

IFT 615 – Intelligence Artificielle

Hiver 2022

Raisonnement probabiliste

Réseaux bayésiens

Professeur: Froduald Kabanza

Assistants: D'Jeff Nkashama & Jean-Charles Verdier

Sujets couverts

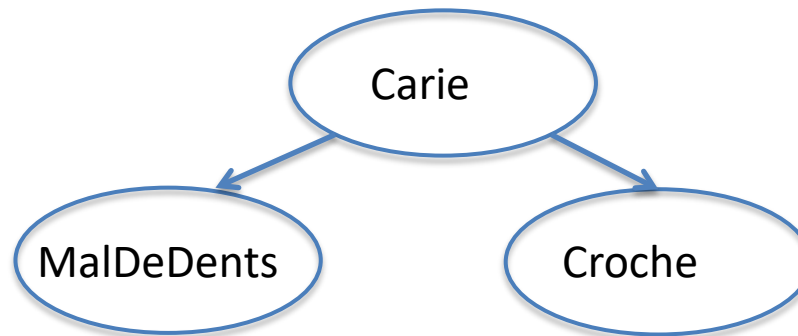
- C'est quoi un réseau bayésien (RB)?
 - ◆ structure d'un RB
 - ◆ calcul de probabilités dans un RB
- Indépendance conditionnelle dans un RB
- Inférence dans un réseau bayésien
 - ◆ inférence exacte
 - ◆ inférence approximative

Réseaux bayésiens

- Les **réseaux bayésiens** (RB) sont un mariage entre la théorie des graphes et la théorie des probabilités
- Un RB permet de représenter les connaissances probabilistes d'une application donnée :
 - ◆ par exemple, les connaissances cliniques d'un médecin sur des liens de causalité entre maladies et symptômes
- Les RB sont utiles pour modéliser des connaissances d'un système expert ou d'un système de support à la décision, dans une situation pour laquelle :
 - ◆ la causalité joue un rôle important (des événements en causent d'autres)
 - ◆ mais notre compréhension de la causalité des événements est incomplète (on doit recourir aux probabilités)

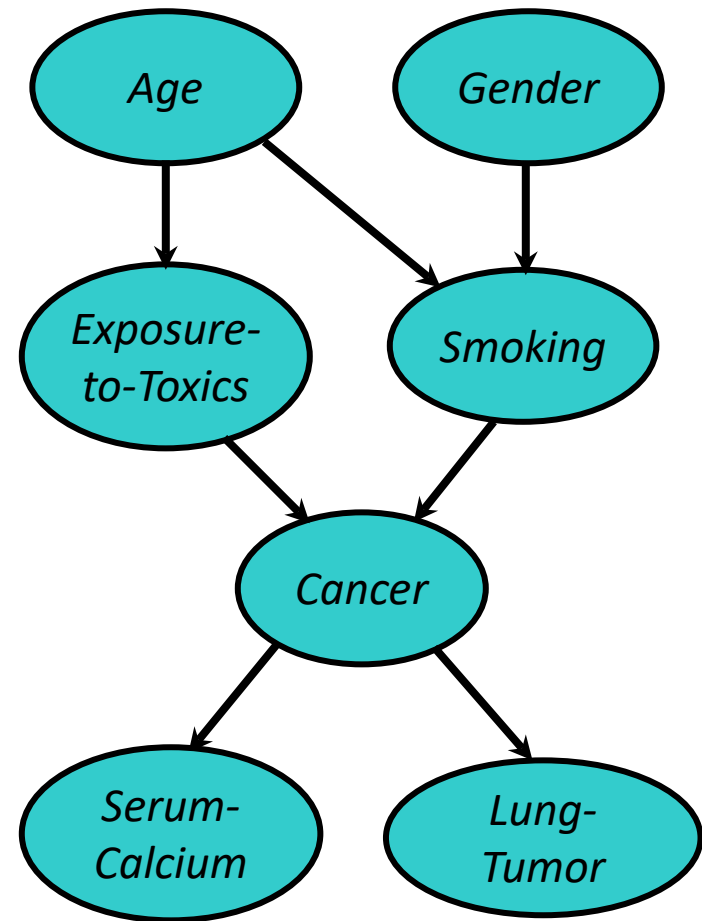
Syntaxe

- Un RB est un **graphe**
 - ◆ **orienté**
 - ◆ **acyclique**
 - ◆ dont les **nœuds** sont des **variables aléatoires** et
 - ◆ les **arcs** représentent
 - » des **dépendances** (par exemple des causalités) probabilistes entre les variables et
 - » des **distributions de probabilités conditionnelles** (locales) pour chaque variable étant donnés ses parents



Application : diagnostique médical

- Déterminer la maladie d'un patient, sachant des symptômes
- On peut avoir une maladie mais montrer seulement un sous-ensemble des symptômes possibles



Applications

- Diagnostique : $P(\text{Causes} \mid \text{Symptômes})$
- Prédiction : $P(\text{Symptômes} \mid \text{Causes})$
- Classification : $\max P(\text{Classe} \mid \text{Données})$

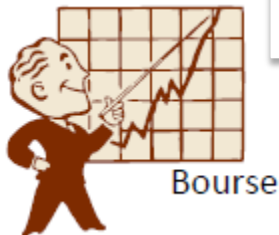
classe



Commerce électronique



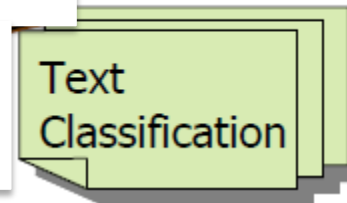
Médecine



Bourse

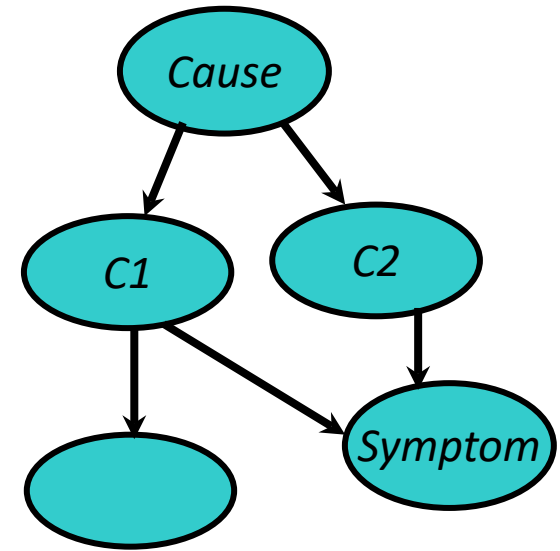


Bio-informatique

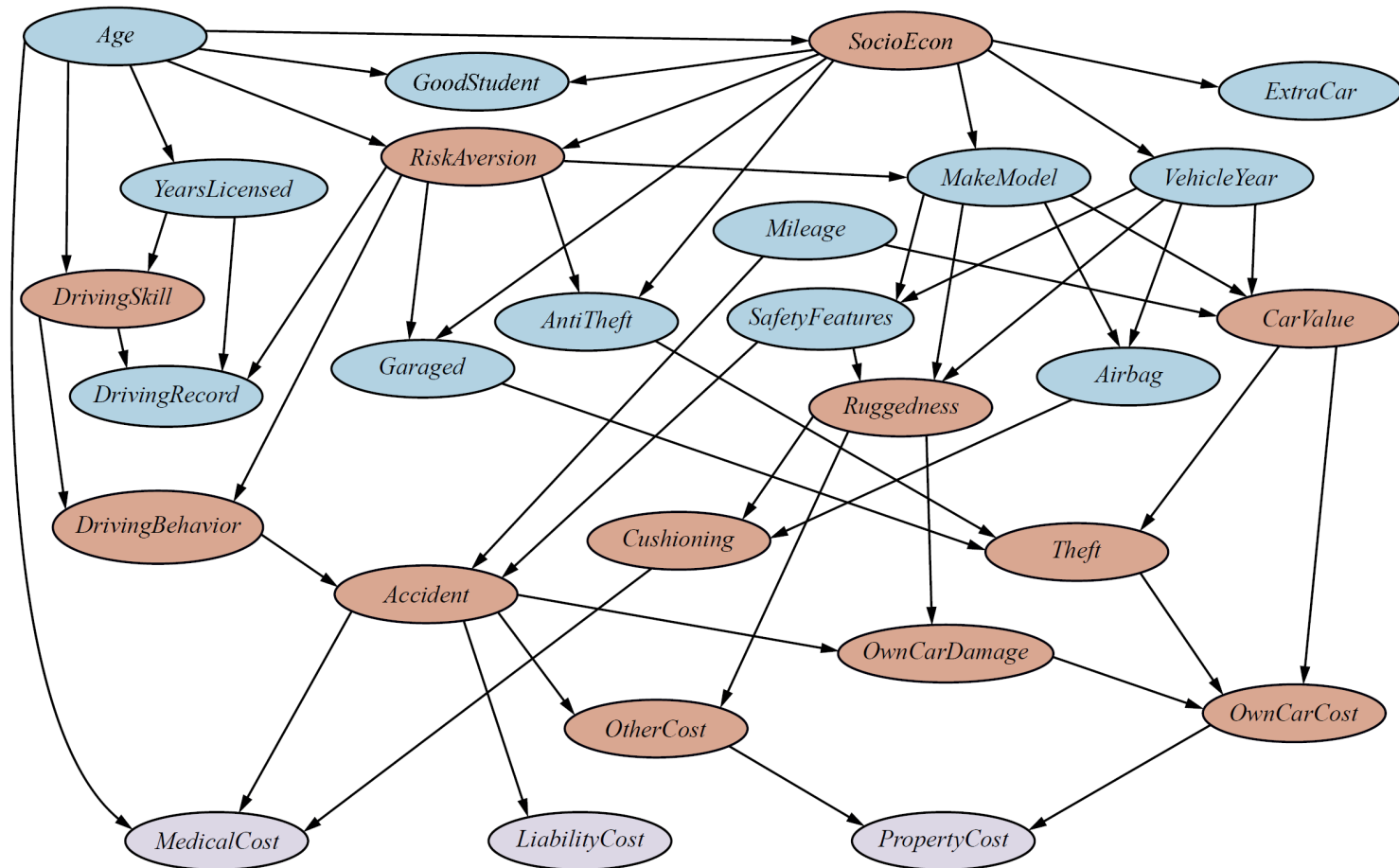


Text
Classification

Diagnostic
informatique



RB pour évaluation des applications pour l'assurance auto



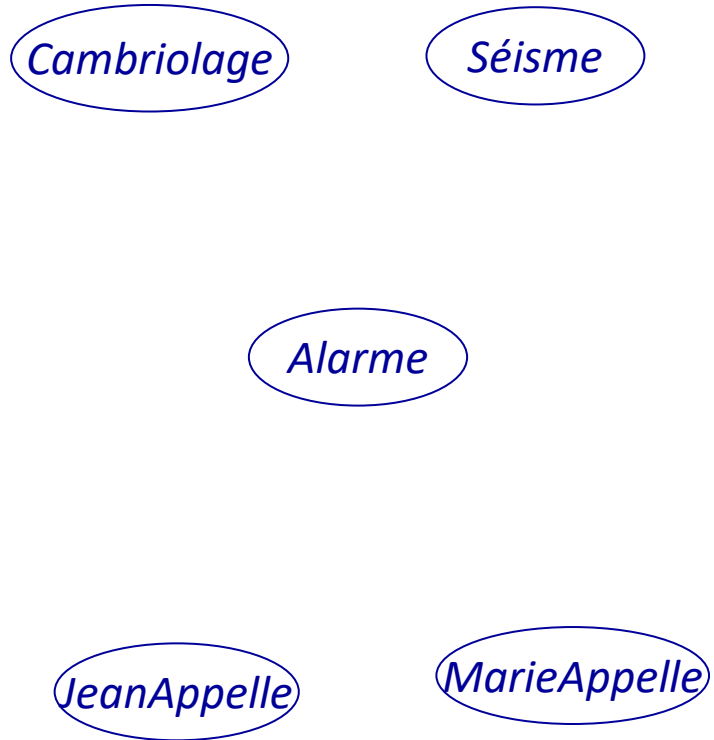
Exemple

- Considérons la situation suivante :
 - ◆ je suis au travail, et mes voisins Marie et Jean m'ont promis de m'appeler chaque fois que mon alarme sonne
 - ◆ mon voisin Jean m'appelle pour me dire que mon alarme sonne
 - » parfois il confond l'alarme avec la sonnerie du téléphone
 - ◆ par contre ma voisine Marie ne m'appelle pas toujours
 - » parfois elle met la musique trop fort
 - ◆ parfois mon alarme se met à sonner lorsqu'il y a de légers séismes
 - ◆ comment conclure qu'il y a ou non un cambriolage chez moi?
- On peut représenter ce problème par un RB

Exemple

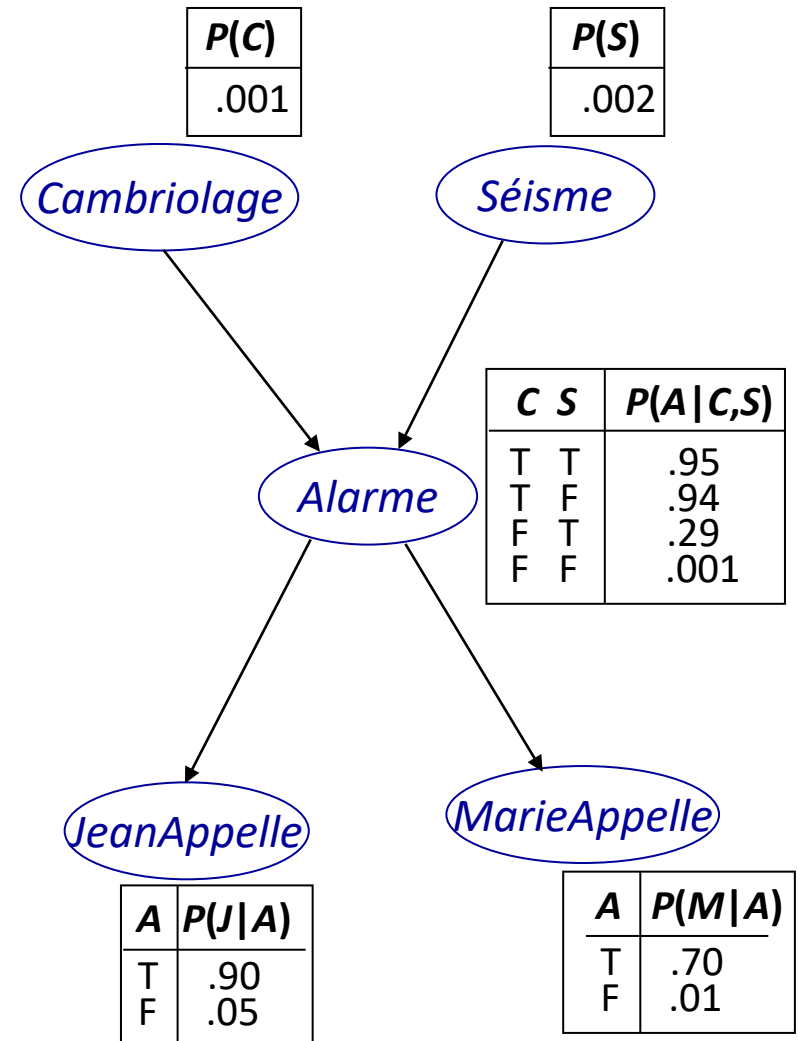
- Variables aléatoires :

- ◆ *Cambriolage*
- ◆ *Séisme*
- ◆ *Alarme*
- ◆ *JeanAppelle*
- ◆ *MarieAppelle*



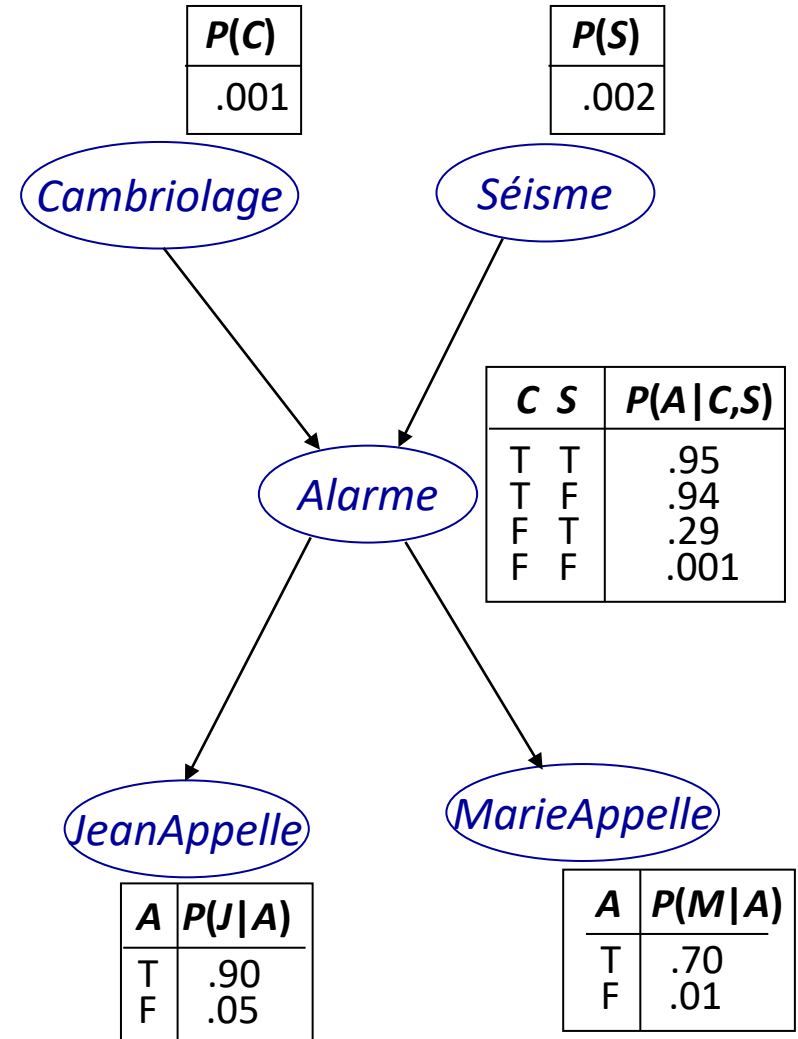
Exemple

- La topologie du RB modélise les relations de causalité
- Un arc d'un nœud X vers un nœud Y signifie que la variable X **influence** la variable Y
 - ◆ un cambriolage peut déclencher l'alarme
 - ◆ un séisme aussi
 - ◆ l'alarme peut inciter Jean à appeler
 - ◆ idem pour Marie
- Une **table de probabilités conditionnelles** (TPC) donne la probabilité pour chaque valeur du nœud étant donnés les combinaisons des valeurs des parents du nœud (c'est l'équivalent d'une **distribution**)



