INF8225

Leçon 2

Christopher Pal
École Polytechnique de Montréal

Au menu

- Éléments de base de la probabilité et la conception d'un modèle probabiliste
- Les réseaux de Bayes
- Raisonnement incertain avec des réseaux de Bayes
- Les graphes de facteurs et les calculs probabilistes -> l'inférence probabiliste utilisant ces graphes

Dans notre plan du cours

- Notez que : on doit remettre le TP1 en utilisant moodle avant le labo de 31 janv.

Période	Date	Activité	Salle
1	17 janvier	Énoncé du TP1	L-3714
2	31 janvier	Énoncé du TP2	L-3714
3	14 février	Énoncé du TP3	L-3714
4	28 février	Énoncé du projet	L-3714
5	21 mars	Projet volet 1	L-3714
6	4 avril	Projet volet 2	L-3714

Les axiomes des probabilités

1. Toutes les probabilités sont comprises entre 0 et 1 pour toute proposition a, $0 \leq P(a) \leq 1$
2. Propositions nécessairement vraies: $P(vrai) = 1$
   Propositions nécessairement fausses: $P(faux) = 0$
3. La probabilité d'une disjonction:

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

Notation: $P(a \wedge b) = P(a, b)$

Pourquoi probabilité?

- Bruno de Finetti 1931: Si l'agent A exprime une série de degrés de croyance qui violent les axiomes de la théorie des probabilités et si il y a une manière de parier sur les résultats, alors il y a une combinaison de paris par l'agent B, qui garantissent que l'agent A va perdre de l'argent <u>à chaque fois</u>.

Limites de la logique classique

- Fastidieux d'énumérer toutes les possibilités
- On n'a pas toute l'information
  - Sur la théorie de manière générale
  - Sur un cas particulier
- Solution: probabilités

Exemple de base de règles

$\text{DentisteCompétent} \wedge \text{Carie} \Rightarrow \text{DétecteAnomalie}$

$\text{Carie} \Rightarrow \text{Douleur}$

Que peut-on déduire à partir de cette base?

Exemple: Traduction en Prolog

```
detecteAnomalie(X,Y):-
 dentisteComp(Y),
 carie(X).


douleur(C) :-
 carie(X).


dentisteComp(jean).
douleur(moi).
```

- Quels types de requêtes sommes-nous vraiment souhaitant à poser?

Exemples:

- Si je vais chez le dentiste à Montréal, quelle est la probabilité qu'il détecte une anomalie
- Si je vais chez un dentiste sans douleur, quelle est la probabilité, qu'il va dire que j'ai une carie:

Probabilités – notions de base

- ***Variables aléatoires:*** chaque variable a un domaine de valeurs possibles (discrète, continue, binaire)
- *Événement atomique:* spécification complète d'un état du monde, qui se résume à une assignation de toutes les variables qui définissent ce monde
- Les événements atomiques sont mutuellement exclusifs
- L'ensemble de tous les événements atomiques est exhaustif

Probabilités – notions de base

- *Probabilité inconditionnelle:* degré de croyance en une proposition *a*, en l'absence de toute autre information
- *Probabilité conditionnelle:* $P(a|b)$ signifie la probabilité de *a* étant donné que tout ce que nous savons sur le monde est *b*
- $P(a|b) = \dfrac{P(a \wedge b)}{P(b)}$

Notation

- La distribution conjointe est la probabilité d'une conjonction d'affectations particulières à chaque variable comme:

$$P(V_1 = v_1, V_2 = v_2, \ldots, V_n = v_n)$$

- Souvent nous employons la notation:

a) $P(V_1, V_2, \ldots, V_n)$ et b) $P(v_1, v_2, \ldots, v_n)$

avec les variables discrètes ou binaires pour indiquer : a) les tables de probabilités (multidimensionnel) et/ou b) la probabilité unique associée à une configuration particulière

Les concepts essentiels de base

- **Règle du produit** (Règle fondamental)

$$P(A,B) = P(A \mid B)P(B)$$

- **Règle de Bayes**

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

- **Règle de la somme** (Marginalisation)

$$P(X_1) = \sum_{\{X\}\backslash X_1} P(X_1, X_2, \ldots, X_N)$$

- Notez : pour obtenir une probabilité conditionnelle on peut utiliser le concept de conditionnement « **conditioning** »

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

Probabilités – notions de base

- Théorème de Bayes avec évidence (**e**), des observations, ou certaines variables observées:

$$\mathbf{P}(X|Y) = \frac{\mathbf{P}(Y|X)\,\mathbf{P}(X)}{\mathbf{P}(Y)}$$

$$\mathbf{P}(X|Y,\mathbf{e}) = \frac{\mathbf{P}(Y|X,\mathbf{e})\,\mathbf{P}(X|\mathbf{e})}{\mathbf{P}(Y|\mathbf{e})}$$

Probabilités – notions de base

- *Indépendance:* deux variables X et Y sont indépendantes si $P(X,Y) = P(X)P(Y)$ ou encore si $P(X|Y) = P(X)$ et $P(Y|X) = P(Y)$
- *Indépendance conditionnelle:*
  - $P(X,Y|A) = P(X|A)P(Y|A)$
  - ou encore: $P(X|Y, A) = P(X|A)$ *et*
    $P(Y|X, A) = P(Y|A)$

Définition: Un réseau de Bayes

- Il représente une factorisation d'une probabilité jointe des variables V_i de la forme:

$$P(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} P(V_i \mid \text{parents}(V_i))$$

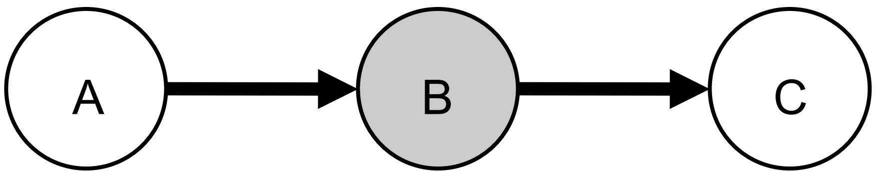
$$\text{si} \qquad \text{parents}(V_i) = \varnothing \quad \text{(il n'a pas de parents)}$$

$$\text{alors,} \quad P(V_i \mid \varnothing) = P(V_i)$$

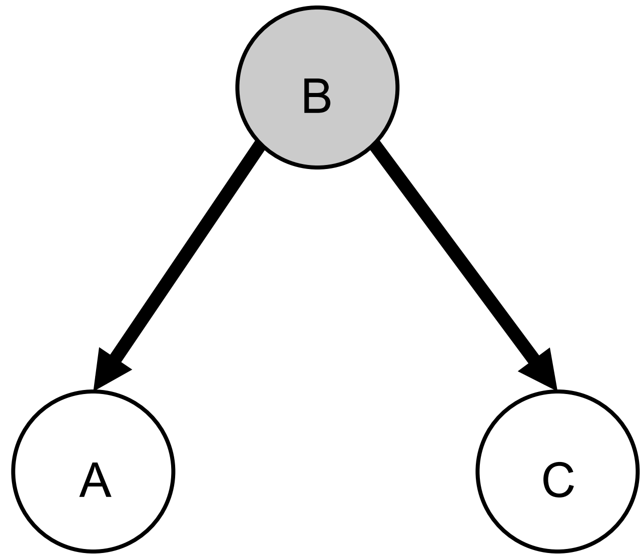
- Pour écrire le réseau: il y a un nœud (cercle) pour chaque variable et une arête orienté entre chaque variable et ses parents.

- Nœud gris implique que la variable est observée avec évidence (exact / dur, ou incertain / douce)

Exemples des modèles ou A et C sont conditionnellement indépendants sachant B

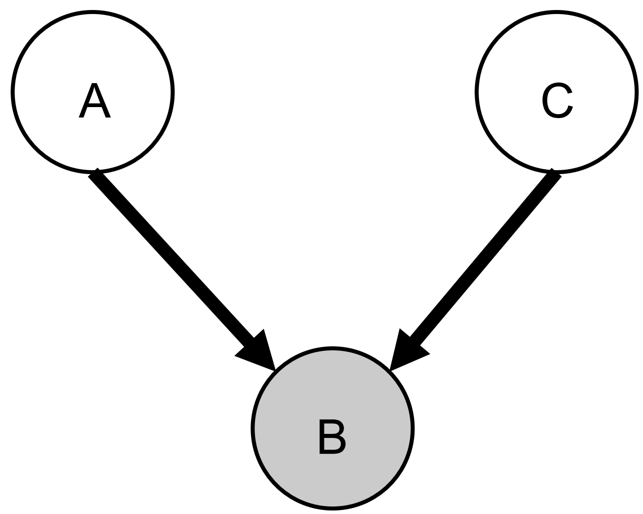


Influence pipelinée



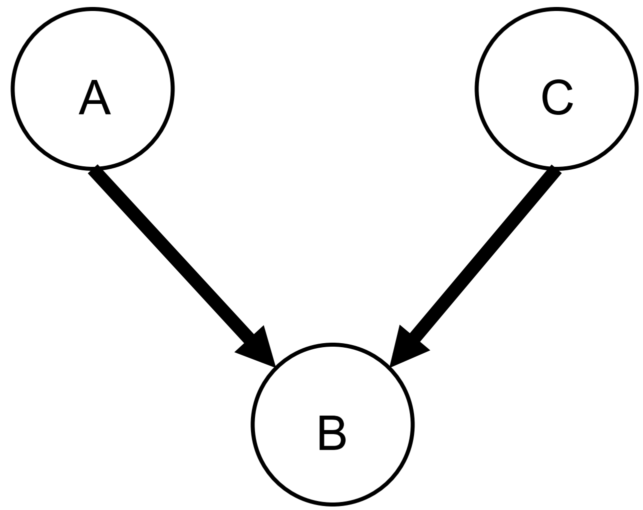
Influence divergente

Exemples d'un modèle ou A et C sont conditionnellement <u>dépendants</u> sachant B

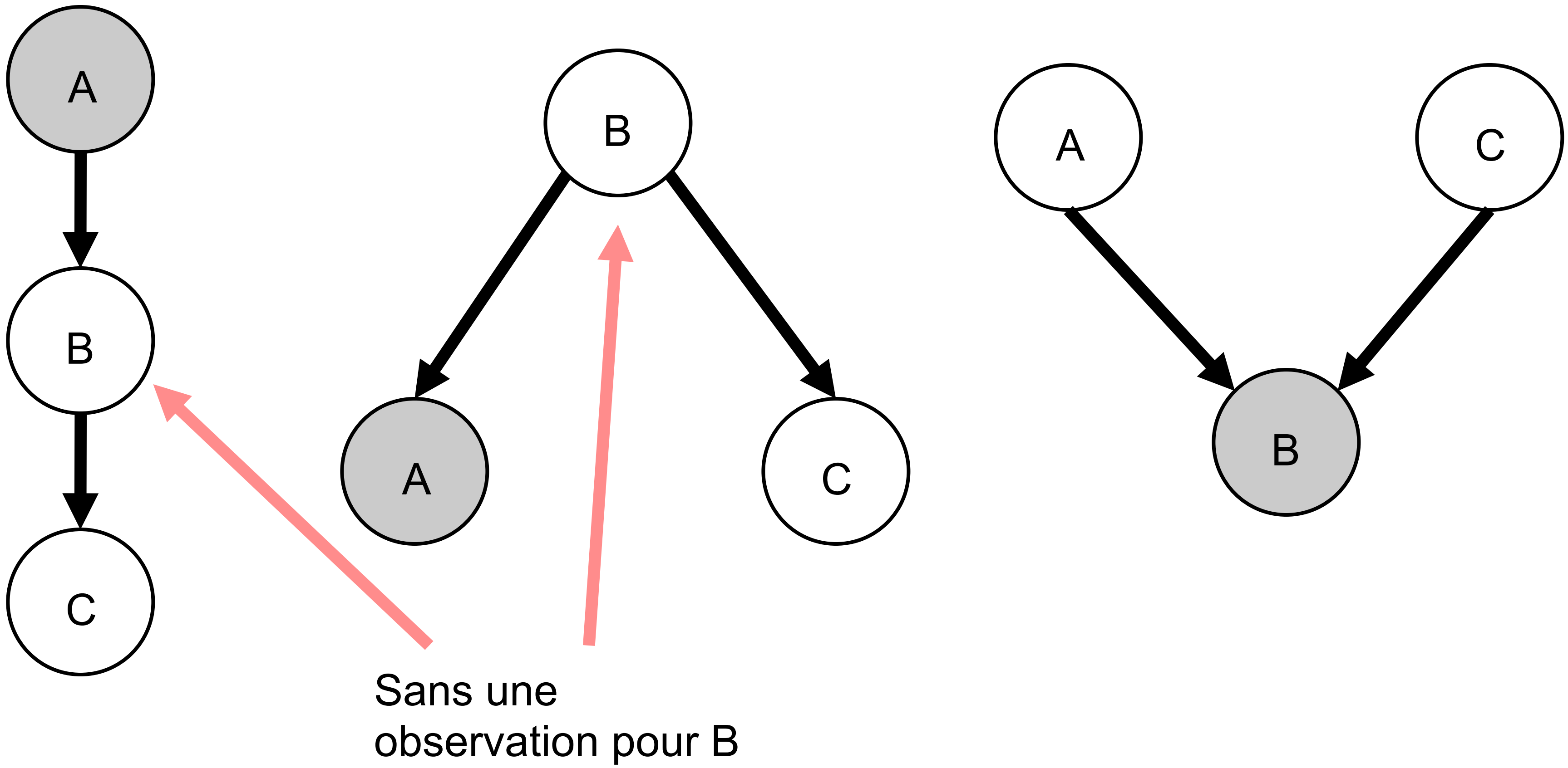


Influence convergente

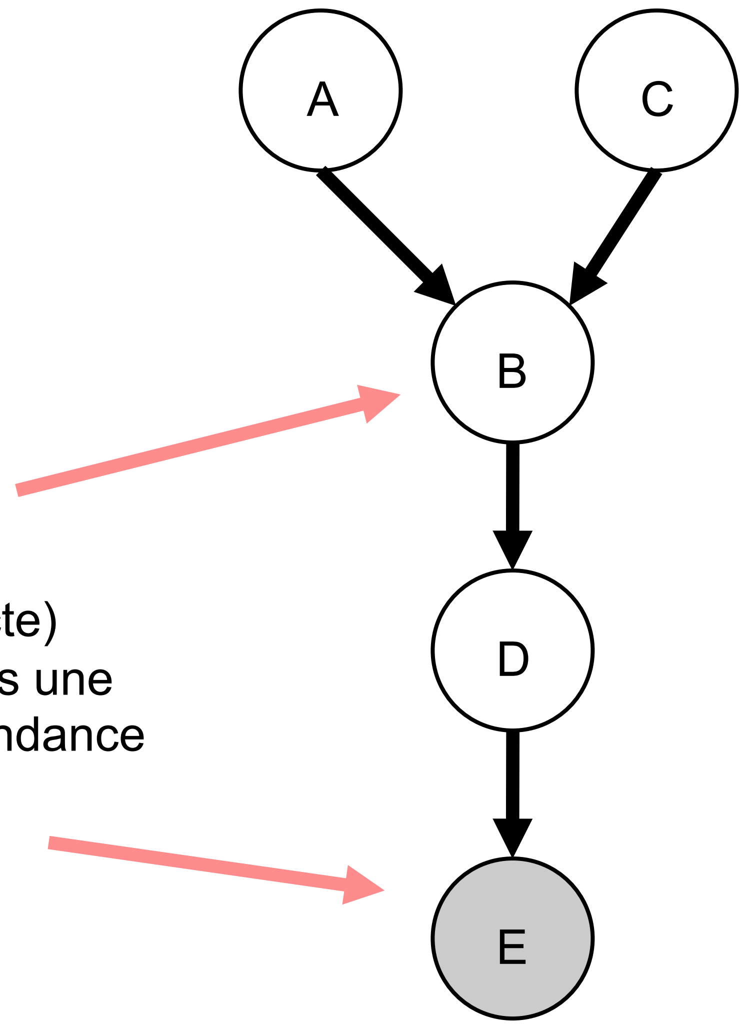
Par contre, sans évidence A et C sont indépendantes



