réseaux de Markov, il faut maintenant relier propriétés de Markov et factorisation de la probabilité. Cette factorisation est étonnamment simple et suffit à elle seule à expliquer une grande partie de l’intérêt porté aux réseaux bayésiens.

► DÉFINITION 4.15 (FACTORISATION RÉCURSIVE)

Soit une loi de probabilité jointe $\mathcal{P}(V)$ et un graphe orienté $\vec{G} = (V, E)$. On dit que \mathcal{P} possède une factorisation récursive selon \vec{G} si et seulement si pour tout nœud X de G , il existe une fonction $k_X(X, \Pi_X)$ telle que :

$$\mathcal{P}(V) = \prod_{X \in V} k_X(X, \Pi_X)$$

De plus, les fonctions k_X (kernels ou noyaux de Markov) sont les lois de probabilité conditionnelles de X étant donné Π_X :

$$\mathcal{P}(V) = \prod_{X \in V} \mathcal{P}(X | \Pi_X)$$

Bien sûr, toutes les lois ne se factorisent pas ainsi. Soit (OF) la propriété « \mathcal{P} se factorise récursivement selon \vec{G} ». La relation entre factorisation récursive et propriétés orientées de Markov peut alors s’énoncer comme suit :

THÉORÈME 4.9

$$(OF) \iff (OG)$$

Par ailleurs, il n'y a pas forcément unicité du graphe \vec{G} permettant la factorisation récursive de \mathcal{P} , on appelle *classe d'équivalence de Markov* l'ensemble des graphes permettant une factorisation récursive de \mathcal{P} , c'est-à-dire l'ensemble des graphes représentant le même modèle d'indépendance.

Bien plus que pour les réseaux de Markov, la factorisation récursive est intéressante aussi pour la représentation de la loi : les noyaux de Markov sont des fonctions dépendant chacune d'un unique nœud du graphe. Ce qui implique que la représentation graphique de la loi peut être améliorée grâce à une localisation des données qui suit exactement le graphe, sans structure complémentaire. Le graphe ainsi augmenté représente qualitativement la loi de probabilité (en indiquant les indépendances conditionnelles) mais aussi quantitativement (en permettant les calculs par factorisation).

EXEMPLE 4.10 Soit le graphe \vec{G} de la figure 4.4 page 86. Une loi \mathcal{P} se factorisant récursivement par rapport à \vec{G} peut s'écrire :

$$\mathcal{P}(1, 2, 3, 4, 5, 6, 7) = \mathcal{P}(1) \cdot \mathcal{P}(3) \cdot \mathcal{P}(2 | 1, 3) \cdot \mathcal{P}(4 | 1, 2) \cdot \mathcal{P}(5 | 3) \cdot \mathcal{P}(6 | 2, 5) \cdot \mathcal{P}(7 | 4)$$

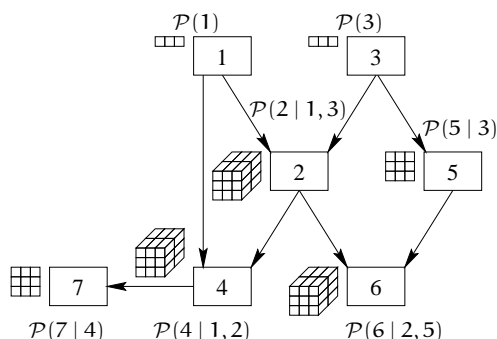


FIG. 4.5 Représentation graphique d'un réseau bayésien – les probabilités conditionnelles (ainsi que leurs dimensions) sont représentées pour chaque nœud.

Ce qui est intéressant à noter est, tout d'abord, la compression que représente une telle factorisation : en supposant que toutes les variables sont binaires, $\mathcal{P}(1, 2, 3, 4, 5, 6, 7)$ représente un tableau comprenant $2^7 = 128$ valeurs. La factorisation, elle, est représentée par un ensemble de tableaux comprenant en tout $2 + 2 + 8 + 8 + 4 + 8 + 4 = 36$ valeurs. On représente donc exactement la même loi avec trois fois moins de valeurs.