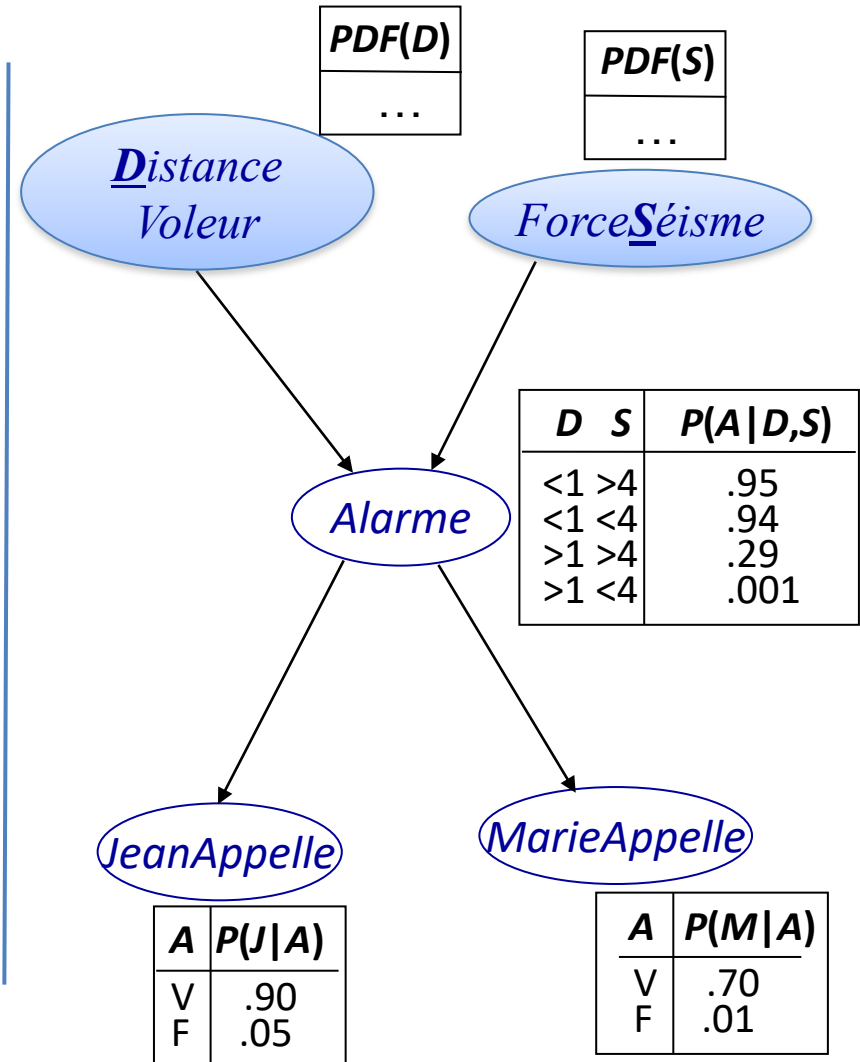
Définitions

- S'il y a un arc d'un nœud Y vers un nœud X , cela signifie que la variable Y influence la variable X
 - ◆ Y est appelé le **parent** de X
 - ◆ $Parents(X)$ est l'ensemble des parents de X
- Si X n'a pas de parents, sa distribution de probabilités est dite **inconditionnelle** ou **a priori**
- Si X a des parents, sa distribution de probabilités est dite **conditionnelle**



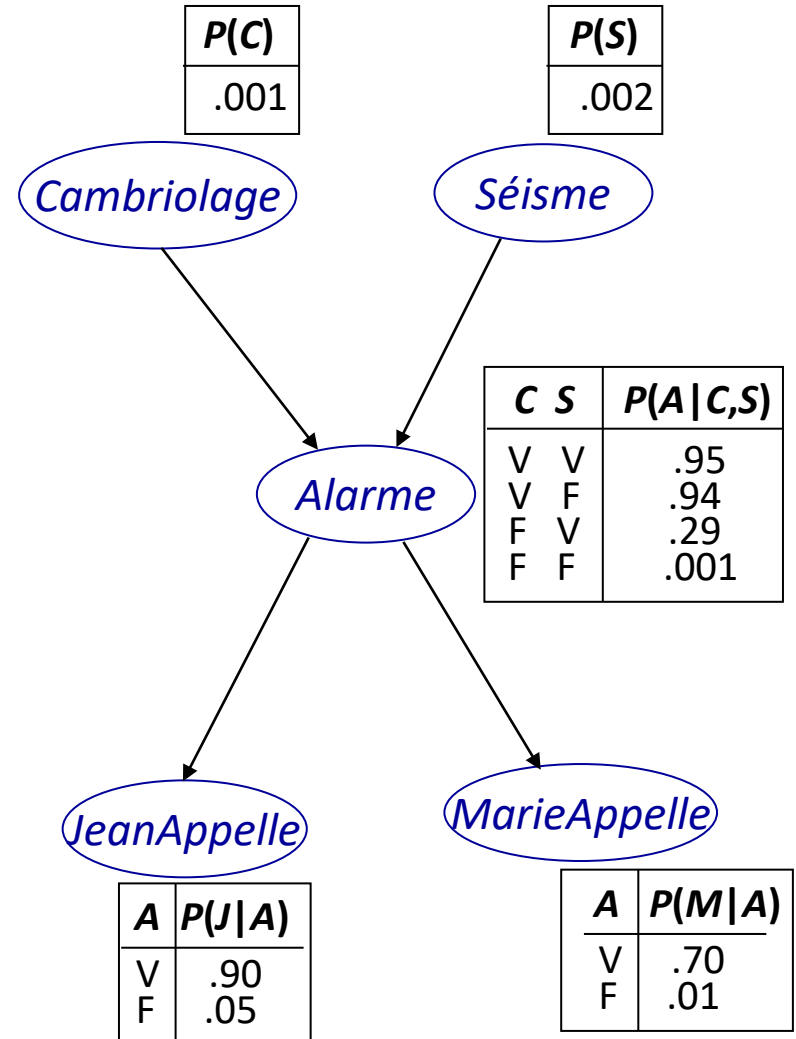
RB avec des variables continues

- Dans ce cours, on considère uniquement des RB avec des variables discrètes :
 - ◆ les TPC sont spécifiées en énumérant toutes les entrées
- Mais les RB peuvent aussi supporter les variables continues :
 - ◆ les probabilités conditionnelles sont spécifiées par des **fonctions de densité de probabilités (PDF)**
 - ◆ exemples :
 - » distance entre voleur et le capteur de mouvement
 - » force du séisme sur l'échelle de Richter



Autres appellations

- Il y a d'autres appellations pour les RB :
 - ◆ réseaux de croyance (*belief networks*)
 - ◆ modèle graphique dirigé acyclique
- Les RB font partie de la classe plus courante des **modèles graphiques**



Sémantique

- Un RB est une façon compacte de représenter des probabilités conjointes
- Par définition, la probabilité conjointe de X_1 et X_2 est donnée par la distribution $\mathbf{P}(X_1, X_2)$, pour une valeur donnée de X_1 et X_2
- La distribution conditionnelle de X_1 sachant X_2 est notée $\mathbf{P}(X_1 | X_2)$
 - ◆ $\mathbf{P}(X_1, X_2) = \mathbf{P}(X_1 | X_2) P(X_2)$
- Soit $X = \{X_1, \dots, X_n\}$, l'ensemble des variables d'un RB :

$$\mathbf{P}(X_1, \dots, X_n) = \prod_{i=1}^n \mathbf{P}(X_i | \text{Parents}(X_i))$$

- En d'autres mots, la distribution conjointe des variables d'un RB est définie comme étant le produit des distributions conditionnelles (locales)

Calcul de probabilité conjointe

- Nous avons vu que, quelque soit l'ensemble de variables $X = \{X_1, \dots, X_n\}$, par définition :

$$\begin{aligned} P(X_1, \dots, X_n) &= P(X_n \mid X_{n-1}, \dots, X_1) P(X_{n-1}, \dots, X_1) \\ &= P(X_n \mid X_{n-1}, \dots, X_1) P(X_{n-1} \mid X_{n-2}, \dots, X_1) \dots P(X_2 \mid X_1) P(X_1) \\ &= \prod_{i=1}^n P(X_i \mid X_{i-1}, \dots, X_1) \end{aligned}$$

- Pour un RB : $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid \text{Parents}(X_i))$
 - ◆ ceci est cohérent avec l'assertion précédente pour autant que $\text{Parents}(X_i)$ soit l'ensemble de $\{X_{i-1}, \dots, X_1\}$
 - ◆ ainsi, un RB est en fait une façon de **représenter les indépendances conditionnelles**

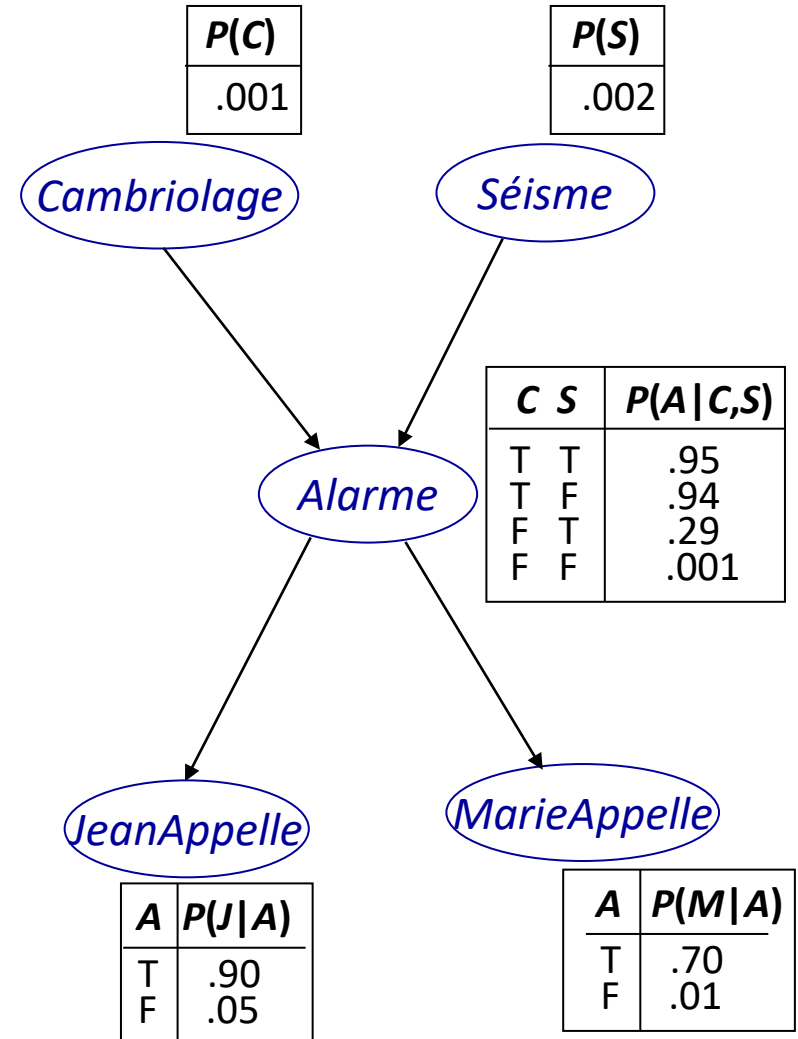
Exemple : probabilité conjointe

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid \text{Parents}(X_i))$$

$$\begin{aligned} P(j, m, a, \neg c, \neg s) \\ &= P(j \mid a) P(m \mid a) P(a \mid \neg c, \neg s) \\ &\quad P(\neg c) P(\neg s) \\ &= .90 * .70 * .001 * \\ &\quad .999 * .998 \\ &\approx .00062 \end{aligned}$$

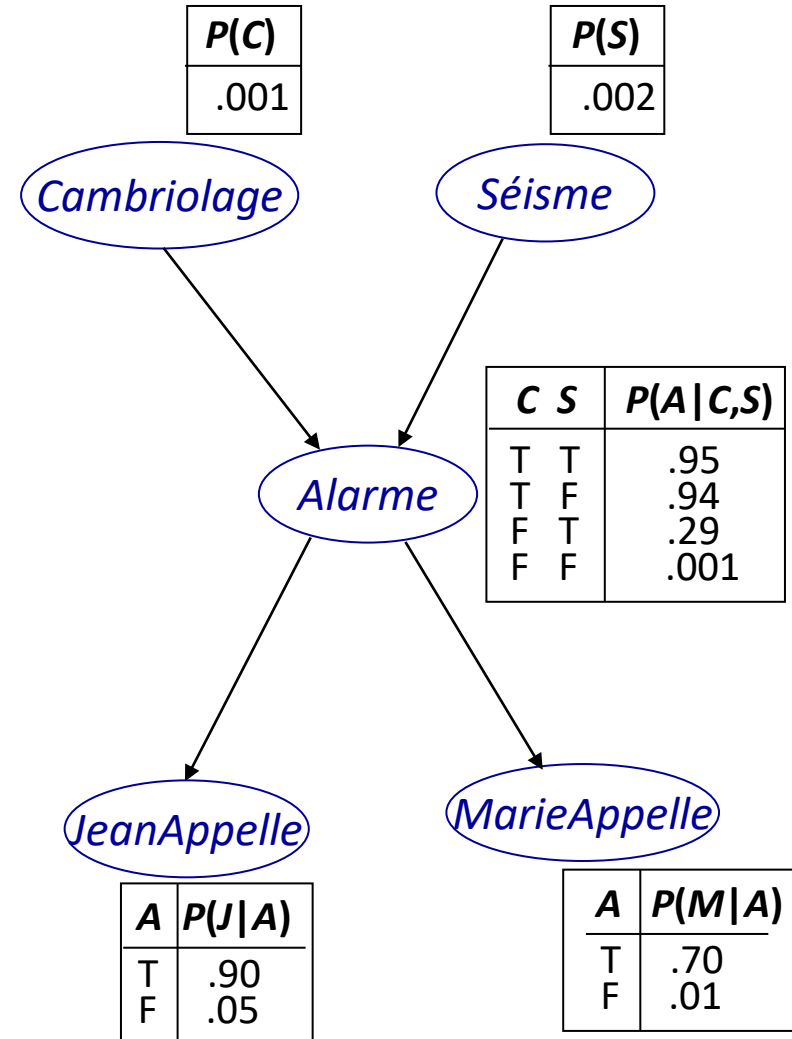
$P(J=T, M=T, A=T, C=F, S=F)$
est aussi noté $P(j, m, a, \neg c, \neg s)$

$P(J=j, M=m, A=a, C=\neg c, S=\neg s)$
est aussi noté $P(j, m, a, \neg c, \neg s)$



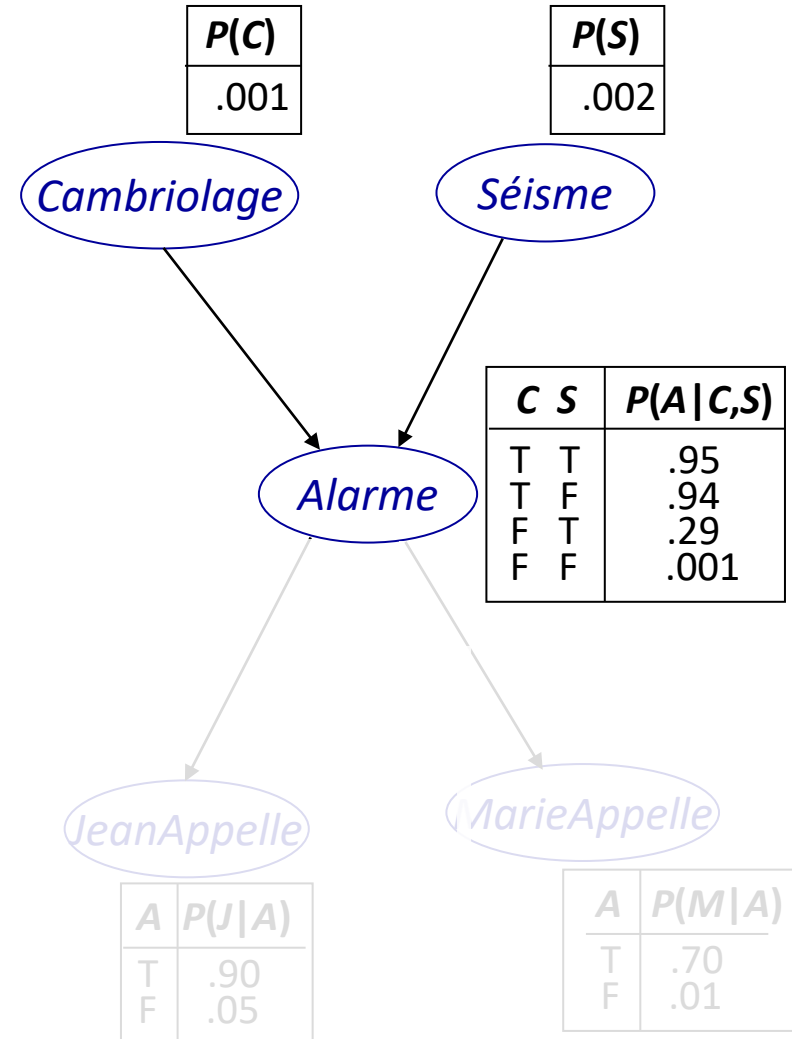
Exemple : probabilité marginale

$$\begin{aligned}
 P(\neg c, a) &= \sum_m \sum_j \sum_s P(J=j, M=m, a, \neg c, S=s) \\
 &= \sum_m \sum_j \sum_s P(j|a) P(m|a) P(a | \neg c, s) P(\neg c) P(s) \\
 &= \sum_s \sum_j \sum_m P(j|a) P(m|a) P(a | \neg c, s) P(\neg c) P(s) \\
 &= \sum_s \sum_j P(j|a) P(a | \neg c, s) P(\neg c) P(s) \underbrace{\sum_m P(m|a)}_{=1} \\
 &= \sum_s P(a | \neg c, s) P(\neg c) P(s) \underbrace{\sum_j P(j|a)}_{=1} \\
 &= P(a | \neg c, s) P(\neg c) P(s) + P(a | \neg c, \neg s) P(\neg c) P(\neg s) \\
 &= .29 * .999 * .002 + .001 * .999 * .998 \\
 &\approx 0.0016
 \end{aligned}$$