Exemples ou A et C sont dépendants



Exemple ou A et C sont dépendants



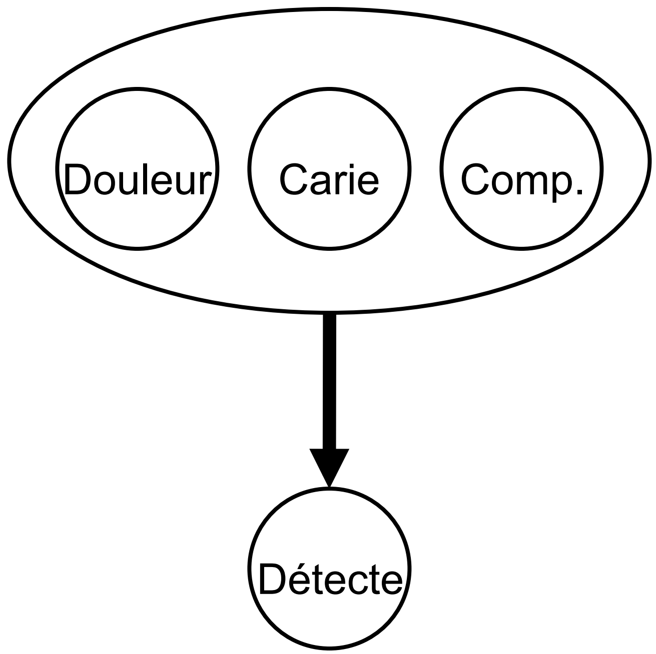
Même sans une observation (directe) pour B nous avons une exemple de dépendance conditionnelle

Exemple - Modélisation

Considérons la modélisation suivant d'un problème simple

P(Détecte,Douleur,Carie,Compétent)

= **P**(Détecte|Douleur,Carie,Compétent) × **P**(Douleur,Carie,Compétent)



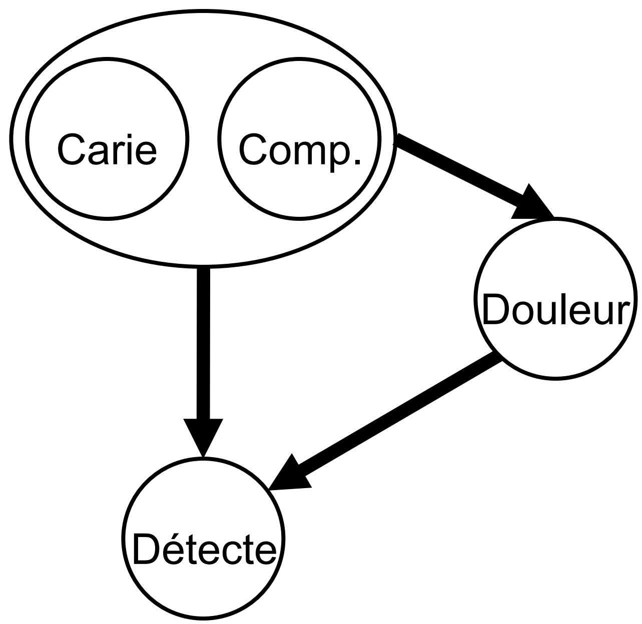
<u>Idée importante</u> : Il est possible d'appliquer nos règles simples récursivement avec les groupes de variables.

Exemple (suite) - Modélisation

P(Détecte,Douleur,Carie,Compétent)

= **P**(Détecte|Douleur,Carie,Compétent) × **P**(Douleur,Carie,Compétent)

= **P**(Détecte|Douleur,Carie,Compétent) × **P**(Douleur|Carie,Compétent) × **P**(Carie,Compétent)



Exemple (suite)

$\mathbf{P}$(Détecte,Douleur,Carie,Compétent)

= $\mathbf{P}$(Détecte|Douleur,Carie,Compétent) × $\mathbf{P}$(Douleur,Carie,Compétent)

= $\mathbf{P}$(Détecte|Douleur,Carie,Compétent) × $\mathbf{P}$(Douleur|Carie,Compétent) × $\mathbf{P}$(Carie,Compétent)

= $\mathbf{P}$(Détecte|Douleur,Carie,Compétent) × $\mathbf{P}$(Douleur|Carie,Compétent) × $\mathbf{P}$(Carie|Compétent) × $\mathbf{P}$(Compétent)

Possible, mais…

Exemple (suite)

$\mathbf{P}(\text{Détecte},\text{Douleur},\text{Carie},\text{Compétent})$

$= \mathbf{P}(\text{Détecte}|\text{Douleur},\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Douleur}|\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Carie}|\cancel{\text{Compétent}}) \times \mathbf{P}(\text{Compétent})$

$= \mathbf{P}(\text{Détecte}|\text{Douleur},\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Douleur}|\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Carie}) \times \mathbf{P}(\text{Compétent})$

La présence de carie ne dépend pas de la compétence du dentiste

Exemple (suite)

$\mathbf{P}$(Détecte,Douleur,Carie,Compétent)

= $\mathbf{P}$(Détecte|Douleur,Carie,Compétent) × $\mathbf{P}$(Douleur|Carie,Compétent) × $\mathbf{P}$(Carie|Compétent) × $\mathbf{P}$(Compétent)

= $\mathbf{P}$(Détecte|Douleur,Carie,Compétent) × $\mathbf{P}$(Douleur|Carie~~,Compétent~~) × $\mathbf{P}$(Carie) × $\mathbf{P}$(Compétent)

= $\mathbf{P}$(Détecte|Douleur,Carie,Compétent) × $\mathbf{P}$(Douleur|Carie) × $\mathbf{P}$(Carie) × $\mathbf{P}$(Compétent)

La présence de douleur de ne dépend pas de la compétence du dentiste

Exemple (suite)

$\mathbf{P}(\text{Détecte},\text{Douleur},\text{Carie},\text{Compétent})$

$= \mathbf{P}(\text{Détecte}|\text{Douleur},\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Douleur}|\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Carie}|\text{Compétent}) \times \mathbf{P}(\text{Compétent})$

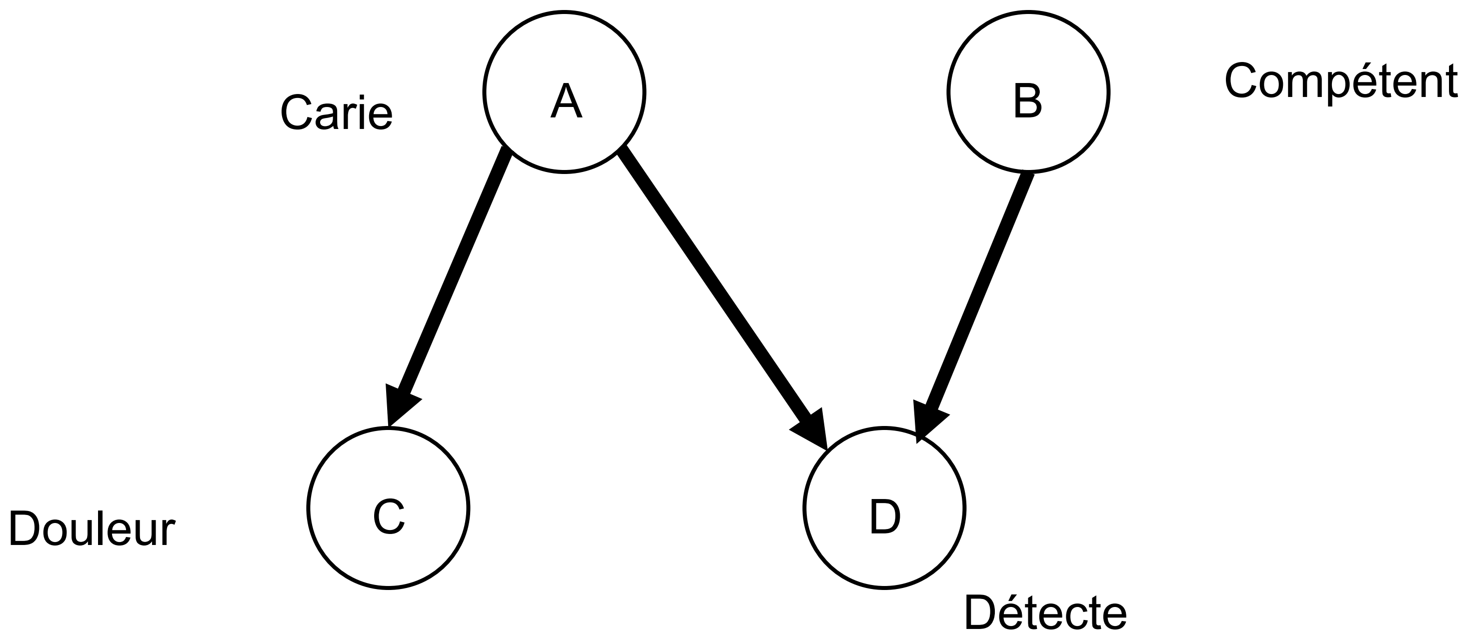
$= \mathbf{P}(\text{Détecte}|\cancel{\text{Douleur}},\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Douleur}|\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Carie}) \times \mathbf{P}(\text{Compétent})$

$= \mathbf{P}(\text{Détecte}|\text{Carie},\text{Compétent}) \times \mathbf{P}(\text{Douleur}|\text{Carie}) \times \mathbf{P}(\text{Carie}) \times \mathbf{P}(\text{Compétent})$

La détection d'une anomalie ne dépend pas de la présence de douleur

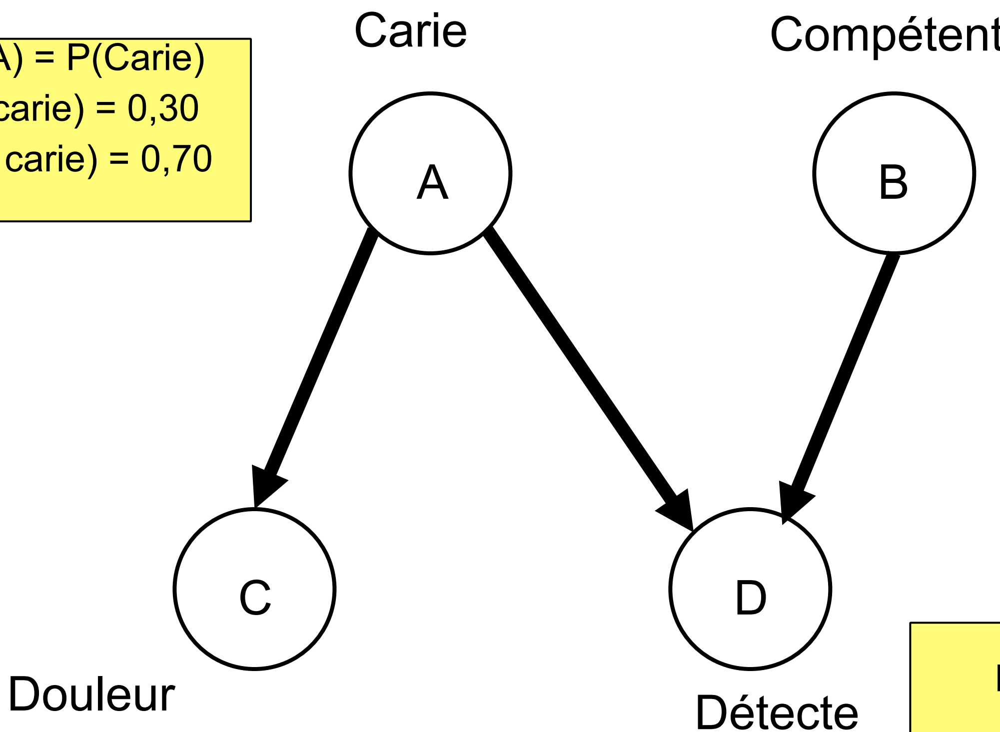
Exemple (suite)

On obtient donc un réseau de Bayes avec la structure suivante :



Exemple (suite)

P(A) = P(Carie)
P(carie) = 0,30
P(¬carie) = 0,70



P(B) = P(Compétent)
P(compétent) = 0,95
P(¬compétent) = 0,05

P(C|A) = P(Douleur | Carie)
P(douleur | carie) = 0,8
P(douleur | ¬carie) = 0,3

P(D | A, B) =
P(Détecte | Carie , Compétent)
P(détecte | compétent ∧ carie) = 0,99
P(détecte | compétent ∧ ¬carie) = 0,05
P(détecte | ¬compétent ∧ carie) = 0,50
P(détecte | ¬compétent ∧ ¬carie) = 0,15

