Introduction aux modèles graphiques	**2009/2010**
Cours 3 — 14 octobre 2009	
Enseignant: Francis Bach	*Scribe: Thierry Guillemot, Amandine Schreck*

3.1 Notations et rappels de théorie des graphes

Un graphe G est la donnée d'un ensemble V de sommets et d'un ensemble E d'arêtes, c'est à dire d'éléments de $V \times V$. Les sommets sont souvent appelés noeuds, en anglais "vertex" ou "vertices". Les arêtes sont quant à elles appelées "edge" en anglais.

Il existe deux types de graphes : les graphes orientés et les graphes non orientés.

3.1.1 Graphes non orientés

Un graphe $G = (V, E)$ non orienté est un graphe vérifiant la propriété suivante :

$$\forall (u, v) \in V \times V \quad (u, v) \in E \Leftrightarrow (v, u) \in E$$

Définitions :

Voisins L'ensemble des voisins d'un noeud v, noté $N(v)$, est par définition

$$N(v) = \{u \in V \mid (v, u) \in E\}$$

Clique Une clique est un sous ensemble de V totalement connecté (ie. $C \subset V$ est une clique si $\forall (u, v) \in C \times C \ (u, v) \in E$)

Clique maximale Les cliques maximales sont des éléments maximaux pour l'inclusion (ie. une clique maximale n'est strictement incluse dans aucune clique)

Chemin Un chemin est une suite d'éléments deux à deux voisins, et globalement distincts (chaque noeud n'apparaît qu'une fois au plus).

Cycle Un cycle est une séquence d'éléments $(v_0, \ldots, v_q)$ deux à deux voisins et tels que :

1. $v_0 = v_q$
2. $\forall\, i, j,\ v_i \neq v_j \quad$ si $(i, j) \neq (0, q)$

Séparation Pour A,B,C $\in C^3$, on dit que C sépare A et B si tous les chemins de A vers B passent par C.

Composante connexe Soit R la relation : u R v si et seulement si il existe un chemin de u vers v. On appelle composante connexe les classes d'équivalence pour R. Les graphes traités dans la suite du cours contiendront une seule composante connexe. Si ils en contiennent plusieurs, chacune pourra être traitée séparément.

3.1.2 Graphes orientés

les définitions en rapport avec les graphes orientés (en anglais "directed graphs") sont les suivantes :

Parent On dit que $u \in V$ est parent de v si $(u, v) \in E$

Enfant On dit que v est enfant de u si u est parent de v

Ancêtre On dit que u est un ancêtre de v si il existe un chemin de u vers v. On dit alors que v est un descendant de u

Cycle Un cycle est un n-uplet $(u_1, \ldots, u_n)$ (avec $n \geq 3$) tel que :

1. $\forall p \in \{1, \ldots, n-1\}, \left(u_p, u_{p+1}\right) \in E$
2. $u_1 = u_n$
3. les noeuds $u_1, \ldots, u_{n-1}$ sont distincts

Graphe acyclique Un graphe acyclique est simplement un graphe sans cycle. Il sera parfois noté DAG par abréviation (de l'anglais "directed acyclic graph")

Ordre topologique Pour les graphes orientés acycliques on appelle ordre topologique un ordre I tel que : pour tout u et v de V, si u est parent de v alors $I(u) \leq I(v)$.

Proposition 3.1 *G est un DAG* $\Longleftrightarrow$ *Il existe au moins un ordre topologique pour G. Le plus souvent, cet ordre topologique n'est pas unique.*

3.2 Notations et rappels de probabilités

Dans la suite, nous considèrerons un ensemble $\{X_1, X_2 \ldots, X_n\}$ de variables aléatoires **discrètes** et nous noterons x_i la réalisation de la variable X_i pour tout $i = 1 \ldots n$. On note $X = (X_1, \ldots, X_n)$, et $X_A = (X_a)_{a \in A}$ où $A \subset \{1 \ldots n\}$. Nous garderons à l'esprit que n est en pratique assez grand.