Exemple : probabilité marginale

$$\begin{aligned}
 P(\neg c, a) &= \sum_m \sum_j \sum_s P(J=j, M=m, a, \neg c, S=s) \\
 &= \sum_m \sum_j \sum_s P(j|a) P(m|a) P(a | \neg c, s) P(\neg c) P(s) \\
 &= \sum_s \sum_j \sum_m P(j|a) P(m|a) P(a | \neg c, s) P(\neg c) P(s) \\
 &= \sum_s \sum_j P(j|a) P(a | \neg c, s) P(\neg c) P(s) \underbrace{\sum_m P(m|a)}_{=1} \\
 &= \sum_s P(a | \neg c, s) P(\neg c) P(s) \underbrace{\sum_j P(j|a)}_{=1} \\
 &= P(a | \neg c, s) P(\neg c) P(s) + P(a | \neg c, \neg s) P(\neg c) P(\neg s) \\
 &= .29 * .999 * .002 + .001 * .999 * .998 \\
 &\approx 0.0016
 \end{aligned}$$

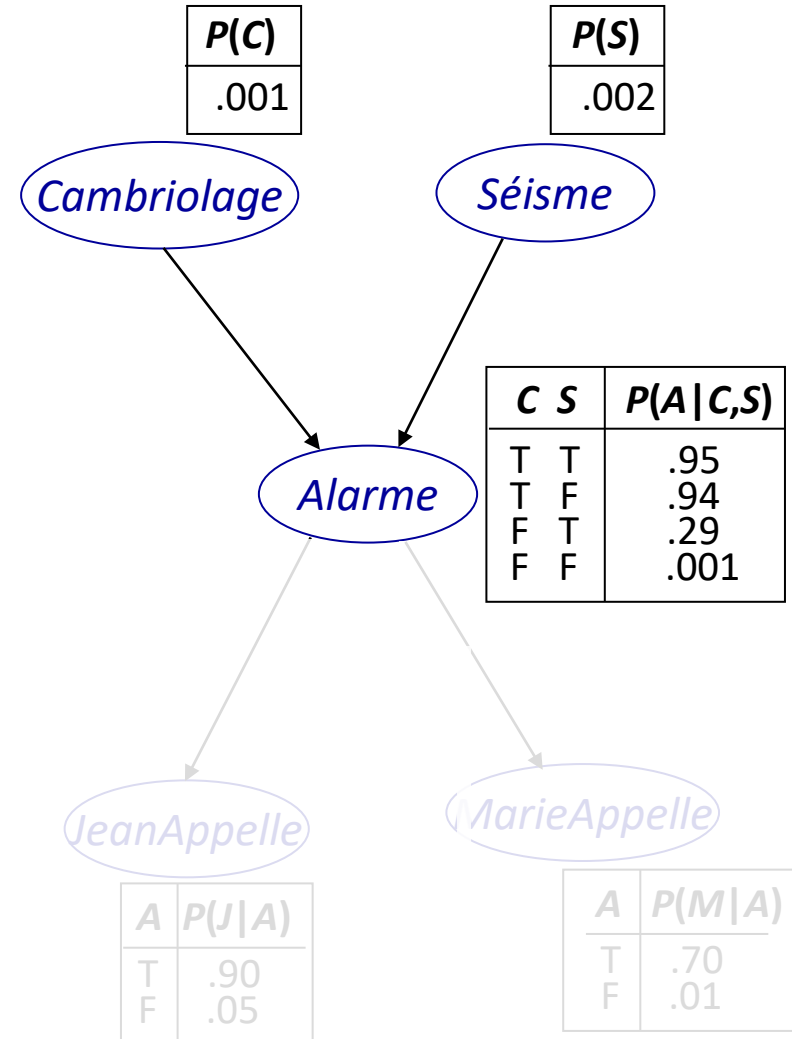


Probabilité marginale

$$P(\neg c, a) = \sum_m \sum_j \sum_s P(J=j, M=m, a, \neg c, S=s) \\ = \sum_s P(a \mid \neg c, S=s) P(\neg c) P(S=s)$$

● Pour les probabilités marginales, on peut ignorer les nœuds **dont les descendants ne sont pas les nœuds observés**

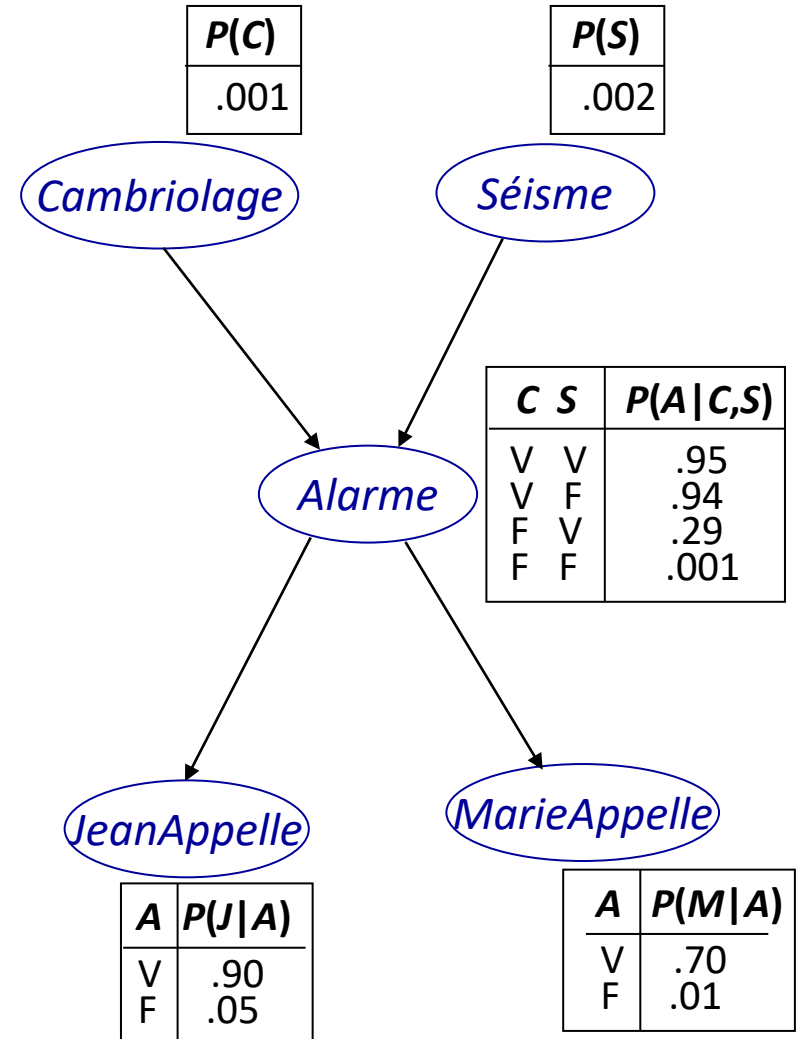
- ◆ *JeanAppelle* et *MarieAppelle* et leurs descendants ne sont pas observés, alors on peut les ignorer
- ◆ *Séisme* est un ancêtre de *Alarme*, alors on doit le marginaliser explicitement



Indépendance conditionnelle dans un RB

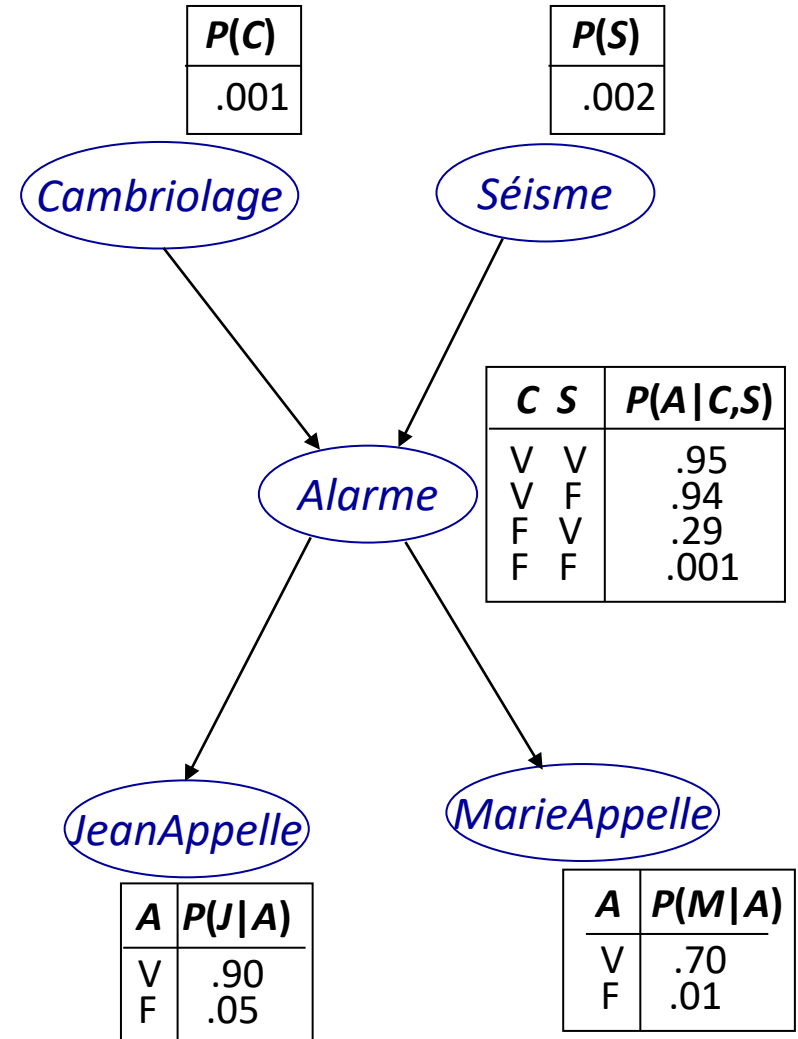
1. Relation entre **grand-parent** et **enfant** étant donné les parents
 - ◆ sont indépendants si tous les parents sont observés.
- Exemples :
 - ◆ *Cambriolage* et *MarieAppelle* sont **dépendants** a priori
 - ◆ mais ils sont **indépendants** étant donné *Alarme* :

$$P(M|A,C) = P(M|A)$$
 - ◆ si *A* est connu, *C* n'intervient pas dans le calcul
 - ◆ **connaître** *A* « bloque » le chemin entre *M* et *C*



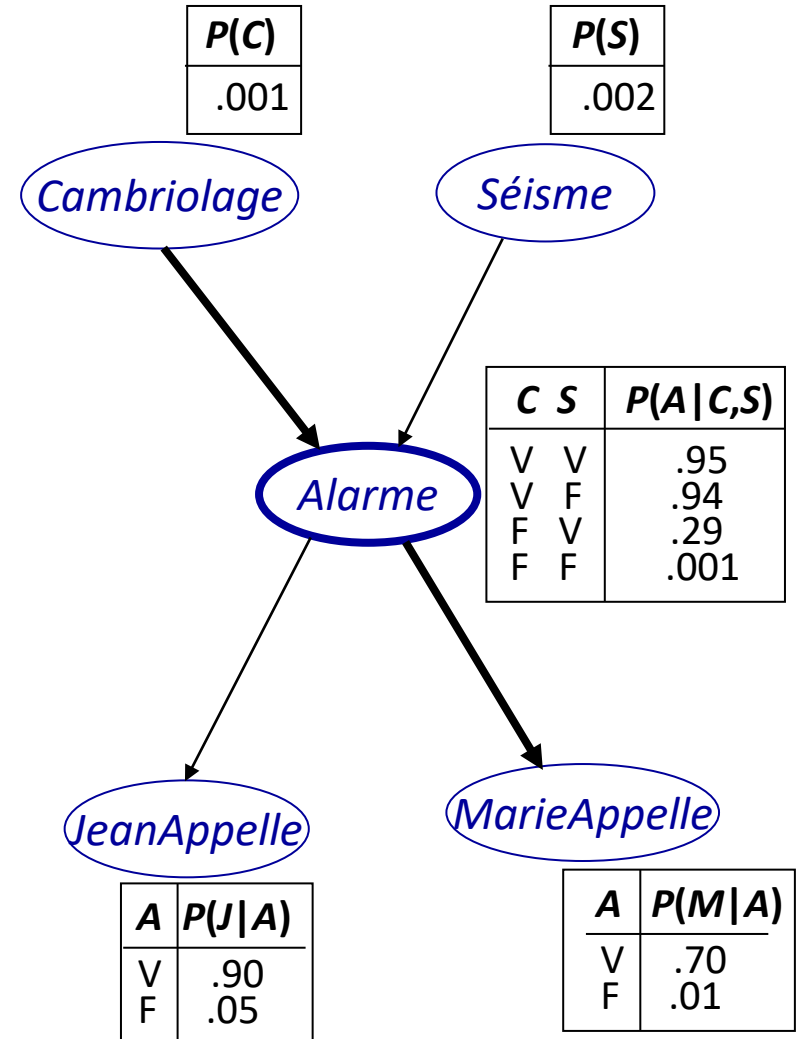
Indépendance conditionnelle dans un RB

$$\begin{aligned}
 P(M|A,C) &= P(M,A,C) / P(A,C) \\
 &= \frac{\sum_s P(M,A,C,S=s)}{\sum_s P(A,C,S=s)} \\
 &= \frac{\sum_s P(M|A) P(A|C,S=s) P(S=s) P(C)}{\sum_s P(A|C,S=s) P(S=s) P(C)} \\
 &= \frac{P(M|A) \cancel{\sum_s P(A|C,S=s) P(S=s) P(C)}}{\cancel{\sum_s P(A|C,S=s) P(S=s) P(C)}} \\
 &= P(M|A)
 \end{aligned}$$



Indépendance conditionnelle dans un RB

$$\begin{aligned}
 P(M|A,C) &= P(M,A,C) / P(A,C) \\
 &= \frac{\sum_s P(M,A,C,S=s)}{\sum_s P(A,C,S=s)} \\
 &= \frac{\sum_s P(M|A) P(A|C,S=s) P(S=s) P(C)}{\sum_s P(A|C,S=s) P(S=s) P(C)} \\
 &= \frac{P(M|A) \cancel{\sum_s P(A|C,S=s) P(S=s) P(C)}}{\cancel{\sum_s P(A|C,S=s) P(S=s) P(C)}} \\
 &= P(M|A)
 \end{aligned}$$



Indépendance conditionnelle dans un RB

2. Relation entre **deux enfants** étant donné un parent:

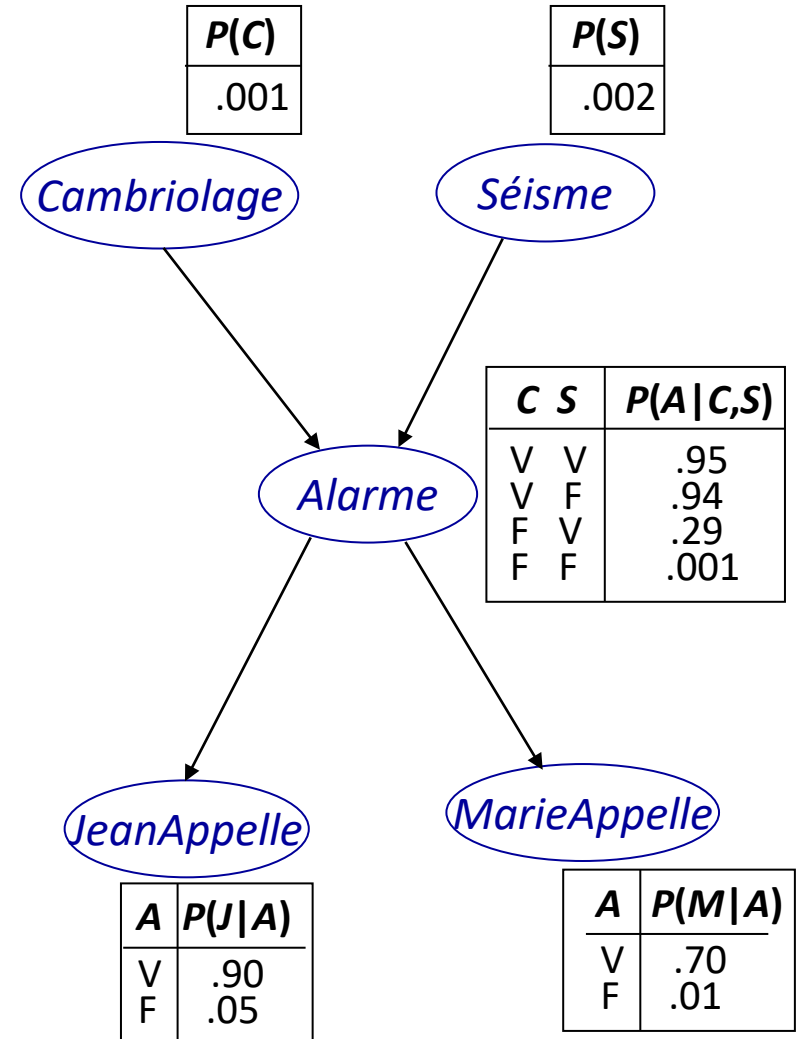
- ◆ sont indépendants si tous leurs parents sont observés