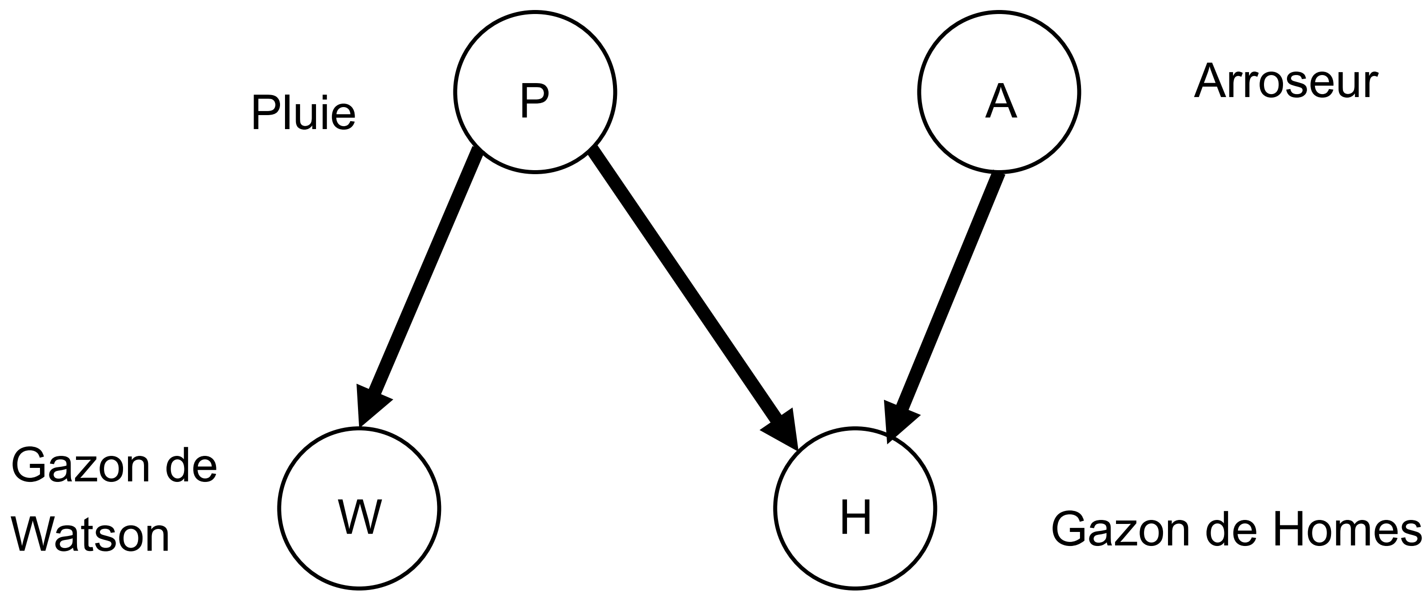
Notre exemple (suite)

- Distribution jointe complète:

[detecte===0, douleur===0, competent===0, carie===0]	0.020825
[detecte===1, douleur===0, competent===0, carie===0]	0.003675
[detecte===0, douleur===1, competent===0, carie===0]	0.008925
[detecte===1, douleur===1, competent===0, carie===0]	0.001575
[detecte===0, douleur===0, competent===1, carie===0]	0.442225
[detecte===1, douleur===0, competent===1, carie===0]	0.023275
[detecte===0, douleur===1, competent===1, carie===0]	0.189525
[detecte===1, douleur===1, competent===1, carie===0]	0.009975
[detecte===0, douleur===0, competent===0, carie===1]	0.0015
[detecte===1, douleur===0, competent===0, carie===1]	0.0015
[detecte===0, douleur===1, competent===0, carie===1]	0.006
[detecte===1, douleur===1, competent===0, carie===1]	0.006
[detecte===0, douleur===0, competent===1, carie===1]	0.00057
[detecte===1, douleur===0, competent===1, carie===1]	0.05643
[detecte===0, douleur===1, competent===1, carie===1]	0.00228
[detecte===1, douleur===1, competent===1, carie===1]	0.22572

Autre exemple avec le même structure (exemple célèbre de « wetgrass »)

On obtient donc un réseau de Bayes avec la structure suivante :



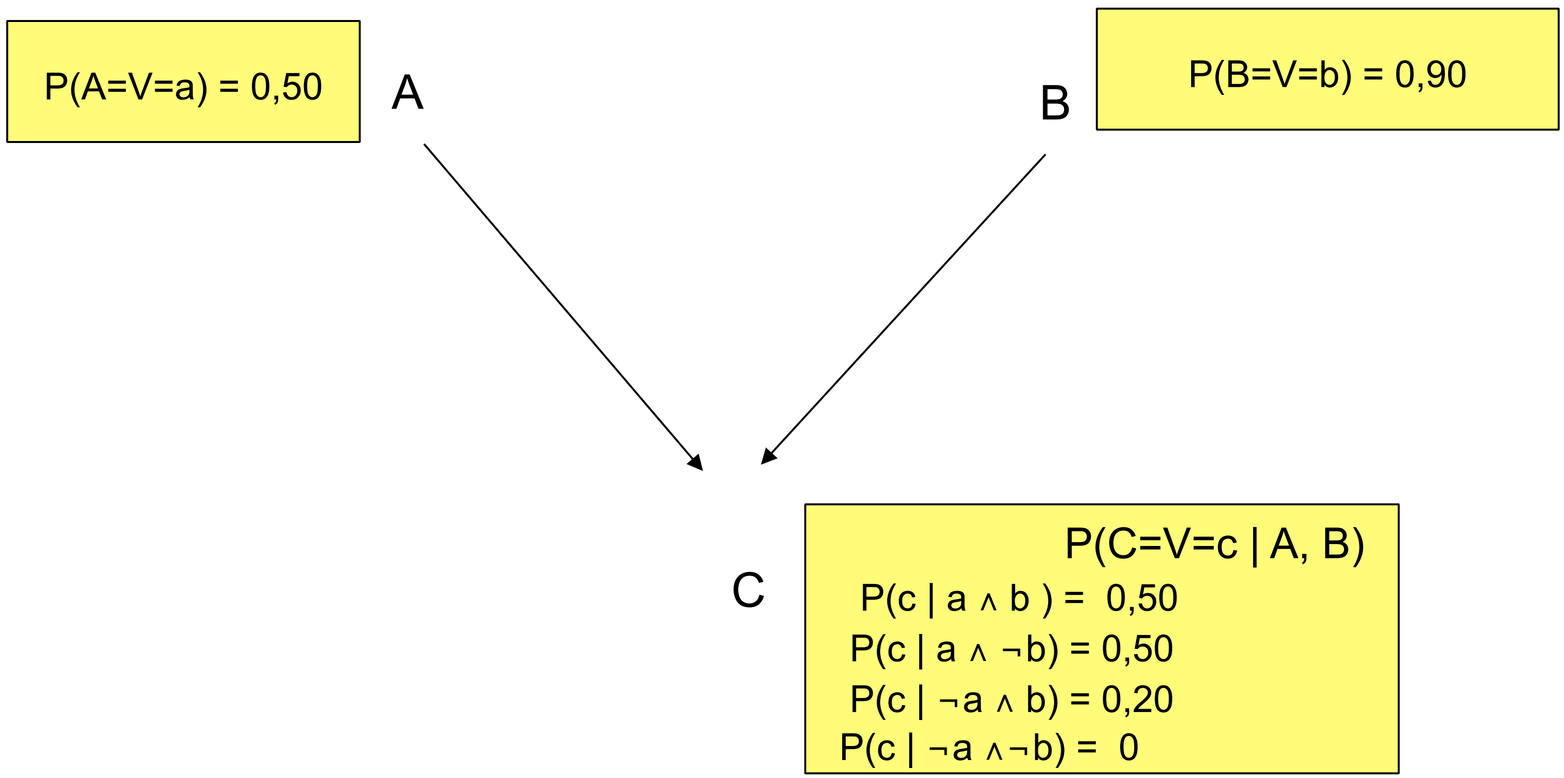
Exemple : « Explaining Away »

Pour un même jeu de données… Plusieurs réseaux possibles

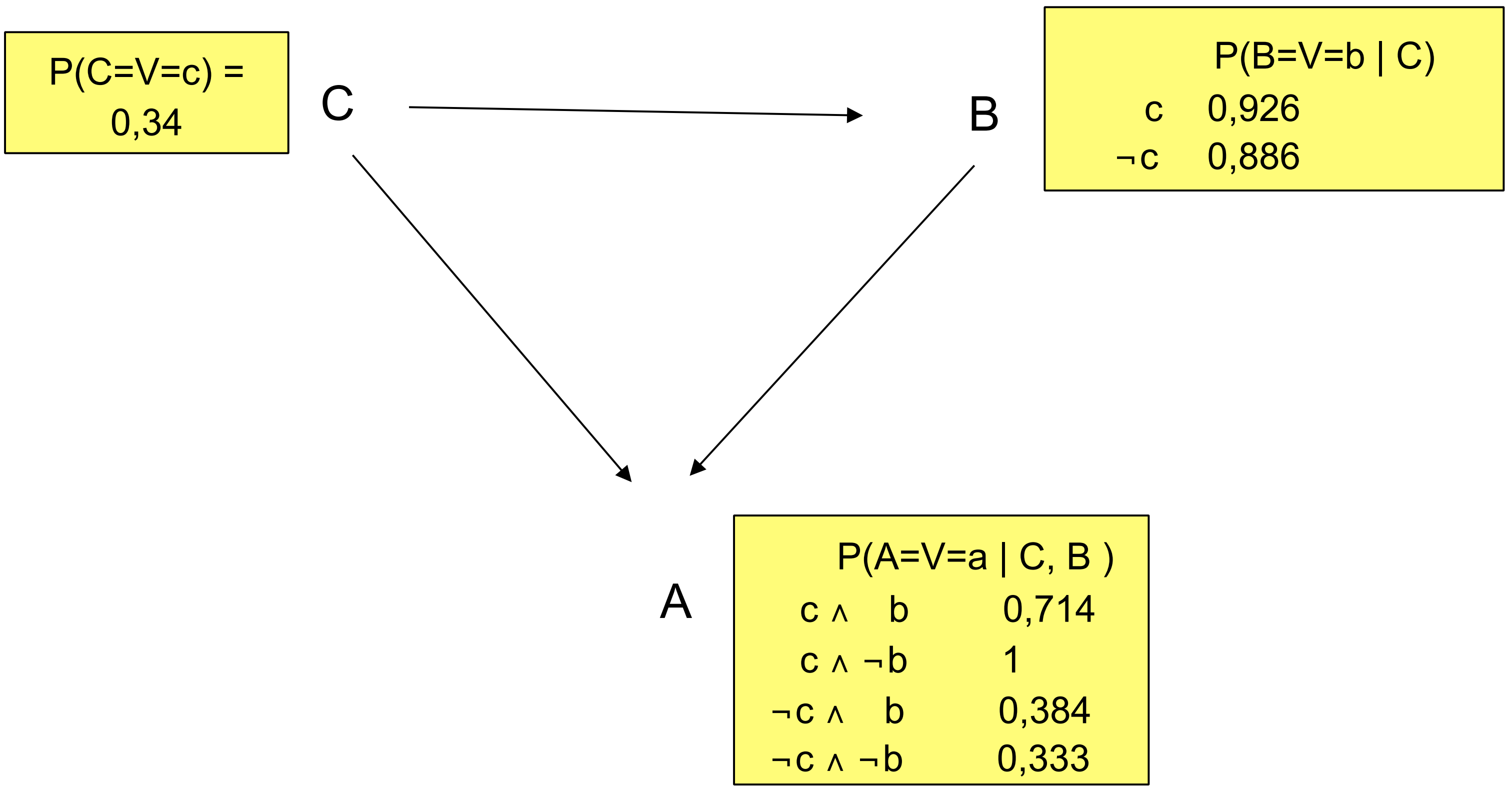
Exemple : Etant donné la distribution jointe complète pour les variables binaires A, B et C,
Rapel – notation : A = Vrai = V = a, A=Faux=F=¬a

$a \wedge b \wedge c$	0.225
$a \wedge b \wedge \neg c$	0.225
$a \wedge \neg b \wedge c$	0.025
$a \wedge \neg b \wedge \neg c$	0.025
$\neg a \wedge b \wedge c$	0.09
$\neg a \wedge b \wedge \neg c$	0.36
$\neg a \wedge \neg b \wedge c$	0
$\neg a \wedge \neg b \wedge \neg c$	0.05

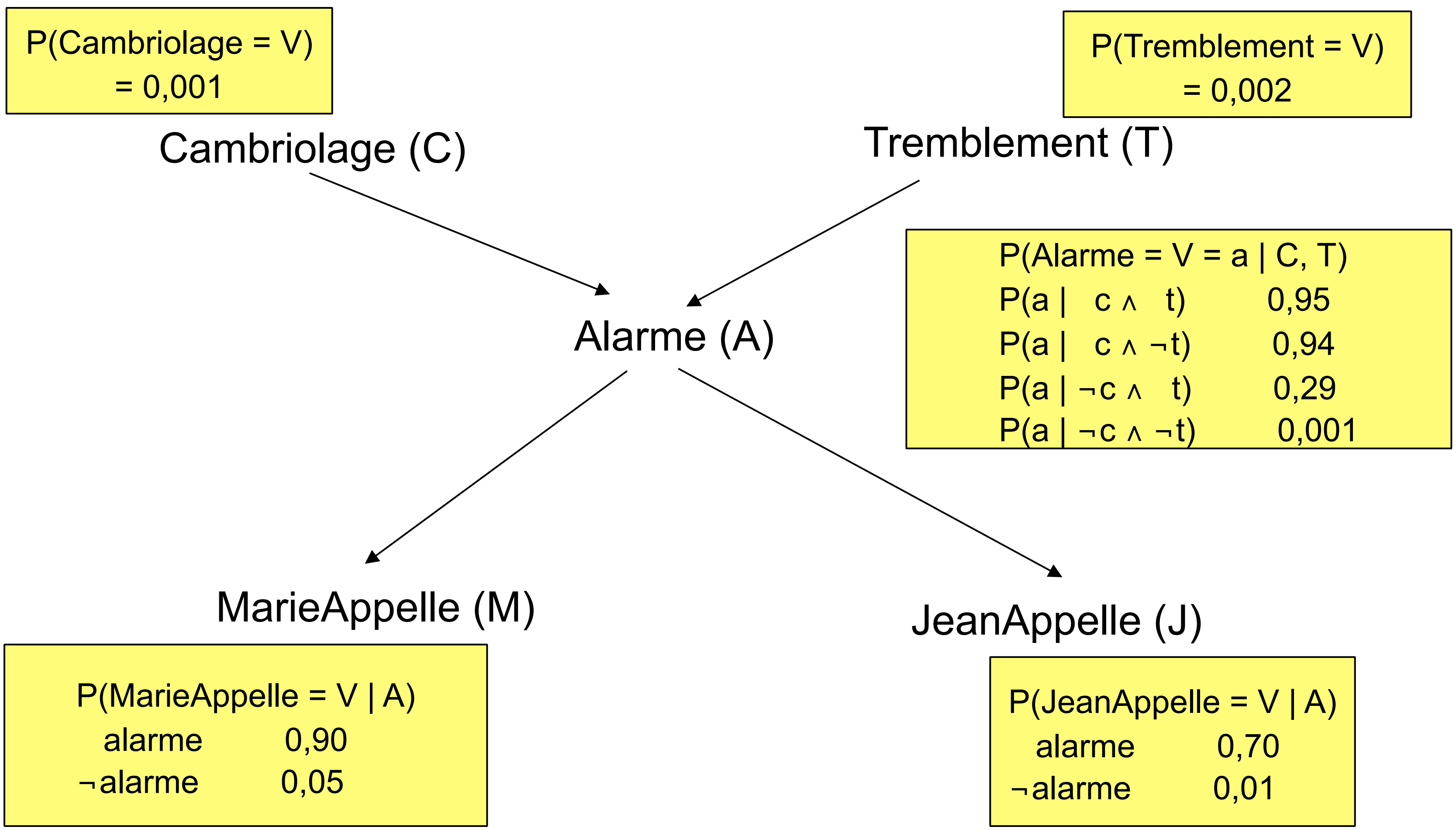
Plusieurs réseaux possibles pour un même jeu de données



