Introduction aux modèles graphiques 2007/2008

Cours 3 — 22 octobre 2008

Enseignant: Francis Bach *Scribe: Florent Couzinie-Devy, Fanny Doré*

3.1 Notations et rappels de probabilités

Dans ce cours, 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 :

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

3.2 Modèles graphiques orientés

3.2.1 Définitions et premières proprietés

Définition 3.1 *Soit un DAG $G = (V, E)$, $V = \{1, ..., n\}$ et n variables aléatoires $X_1, ..., X_n$ discrètes.*

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

avec :

1. $\forall i, f_i \geq 0$
2. $\forall i, \sum_{x_i} f_i(x_i, x_{\pi_i}) = 1$

NB : Conséquence directe : $E = \emptyset \Rightarrow L(G) = \{loi \mid X_1 \perp \ldots \perp X_n\}$

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

Proposition 3.2 $p(x) \in L(G) \Rightarrow \forall i, f_i(x_i, x_{\pi_i}) = p(x_i|x_{\pi_i})$

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 le noeud n est sans enfant (cela est possible car le graphe est acyclique). On a :

$$p(x_n|x_{\pi_n}) = \sum_{x_i, i \ni \{n\} \cup \pi_n} p(x_1, x_2, ..., x_n)$$

$$p(x_n|x_{\pi_n}) = \left(\sum_{x_i, i \notin \{n\} \cup \pi_n} \prod_{i=1}^{n-1} f_i(x_i, x_{\pi_i}) \right) f_i(x_i, x_{\pi_i})$$

$$p(x_n|x_{\pi_n}) = g(x_{\pi_n}) f_i(x_i, x_{\pi_i})$$

Or $p(x_{\pi_n}) = \sum_{x_n} p(x_n|x_{\pi_n}) = g(x_{\pi_n}) \times 1$.
Et $p(x_n|x_{\pi_n}) = \frac{p(x_n, x_{\pi n})}{p(x_{\pi n})}$.
Donc :

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

■

La proposition précédente permet la nouvelle définition suivante :

Définition 3.3

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

définit un modèle graphique orienté.

Graphe complet Dans le cas d'un graphe orienté, on dit qu'il est complet si $\forall i, \pi_i = \{1, \ldots, i-1\}$. On a alors d'après la définition précédente :

$$L(G) = \{toutes\ les\ lois\}$$

3.2.2 Exemples canoniques de graphes à trois noeuds

– **"chaîne de Markov"**
  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 :

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

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

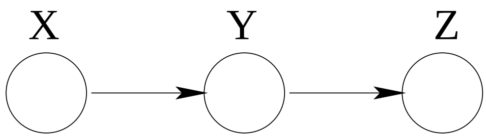

Fig. 3.1. Configuration 1.

– **"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.

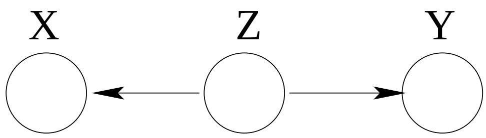

Fig. 3.2. Configuration 2.

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"**

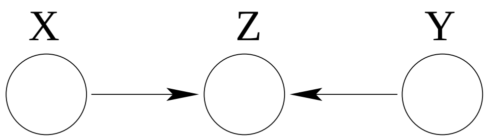

Fig. 3.3. Configuration 3.

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

3.2.3 Marginalisation dans un modèle graphique orienté (MGO)

Proposition 3.5 *Soit $p(x) \in L(G)$,*

$$i \text{ est une feuille } \Rightarrow \sum_{x_i} p\left(x_i, x_{\{i\}^c}\right) \in L(\tilde{G})$$

avec $\tilde{V} = V \setminus \{i\}$ 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 dont on a enlevé la feuille.

Démonstration

$$p(x_{V \setminus feuille}) = \sum_{x_{feuille}} \left(\prod_{i \neq feuille} p(x_i|x_{\pi_i}) \right) p(x_{feuille}|x_{\pi_{feuille}})$$

$$p(x_{V \setminus feuille}) = \prod_{i \neq feuille} p(x_i|x_{\pi_i})$$

■

Remarque Avec un noeud quelconque, c'est plus compliqué. La marginalisation n'est pas "close", *i.e.*, si i 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 3.6 *Si G est un DAG muni d'un ordre topologique (c'est-à -dire $j \in \pi_i \Rightarrow j \leq i$), alors*

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

Démonstration On démontre ce résultat par récurrence sur le nombre de nœuds, en considérant une feuille (ceci est possible ca on a un DAG). ■

Corollaire 3.7 *Si G est un DAG muni d'un ordre topologique,*

$$X_i \perp\!\!\!\perp X_{\{1,\ldots,i-1\} \setminus \pi_i} \mid X_{\pi_i}$$

Corollaire 3.8 *Si G est un DAG, alors*

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

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). ■

La réciproque sera montrée en exercice :

Corollaire 3.9 *Si G est un DAG et une loi $p(x)$ vérifie pour tout i*

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

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.