● Exemples :

- ◆ *JeanAppelle* et *MarieAppelle* sont **dépendants** à priori
- ◆ mais ils sont **indépendants** étant donné *Alarme* :

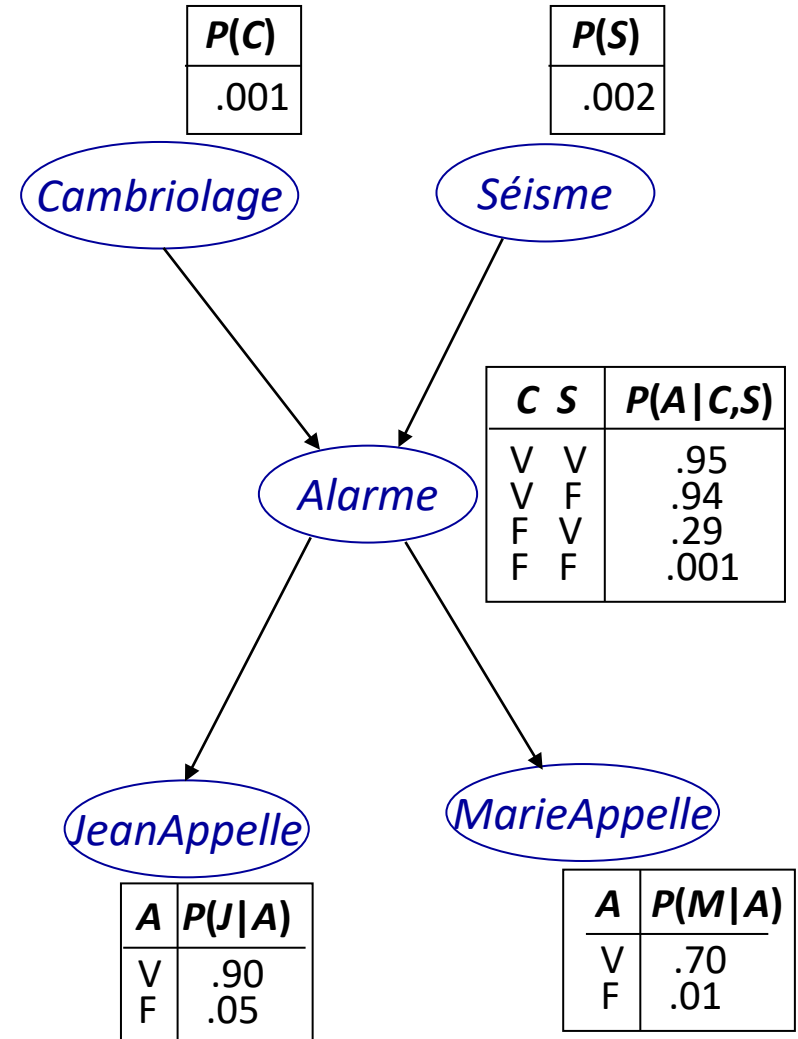
$$P(M|A,J) = P(M|A)$$

- ◆ si *A* est connu, *J* n'intervient pas dans le calcul
- ◆ **connaître** *A* « bloque » le chemin entre *J* et *M*



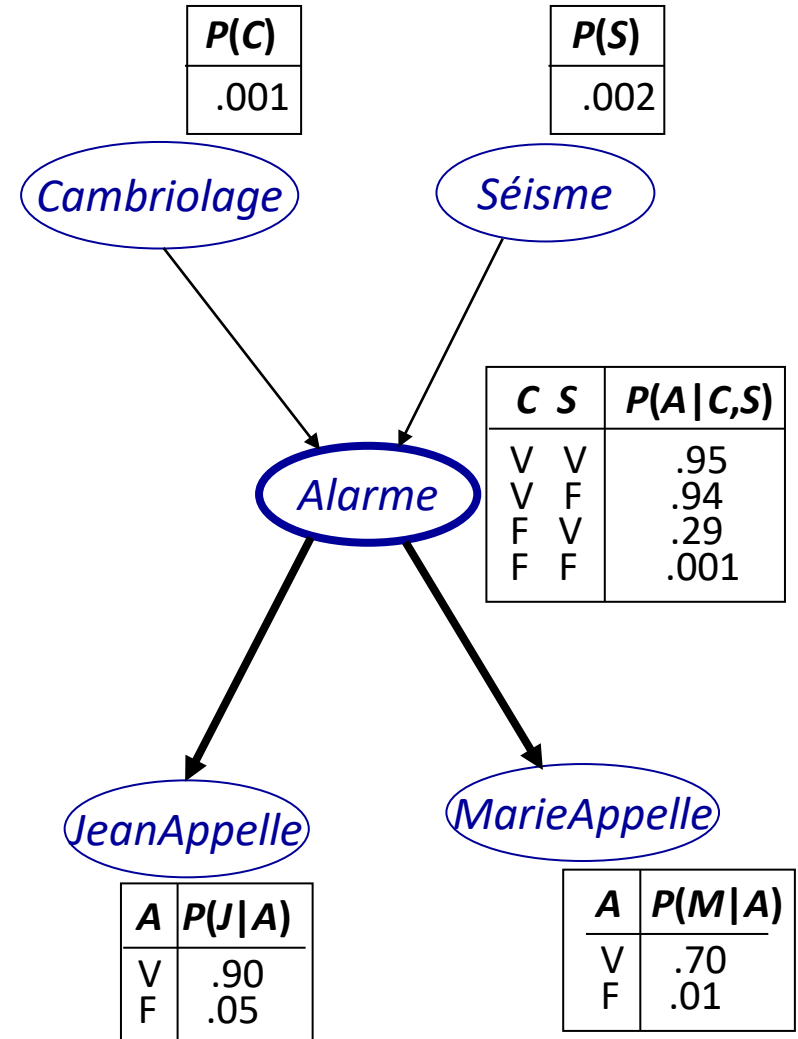
Indépendance conditionnelle dans un RB

$$\begin{aligned}
 P(M|A,J) &= P(M,A,J) / P(A,J) \\
 &= \frac{\sum_s \sum_c P(M,A,J,S=s,C=c)}{\sum_s \sum_c P(A,J,S=s,C=c)} \\
 &= \frac{\sum_s \sum_c P(J|A) P(M|A) P(A,S=s,C=c)}{\sum_s \sum_c P(J|A) P(A,S=s,C=c)} \\
 &= \frac{P(M|A) \sum_s \sum_c P(J|A) P(A,S=s,C=c)}{\sum_s \sum_c P(J|A) P(A,S=s,C=c)} \\
 &= P(M|A)
 \end{aligned}$$



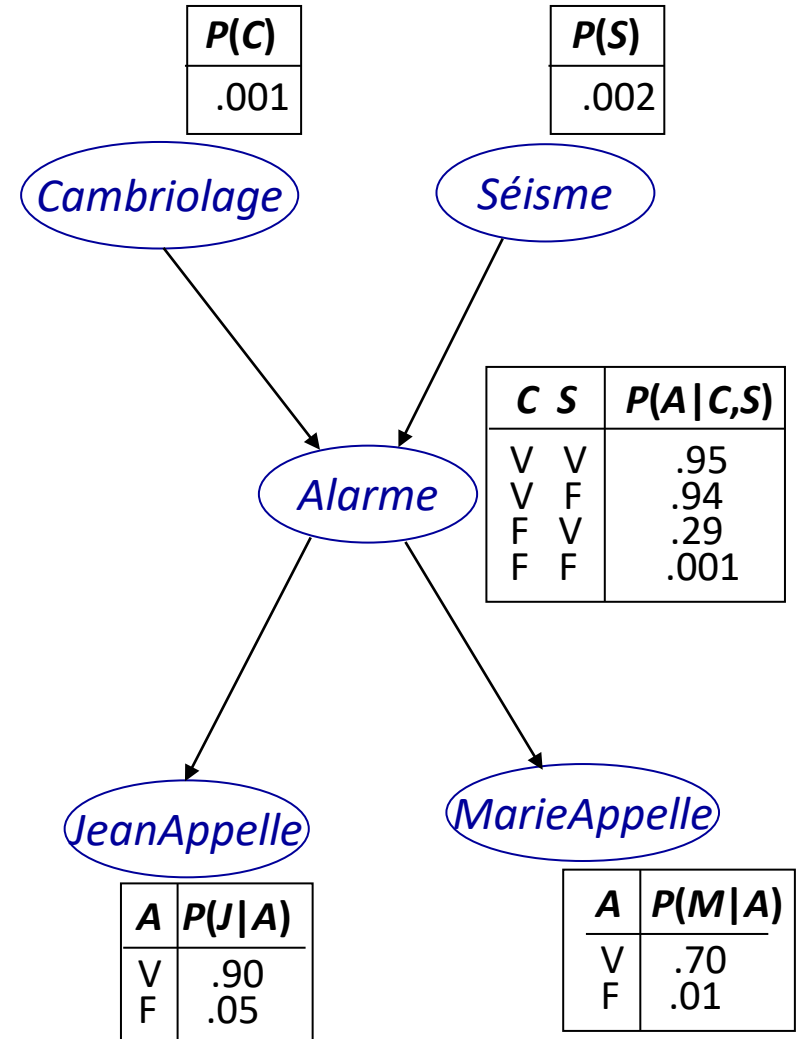
Indépendance conditionnelle dans un RB

$$\begin{aligned}
 P(M|A,J) &= P(M,A,J) / P(A,J) \\
 &= \frac{\sum_s \sum_c P(M,A,J,S=s,C=c)}{\sum_s \sum_c P(A,J,S=s,C=c)} \\
 &= \frac{\sum_s \sum_c P(J|A) P(M|A) P(A,S=s,C=c)}{\sum_s \sum_c P(J|A) P(A,S=s,C=c)} \\
 &= \frac{P(M|A) \sum_s \sum_c P(J|A) P(A,S=s,C=c)}{\sum_s \sum_c P(J|A) P(A,S=s,C=c)} \\
 &= P(M|A)
 \end{aligned}$$



Indépendance conditionnelle dans un RB

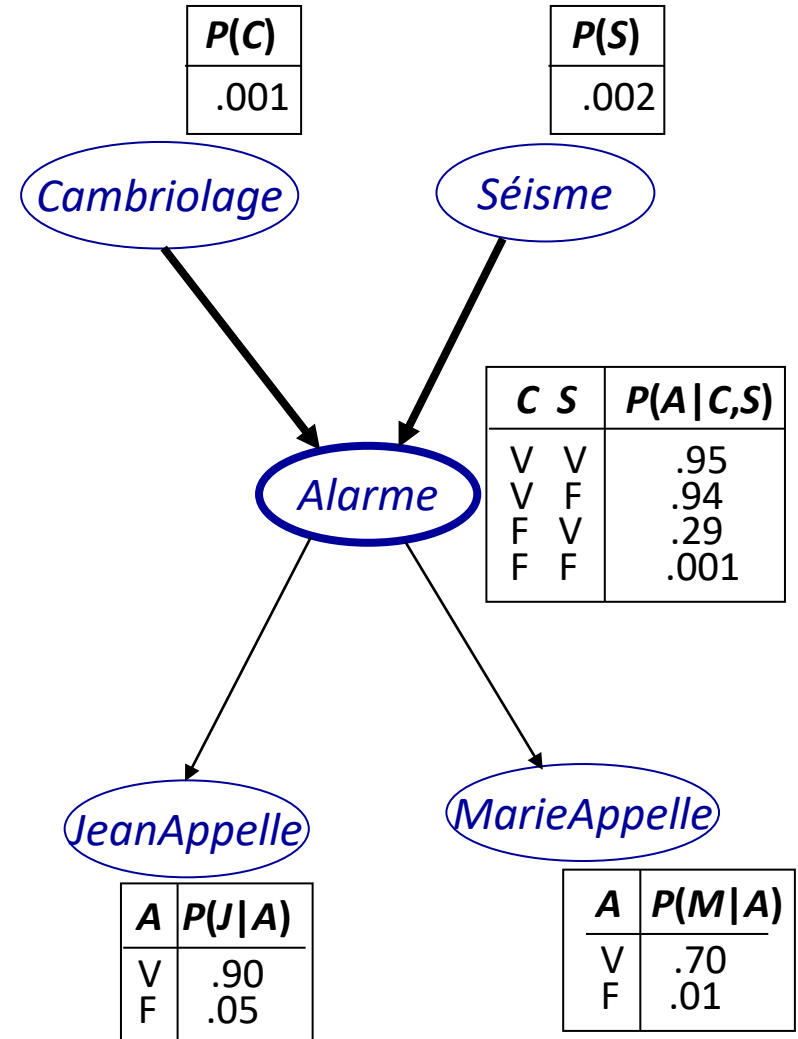
3. Relation entre **deux parents** étant donné un enfant
- ◆ sont indépendants si enfant **non** observé
 - Exemples :
 - ◆ *Cambriolage* et *Séisme* sont **indépendants** à priori
 - ◆ mais ils sont **dépendants** étant donné *Alarme*
 - » $P(C|A,S)$ n'est pas simplifiable, parce que $P(A|C,S)$ n'est pas simplifiable
 - ◆ **ne pas connaître** A « bloque » le chemine entre C et S



Indépendance conditionnelle dans un RB

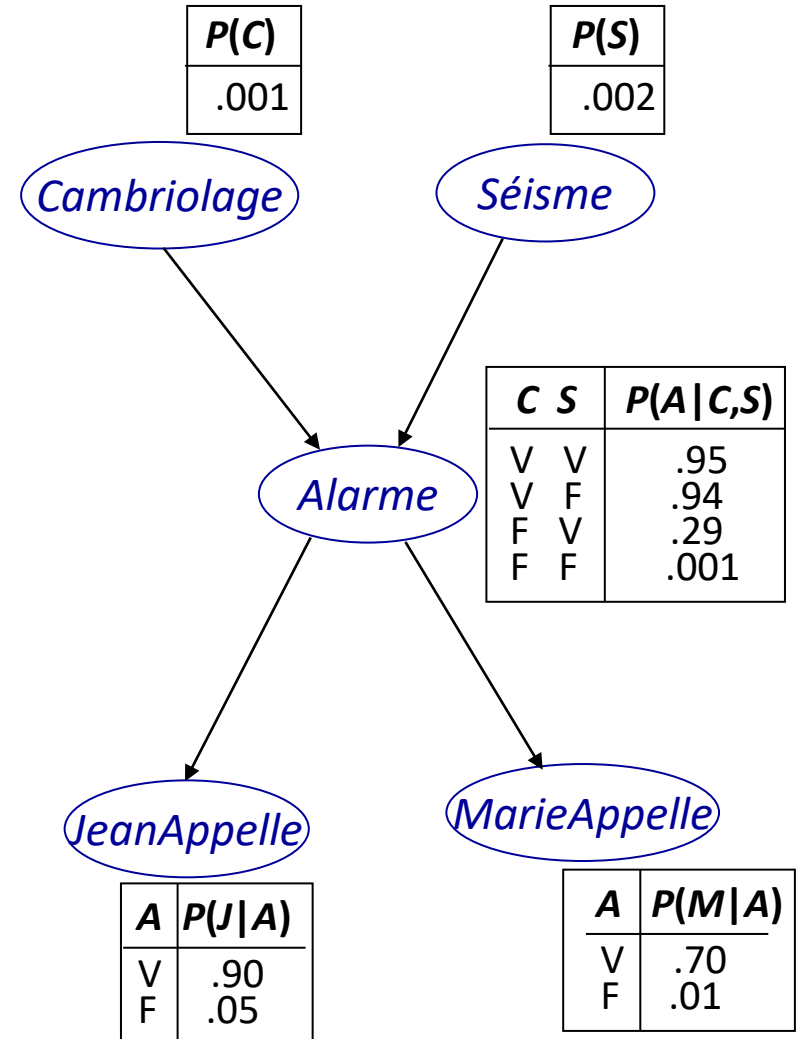
$$P(C|A,S) = P(C,A,S) / P(A,S)$$

$$= \frac{P(A|S,C) P(S) P(C)}{\sum_c P(A|S) P(S) P(C)}$$



Indépendance conditionnelle dans un RB

- De ces trois dernières règle, émane une **règle plus générale**:
 - ◆ un nœud est indépendant de ses **non-descendants**, étant donné ses parents
 - ◆ exemples :
 - » *Cambriolage* et *Séisme* sont **indépendants** a priori
 - » mais ils sont **dépendants** étant donné *Alarme*
 - $P(C|A,S)$ n'est pas simplifiable, parce que $P(A|C,S)$ n'est pas simplifiable
 - » **ne pas connaître A** « bloque » le chemine entre C et S

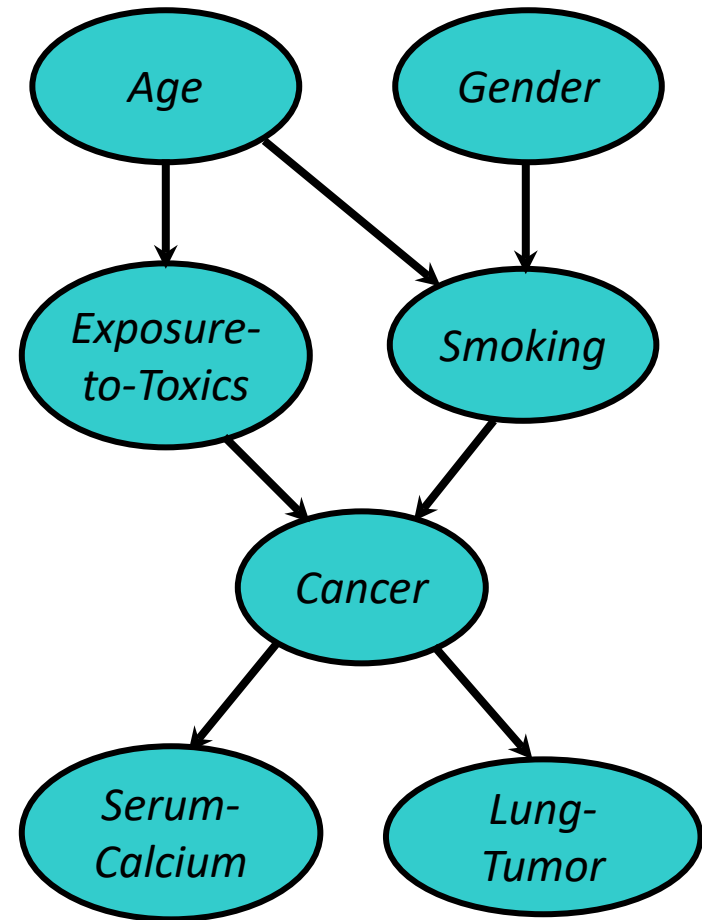


Indépendance conditionnelle dans un RB : D-séparation

- **D-séparation** : critère général pour décider si un nœud X est indépendant d'un nœud Y , étant donnés d'autres nœuds $Z = \{Z_1, \dots, Z_m\}$
- X est indépendant de Y sachant Z si tous **les chemins non-dirigés** entre X et Y sont **bloqués** par Z
- Un **chemin est bloqué** s'il contient au moins un nœud N qui satisfait une ou l'autre des conditions suivantes :
 1. il inclue un nœud $\rightarrow \textcircled{N} \rightarrow$ ou $\leftarrow \textcircled{N} \rightarrow$, où $N \in \{Z_1, \dots, Z_m\}$
 2. il inclue un nœud $\rightarrow \textcircled{N} \leftarrow$ et $N \notin \{Z_1, \dots, Z_m\}$, et **aucun des descendants** de N appartient $\{Z_1, \dots, Z_m\}$.

