

Cours 2 — 10 octobre 2007

Enseignant: Francis Bach

Scribe: Anne-Laure Fouque, Fabrice Michel

Pour information

- Page web du cours
[http ://www.di.ens.fr/~fbach/courses/fall2007/](http://www.di.ens.fr/~fbach/courses/fall2007/)
- Les preuves non développées ici sont présentes dans le livre *Graphical Models* de Steffen L. Lauritzen (Oxford University Press, 1996).

2.1 Modèles graphiques orientés (suite)



Les modèles graphiques orientés sont souvent appelés réseaux Bayésiens. Cette appellation peut induire en erreur lorsque ces modèles sont utilisés dans un cadre statistique (voir cours suivants).

2.1.1 Rappel du cadre

$G = (\{1, \dots, n\}, E)$ est un graphe orienté acyclique (DAG).

$X_1, \dots, X_n$ sont des variables aléatoires.

$L(G) = \{loip(x) \text{ t.q. } p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{\pi_i})\}$

Relations d'indépendance dans les trois graphes canoniques (voir cours précédent)

- "Chaîne de Markov" $X \perp\!\!\!\perp Y \mid Z$
- "Cause commune" $X \perp\!\!\!\perp Y \mid Z$
- "Explaining away" $X \perp\!\!\!\perp Y$

2.1.2 Marginalisation

Proposition 2.1 Soit $p(x) \in L(G)$,

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

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.

Remarque 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 2.2 Si G est un DAG muni d'un ordre topologique (c'est-à-dire $j \in \pi_i \Rightarrow j \leq i$), alors

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

Démonstration On démontre ce résultat par récurrence sur le nombre de nœuds, en considérant une feuille. ■

Corollaire 2.3 Si G est un DAG muni d'un ordre topologique,

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

Corollaire 2.4 Si G est un DAG, alors

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

Démonstration Quelque soit i , il existe un ordre topologique tel que les non-descendants de i sont inférieurs à i . ■

La réciproque sera montrée en exercice :

Corollaire 2.5 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.

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

On considère maintenant le cas général de trois sous-ensembles disjoints A, B, C pour lesquels $X_A \perp\!\!\!\perp X_B \mid X_C$ si $p(x) \in L(G)$.

Définition 2.6 Soit $a, b \in V$. Une *chaîne* de a vers b est une séquence de nœuds $a = v_1, \dots, v_k = b$ 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.

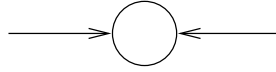


FIG. 2.1. v-structure

Définition 2.7 Si i est un nœud, il y a une **v-structure** en i s'il y a deux flèches arrivant sur i .

Définition 2.8 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 2.9 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 2.10 Soit $p(x) \in L(G)$, alors 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.

2.1.4 Modèles courants

Chaîne de Markov

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

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

Modèle de Markov caché ou HMM (Hidden Markov Model)

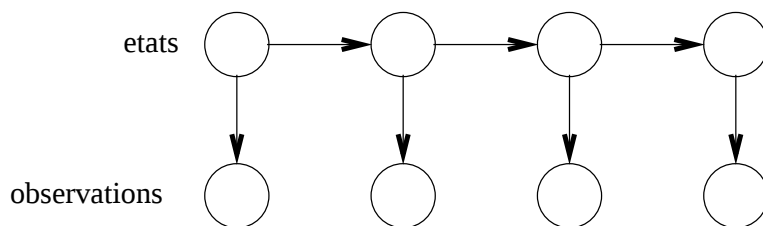


FIG. 2.2. Modèle de Markov caché

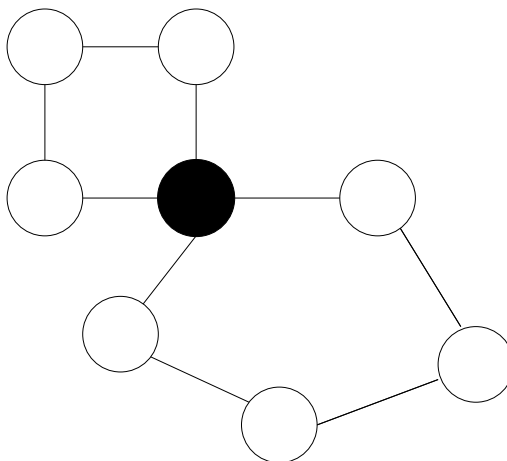


FIG. 2.3. Le nœud grisé sépare les deux ensembles de nœuds

2.2 Modèles graphiques non-orientés

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

Définition 2.11 Soit $G = (\{1, \dots, n\}, E)$ un graphe non orienté.

Soit C l'ensemble des cliques maximales de G .

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

$$\mathcal{L}(G) = \left\{ p \mid p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C), \psi_C \geq 0 \right\} \quad (2.1)$$

où Z est une constante de normalisation :

$$Z = \sum_x \prod_{C \in \mathcal{C}} \psi_C(x_C) \quad (2.2)$$

remarque :

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

Proposition 2.12 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 » (au sens de la figure 2.3) A et B alors $X_A \perp\!\!\!\perp X_B \mid X_C$.