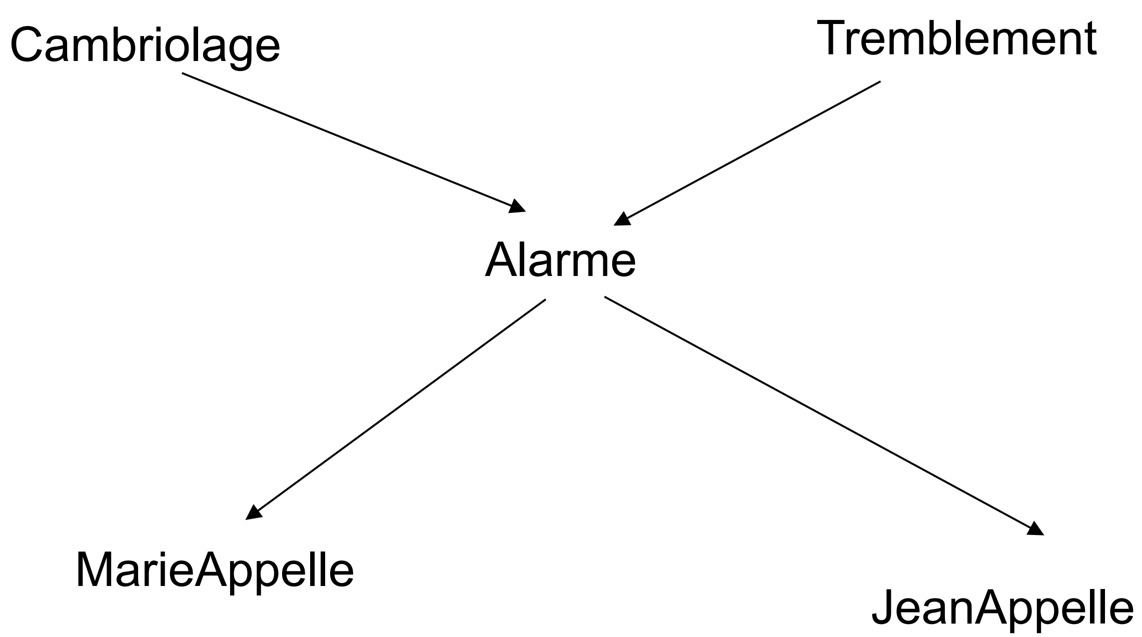
Plusieurs réseaux possibles…



Exemple (célèbre) « earthquake »

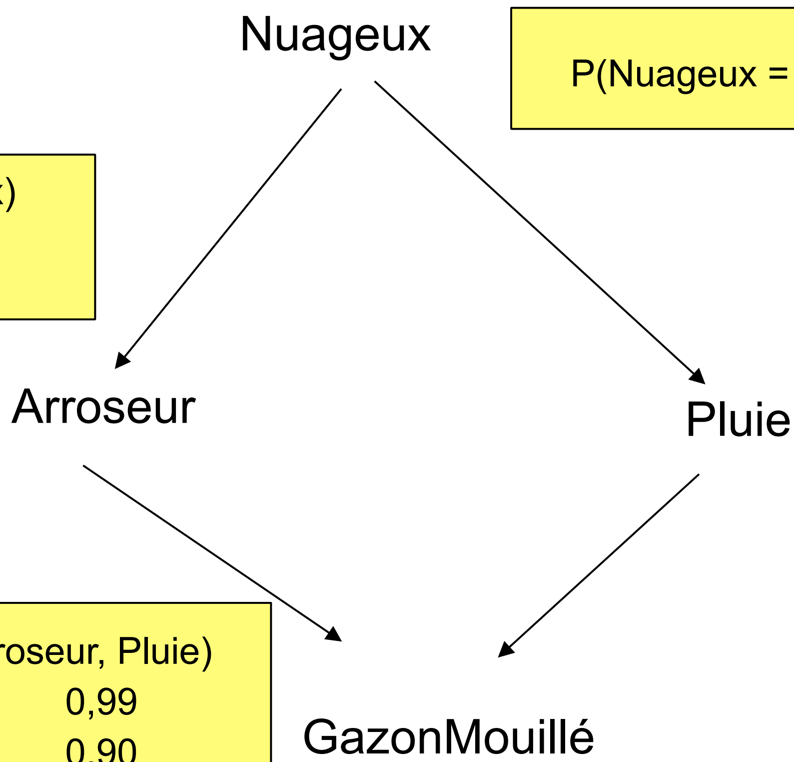


Exemple



Cambriolage
Tremblement
Alarme
MarieAppelle
JeanAppelle

Exemple



P(Nuageux = V) = 0,5

P(Arroseur = V | Nuageux)	
nuageux	0,10
¬nuageux	0,50

P(Pluie = V | Nuageux)	
nuageux	0,80
¬nuageux	0,20

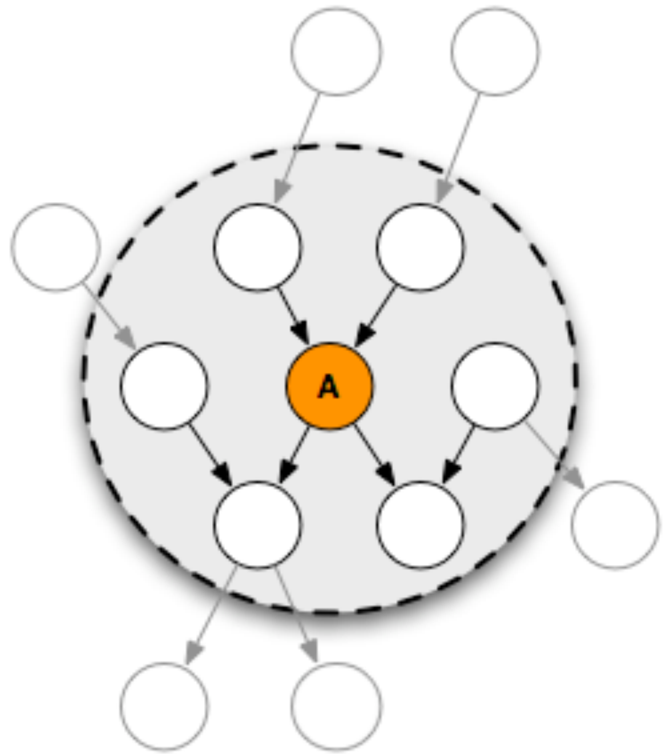
P(GazonMouillé = V | Arroseur, Pluie)	
arroseur ∧ pluie	0,99
arroseur ∧ ¬pluie	0,90
¬arroseur ∧ pluie	0,90
¬arroseur ∧ ¬pluie	0,00

Indépendance conditionnelle dans les réseaux bayésiens

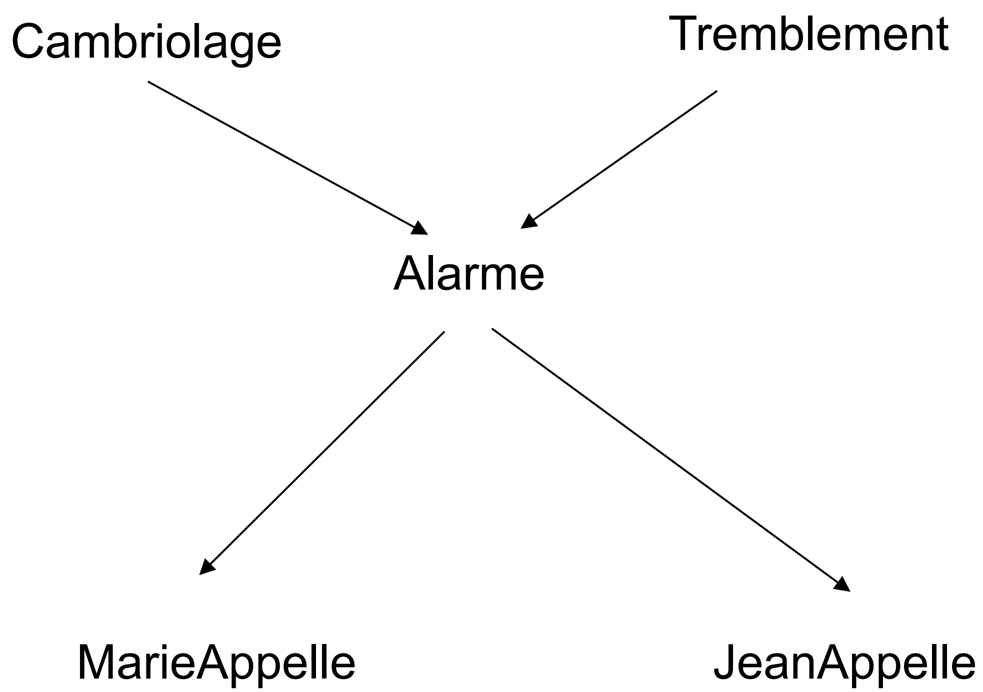
- Un noeud est conditionnellement indépendant de ses non-descendants, étant donné ses parents
- Un noeud est conditionnellement indépendant de tous les autres noeuds du réseau, étant donné sa couverture de Markov (c'est-à-dire ses parents, ses enfants et les parents de ses enfants)

Couverture de Markov

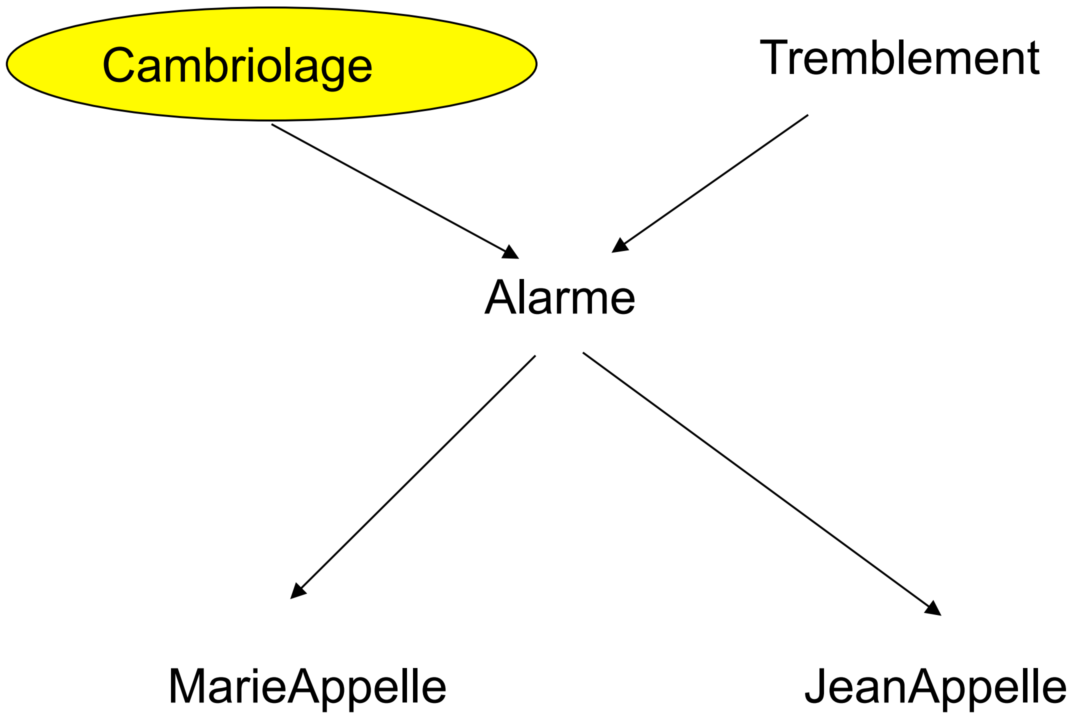
- Etant donné un nœud, la couverture de Markov consiste de ses parents, ses enfants et les autres parents de ses enfants.