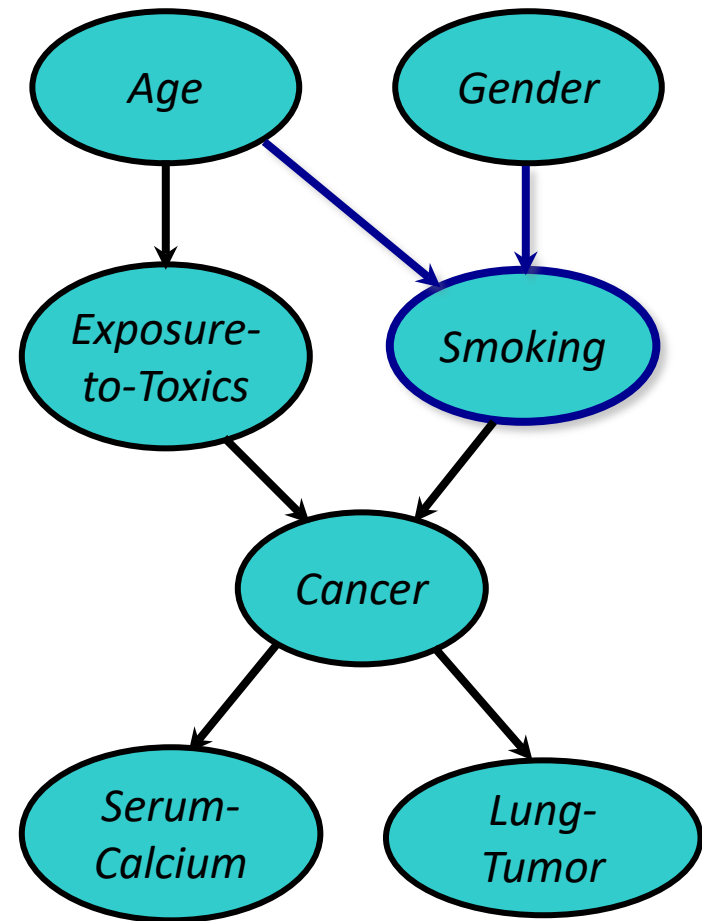
Exemple

- Est-ce que *Age* et *Gender* sont indépendants ?



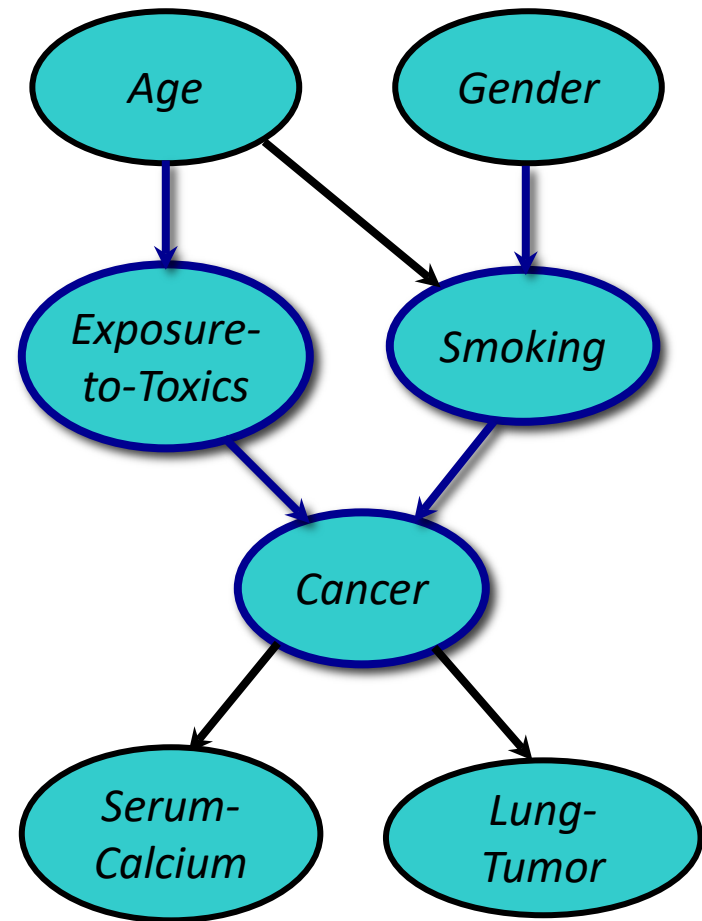
Exemple

- Est-ce que *Age* et *Gender* sont indépendants ?
- ◆ chemin 1 est bloqué au niveau de *Smoking* → (N) ←
 - » *Smoking* et ses descendants *Cancer*, *Serum-Calcium* et *Lung-Tumor* ne sont pas observés



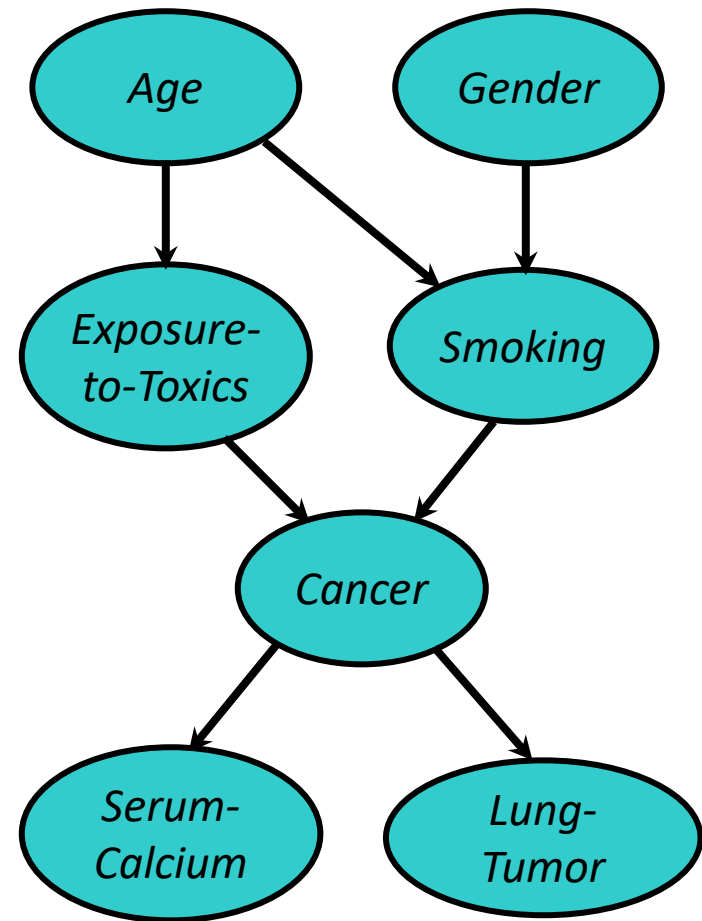
Exemple

- Est-ce que *Age* et *Gender* sont indépendants ?
 - ◆ chemin 1 est bloqué au niveau de *Smoking* → (N) ←
 - » *Smoking* et ses descendants *Cancer*, *Serum-Calcium* et *Lung-Tumor* ne sont pas observés
 - ◆ chemin 2 est aussi *Cancer*
 - » même raisons → (N) ←
- Réponse : **oui**



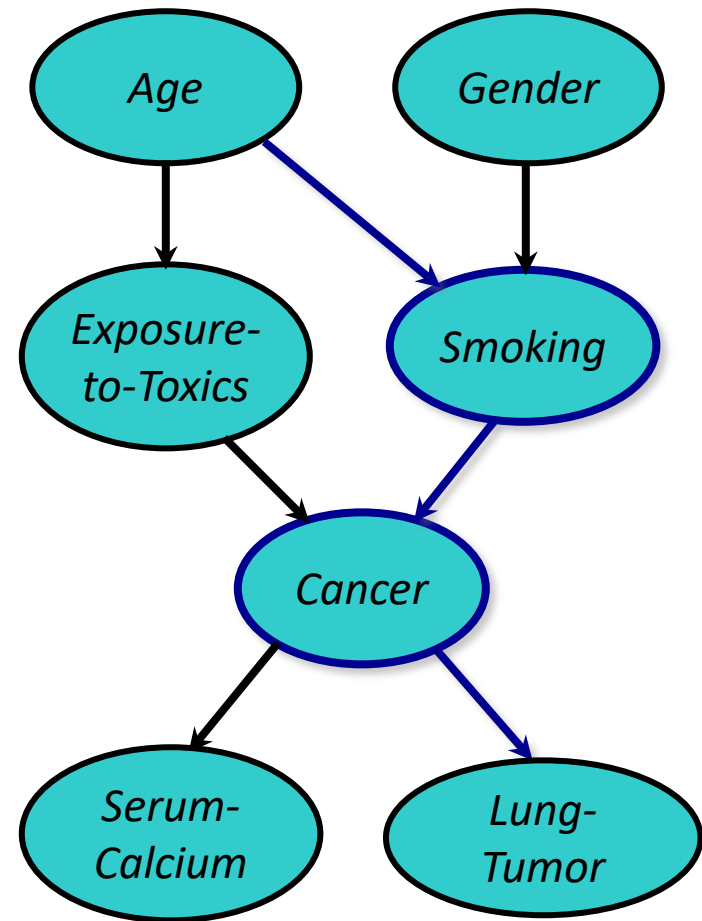
Exemple

- Est-ce que *Age* et *Lung-Tumor* sont indépendants sachant *Smoking* ?



Exemple

- Est-ce que *Age* et *Lung-Tumor* sont indépendants sachant *Smoking* ?
- ◆ chemin 1 est bloqué au niveau de *Smoking* → (N) →
» *Smoking* est observé



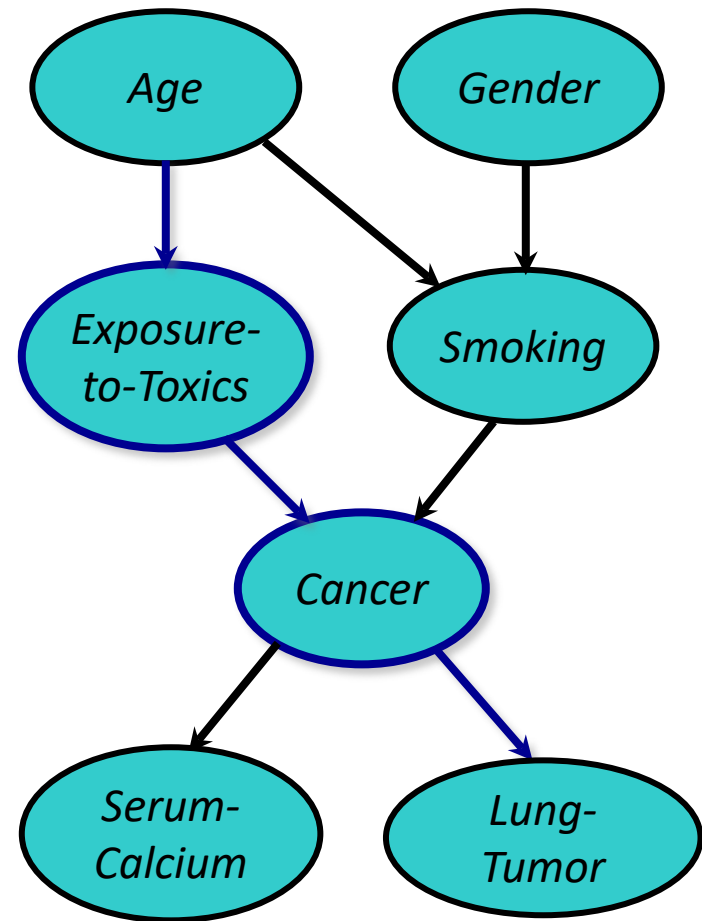
Exemple

- Est-ce que *Age* et *Lung-Tumor* sont indépendants sachant *Smoking* ?

- ◆ chemin 1 est bloqué au niveau de *Smoking* → (N) →
» *Smoking* est observé

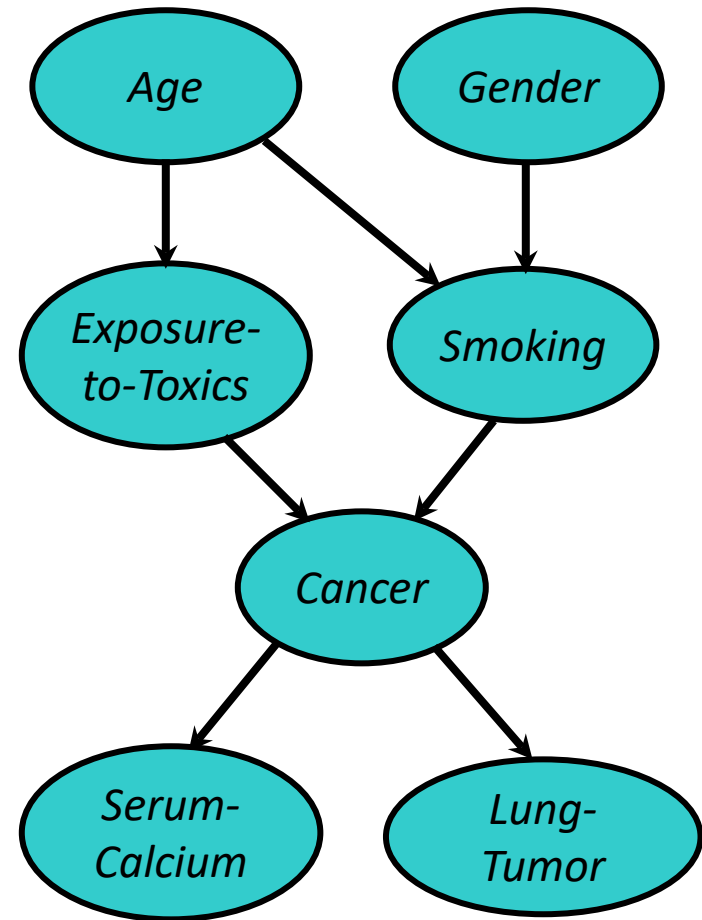
- ◆ chemin 2 n'est pas bloqué
» *Exposure-to-Toxics* → (N) → n'est pas observé
» *Cancer* → (N) → n'est pas observé

- Réponse : **non**



Exemple

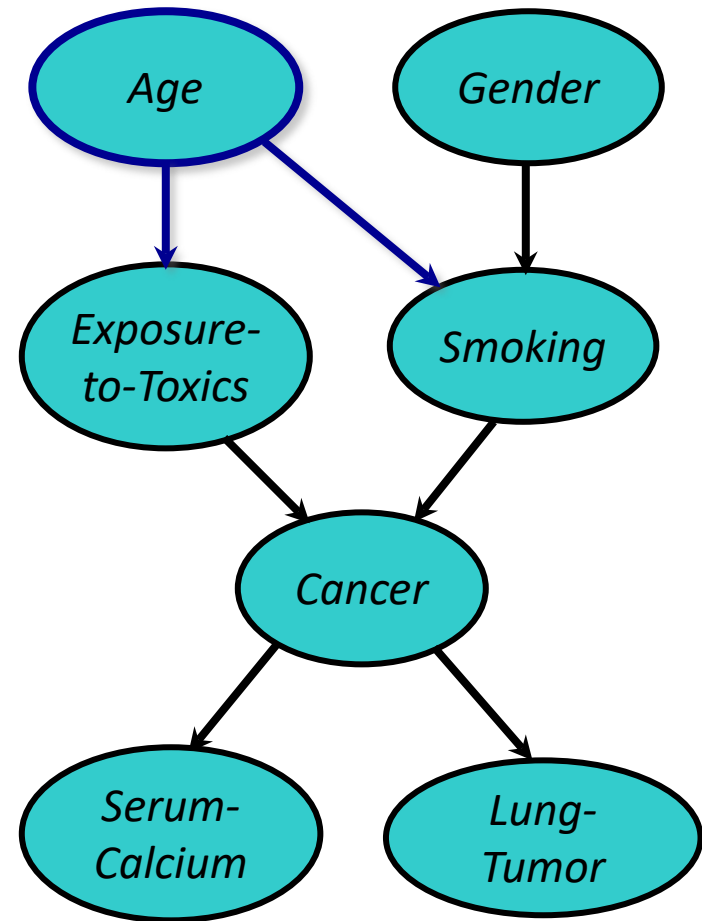
- Est-ce que *Exposure-to-Toxics* et *Smoking* sont indépendants sachant *Age* et *Lung-Tumor*?