Démonstration On supposera, sans perte de généralité que A et B n'ont qu'une seule composante connexe. On a immédiatement l'équivalence :

S sépare A et $B \iff$ tous les chemins de $a \in A$ vers $b \in B$ passent par S

Il en résulte que $V \setminus S$ a au moins deux composantes connexes. Alors il existe une composante connexe $\tilde{A}$ telle que $A \subset \tilde{A}$. Définissons maintenant $\tilde{B}$ tel que $\tilde{B} = V \setminus \{\tilde{A} \cup S\}$. On a donc $\forall C \in \mathcal{C}$, $C \subset \tilde{A} \cup S$ ou $C \subset \tilde{B} \cup S$. Car si on prend $C \subset \tilde{A} \cup S \cap \tilde{B} \cup S$, alors $\forall a \in C$, si $a \notin S$ il existe un chemin de A vers B ne passant pas par S , ce qui est impossible. On a donc :

$$C \subset \tilde{A} \cup S \cap \tilde{B} \cup S \Rightarrow C \subset S$$

donc :

$$\begin{aligned} p(x) &= \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) \\ &= \frac{1}{Z} \prod_{C \subset \tilde{A} \cup S} \psi_C(x_C) \prod_{C \subset \tilde{B} \cup S} \psi_C(x_C) \\ &= f(x_{\tilde{A} \cup S}) \times g(x_{\tilde{B} \cup S}) \end{aligned}$$

D'où $X_{\tilde{A}} \perp\!\!\!\perp X_{\tilde{B}} \mid X_S$ et finalement $X_A \perp\!\!\!\perp X_B \mid X_S$

■

La réciproque est admise :

Théorème 2.13 Hammersley-Cliford

Si $\forall x, p(x) > 0$ alors :

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

2.2.1 Marginalization

Proposition 2.14 Soit $G = (\{1, \dots, n\}, E)$ un graphe non orienté et $p(x) \in \mathcal{L}(G)$

Alors $p(x_1, \dots, x_{n-1}) \in \mathcal{L}(\tilde{G})$ avec :

$$\begin{aligned} \tilde{V} &= \{1, \dots, n-1\} \\ \tilde{E} &= [E \cap (\tilde{V} \times \tilde{V})] \cup [N_G(n) \times N_G(n)] \end{aligned}$$

Où $N_G(i)$ est le voisin du nœud i . Cela revient finalement à connecter les voisins du nœud qu'on élimine.

Démonstration

$$\begin{aligned}
\sum_{x_n} \prod_{C \in \mathcal{C}} \psi_C(x_C) &= \sum_{x_n} \prod_{C \ni \{n\}} \psi_C(x_C) \prod_{C \not\ni \{n\}} \psi_C(x_C) \\
&= \prod_{C \ni \{n\}} \psi_C(x_C) \underbrace{\sum_{x_n} \prod_{C \not\ni \{n\}} \psi_C(x_C)}_{f(x_n, x_{N_G(n)})} \\
&= \prod_{C \ni \{n\}} \psi_C(x_C) g(x_{N_G(n)})
\end{aligned}$$

■

2.2.2 Comparaison des graphes orientés et des graphes non-orientés

Listons toutes les propriétés que nous avons vues jusqu'à présent dans le tableau 2.1. On peut se poser la question de savoir si une famille de loi caractérisée par un graphe

Graphe non-orienté	Graphe orienté
$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$	$p(x) = \prod_{i=1}^n p(x_i x_{\pi_i})$
ψ_C n'est pas une probabilité	$p(x_i x_{\pi_i})$ est une loi
[A et B séparés par C] $\Rightarrow X_A \perp\!\!\!\perp X_B X_C$	[A et B d-séparés par C] $\Rightarrow X_A \perp\!\!\!\perp X_B X_C$
Marginalisation close	Marginalisation non close

TAB. 2.1. Tableau récapitulatif graphe orienté / non-orienté

orienté peut être représentée par un graphe non-orienté et vice-versa *i.e.* si on se donne un graphe G_o orienté, est-ce qu'il existe un graphe G , tel que :

$$p(x) \in \mathcal{L}(G_o) \iff p(x) \in \mathcal{L}(G)$$

En règle générale on se rend compte que les notions de graphe orienté et de graphe non-orienté sont deux notions différentes.

Cependant il existe un cas où l'on peut dire quelque chose. Définissons d'abord un graphe symétrisé :

Définition 2.15 Soit $G = (V, E)$ un graphe orienté.

