Introduction aux modèles graphiques **2007/2008**

Cours 1 — 3 octobre 2007

Enseignant: Francis Bach *Scribe: Edouard d'Archimbaud, Nicolas Honnorat*

Pour information

- Page web du cours
  `http://www.di.ens.fr/~fbach/courses/fall2007/`

1.1 Notations et rappels de probabilités

Dans ce cours, nous considérerons 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 $i = 1 \ldots n$. nous garderons à l'esprit que n est en pratique assez grand.
Les X_i peuvent être définies simplement par la donnée de leur loi jointe $P(X_1 = x_1, \ldots, X_n = x_n)$ (nous verrons que ce n'est pas la meilleure manière de procéder en particulier lorsque n est grand)

1.1.1 Définitions

Indépendance

Deux variables aléatoires X et Y sont indépendantes si quelles que soient les valeurs x et y prises par X et Y, on a :

$$P(X = x, Y = y) = P(X = x)\, P(Y = y)$$

On notera alors $X \perp\!\!\!\perp Y$.

indépendance conditionnelle

Soient X, Y, et Z trois variables aléatoires. On dit que X est indépendant de Y sachant Z si X, Y et Z vérifient l'une des deux définitions équivalentes suivantes :

$$\forall x\ \forall y\ \forall z,\ P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z)\, P(Y = y \mid Z = z)$$

$$\forall x\ \forall y\ \forall z,\ P(X = x \mid Y = y, Z = z) = P(X = x \mid Z = z)$$

On exprimera ceci en écrivant $X \perp\!\!\!\perp Y \mid Z$

1.1.2 Notations

Afin d'alléger les écritures, nous adopterons les notations suivantes dans la suite du cours :

$$P(X = x) = p(x)$$

De même, nous poserons pour les probabilités jointes et pour les probabilités discrètes :

$$P(X = x, Y = y) = p(x, y)$$

$$P(X = x \mid Y = y) = p(x \mid y)$$

Soit $A = \{a_1, \ldots, a_k\}$ une partie de $\{1, \ldots, n\}$. Nous utiliserons dans la suite du cours les abréviations suivantes pour la marginalisation de variables :

$$P(X_A = x_A) = P(X_{a_1} = x_{a_1}, \ldots, X_{a_k} = x_{a_k}) = p(x_A)$$

$$\sum_{x_{a_1}} \sum_{x_{a_2}} \cdots \sum_{x_{a_k}} p(x_{a_1}, x_{a_2}, \ldots, x_{a_k}) = \sum_{x_A} p(x_A)$$

1.1.3 Rappels

Formule de Bayes

Soient A et B deux événements, alors

$$P(A \mid B) = \frac{P(B \mid A) P(A)}{P(B)}$$

Marginalisation

on calcule en pratique les probabilités de la manière suivante :

$$p(x_1) = \sum_{x_2} \sum_{x_2} \cdots \sum_{x_n} p(x_1, x_2, \ldots, x_n)$$

et on a ainsi pour toute partie A de $\{1, \ldots, n\}$

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

1.2 Notations et rappels de théorie des graphes

Un graphe G est la donnée d'un ensemble V de sommets et d'un ensemble d'arêtes E inclu dans $V \times V$. Les sommets sont souvent appelés noeuds, et 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.

1.2.1 graphes non orientés

les graphes $G = (V, E)$ non orientés vérifient la propriété suivante :

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

Notions importantes

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 est incluse dans aucune clique strictement plus grande qu'elle)

composante connexe soit $\mathfrak{R}$ la relation : u $\mathfrak{R}$ v si et seulement si il existe un chemin de u vers v. On appelle composante connexe les classes d'équivalence pour $\mathfrak{R}$. Les graphes traités dans la suite du cours contiendront une seule composante connexe, car si ils en ont plusieurs, elles pourront être traités séparément.

1.2.2 graphes orientés

les notions importantes concernant les graphes orientés (appelés 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 $(u_1, \ldots, u_n)$ (pour $n \geq 3$) tel que
$\forall p \in \{1, \ldots, n-1\} \; \left(u_p, u_{p+1}\right) \in E \; et \; u_1 = u_n$

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 $I(u) \leq I(v)$. Pour tout DAG, il existe forcément un ordre topologique (le plus souvent non unique). Voir "Introduction to Algorithms" (Second Edition), de Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Cliff Stein (MIT Press,McGraw-Hill).

1.3 Modèles graphiques orientés

1.3.1 Définitions

Définition 1.1 *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'indépance jointe.

Proposition 1.2 $p(x) \in L(G) \Rightarrow \forall i, f_i(x_i, x_{\pi_i}) = p(x_i|x_{\pi_i})$

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 1.3

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

définit un modèle graphique orienté.

1.3.2 Indépendances conditionnelles

Nous allons montrer les propositions suivantes (cours suivant) :

Proposition 1.4 *Soit* $p(x) \in L(G)$. $E = \emptyset \Rightarrow \forall (i, j)\ X_i \perp X_j$

Proposition 1.5 *Soit* $p(x) \in L(G)$. $\forall i, X_i \perp X_{non-descendant(i)} \mid X_{\pi_i}$

Avant cela, considérons les trois graphes canoniques à trois sommets :
– **"chaîne de Markov"**

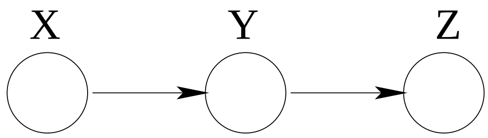

Fig. 1.1. Configuration 1.

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 :

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

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

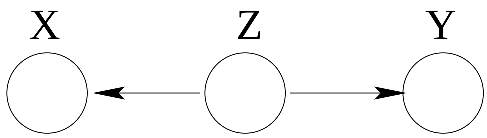

Fig. 1.2. Configuration 2.

- **"Cause commune"**
  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 :

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

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

- **"Explaining away"**

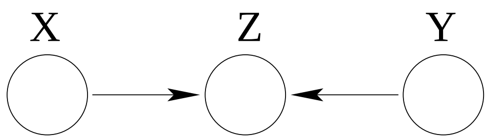

Fig. 1.3. Configuration 3.

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

Pour ces trois graphes canoniques, les relations d'indépendances démontrées sont les seules qui sont toujours vrais, ce qui n'empêche pas que d'autres soient vérifiées dans certains cas