Exemple

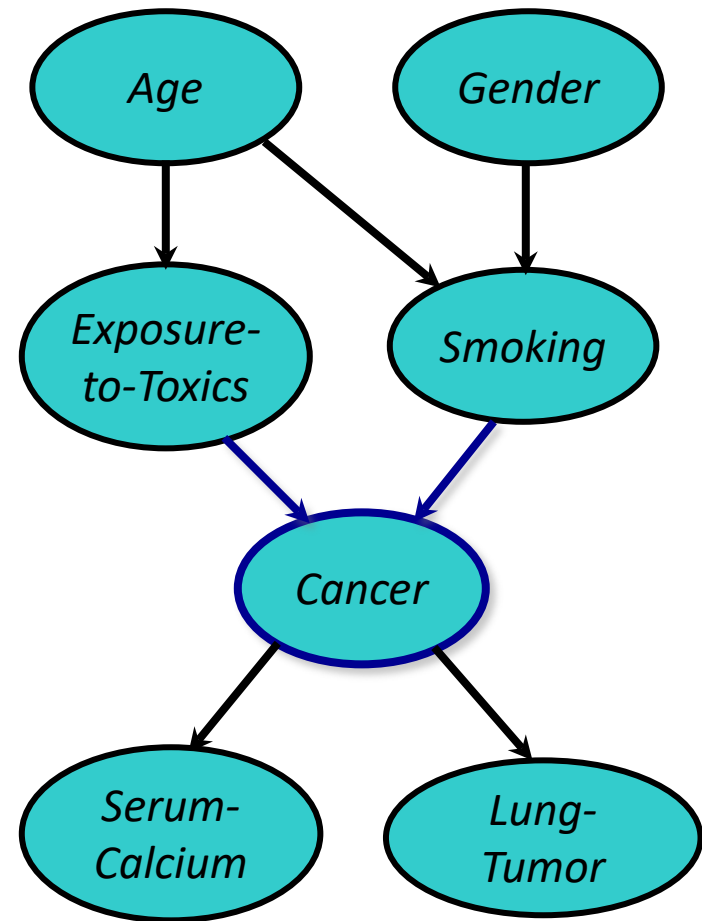
- Est-ce que *Exposure-to-Toxics* et *Smoking* sont indépendants sachant *Age* et *Lung-Tumor*?

- ◆ chemin 1 est bloqué au niveau de *Age* ← (N) →
» *Age* est observé



Exemple

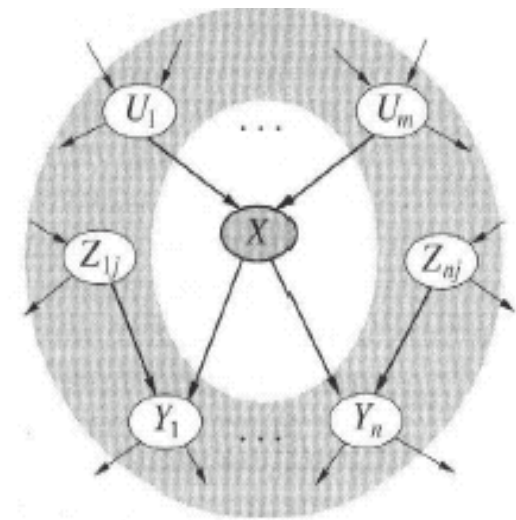
- Est-ce que *Exposure-to-Toxics* et *Smoking* sont indépendants sachant *Age* et *Lung-Tumor*?
 - ◆ chemin 1 est bloqué au niveau de *Age* $\leftarrow (N) \rightarrow$
 - » *Age* est observé
 - ◆ chemin 2 n'est pas bloqué
 - » $Cancer \rightarrow (N) \leftarrow$
ne bloque pas le chemin puisque *Lung-Tumor*, un de ses descendants, est observé
- Réponse : **non**



Indépendance conditionnelle dans un RB : Couverture de Markov

- Soit la **couverture de Markov** (**Markov blanket**) $MB(X)$ d'un nœud X , c'est à dire :
 - ◆ les parents de X
 - ◆ les enfants de X
 - ◆ et les parents des enfants de X
- Le nœud X est conditionnellement indépendant des autres nœuds (hors de la couverture de Markov), étant donné les nœuds de la couverture de Markov :
$$P(X|MB(X), Others) = P(X|MB(X))$$
- La **couverture de Markov** est décrite dans le manuel du cours mais elle **est moins générale que la D-séparation**.

Couverture de Markov de X

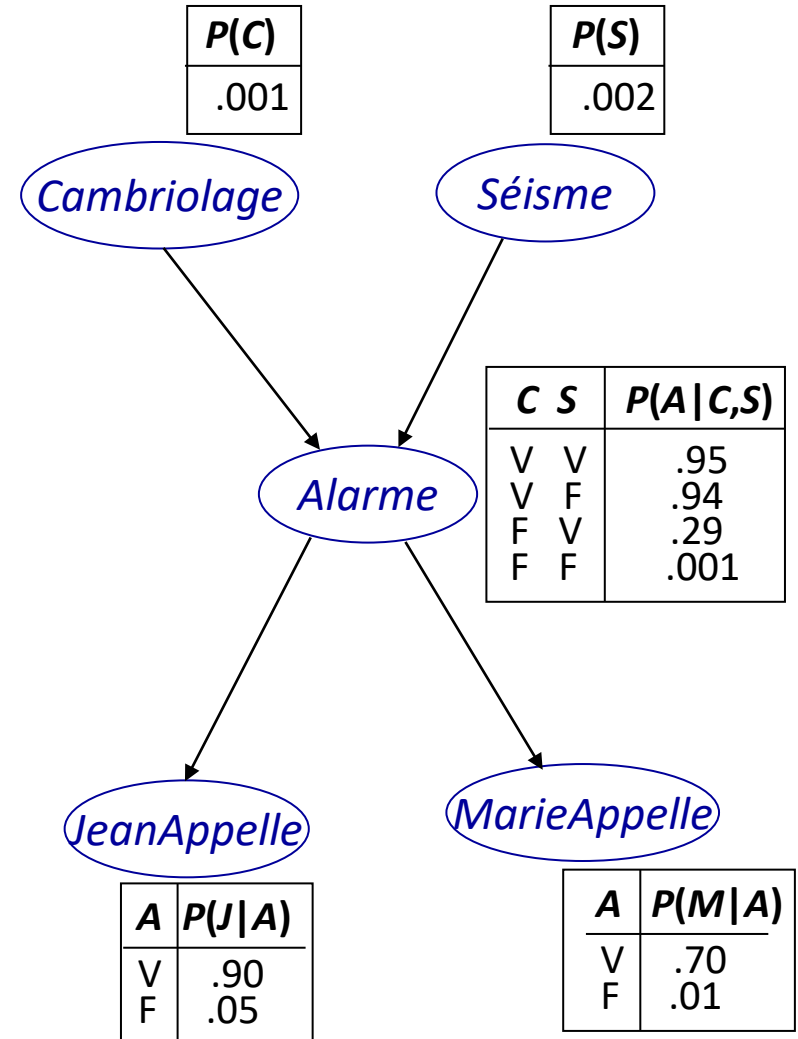


Requête dans un RB

- L'usage principal d'un RB est de calculer les probabilités à posteriori, étant donné un événement observé
 - ◆ un événement est une assignation de valeurs à certaines variables d'observation
 - ◆ ex. : sachant le résultat d'une batterie de test, quelle est maintenant la probabilité qu'un patient ait une maladie X ?
- On va noter
 - ◆ X l'ensemble de variables pour lesquelles on fait une requête
 - » ex. : la patient a la maladie X
 - ◆ E l'ensemble des variables d'observation et e les valeurs observées
 - » ex. : $E_i = e_i$ est le résultat d'un test
 - ◆ Y l'ensemble des variables cachées (qui ne sont pas observées)
 - » ex. : Y_i est le résultat de tests qui n'ont pas été faits
- Une **requête** est l'inférence de $\mathbf{P}(X|e)$, où e est une assignation de valeurs aux variables dans E

Types d'interrogations d'un RB

- **Diagnostic** (on connaît les effets, on cherche les causes)
 - ◆ $P(\text{Cambriolage} | \text{JeanAppelle}=\text{vrai})$
 - ◆ garder à l'esprit qu'on a des arcs « causes / effets ».
- **Prédiction** (étant données les causes, quels sont les effets)
 - ◆ $P(\text{JeanAppelle} | \text{Cambriolage}=\text{vrai})$
- **Probabilité conjointe ou marginale**
 - ◆ $P(\text{Alarme})$



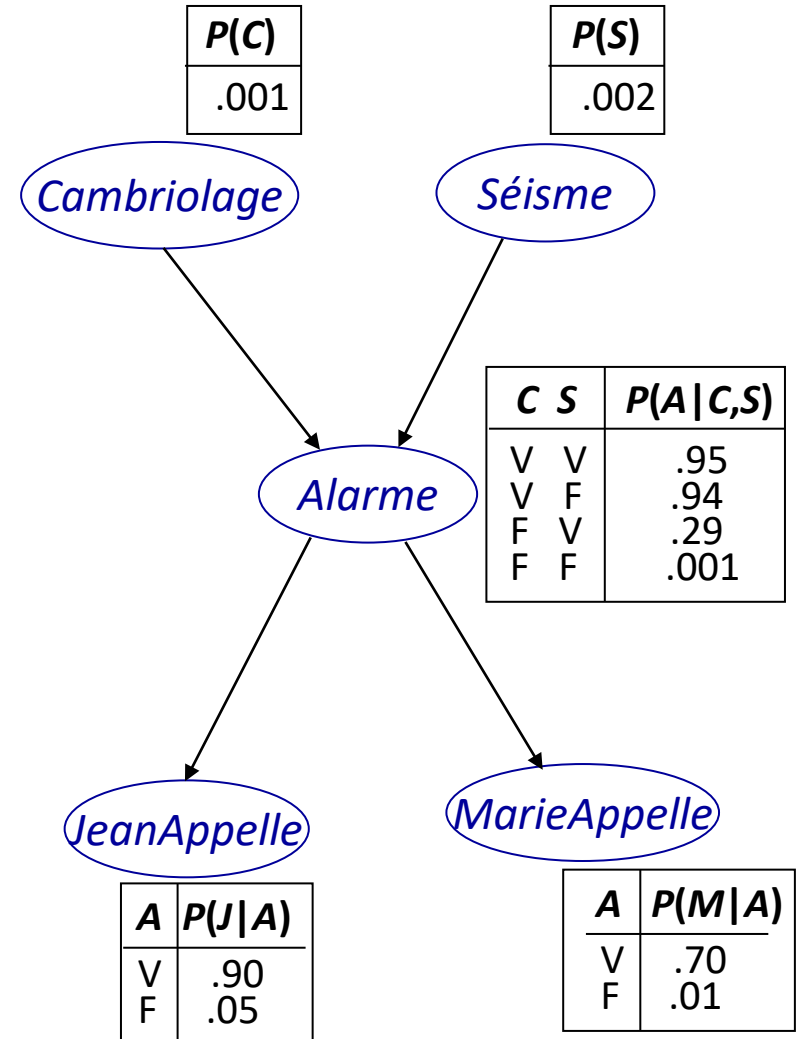
Requête dans un RB

- Exemple :

$P(\text{Cambriolage} \mid \text{JeanAppelle} = \text{vrai}, \text{MarieAppelle} = \text{vrai})$
 $= [0.284, 0.716]$

- Comment fait-on un tel calcul?

- ◆ **inférence exacte** (prohibitif)
 - » par énumération
- ◆ **inférence approximative par échantillonnage** avec les méthodes Monte-Carlo (plus efficace)
 - » méthode de rejet



Inférence par énumération

- On veut calculer la distribution sur les variables de requêtes **sachant** les observations

$$\mathbf{P}(X|e) = \alpha \mathbf{P}(X,E=e) = \alpha \sum_y \mathbf{P}(X, e, y)$$

- Les termes $P(X, e, y)$ peuvent s'écrire comme le produit des probabilités conditionnelles du réseau
- On peut donc calculer la réponse à une requête $P(X|e)$ dans un RB, simplement en
 1. calculant les sommes des produits des probabilités conditionnelles du RB
 2. normalisant ces sommes de façon à obtenir une distribution qui somme à 1
- Les ensembles des variables X , E et Y couvrent ensemble tous les noeuds
 - ◆ complexité en temps : $O(d^{|X|+|Y|})$, avec d la taille du plus grand domaine
 - ◆ complexité en espace : $O(d^{|X|})$, pour stocker la distribution

Exemple 1

- $P(\text{Cambriolage} \mid \text{JeanAppelle} = \text{vrai}, \text{MarieAppelle} = \text{vrai})$

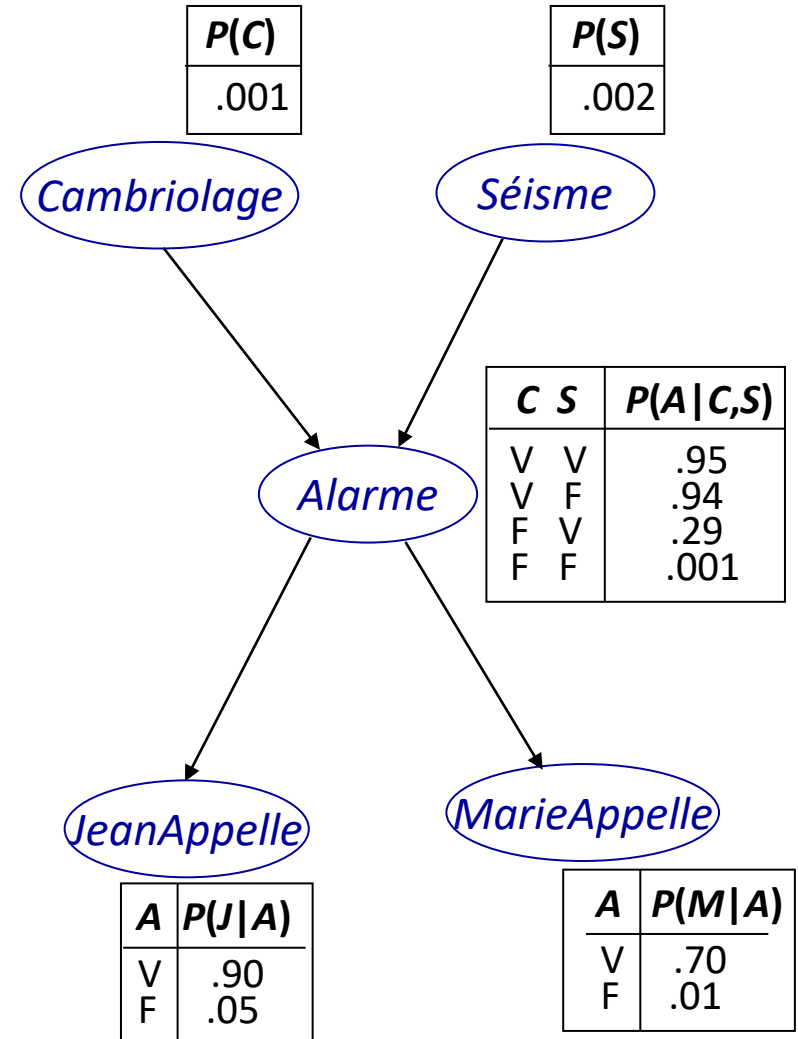
◆ noté $P(C \mid j, m)$

- Les variables cachées sont *Séisme* et *Alarme*

$$\begin{aligned}
 P(C \mid j, m) &= \alpha P(C, j, m) \\
 &= \alpha \sum_s \sum_a P(C, s, a, j, m) \\
 &= P(C) \sum_s \sum_a P(s) P(a \mid C, s) P(j \mid a) P(m \mid a) \\
 &= P(C) \sum_s P(s) \sum_a P(a \mid C, s) P(j \mid a) P(m \mid a)
 \end{aligned}$$

Note :

- ◆ s et a prennent toutes les valeurs possibles pour $S=s$ et $A=a$
- ◆ ne pas confondre avec j et m qui sont des observations fixes ($J=\text{vrai}$ et $M=\text{vrai}$)