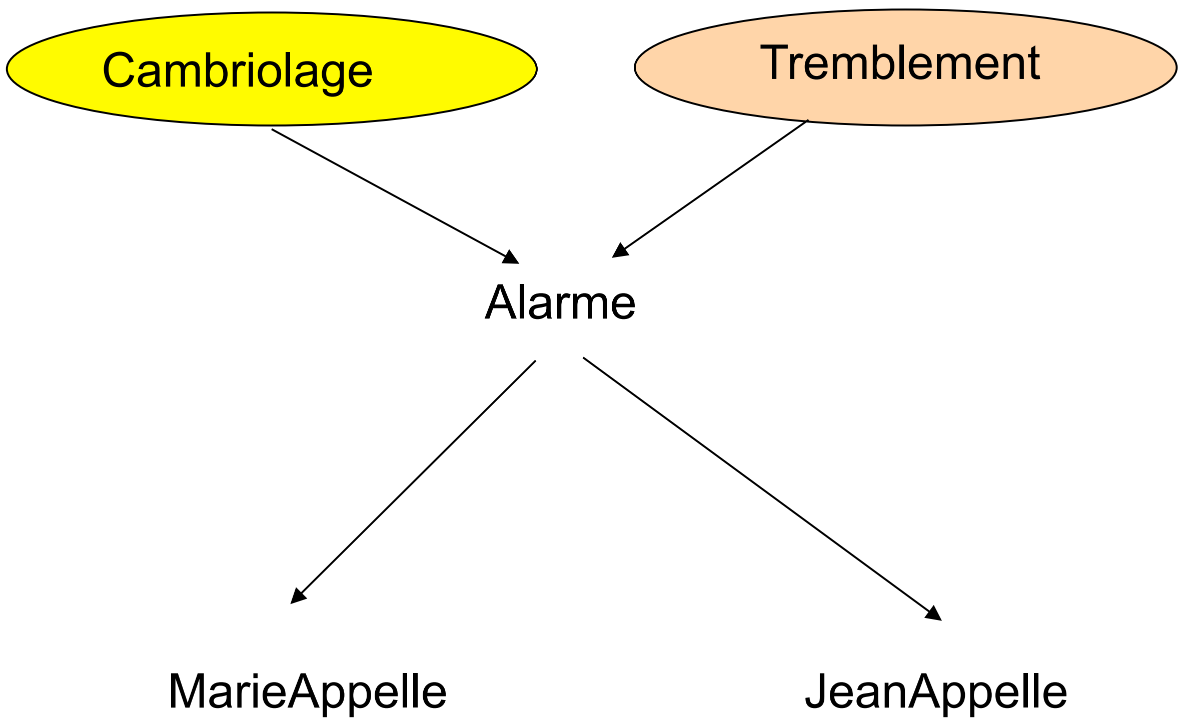
Exemple – indép. conditionnelle



Exemple – indép. conditionnelle

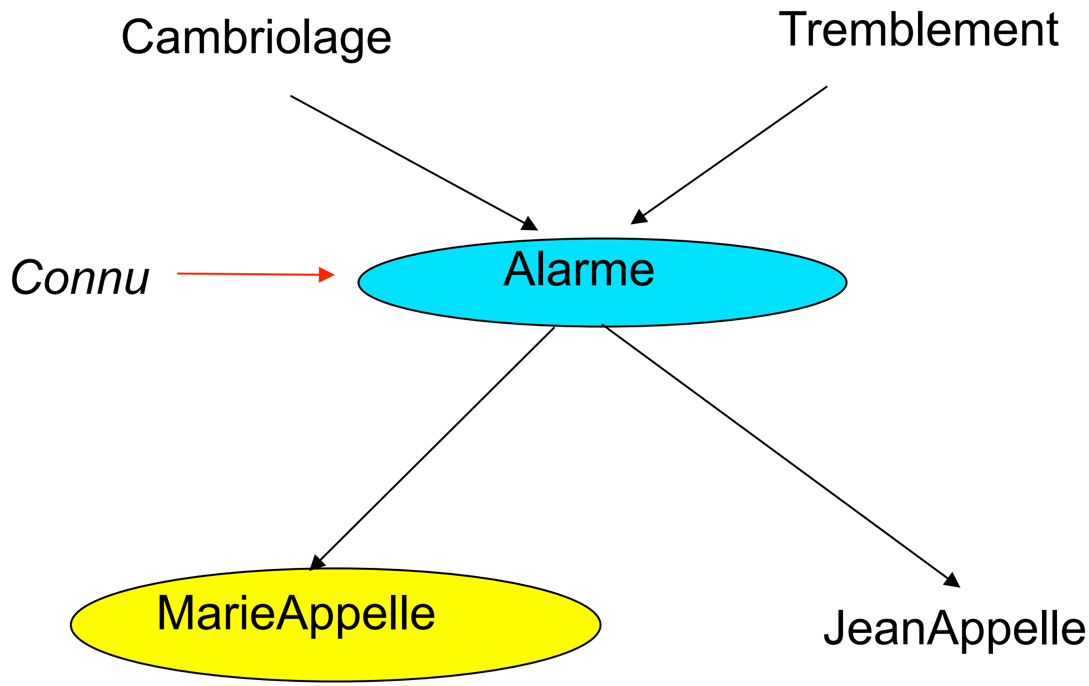


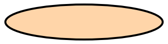
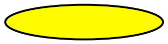
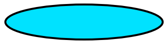
Exemple – indép. conditionnelle



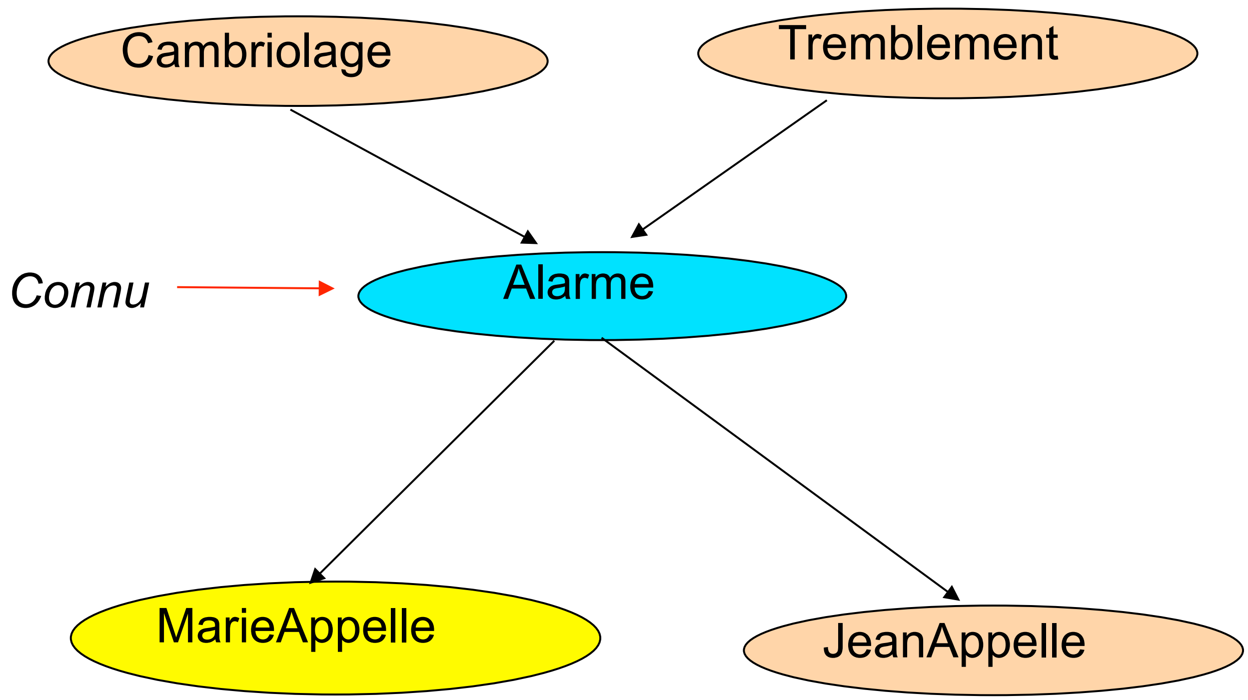
Les noeuds sont indépendants du noeud étant donné le noeud

Exemple – indép. conditionnelle



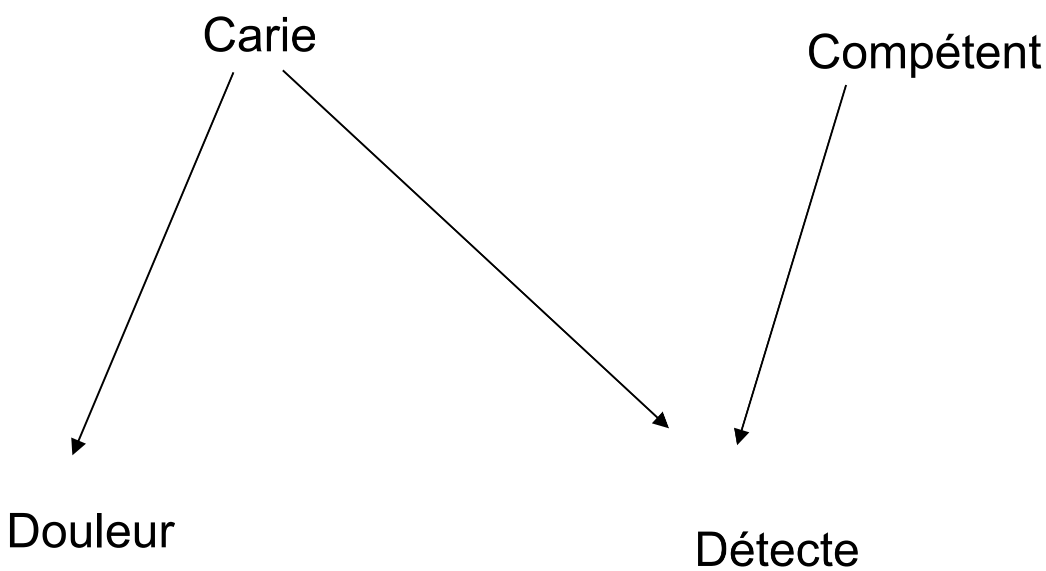
Les noeuds  *sont indépendants du noeud*  *étant donné le noeud* 