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

Cette notion est introduite pour répondre à la question suivante :

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

La notion classique de séparation n'étant pas suffisante dans le cadre de graphes orientés, une séparation plus complexe est introduite, la d-séparation :

Définition 3.10 *Soit $a, b \in V$. Une **chaîne** de a vers b est une séquence de noeuds $a = v_1, \ldots, v_k = b$ telle que*

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

*C'est un **chemin** de a vers b en ignorant la direction des flèches.*

Définition 3.11 *Si i est un nœud, il y a une **v-structure** en i s'il y a deux flèches arrivant sur i.*

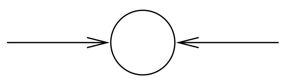

FIG. 3.4. v-structure

Définition 3.12 *Une chaîne γ de a vers b est **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 3.13 *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 $X_A \perp\!\!\!\perp X_B \mid X_C$.*

Proposition 3.14 *Soit $p(x) \in L(G)$, si C d-sépare A et B, alors $X_A \perp\!\!\!\perp X_B \mid X_C$*

Si A et B ne sont pas d-séparés par C, $X_A \perp\!\!\!\perp X_B \mid X_C$ est faux en général mais peut être vrai, pour certains cas particuliers.

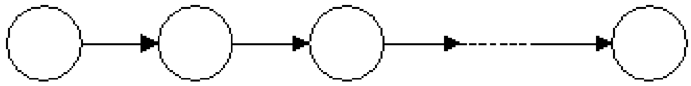

FIG. 3.5. Modèle de chaîne de Markov

3.2.5 Modèles classiques simples

Chaîne de Markov

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

On dit souvent que le futur est indépendant du passé sachant le présent, résultat non trivial à montrer sans la notion de d-séparation.

Modèle de Markov caché ou HMM (Hidden Markov Model)

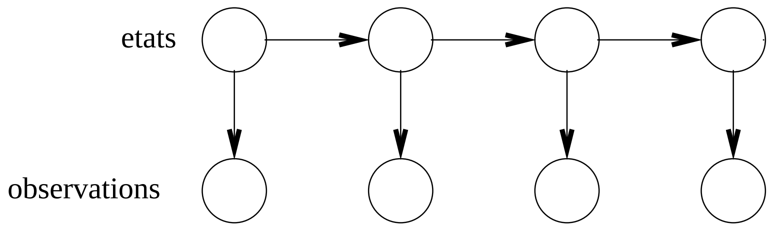

FIG. 3.6. Modèle de Markov caché

Dans ce cas, la chaîne de Markov n'est pas directement observable. On observe 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 imprécisions).

Encore un peu plus compliqué : le modèle de Markov caché factoriel. Ici la variable observée est dûe non pas à une chaîne de Markov mais à plusieurs (exemple de deux personnes parlant en même temps).

3.3 Modèles graphiques non-orientés

Les Graphes non orientés sont aussi appelés « Champs de Markov ».

Définition 3.15 *Soit $G = (\{1, \ldots, n\}, E)$ un graphe non orienté.*
Soit $\mathcal{C}$ l'ensemble des cliques maximales de G.

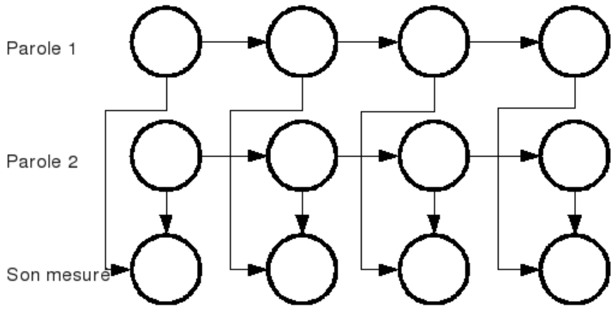

FIG. 3.7. Modèle de Markov caché factoriel

Alors la famille des lois associées à G, $\mathcal{L}(G)$ s'écrit :

$$\mathcal{L}(C) = \left\{ p \mid p(x) = \frac{1}{Z} \prod_{C \in C} \psi_C(x_C), \psi_C \geqslant 0 \right\} \tag{3.1}$$

où Z est une constante de normalisation :

$$Z = \sum_x \prod_{C \in C} \psi_C(x_C) \tag{3.2}$$

Remarque

- Dans le cas d'un DAG, nous avions $p(x) = \prod_{i=1}^{n} p(x_i | x_{\pi_i})$.
- Contrairement au cas du DAG, il n'y a pas d'interprétation directe pour $\psi_C(x_C)$.

Proposition 3.16 *Soit $G = (V, E)$ un graphe non orienté.*
Si $p(x) \in \mathcal{L}(G)$, alors pour tout $A, B, C \subset V$ tels que C sépare A et B alors $X_A \perp\!\!\!\perp X_B \mid X_C$.

La réciproque est admise :

Théorème 3.17 ***Hammersley-Cliford***
Si $\forall x, p(x) > 0$ alors :

$$\left[\forall A, B, C \subset V \mid C \text{ sépare } A \text{ et } B \Rightarrow X_A \perp\!\!\!\perp_p X_B \mid X_C \right] \Longrightarrow p(x) \in \mathcal{L}(G)$$