Loi jointe

Les X_i peuvent être définies simplement par la donnée de leur loi jointe :

$$p(x) = P(X = x) = P(X_1 = x_1, \ldots, X_n = x_n)$$

Loi marginale

Par abus de notation on utilisera p pour désigner :

$$p(x_A) = \sum_{x_{A^c}} p(x_A, x_{A^c})$$

Propriété

Pour toute loi p, on a le résultat :

$$p(x_1, \ldots, x_n) = p(x_1)\, p(x_2|x_1) \ldots p(x_n|x_1, \ldots, x_{n-1})$$

3.3 Modèles graphiques orientés

3.3.1 Définitions et premières proprietés

Définition 3.2 *Soient $G = (V, E)$ un DAG avec $V = \{1, ..., n\}$, et $X_1, ..., X_n$ n variables aléatoires discrètes.*

$$L(G) = \{\text{loi } p \text{ telles que } p(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f_i(x_i, x_{\pi_i})\}$$

avec :

1. *π_i est l'ensemble des parents de i*
2. $\forall i, f_i \geq 0$
3. $\forall i, \sum_{x_i} f_i(x_i, x_{\pi_i}) = 1$

Si $p \in L(G)$, on dit que p se factorise dans G.

ATTENTION : l'indépendance par paires n'implique pas forcément l'indépendance jointe.

Proposition 3.3 $p(x) \in L(G) \Rightarrow \forall i, f_i(x_i, x_{\pi_i}) = p(x_i|x_{\pi_i})$

Démonstration Elle se fait par récurrence sur le nombre de sommets. Le cas $n = 1$ est trivial. On passe de $n-1$ à n en considérant que comme G est acyclique, il contient une feuille. Sans perte de généralité, on peut supposer que cette feuille est le noeud n. Si $p \in L(G)$,on a alors :

$$\begin{aligned}
p(x_1, \ldots, x_{n-1}) &= \sum_{x_n} p(x) \\
&= \sum_{x_n} \left(\prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i}) f_n(x_n, x_{\pi_n}) \right) \\
&= \prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i}) \sum_{x_n} f_n(x_n, x_{\pi_n}) \\
&= \prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i})
\end{aligned}$$

On obtient donc que la loi de $(X_1, \ldots, X_{n-1})$ se factorise dans le sous-graphe constitué de G privé de la feuille n. Alors, par hypothèse de récurrence, on a pour tout $i \leq n-1$, $f(x_i, x_{\pi_i}) = p(x_i|x_{\pi_i})$. En notant :

$$g(x) = \prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i})$$

on a :

$$\begin{aligned}
p(x_n, x_{pi_n}) &= \sum_{x_i, i \notin \{n\} \cup \pi_n} p(x) \\
&= \sum_{x_i, i \notin \{n\} \cup \pi_n} g(x) f_n(x_n, x_{\pi_n}) \\
&= h(x_{\pi_n}) f_n(x_n, x_{\pi_n})
\end{aligned}$$

avec

$$h(x_{\pi_n}) = \sum_{x_i, i \notin \{n\} \cup \pi_n} g(x)$$

qui ne dépend que de x_{π_n}. Puis :

$$p(x_{\pi_n}) = \sum_{x_n} p(x_n, x_{\pi_n}) = h(x_{\pi_n})$$

Et finalement :

$$p(x_n|x_{\pi_n}) = \frac{p(x_n, x_{\pi_n})}{p(x_{\pi_n})} = \frac{h(x_{\pi_n})f_n(x_n, x_{\pi_n})}{h(x_{\pi_n})} = f_n(x_n, x_{\pi_n})$$

■

Corollaire 3.4 *Si $p(x) \in L(G)$, alors $\sum_x p(x) = 1$*

Démonstration Supposons sans perte de généralité que 1 est racine du DAG. Alors :

$$f_1(x_1) = \sum_{x_2,\dots,x_n} p(x)$$

et :

$$\sum_x p(x) = \sum_{x_1} f_1(x_1) = 1$$

■

Proposition 3.5 $p(x) \in L(G) \iff \forall x,\ p(x) = \prod_{i=1}^n p(x_i|x_{\pi_i})$

3.3.2 Exemples

1. $E = \emptyset$
   Alors $p(x) \in L(G) \iff p(x) = \prod_{i=1}^n p(x_i) \iff (X_1, \dots, X_n)$ sont indépendantes.
2. $\forall\ i \leq j,\ (i, j) \in E$
   Alors pour toute loi p, $p(x) \in L(G)$.

3.3.3 Exemples canoniques de graphes à trois noeuds

– **"chaîne de Markov"**

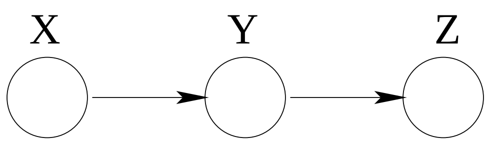

Figure 3.1. Configuration 1.