Exemple – indép. conditionnelle

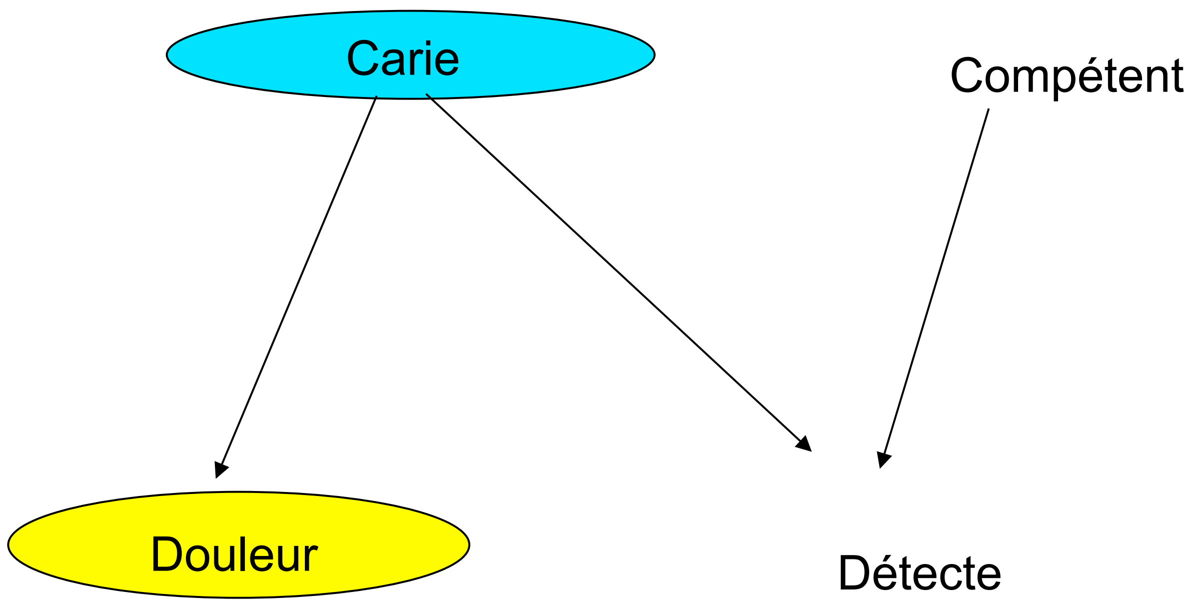


Les noeuds sont indépendants du noeud étant donné le noeud

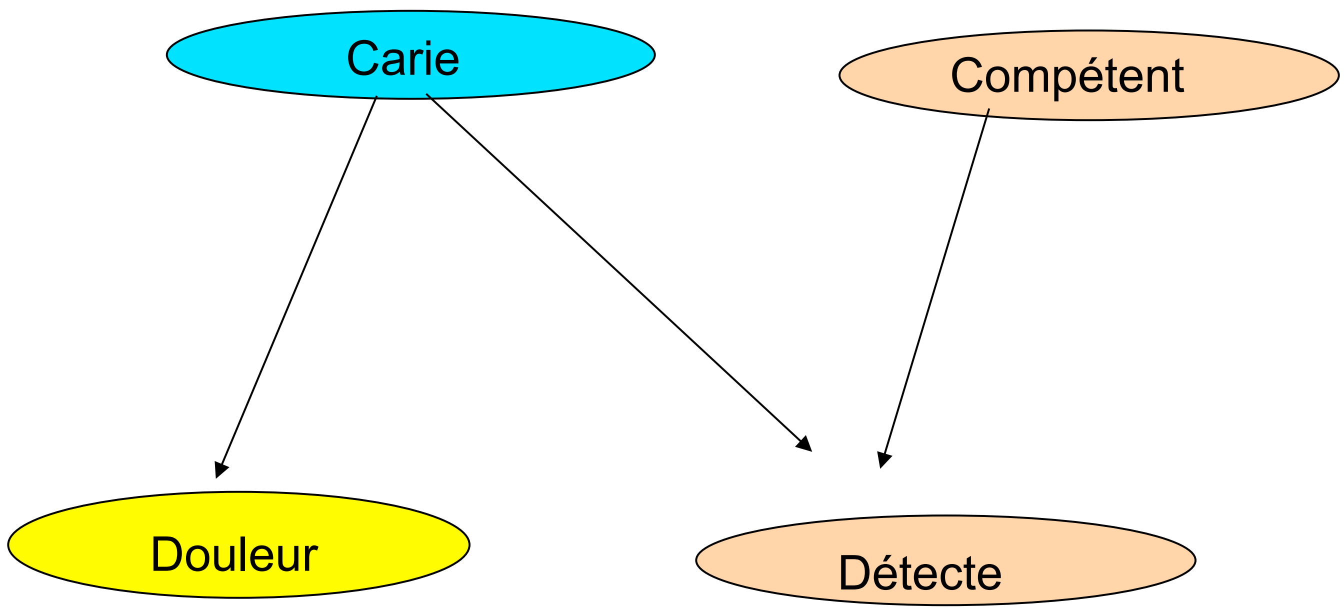
Exemple – indép. conditionnelle



Exemple – indép. conditionnelle

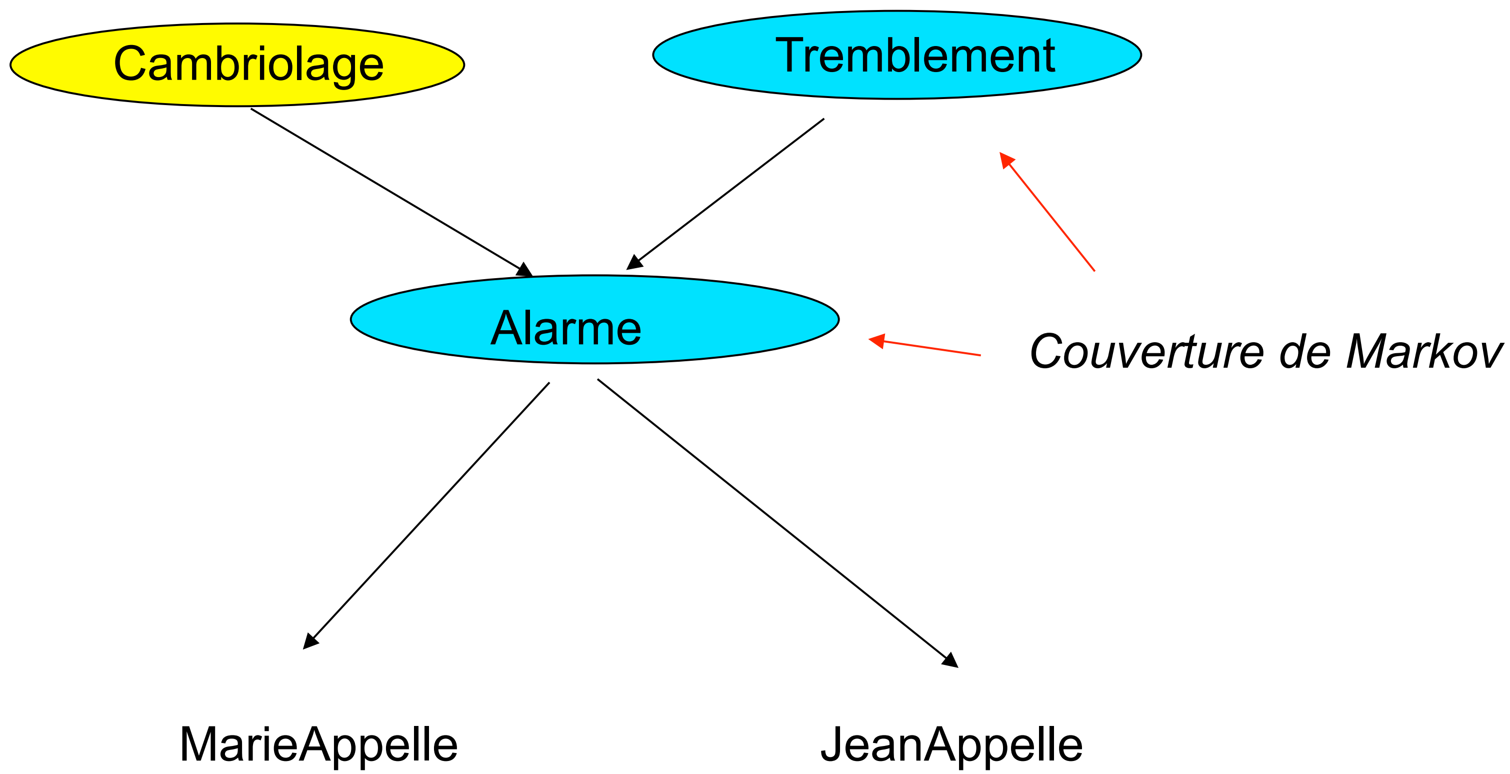


Exemple – indép. conditionnelle

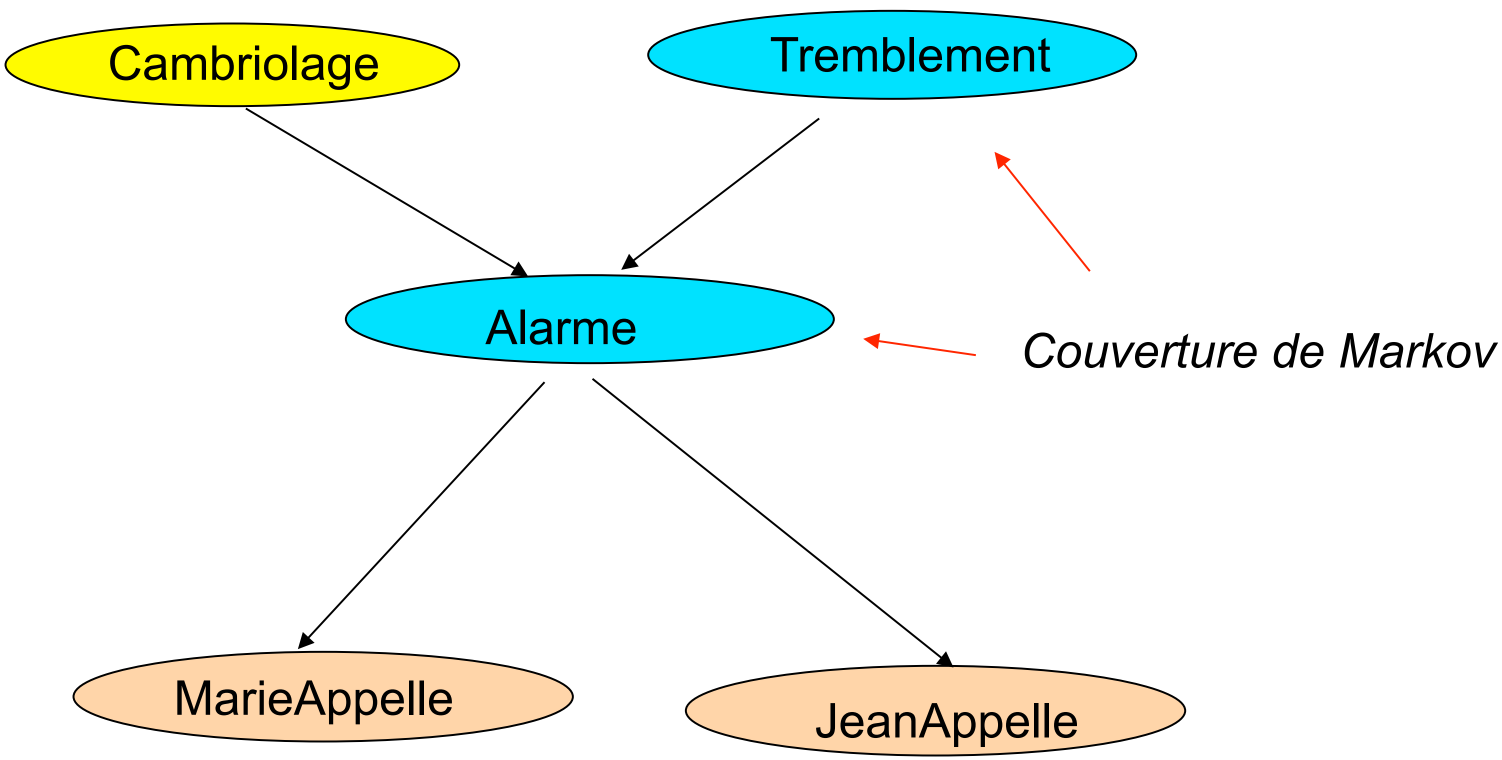


Les noeuds sont indépendants du noeud étant donné le noeud

Exemple – indép. conditionnelle



Exemple – indép. conditionnelle



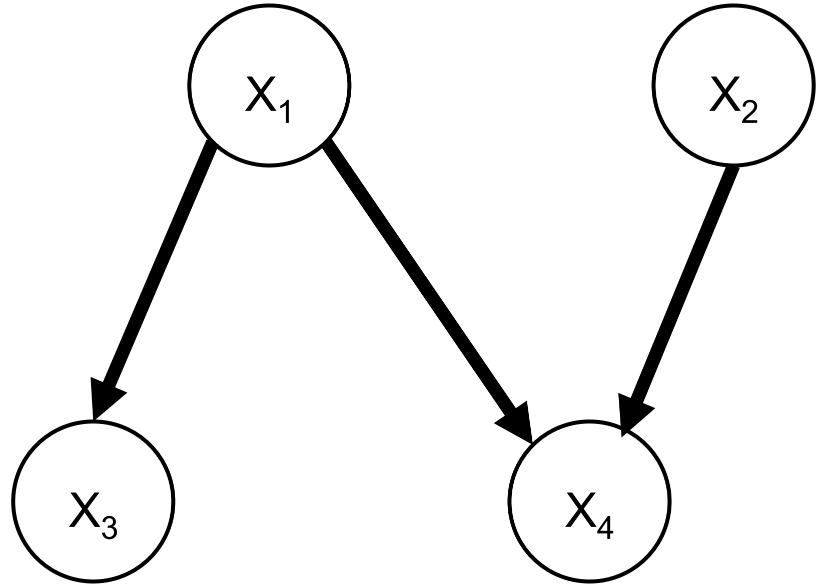
Les noeuds sont indépendants du noeud étant donné le noeud

Calcul des probabilités

- Technique du OU bruité
- On calcule individuellement la probabilitié d'une des causes
- On suppose
  - que toutes les causes sont connues
  - que l'inhibition d'une cause est indépendante de celle des autres causes

Considérons

La marginalisation nécessaire pour calculer :



$$P(X_4 = x_4)$$

$$= \sum_{x_1} \sum_{x_2} \sum_{x_3} P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)$$

$$= \sum_{x_1} \sum_{x_2} \sum_{x_3} P(X_1 = x_1)\ P(X_2 = x_2)\ P(X_3 = x_3 | X_1 = x_1)$$
$$P(X_4 = x_4 | X_1 = x_1, X_2 = x_2)$$

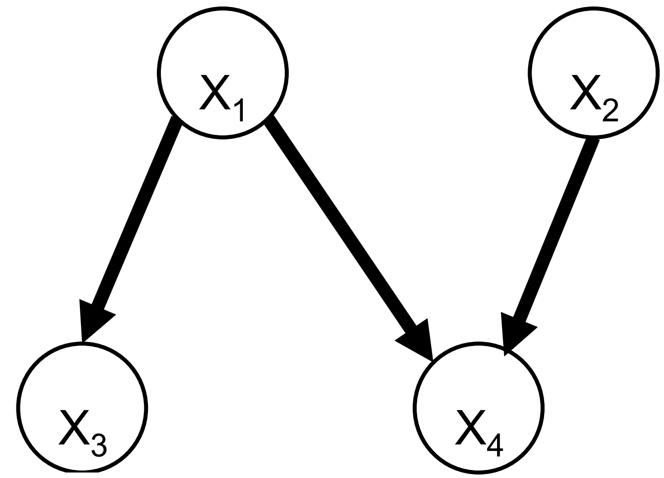
Exercice : Combien multiplications et additions avec des variables binaires?

Inférence exacte

- On calcule les probabilités en utilisant directement la formule, en sommant sur les variables cachées
- Si le réseau est un poly-arbre, le temps est proportionnel à la taille du réseau
- Puisque la logique propositionnelle est un cas particulier des réseaux bayésiens, l'inférence est donc NP-difficile de manière générale

Considérons

$$\begin{aligned}
f_1(x_1) &= P(X_1 = x_1) \\
f_2(x_2) &= P(X_2 = x_2) \\
f_3(x_1, x_3) &= P(X_3 = x_3 | X_1 = x_1) \\
f_4(x_1, x_2, x_4) &= P(X_4 = x_4 | X_1 = x_1, X_2 = x_2)
\end{aligned}$$



$P(X_4 = x_4) =$

$$= \sum_{x_1} \sum_{x_2} \sum_{x_3} f_1(x_1)\ f_2(x_2)\ f_3(x_1, x_3)\ f_4(x_1, x_2, x_4)$$

$$= \sum_{x_1} f_1(x_1) \sum_{x_3} f_3(x_1, x_3) \underbrace{\sum_{x_2} f_2(x_2)\ f_4(x_1, x_2, x_4)}_{g_2(x_1, x_4)}$$

$$= \sum_{x_1} f_1(x_1) \left(\sum_{x_2} f_2(x_2)\ f_4(x_1, x_2, x_4) \right) \sum_{x_3} f_3(x_1, x_3)$$

Exercice : Combien multiplications et additions avec des variables binaires?

Idée fondamentale : Factor Graphs

Kschischang, Frank R.; Brendan J. Frey and Hans-Andrea Loeliger (2001), "Factor Graphs and the Sum-Product Algorithm", *IEEE Transactions on Information Theory* 47 (2): pp. 498–519.

Définition: Un graphe de facteurs

- Il est possible de créer une fonction à partir des produits de *s* fonctions f_i de sous-ensembles de variables $\{X\}_i$

$$F(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{s} f_i(\{X\}_i)$$

- Exemple : notre modèle recent, où nous avons

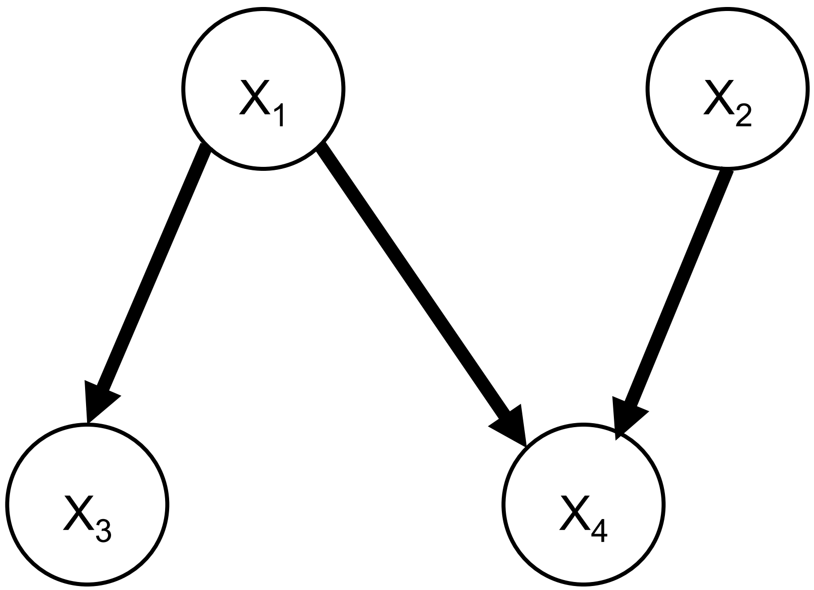
$$\begin{aligned} P(A,B,C,D) &= f_1(\{X\}_1) f_2(\{X\}_2) f_3(\{X\}_3) f_4(\{X\}_4) \\ &= P(C \mid A) P(D \mid A,B) P(A) P(B) \end{aligned}$$

Un graphe de facteurs

Il se compose de (simplement) des :

- Rectangles pour chaque fonction, et des
- Liens, sans orientation entre chaque fonction et chaque variable dans la fonction

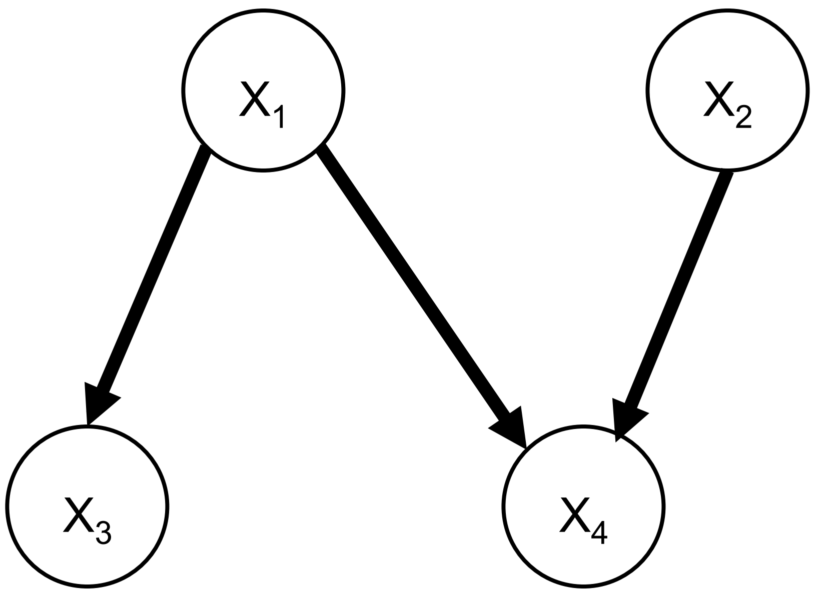
Considérons



Exercice : Ecrivez un graphe de facteurs.