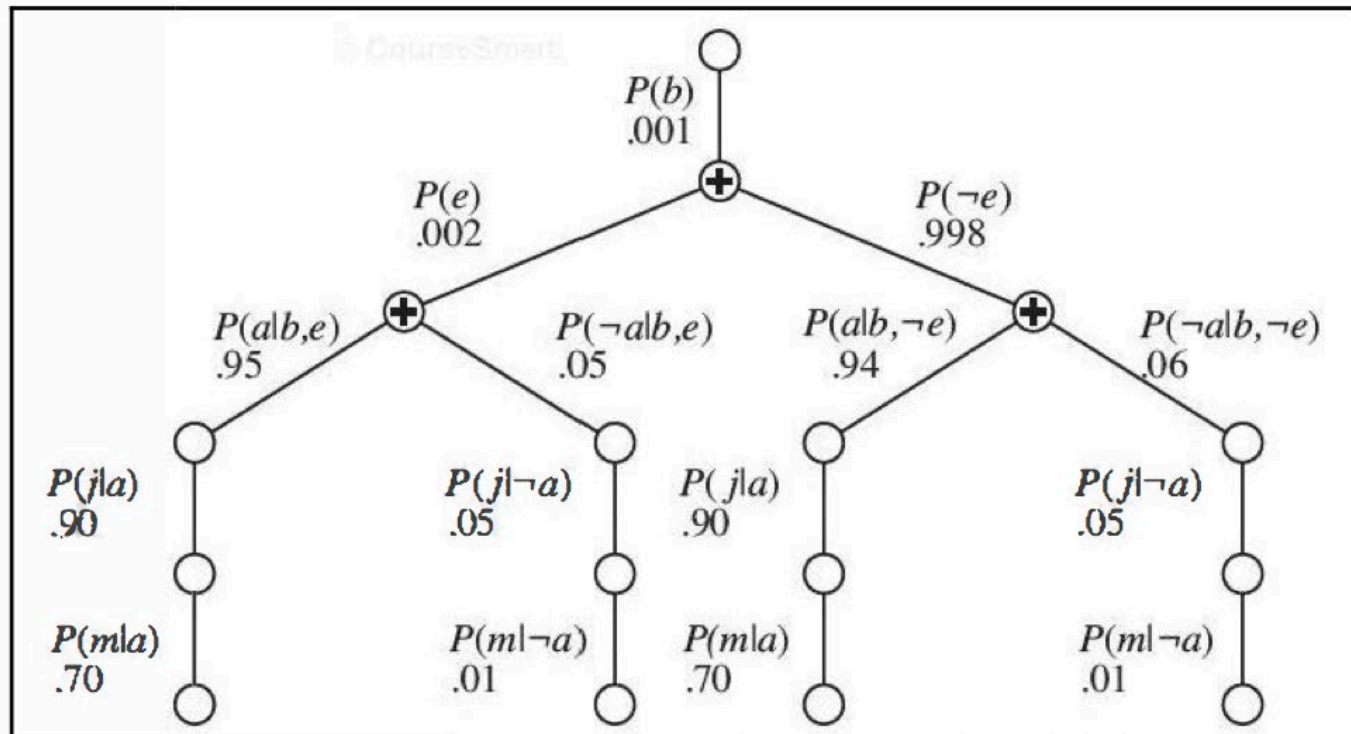


Exemple 1 (suite)

- Structure de l'expression représentée par l'équation

$$P(c \mid j, m) = \alpha * P(c) \sum_s P(s) \sum_a P(a \mid c, s) P(j \mid a) P(m \mid a)$$

$$P(b \mid j, m) = \alpha * P(b) \sum_s P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a) \text{ - En anglais}$$



Exemple 1 (suite)

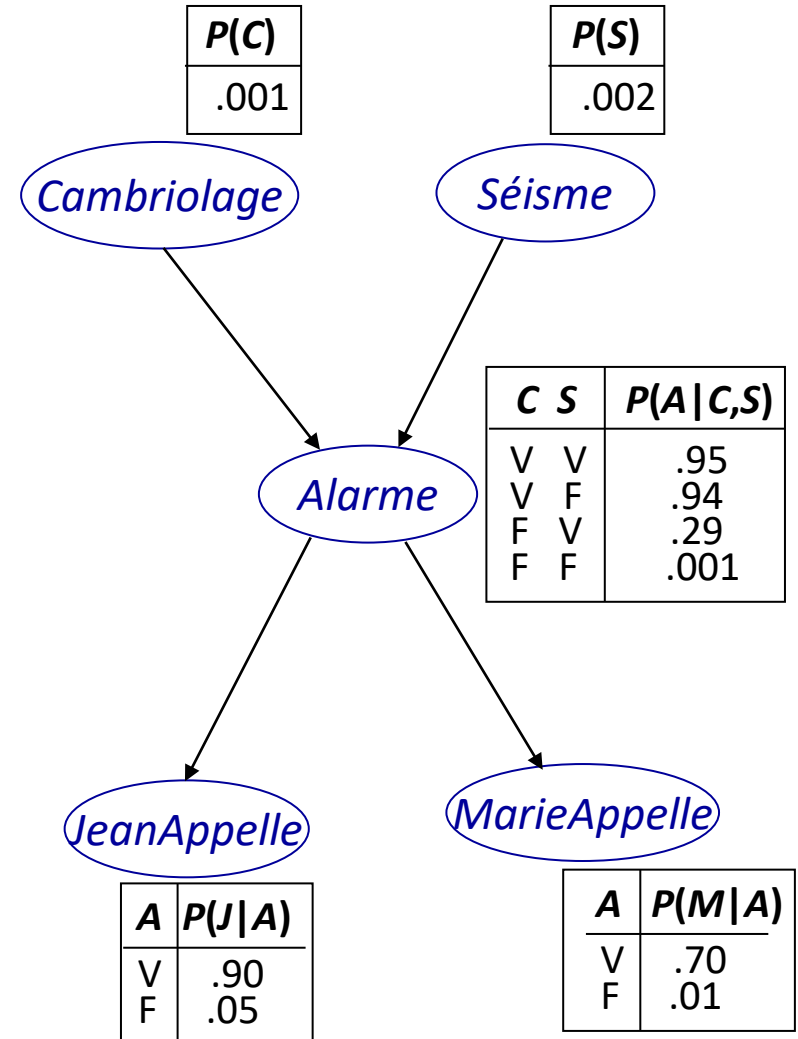
- On calcule pour $C = \text{vrai}$

$$\begin{aligned}
 P(c \mid j, m) &= \alpha P(c) \sum_s P(s) \sum_a P(a \mid c, s) P(j \mid a) P(m \mid a) \\
 &= \alpha * 0.001 * (0.002 * (0.95 * 0.90 * 0.70 + \\
 &\quad 0.05 * 0.05 * 0.01) + \\
 &\quad 0.998 * (0.94 * 0.90 * 0.70 + \\
 &\quad 0.06 * 0.05 * 0.01)) \\
 &= \alpha * 0.00059224
 \end{aligned}$$

- Et $C = \text{faux}$

$$\begin{aligned}
 P(\neg c \mid j, m) &= \alpha P(\neg c) \sum_s P(s) \sum_a P(a \mid \neg c, s) P(j \mid a) P(m \mid a) \\
 &= \alpha * 0.0014919 \\
 \alpha &= 1 / (0.00059224 + 0.0014919)
 \end{aligned}$$

- Donc, $\mathbf{P}(C \mid j, m) = [0.284, 0.716]$



Exemple 2 : Évaluation par énumération

Requête :

Calculer $P(T=vrai | F=faux, M=vrai)$

Variables connues :

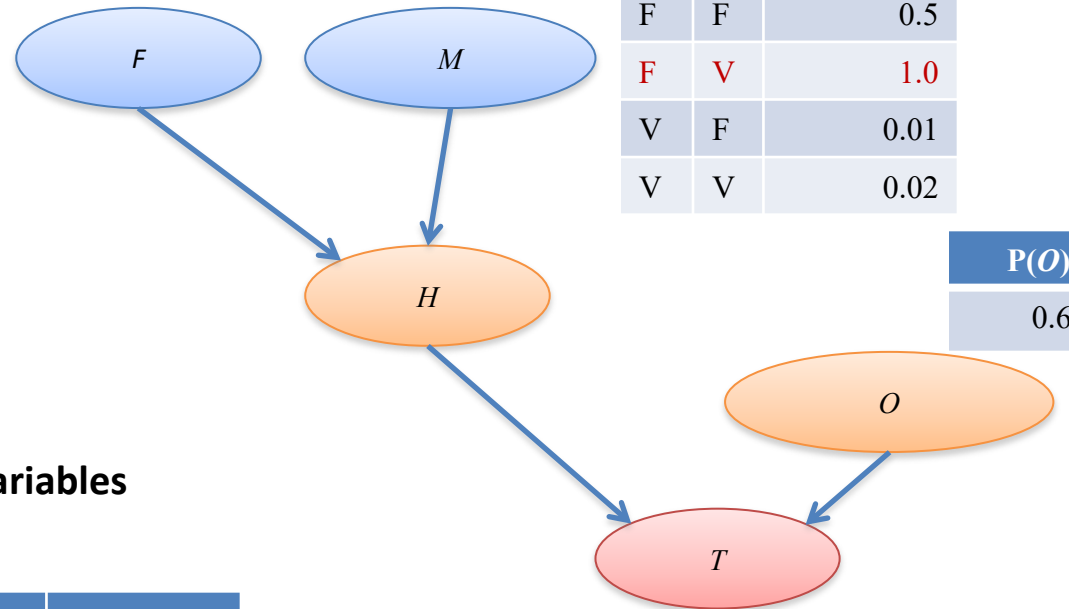
$F = faux$

$M = vrai$

Variables inconnues :

H

O



F	M	$P(H F,M)$
F	F	0.5
F	V	1.0
V	F	0.01
V	V	0.02

$P(O)$
0.6

Énumération des valeurs possibles des variables cachées (2*2)

H	O	$P(H F,M) * P(O) * P(T H,O)$	=
F	F	$0.0 * 0.4 * 0.1$	0
F	V	$0.0 * 0.6 * 0.5$	0
V	F	$1.0 * 0.4 * 0.5$	0.20
V	V	$1.0 * 0.6 * 1.0$	0.60
TOTAL			0.80

H	O	$P(T H,O)$
F	F	0.1
F	V	0.5
V	F	0.5
V	V	1.0

Exemple 3 : Évaluation par énumération

Requête :

Calculer $P(T=vrai | M=vrai)$

Variables connues :

$M = vrai$

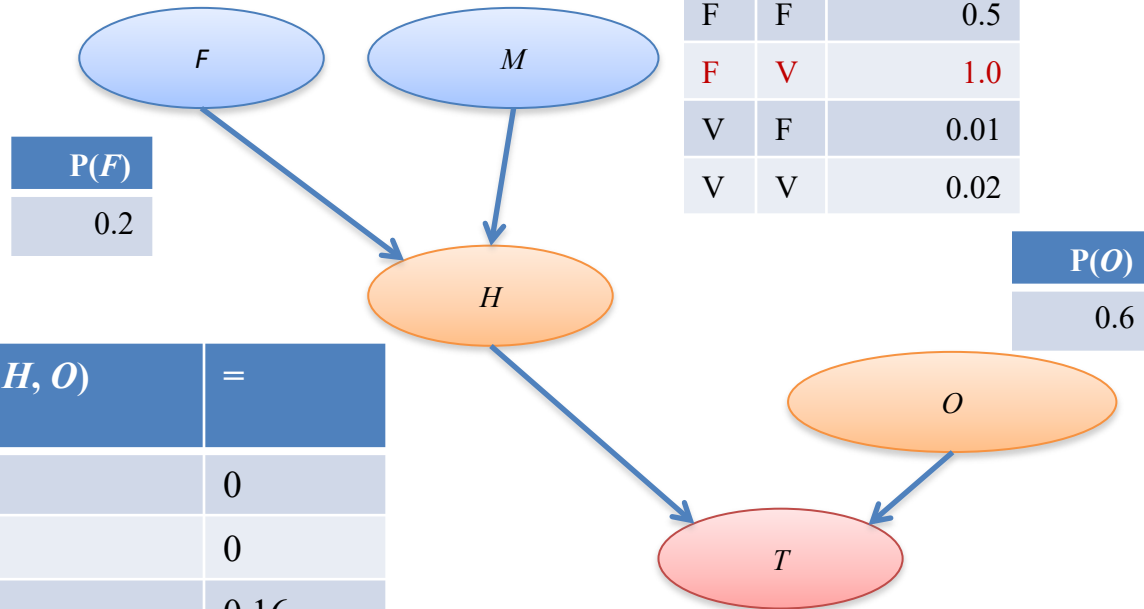
Variables inconnues :

H

O

F

F	H	O	$P(F) * P(H F,M) * P(O) * P(T H, O)$	=
F	F	F	$0.8 * 0.0 * 0.4 * 0.1$	0
F	F	V	$0.8 * 0.0 * 0.6 * 0.5$	0
F	V	F	$0.8 * 1.0 * 0.4 * 0.5$	0.16
F	V	V	$0.8 * 1.0 * 0.6 * 1.0$	0.48
V	F	F	$0.2 * 0.98 * 0.4 * 0.1$	0.00784
V	F	V	$0.2 * 0.98 * 0.6 * 0.5$	0.0588
V	V	F	$0.2 * 0.02 * 0.4 * 0.5$	0.0008
V	V	V	$0.2 * 0.02 * 0.6 * 1.0$	0.0024
			TOTAL	0.71



Exercice

- Soit le RB ayant les tables de probabilités conditionnelles suivantes

◆ Dessinez le graphe du RB.

◆ Calculez $P(A=\text{faux} \mid E=\text{vrai})$.

◆ Dites si B et E sont indépendants sachant F . Pourquoi?

◆ Dites si E et F sont indépendants sachant A et C . Pourquoi?

A	C	$F=\text{vrai}$
<i>faux</i>	<i>faux</i>	0.1
<i>faux</i>	<i>vrai</i>	0.2
<i>vrai</i>	<i>faux</i>	0.8
<i>vrai</i>	<i>vrai</i>	0.7

E	$C=\text{vrai}$
<i>faux</i>	0.2
<i>vrai</i>	0.4

$D=\text{vrai}$
0.2

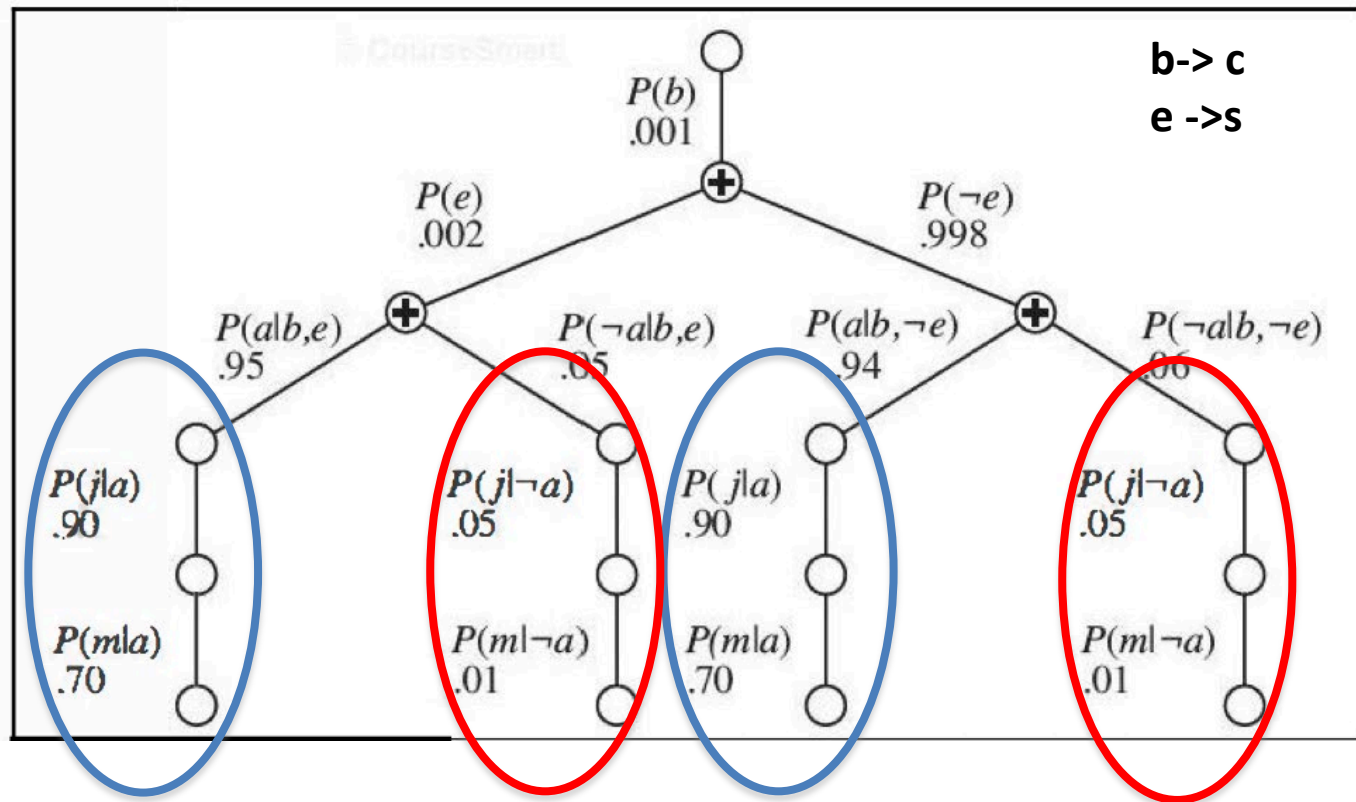
B	D	E	$A=\text{vrai}$
<i>faux</i>	<i>faux</i>	<i>faux</i>	0.7
<i>faux</i>	<i>faux</i>	<i>vrai</i>	0.2
<i>faux</i>	<i>vrai</i>	<i>faux</i>	0.5
<i>faux</i>	<i>vrai</i>	<i>vrai</i>	0.1
<i>vrai</i>	<i>faux</i>	<i>faux</i>	0.2
<i>vrai</i>	<i>faux</i>	<i>vrai</i>	0.9
<i>vrai</i>	<i>vrai</i>	<i>faux</i>	0.8
<i>vrai</i>	<i>vrai</i>	<i>vrai</i>	0.6

$E=\text{vrai}$
0.9

$B=\text{vrai}$
0.7

Inférence par élimination des variables

- Même principe que l'inférence par énumération, mais on évite les répétitions de calculs déjà faits (comme en programmation dynamique)
- Voir section 14.4.2 du livre: $P(b \mid j, m) = \alpha * P(b) \sum_s P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a)$