La seconde propriété intéressante d'une telle décomposition est que les probabilités conditionnelles impliquées dans la factorisation sont très faciles à lire à partir du graphe puisqu'elles sont toutes liées à un nœud particulier. En fait, dans la représentation graphique, on peut joindre à chaque nœud sa probabilité conditionnelle suivant ses parents (voir figure 4.5 page précédente), ce qui augmente l'aspect synthétique de la représentation.

4.5.4 Limites

Tout comme pour les réseaux de Markov, il existe des limites à la capacité d'expressivité du modèle des réseaux bayésiens. Les questions sont toujours : existe-t-il des lois de probabilité (des modèles d'indépendance) qui n'étaient pas représentables par un réseau de Markov mais qui le soient par un réseau bayésien ? Existe-t-il des lois de probabilité qui ne soient pas représentables par un réseau bayésien⁵ ?

En ce qui concerne la première question, l'exemple de la sous-section 4.4.4 page 83 — dont le modèle d'indépendance est donné par le tableau 4.1 page 84 — est bien représentable par un réseau bayésien (voir la figure 4.6). En effet, le puits S de l'unique chaîne entre D₁ et D₂ assure la dépendance de D₁ et D₂ étant donné S et l'indépendance marginale de D₁ et D₂.

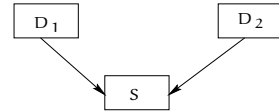
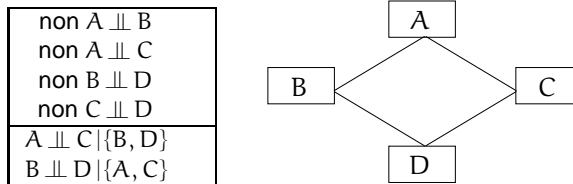


FIG. 4.6 Résolution orientée de la figure 4.3 page 84

Certains modèles sont cependant problématiques. Soit une probabilité sur les quatre variables A, B, C, D vérifiant le modèle d'indépendance conditionnelle décrit dans le tableau 4.2. Ce tableau présente aussi la forme d'un réseau de Markov pour cette loi.



TAB. 4.2 Modèle d'indépendance de {A, B, C, D} et un réseau de Markov le représentant

Il n'est pas possible de trouver un réseau bayésien représentant ce modèle d'indépendance : la structure de base (due aux indépendances margi-

⁵ De même que plus haut, une loi est représentable si elle possède une P-map.

nales) doit être celle du réseau de Markov présenté (le réseau bayésien doit avoir comme graphe non orienté sous-jacent ce réseau de Markov). Mais étant donné qu'un réseau bayésien doit être un DAG, l'un de ces nœuds doit être un puits. Il est alors facile de vérifier que l'une des indépendances conditionnelles ($A \perp\!\!\!\perp C \mid \{B, D\}$ ou $B \perp\!\!\!\perp D \mid \{A, C\}$) sera ainsi nécessairement violée.

4.6 Pourquoi des arcs plutôt que des arêtes ?

Les deux modèles – orienté et non orienté – ont montré leurs imperfections dans le sens où aucun n'a un pouvoir de représentation au moins égal à celui de l'autre. Comment choisir dans ces conditions entre une modélisation orientée et une modélisation non orientée ? Ou plutôt, puisque le suspense n'est pas de mise, pourquoi choisir la représentation sous forme de réseau bayésien ? Il s'agit ici d'essayer de lister l'ensemble des raisons qui participent à ce choix.