Exercice :

Écrivez un graphe de facteurs.

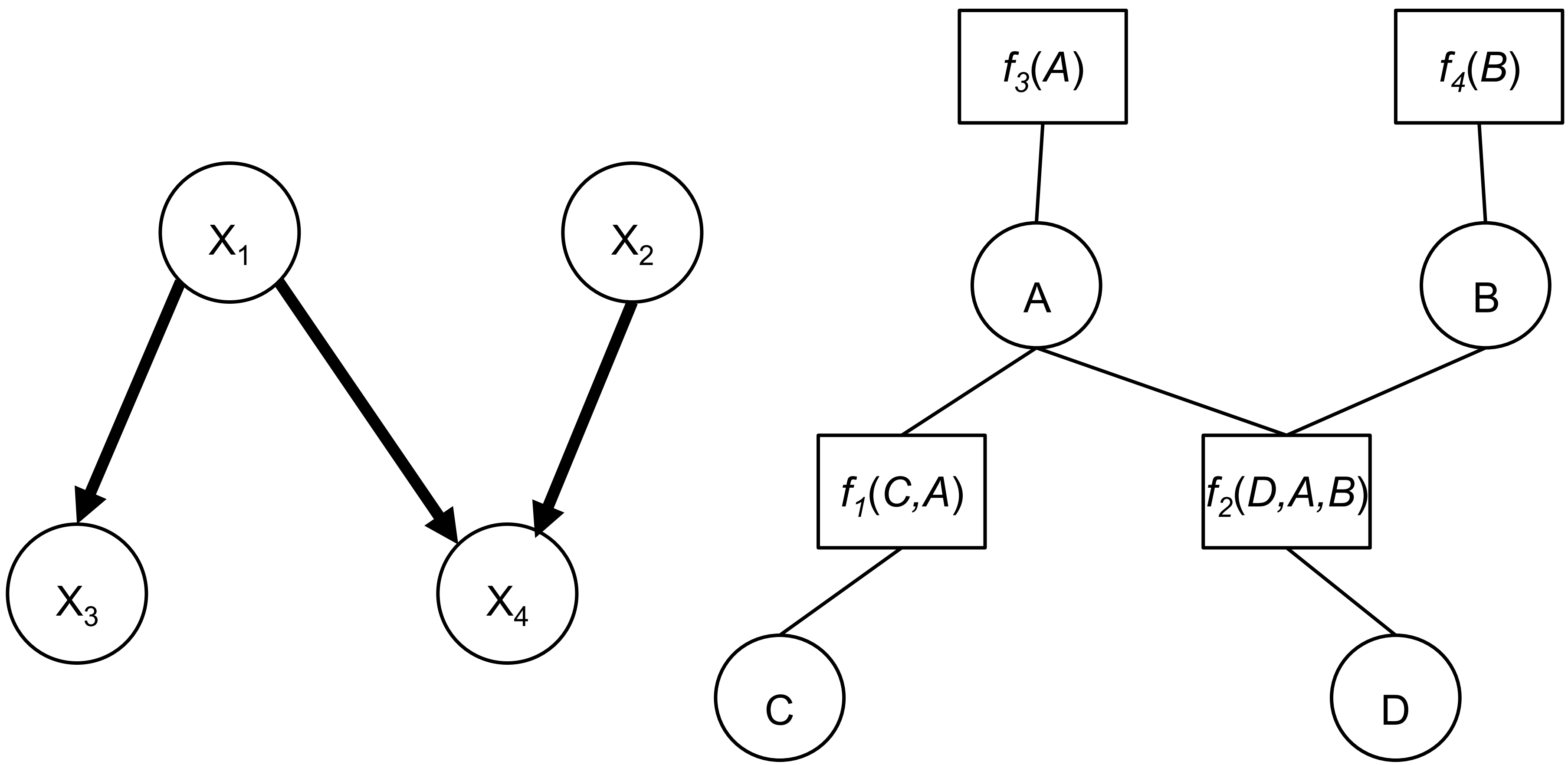


$$f_1(C,A) = P(C \mid A)$$
$$f_2(D,A,B) = P(D \mid A,B)$$
$$f_3(A) = P(A)$$
$$f_4(B) = P(B)$$

Il se compose de :

- Rectangles pour chaque fonction
- Liens, sans orientation entre la fonction et chaque variable dans la fonction

Un graphe de facteurs



Un autre exemple

▶ *Example 1 (A Simple Factor Graph):* Let $g(x_1, x_2, x_3, x_4, x_5)$ be a function of five variables, and suppose that g can be expressed as a product

$$g(x_1, x_2, x_3, x_4, x_5) = f_A(x_1) f_B(x_2) f_C(x_1, x_2, x_3) f_D(x_3, x_4) f_E(x_3, x_5) \quad (2)$$

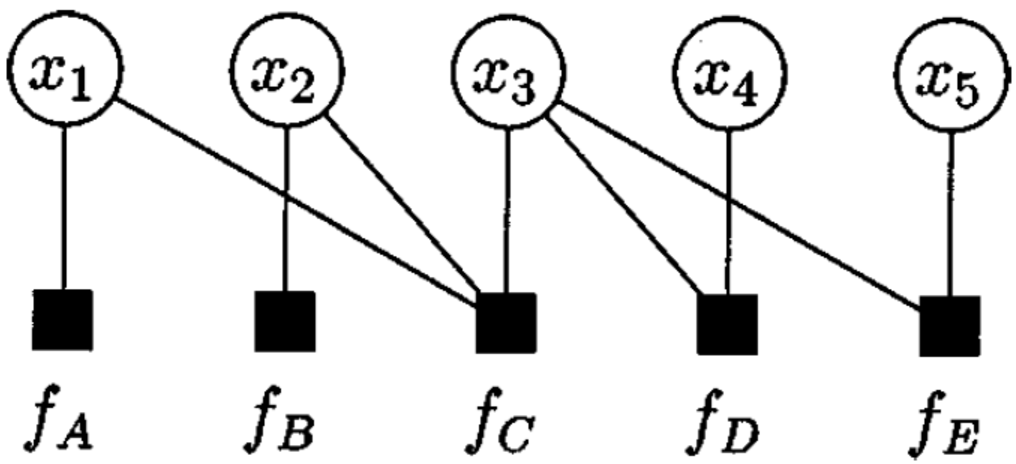


Fig. 1. A factor graph for the product $f_A(x_1) f_B(x_2) f_C(x_1, x_2, x_3) \cdot f_D(x_3, x_4) f_E(x_3, x_5)$.

Une méthode de calculer des probabilités par le passage de messages

- Les messages sont des vecteurs de nombres réels
- Les messages sont transmis de noeuds variables à des nœuds fonction, et de nœuds fonction à nœuds variables
- Un noeud peut envoyer un message à son voisin seulement si elle a reçu tous les messages de ses autres voisins.
- Étant donné un arbre, l'algorithme peut commencer par envoyant des messages de chaque des feuilles, et s'arrête une fois que chaque nœud a transmis un message à tous voisin.

Les messages fonction à variable

- Les messages entre une fonction f et une variable X sont calculé par

$$m_{f\to X}(x) = \sum_{x_1,\ldots,x_k} f(x, x_1, \ldots, x_k)\, m_{X_1\to f}(x_1)\ \cdots m_{X_k\to f}(x_k)$$

- Avec une somme sur toutes les autres variables $X_1,\ldots,X_k$ (sauf X) avec un arc à f et
sur le produit de la fonction et tous les messages variable-fonction à partir de ces autres variables

- Si f contient seulement X, alors $m_{f\to X}(x) = f(x)$.
(on a un nœud fonction feuille)

Les messages variable-fonction

- Les messages entre une variable X et une fonction f sont calculé par

$$m_{X\to f}(x) \;=\; \begin{cases} 1 & \text{, nœud (variable) feuille} \\ m_{f_1\to X}(x)\cdots m_{f_k\to X}(x) & \text{autrement} \end{cases}$$

- Ou simplement un produit sur toutes les messages d'autres fonctions $f_1,\ldots, f_k$ (sauf f) contenant X

À la fin

- Une fois que tous les messages ont été transmis,
- alors le marginal final pour toute variable X peut être calculé par

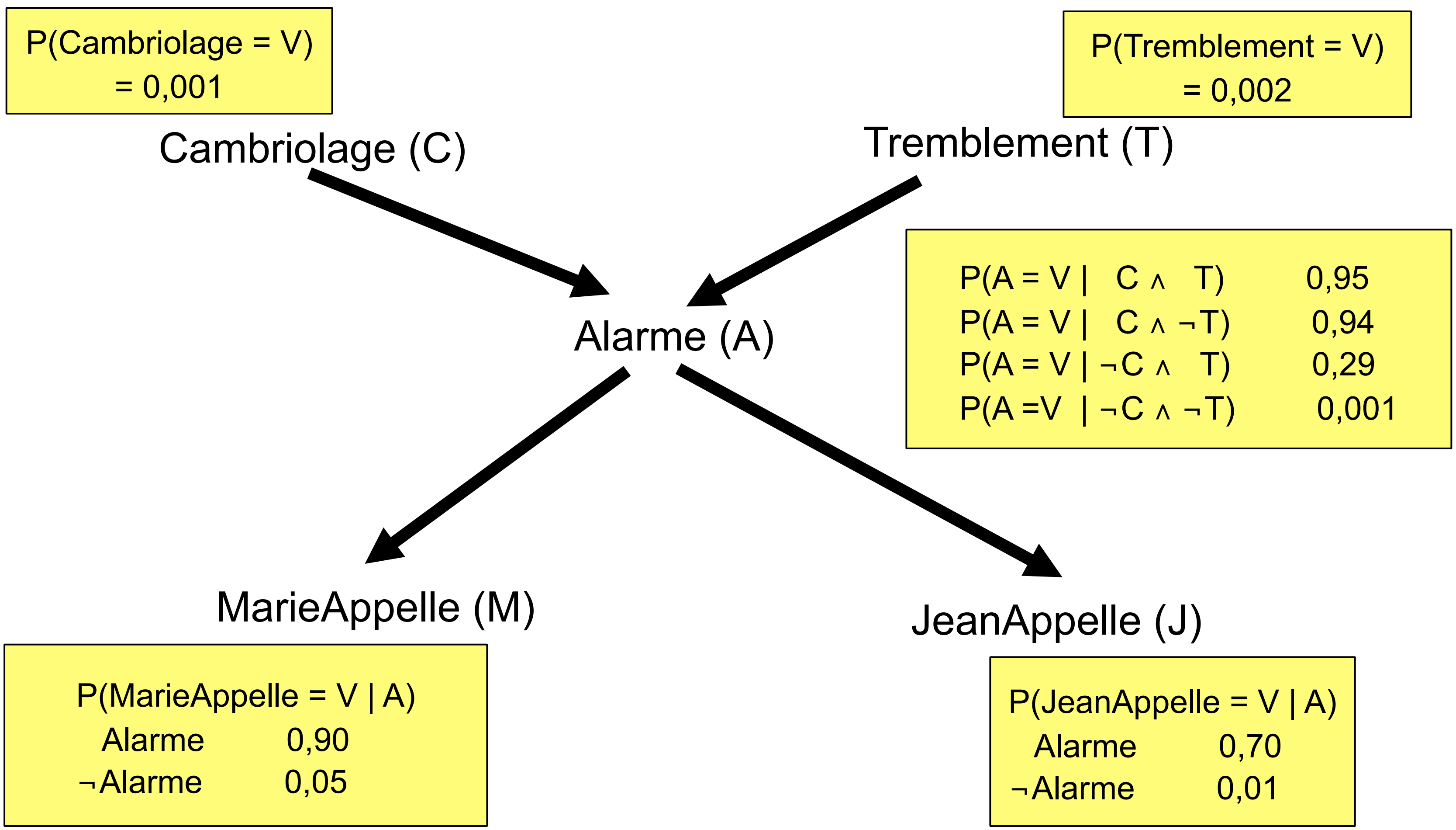
$$P(X_i = x_i) \quad = \quad m_{f_1 \rightarrow X_i}(x_i) \cdots m_{f_k \rightarrow X_i}(x_i)$$