Le graphe **symétrisé** de G : $\tilde{G} = (V, \tilde{E})$ est tel que :

$$\forall (u, v) \in V \times V, (u, v) \in E \Rightarrow (u, v) \in \tilde{E} \text{ et } (v, u) \in \tilde{E}$$

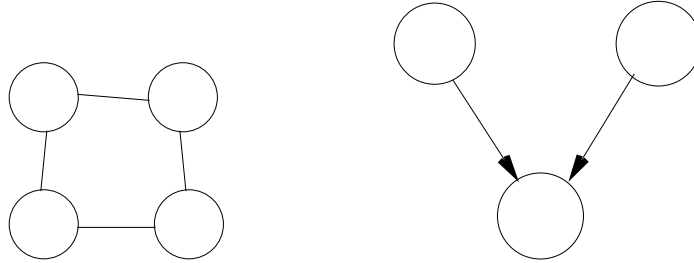


FIG. 2.4. Contre-exemples

Proposition 2.16 Soit G un DAG et $\tilde{G}$ son symétrisé.
Si G n'a pas de v -structure (i.e. $\forall i, \text{card}(\pi_i) \leq 1$) alors $\mathcal{L}(G) = \mathcal{L}(\tilde{G})$.

Démonstration Il est immédiat que sans v -structure, d -séparation $\Leftrightarrow$ séparation. ■

Exemple : La chaîne de markov n'a pas de v -structure, il en résulte que les deux graphes de la figure 2.5 sont équivalents.

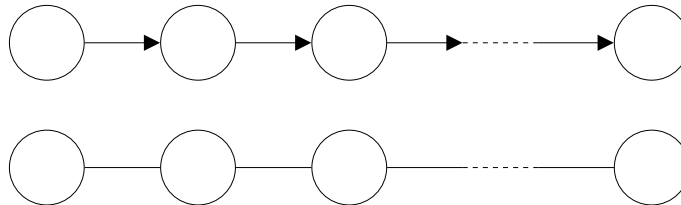


FIG. 2.5. Chaîne de markov, les deux graphes sont équivalents

Définition 2.17 Soit G un graphe orienté avec des v -structures.
Le graphe **moralisé** de G (voir figure 2.6), est le graphe symétrisé de G dont on a connecté tous les parents d'un même enfant (v -structure).

Proposition 2.18

$$p(x) \in \mathcal{L}(G) \Rightarrow p(x) \in \mathcal{L}(\text{moralisé}(G))$$

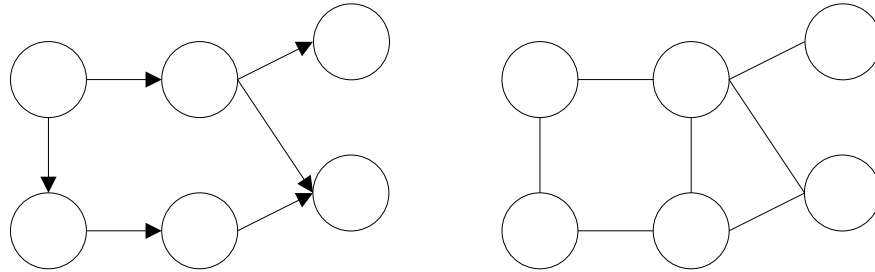


FIG. 2.6. Exemple d'un graphe moralisé

2.2.3 Calcul du coefficient de normalisation pour un graphe non-orienté

Lorsque nous avons $u(x) = \prod_{C \in \mathcal{C}} u_C(x_C)$ nous voudrions être capable de calculer $\sum_x u(x)$

La première approche, naïve, pourrait être d'évaluer u en tout x , cette méthode est cependant exponentielle en n le nombre de variables. Regardons maintenant le cas où l'ensemble des arrêtes est vide. Alors :

$$u(x) = \prod_i u_i(x_i)$$

et

$$\begin{aligned} \sum_x u(x) &= \sum_{x_1, \dots, x_n} \prod_i u_i(x_i) \\ &= \prod_i \sum_{x_i} u_i(x_i) \end{aligned}$$

Ce qui est beaucoup plus économique en calcul

Nous voudrions pouvoir factoriser ce produit dans un cas général. Par exemple si nous avons un graphe tel que :

$$u(x) = u_{12}(x_1, x_2)u_{14}(x_1, x_4)u_{23}(x_2, x_3)u_{34}(x_3, x_4)u_{45}(x_4, x_5)$$

alors

$$\begin{aligned}
 \sum_{x_1, \dots, x_5} u(x_1, \dots, x_5) &= \sum_{x_1, \dots, x_4} u_{12}(x_1, x_2) u_{14}(x_1, x_4) u_{23}(x_2, x_3) u_{34}(x_3, x_4) \underbrace{\sum_{x_5} u_{45}(x_4, x_5)}_{=m_4(x_4)} \\
 &= \sum_{x_1, x_2, x_3} u_{12}(x_1, x_2) u_{23}(x_2, x_3) \underbrace{\sum_{x_4} u_{14}(x_1, x_4) u_{34}(x_3, x_4) m_4(x_4)}_{=m_{13}(x_1, x_3)} \\
 &= \sum_{x_1, x_2} u_{12}(x_1, x_2) \underbrace{\sum_{x_3} u_{23}(x_2, x_3) m_{13}(x_1, x_3)}_{=m_{12}(x_1, x_2)} \\
 &= \sum_{x_1, x_2} u_{12}(x_1, x_2) m_{12}(x_1, x_2)
 \end{aligned}$$

où uniquement des sommes sur des nombres d'indices faibles ont été effectuées.