4.6.1 Factorisation

Comme simple rappel, un réseau de Markov permet de factoriser la loi de probabilité jointe comme suit :

$$\mathcal{P}(V) = \prod_{C \in \mathcal{C}} \Psi_C(V)$$

où \mathcal{C} est l'ensemble des cliques du graphe non orienté. Cette factorisation est à comparer à la factorisation récursive des réseaux bayésiens :

$$\mathcal{P}(V) = \prod_{X \in V} \mathcal{P}(X \mid \Pi_X)$$

Il est alors aisé de voir les avantages de la seconde représentation :

- **Unicité** : la factorisation récursive a le grand mérite d'être unique (étant donné le graphe) alors que la factorisation en potentiels ne l'est pas. Il existe une famille de potentiels qui vérifient cette égalité.
- **Localisation** : alors que dans un réseau bayésien, il semble clair et figuratif de localiser l'information dans chaque nœud du graphe, la factorisation dans un réseau de Markov localise l'information dans chaque clique ; ceci nécessite l'utilisation de la structure seconde qu'est le graphe de jonction pour pouvoir localiser aisément cette information.

4.6.2 Sémantique et causalité

L'orientation permet aussi de garantir une certaine lisibilité du graphe. Même si le sens des arcs peut être illusoire, ou si la désorientation a le mérite d'être plus proche des données (puisque une corrélation en statistique est une opération symétrique), il n'en reste pas moins que l'orientation est une aide précieuse pour la lecture et la compréhension d'un réseau.

Par exemple, la figure 4.7 montre assez clairement comment le théorème de Bayes est représentable uniquement grâce à l'orientation d'une liaison entre deux nœuds.

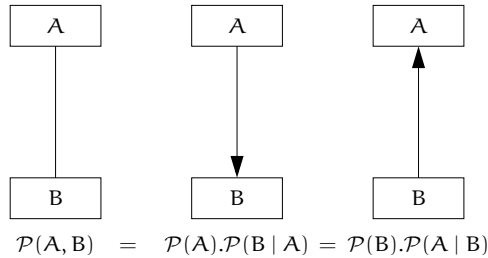


FIG. 4.7 Le théorème de Bayes comme inversion d'arcs

Il n'y a pas lieu ici de parler très précisément de *causalité*. Ce domaine reste un sujet polémique où tous les avis sont encore représentés. La question de savoir si la causalité est une notion mathématiquement représentable ou non n'est pas tranchée ; de même, celle de savoir si une quelconque causalité peut être retrouvée statistiquement à partir d'une base de données. Disons simplement que pour l'instant, le domaine est ouvert⁶.

Toutefois, la *causalité* est une notion intuitive, qu'un humain comprend finalement plus naturellement que la corrélation statistique. Dans le cadre d'un processus de construction d'un modèle par un expert, cette causalité a donc un rôle important. Pour plus de précisions, on pourra se référer, par exemple, à [SGS00] et [Pea99].

D'un point de vue pratique, pour relier la causalité au théorème de Bayes représenté dans la figure 4.7, il semble opportun de représenter la loi jointe d'une cause A et de sa conséquence B par la factorisation $\mathcal{P}(A) \cdot \mathcal{P}(B | A)$: la probabilité que la conséquence se produise – sachant que la cause s'est déjà produite – a bien un sens causal. Seuls les réseaux bayésiens, par opposition aux réseaux markoviens, sont capables d'une telle repré-

⁶ Pour plus de détails, voir principalement [Pea01].

sentation de la causalité. Un réseau bayésien qui respecte cette causalité est nommé *réseau causal*.

4.6.3 Pragmatisme

La phase de construction d'un modèle est bien sûr une phase sensible. Il y a principalement deux méthodes pour le construire : soit utiliser l'apprentissage automatique (et toutes les méthodes qui s'y réfèrent), soit être aidé d'experts qui seront capables de transposer leurs connaissances du domaine dans la formalisation du modèle. Cette transposition n'est pas une tâche facile. Il est donc important de faciliter le plus possible le travail de l'expert. Lui demander de décrire des potentiels de cliques de variables semble vraiment difficile. Alors que l'estimation localisée nœud par nœud, et si possible utilisant la causalité, est beaucoup plus du domaine de l'accessible. Cependant, il ne s'agit pas de se leurrer. Ce travail reste souvent très difficile.