- sur toutes les fonctions $f_1, \ldots, f_k$ contenant X

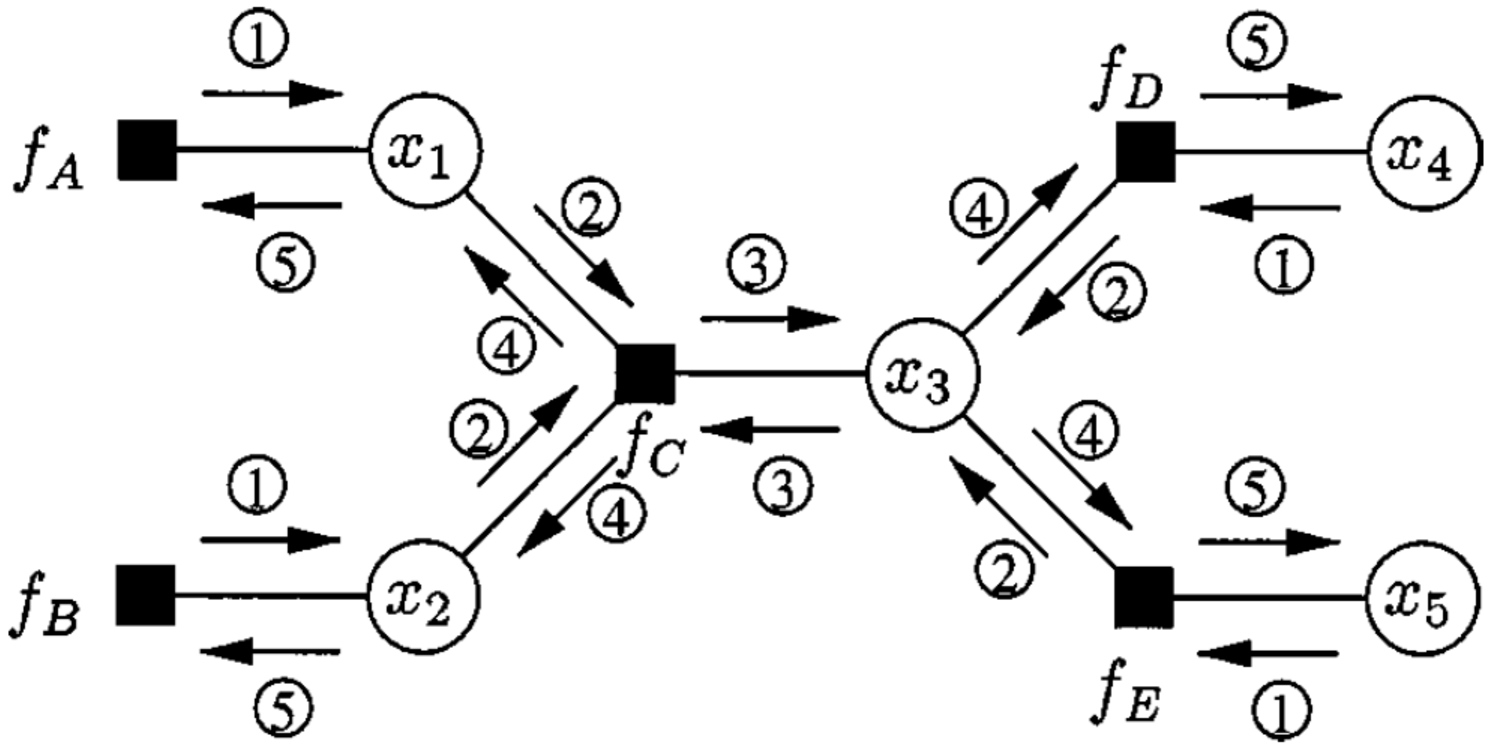
Exemple concrète

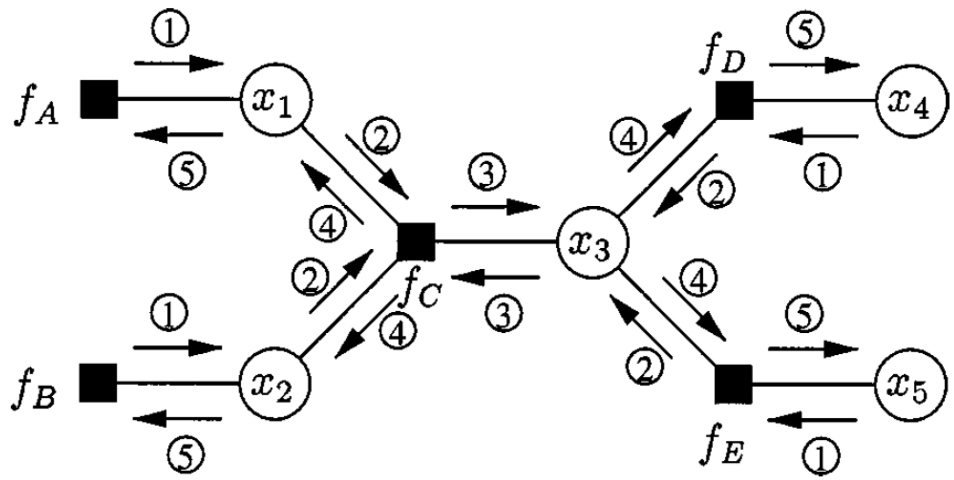
Des graphes de facteurs et
l'algorithme somme - produit

Exemple : Le réseau « Alarme »



Un exemple de passage de message, ou un méthode pour « compiler » le réseau (dans le sens d'un logiciel comme Netica)





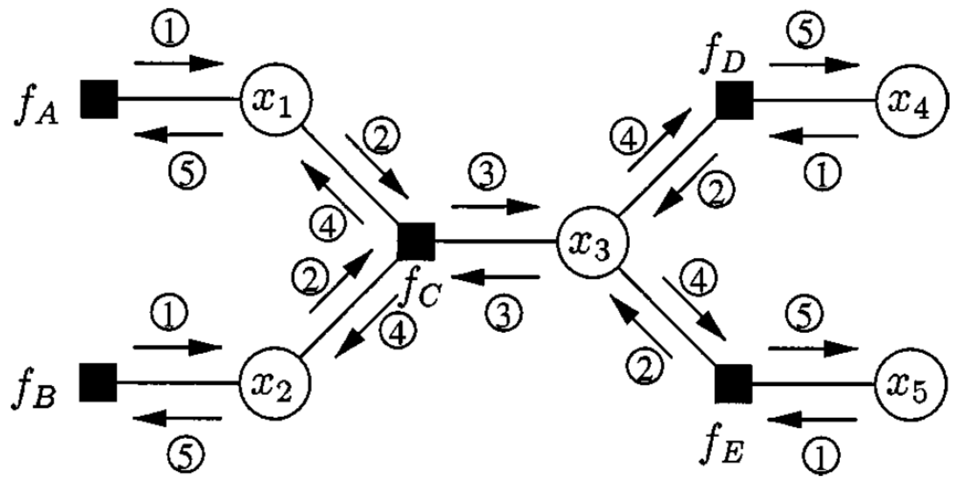
Step 1:

$$\mu_{f_A \to x_1}(x_1) = \sum_{\sim\{x_1\}} f_A(x_1) = f_A(x_1)$$

$$\mu_{f_B \to x_2}(x_2) = \sum_{\sim\{x_2\}} f_B(x_2) = f_B(x_2)$$

$$\mu_{x_4 \to f_D}(x_4) = 1$$

$$\mu_{x_5 \to f_E}(x_5) = 1.$$



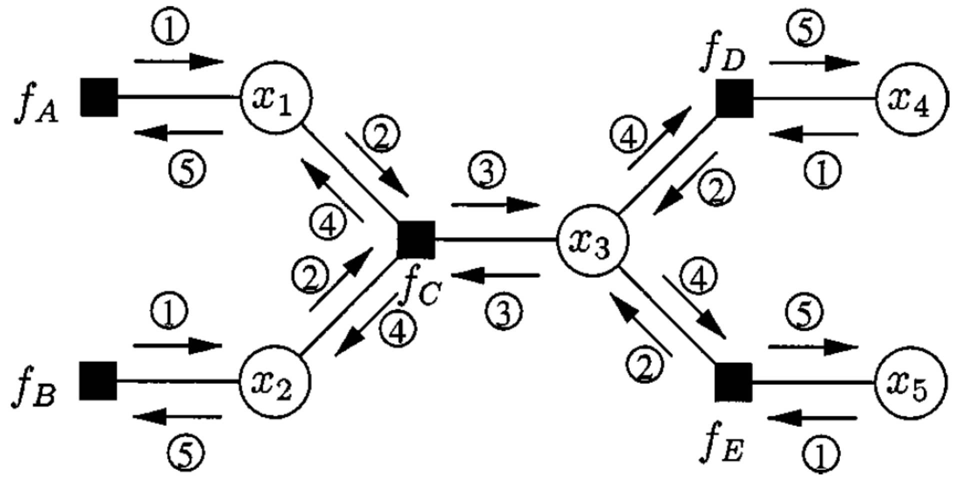
Step 2:

$$\mu_{x_1 \to f_C}(x_1) = \mu_{f_A \to x_1}(x_1)$$

$$\mu_{x_2 \to f_C}(x_2) = \mu_{f_B \to x_2}(x_2)$$

$$\mu_{f_D \to x_3}(x_3) = \sum_{\sim\{x_3\}} \mu_{x_4 \to f_D}(x_4) f_D(x_3, x_4)$$

$$\mu_{f_E \to x_3}(x_3) = \sum_{\sim\{x_3\}} \mu_{x_5 \to f_E}(x_5) f_E(x_3, x_5).$$



Step 3:

$$\mu_{f_C \to x_3}(x_3) = \sum_{\sim\{x_3\}} \mu_{x_1 \to f_C}(x_1)\mu_{x_2 \to f_C}(x_2) f_C(x_1, x_2, x_3)$$

$$\mu_{x_3 \to f_C}(x_3) = \mu_{f_D \to x_3}(x_3)\mu_{f_E \to x_3}(x_3).$$

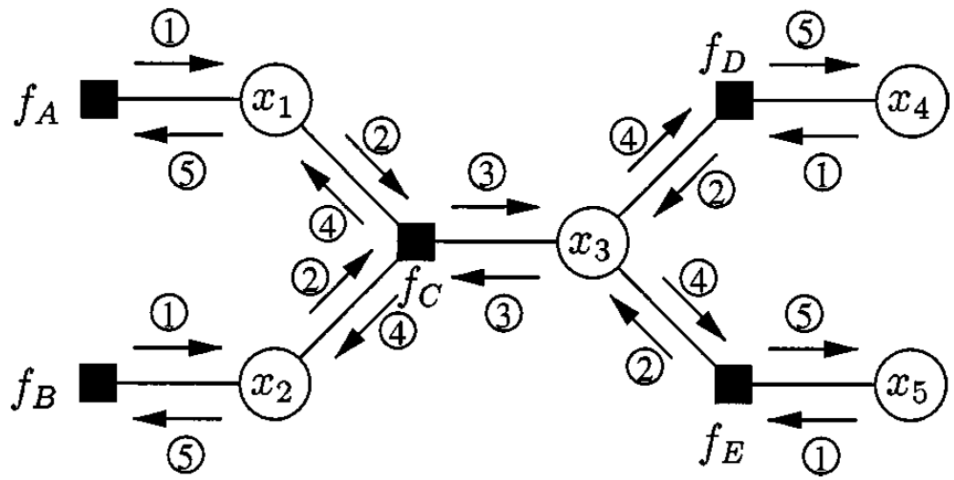
Step 4:

$$\mu_{f_C \to x_1}(x_1) = \sum_{\sim\{x_1\}} \mu_{x_3 \to f_C}(x_3)\mu_{x_2 \to f_C}(x_2) f_C(x_1, x_2, x_3)$$

$$\mu_{f_C \to x_2}(x_2) = \sum_{\sim\{x_2\}} \mu_{x_3 \to f_C}(x_3)\mu_{x_1 \to f_C}(x_1) f_C(x_1, x_2, x_3)$$

$$\mu_{x_3 \to f_D}(x_3) = \mu_{f_C \to x_3}(x_3)\mu_{f_E \to x_3}(x_3)$$

$$\mu_{x_3 \to f_E}(x_3) = \mu_{f_C \to x_3}(x_3)\mu_{f_D \to x_3}(x_3).$$



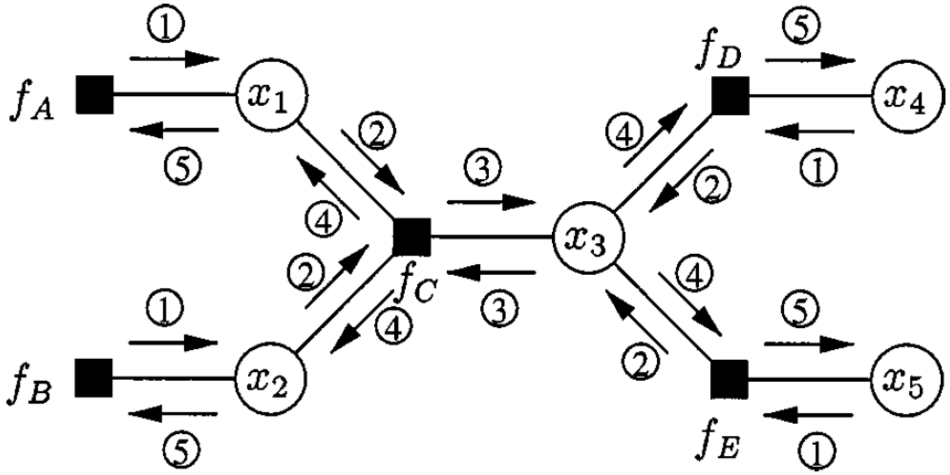
Step 5:

$$\mu_{x_1 \to f_A}(x_1) = \mu_{f_C \to x_1}(x_1)$$

$$\mu_{x_2 \to f_B}(x_2) = \mu_{f_C \to x_2}(x_2)$$

$$\mu_{f_D \to x_4}(x_4) = \sum_{\sim\{x_4\}} \mu_{x_3 \to f_D}(x_3) f_D(x_3, x_4)$$

$$\mu_{f_E \to x_5}(x_5) = \sum_{\sim\{x_5\}} \mu_{x_3 \to f_E}(x_3) f_E(x_3, x_5).$$



Termination:

$$g_1(x_1) = \mu_{f_A \to x_1}(x_1)\mu_{f_C \to x_1}(x_1)$$
$$g_2(x_2) = \mu_{f_B \to x_2}(x_2)\mu_{f_C \to x_2}(x_2)$$
$$g_3(x_3) = \mu_{f_C \to x_3}(x_3)\mu_{f_D \to x_3}(x_3)\mu_{f_E \to x_3}(x_3)$$
$$g_4(x_4) = \mu_{f_D \to x_4}(x_4)$$
$$g_5(x_5) = \mu_{f_E \to x_5}(x_5).$$

Probabilités marginales et l'algorithme de somme-produit

- Dans les graphes avec quelques observations inconnues (ou des variables cachées) nous avons besoin de distributions marginales.
- Pour apprentissage avec des variables inconnues, les algorithmes tels que la maximisation d'espérance (EM) sont typiquement utilisées pour estimer des paramètres
- Ces algorithmes exigeant de nous de calculer des distributions marginales locales
- Donc, dans beaucoup de circonstances importantes on a besoin de marginales

Inférence exacte vs. approximative

- Pour les graphes de facteurs à structure arborescente, nous pouvons calculer toutes les probabilités marginales sans approximation avec l'algorithme de somme-produit (défini avec des graphes de facteurs)
- On parle souvent d'inférence exacte dans un arbre
- Quand il existe un ou plusieurs cycles dans un graphe de facteurs, il est possible d'appliquer l'algorithme de somme-produit, toutefois l'inférence sera une approximation

Calculant la « meilleure configuration » et l'algorithme maximum - produit

- Un marginal est une distribution locale de probabilité dans un graphique après avoir intégré sur d'autres variables
- La meilleure configuration d'un graphe est les meilleures valeurs des variables
- Pour un modèle de probabilité, c'est l'explication la plus probable (l'EPP ou MPE en anglais)
- Pour un réseau bayésien ou un modèle avec une distribution a priori, une vraisemblance et un postérieur, nous référons souvent à la tâche d'inférence maximum a posteriori (MAP) pour : a) des variables aléatoires ou b) pour l'estimation des paramètres en utilisant une probabilité a priori