Montrons que pour toute loi se factorisant dans G, alors, $X \perp Z|Y$, en montrant que $p(z|x, y) = p(z|y)$. On a :

$$p(z|x, y) = \frac{p(x, y, z)}{p(x, y)} = \frac{p(x)p(y|x)p(z|y)}{p(x)p(y|x)}$$

$$p(z|x, y) = p(z|y)$$

– **"Cause commune"**
ATTENTION : le terme de "cause" est à bannir en statistiques car celle-ci est toujours difficile à mettre en évidence. En particulier, ce cours étudie les corrélations entre différentes variables aléatoires, à ne pas confondre avec des potentielles relations de causalité.

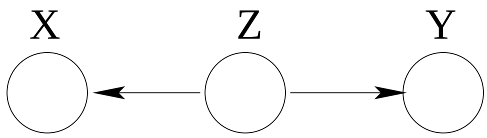

Figure 3.2. Configuration 2.

Montrons que pour toute loi se factorisant dans G, alors, $X \perp Y | Z$, en montrant que $p(x, y|z) = p(x|z)p(y|z)$. On a :

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(z)p(x|z)p(y|z)}{p(z)}$$

$$p(x, y|z) = p(x|z)p(y|z)$$

– **"Explaining away"**

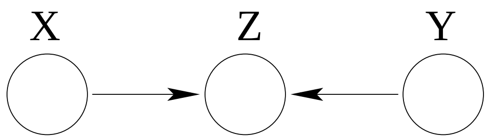

Figure 3.3. Configuration 3.

Montrons que pour toute loi se factorisant dans G, alors, $X \perp Y$, en montrant que $p(x, y) = p(x)p(y)$. On a :

$$p(x, y) = \sum_z p(x, y, z) = \sum_z p(x)p(y)p(z|x, y) = p(x)p(y)$$

$$p(x, y) = p(x)p(y)$$

Proposition 3.6 *$G = (V, E)$ et $G = (V, \tilde{E})$ sont deux DAG, si $E \subset \tilde{E}$ alors $L(G) \subset L(\tilde{G})$.*

3.3.4 Marginalisation dans un modèle graphique orienté (MGO)

Proposition 3.7 *Soit $p(x) \in L(G)$. Si n est une feuille alors $p(x_1, \ldots, x_n)$ se factorise dans le sous-graphe $\tilde{G} = (\tilde{V}, \tilde{E})$, $\tilde{V} = V \setminus \{n\}$ et $\tilde{E} = E \cap (\tilde{V} \times \tilde{V})$.*

En d'autres termes, en marginalisant par rapport à une feuille, on obtient un nouveau modèle graphique à partir de l'ancien modèle dont on a enlevé la feuille.
Démonstration

$$\begin{aligned}
p(x_1, ...x_n - 1) &= \sum_{x_n} p(x_1, \ldots, x_n) \\
&= \sum_{x_n} \left(\prod_{i=1}^{n-1} p(x_i|x_{\pi_i}) p(x_n|x_{\pi_n}) \right) \\
&= \prod_{i=1}^{n-1} p(x_i|x_{\pi_i}) \sum_{x_n} p(x_n|x_{\pi_n}) \\
&= \prod_{i=1}^{n-1} p(x_i|x_{\pi_i})
\end{aligned}$$

■

Remarque Avec un nœud quelconque, c'est plus compliqué. La marginalisation n'est pas "close", *i.e.*, si n n'est pas une feuille, on n'obtient pas forcément une distribution qui se factorise dans un modèle graphique simple. Ceci ne sera pas vrai pour les modèles non orientés.

Proposition 3.8 *Si G est un DAG muni d'un ordre* $\{1, ..., n\}$ *topologique, alors*

$$\forall i : p(x_i|x_{\pi_i}) = p(x_i|x_1, \ldots, x_{i-1})$$

Démonstration On démontre ce résultat par récurrence sur le nombre de nœuds, en considérant une feuille (G étant un DAG, il contient forcement une feuille). ■

Corollaire 3.9 *Si G est un DAG,*

$$p(x) \in L(G) \Rightarrow \forall i, X_i \perp\!\!\!\perp X_{non\text{-}descendants(i)} \mid X_{\pi_i}$$

Corollaire 3.10 *Si G est un DAG muni d'un ordre topologique,*

$$p(x) \in L(G) \Rightarrow \forall i, X_i \perp\!\!\!\perp X_{\{1,\ldots,i-1\}\setminus\pi_i} \mid X_{\pi_i}$$

Démonstration Quelque soit i, il existe un ordre topologique tel que les non-descendants de i sont inférieurs à i (ces ordres sont différents pour chaque i). ■

Corollaire 3.11 *Si G est un DAG et si une loi $p(x)$ vérifie*

$$\forall i, X_i \perp\!\!\!\perp X_{\text{non-descendants}(i)} \mid X_{\pi_i}$$

alors, $p(x) \in L(G)$.

Un modèle graphique orienté peut donc se définir comme un ensemble de lois factorisées ou comme des lois pour lesquelles certaines relations d'indépendances conditionnelles sont vérifiées.

3.3.5 Notion de d-separation *(directed separation)*

La séparation classique n'est pas suffisante dans le cadre de graphes orientés. Une séparation plus complexe a donc du être introduite pour répondre à la question suivante :

Pour $p(x) \in L(G)$, étant donné A, B et C trois nœuds, sous quelles conditions a t-on $X_A \perp\!\!\!\perp X_B \mid X_C$?

Cette séparation s'appelle la **d-separation**.

Définition 3.12 *Soit $a, b \in V$. Une **chaîne** de a vers b est une séquence de nœuds $a = v_1, \ldots, v_k = b$ telle que*

$$\forall k, (v_{k-1}, v_k) \in E \text{ ou } (v_k, v_{k-1}) \in E$$

*Une **chaîne** de a vers b est un **chemin** de a vers b dans le graphe symétrisé (c'est-à-dire en supprimant le sens des flèches).*

Définition 3.13 *Soit i un nœud, on parle de **v-structure** si i possède deux parents au moins (il y a au moins deux flèches arrivant sur i).*

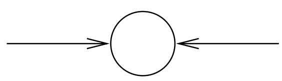

FIGURE 3.4. v-structure

Définition 3.14 *Une chaîne γ de a vers b est dites **bloquée** par $C \subset V$ si et seulement si il existe $d \in \gamma$ tel que :*

- *si $d \in C$, il n'y a pas de v-structure en d*
- *si $d \notin C$, aucun de ses descendants n'appartient à C et il y a une v-structure en d*

Définition 3.15 *Soient $A, B, C \subset V$. A et B sont **d-séparés** par C si et seulement si $\forall a \in A, \forall b \in B$, toutes les chaînes de a vers b sont bloquées. Alors, on a $X_A \perp\!\!\!\perp X_B \mid X_C$.*

Remarque Pour tester si A et B sont d-séparés par C, on peut utiliser l'algorithme des *Bayes Balls* qui consiste à lancer de balles dans le graphe à partir de A et de regarder si elles arrivent dans B.

3.3.6 Modèles classiques simples

Chaîne de Markov

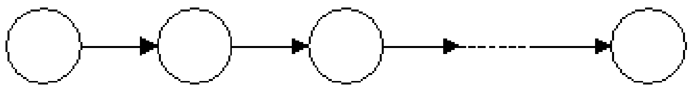

FIGURE 3.5. Modèle de chaîne de Markov

$$p(x_1, \dots, x_T) = p(x_1) \prod_{t=2}^{T} p(x_t | x_{t-1})$$

Avec la notion de d-séparation, il devient plus facile de démontrer que le futur est indépendant du passé sachant le présent.

Modèle de Markov caché ou HMM (Hidden Markov Model)

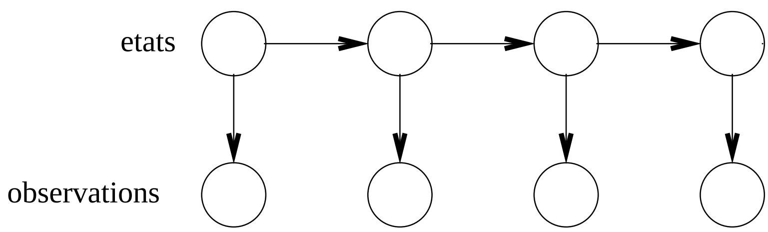

FIGURE 3.6. Modèle de Markov caché

Dans le cas d'un modèle de Markov caché, la chaîne de Markov n'est pas directement observable. On observe seulement des variables aléatoires reliées à la chaîne. Par exemple, la chaîne de Markov peut être la position d'un mobile au cours du temps et on observe sa position sur un écran radar (donc avec des erreurs et des imprécisions).