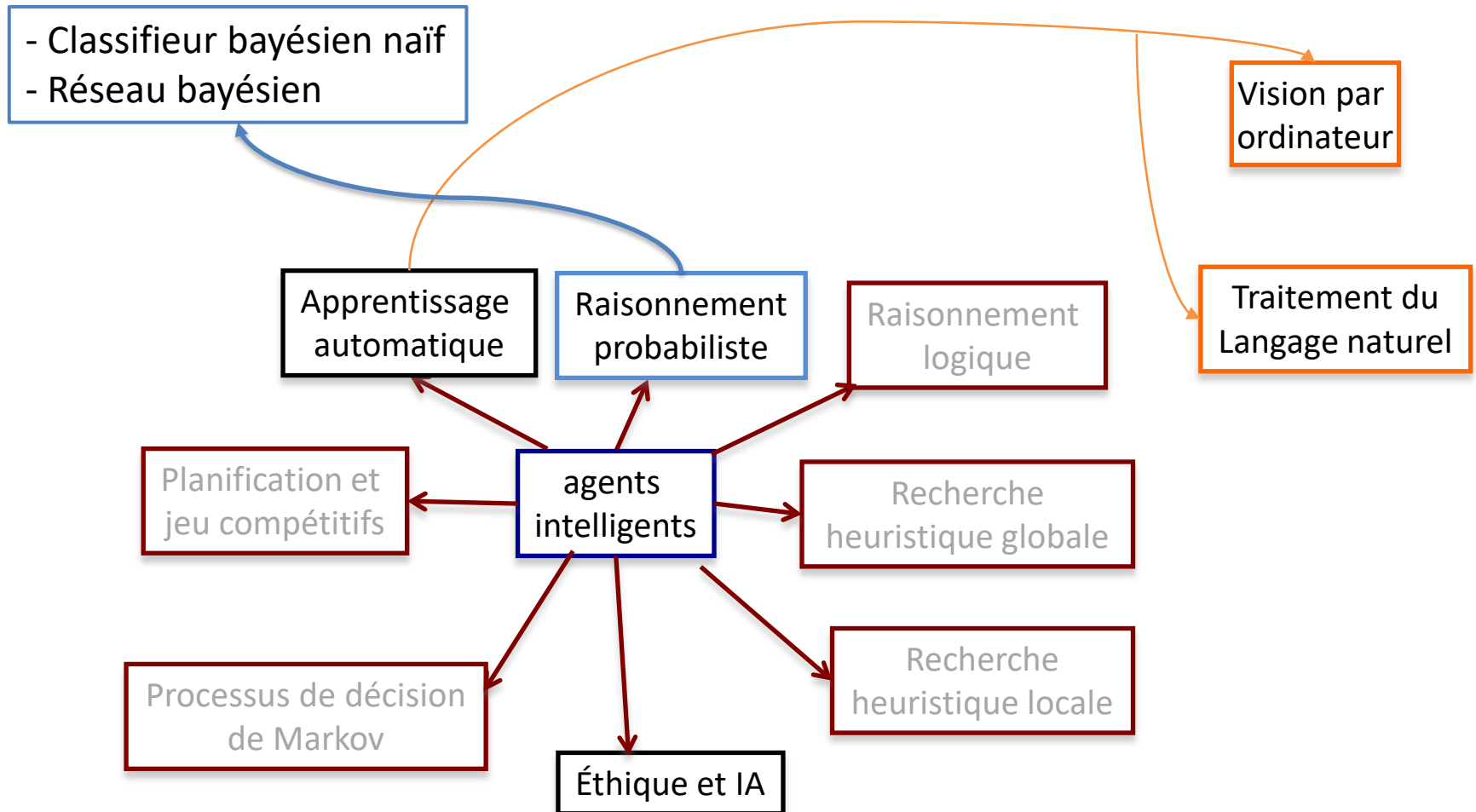
Résumé

- Un RB est un graphe orienté, acyclique, représentant des connaissances causales, et reflétant les dépendances conditionnelles entre des variables
- La topologie du réseau (arcs entre les variables) et les TPC donnent une représentation compacte de la distribution conjointe des probabilités
- Les connaissances du réseau (liens de causalité et probabilités) sont généralement obtenus avec l'aide d'un expert
 - ◆ pour des applications concrètes, ceci peut être très laborieux

Sujets couverts par le cours

Concepts et algorithmes

Applications



Vous devriez être capable de...

- Décrire ce qu'est un réseau bayésien :
 - ◆ qu'est-ce que la topologie représente
 - ◆ quelle est la distribution conjointe associée à un réseau bayésien
- Étant donné un réseau bayésien :
 - ◆ calculer une probabilité conjointe, marginale, conditionnelle
 - ◆ dire si deux variables sont (conditionnellement) indépendantes
- Décrire l'inférence par énumération exacte
- Savoir qu'en pratique c'est l'inférence approximative qui est utilisée parce que plus efficace.
 - ◆ Pas besoin de connaître les méthodes pour l'examen. Savoir simplement qu'il en existe.

Inférence approximative

- Les méthodes d'inférence exactes sont inefficaces
 - ◆ le problème d'inférence dans un RB est NP-Complet
- Les méthodes d'inférence approximatives sont plus pratiques
 - ◆ en général, on n'a pas besoin d'un calcul exact des probabilités pour qu'une conclusion tirée d'un RB soit correcte
 - ◆ les méthodes approximatives assignent des valeurs aux variables aléatoires en fonction des TPC associées à ces variables
 - ◆ ces assignations sont basées sur des simulations stochastiques, plutôt que des observations réelles

Méthode de rejet (*rejection sampling*)

- Pour estimer $P(X=x|e)$
 - ◆ Générer des échantillons complets à partir de la distribution spécifiée par le RB
 - ◆ Rejeter tous les échantillons qui ne correspondent pas à l'observation e .
 - ◆ Estimer $P(X=x|e)$ en comptant combien de fois $X=x$ se produit dans les échantillons restant.
- Autrement dit:
$$P(X=x|e) = \alpha \sum_y P(X=x, e, y) \approx \text{freq}(x,e) / \sum_{x'} \text{freq}(x',e) = \text{freq}(x,e) / \text{freq}(e)$$
- Cette technique est appelée **méthode de rejet** (*rejection sampling*)
 - ◆ le problème avec cette méthode est que si $E=e$ est très rare selon le RB, il y aura peu d'échantillons qui correspondront à cette observation
 - ◆ d'autres méthodes sont plus efficaces et nécessitent moins d'échantillons pour obtenir une bonne estimation
 - ◆ voir la section 14.5 dans le livre

Exemple



Requête :

Calculer $P(T=vrai | M=vrai)$

Variables connues :

$M = vrai$

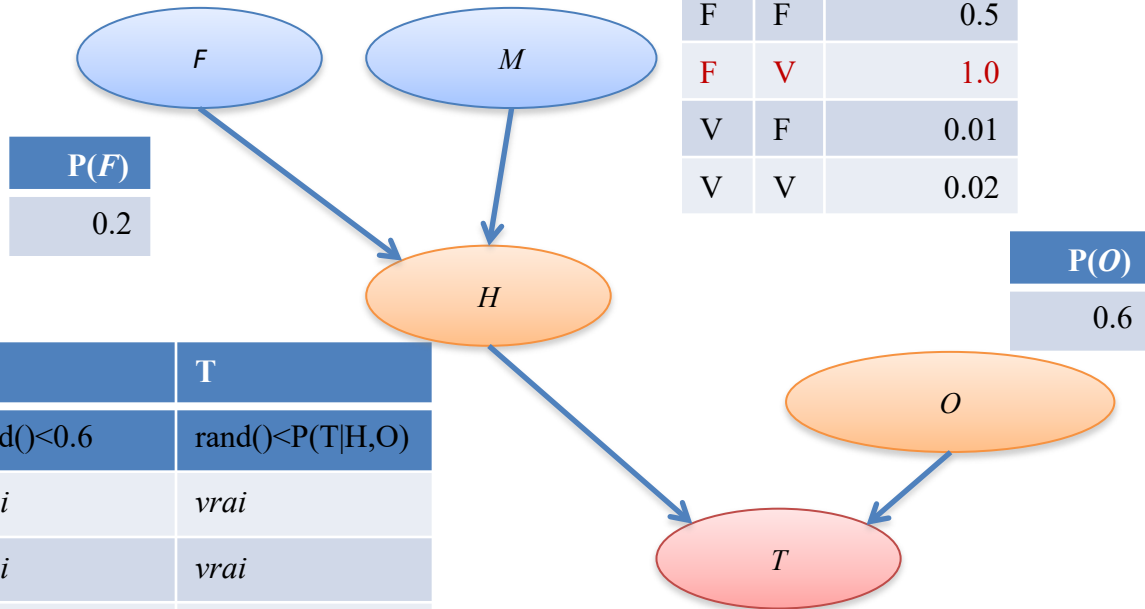
Variables inconnues :

H

O

T

	F	H	O	T
#	rand() <0.2	rand() $<P(H F,M)$	rand() <0.6	rand() $<P(T H,O)$
1	<i>faux</i>	<i>vrai</i>	<i>vrai</i>	<i>vrai</i>
2	<i>faux</i>	<i>vrai</i>	<i>vrai</i>	<i>vrai</i>
3	<i>faux</i>	<i>vrai</i>	<i>faux</i>	<i>faux</i>
4	<i>vrai</i>	<i>faux</i>	<i>faux</i>	<i>faux</i>
5	<i>faux</i>	<i>vrai</i>	<i>vrai</i>	<i>vrai</i>
6	<i>faux</i>	<i>vrai</i>	<i>vrai</i>	<i>vrai</i>
7	<i>faux</i>	<i>vrai</i>	<i>vrai</i>	<i>vrai</i>
8	<i>faux</i>	<i>vrai</i>	<i>faux</i>	<i>vrai</i>
Average of $T=vrai$				6/8 = 0.75



Plus qu'il y a d'échantillons, plus l'erreur d'estimation est faible.

(Vrai réponse : 0.71)

Construction d'un RB

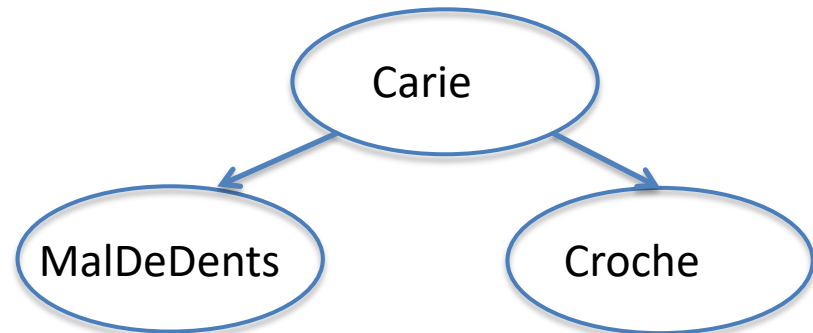
- Comment bâtir un RB afin de modéliser un environnement/problème donné?
- On a besoin de spécifier 2 choses:
 1. La structure (graphe) du réseau (quelles indépendances peut-on supporter)
 2. Les table de probabilités (quelles relations entre les variables de l'environnement?)

Spécifier les tables de probabilités d'un RB

- Si on a un ensemble de données où tous les nœuds X_i sont observés, c'est facile :

$$P(X_i = x \mid \text{Parents}(X_i) = p) \approx \text{freq}(x, p) / \sum_{x'} \text{freq}(x', p)$$

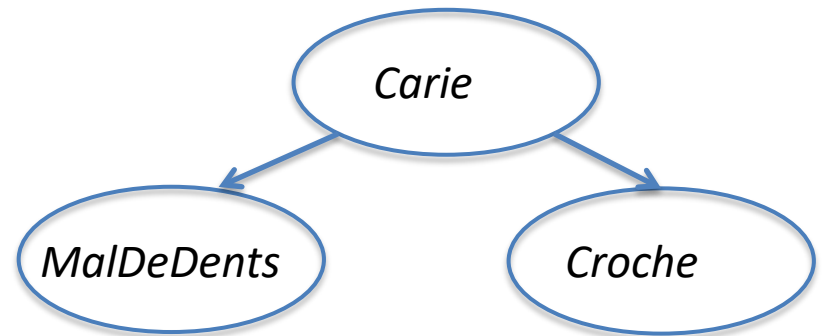
- On fait ce calcul pour toutes les valeurs x de X_i et toutes les valeurs p de ses parents possibles
 - ◆ pour éviter d'avoir de probabilités à 0, on peut initialiser $\text{freq}(x, p) = \delta$ à une petite constante δ (ex. : $\delta=1$)



Exemple

- Supposons que l'on souhaite diagnostiquer une carie à l'aide du RB suivant:
 - ◆ **MalDeDents**: le patient a mal aux dents
 - ◆ **Croche**: la sonde s'accroche dans les dents
 - ◆ **Carie**: le patient a une carie
- Supposons qu'on collecte un ensemble de données sur 120 patients où
 - ◆ 70 des 120 patients avaient une carie

$$P(\text{Carie}=\text{vrai})=(70+1)/(70+1+50+1)=0.52$$

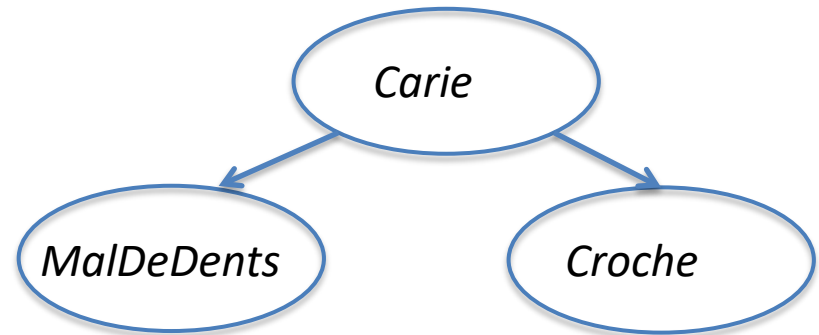


Exemple

- Supposons maintenant que dans mes données de 120 patients,
 - ◆ Parmi les 70 qui avaient une carie
 - » 65 avaient mal aux dents
 - » 51 ont la sonde qui s'accrochait

$$P(\text{MalDeDents}=\text{vrai} \mid \text{Carie}=\text{Vrai}) = (65+1)/(65+1+5+1) = 0.92$$

$$P(\text{Croche}=\text{vrai} \mid \text{Carie}=\text{vrai}) = (51+1)/(51 + 1 + 19 + 1) = 0.72$$

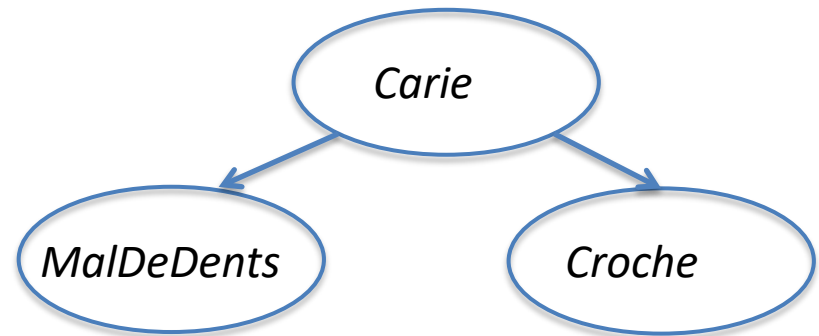


Exemple

- Supposons que dans mes données de 120 patients,
 - ◆ Parmi les 50 qui n'avaient pas de carie
 - » 10 avaient mal aux dents
 - » La sonde ne s'accrochait pour aucun d'eux

$$P(\text{MalDeDents}=\text{vrai} \mid \text{Carie}=\text{faux}) = (10+1)/(10+1+42+1) = 0.20$$

$$P(\text{Croche}=\text{vrai} \mid \text{Carie}=\text{faux}) = (0+1)/(0 + 1 + 50 + 1) = 0.02$$



Spécifier les tables de probabilités d'un RB

- Si on a un ensemble de données où certaines des variables ne sont pas observées, on doit utiliser des méthodes plus sophistiquées.
 - ◆ Algorithme EM (voir section 20.3.2)

Construction de la structure du RB

1. Choisir un ordre des variables X_1, \dots, X_n
 2. Pour $i = 1$ to n :
 - ◆ ajouter X_i au réseau
 - ◆ choisir les parents X_1, \dots, X_{i-1} tels que $P(X_i \mid \text{Parents}(X_i)) = P(X_i \mid X_1, \dots, X_{i-1})$
 - ◆ ce choix garantit que :
$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid X_1, \dots, X_{i-1}) \quad (\text{chain rule})$$
$$= \prod_{i=1}^n P(X_i \mid \text{Parents}(X_i)) \quad (\text{par construction})$$
- Pour construire un bon RB, sa structure doit refléter les indépendances conditionnelles du problème
 - Dans quel ordre ajouter les nœuds au réseau?
 - ◆ mettre les « causes racines » d'abord, ensuite les nœuds qu'ils influencent directement

Exemple

- Supposons qu'on ordonne les variables comme suit : M, J, A, C, S

MarieAppelle

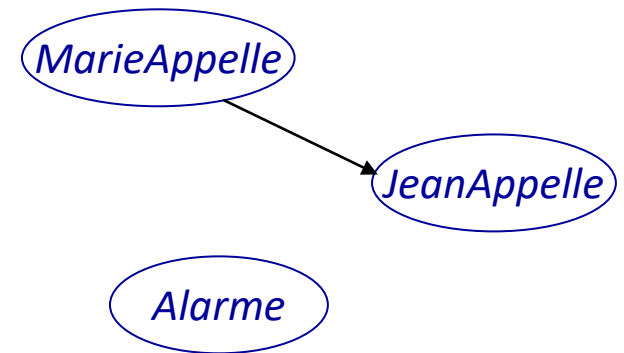
JeanAppelle

- $P(J|M) = P(J)$?
-

Exemple

- Supposons qu'on ordonne les variables comme suit : M, J, A, C, S

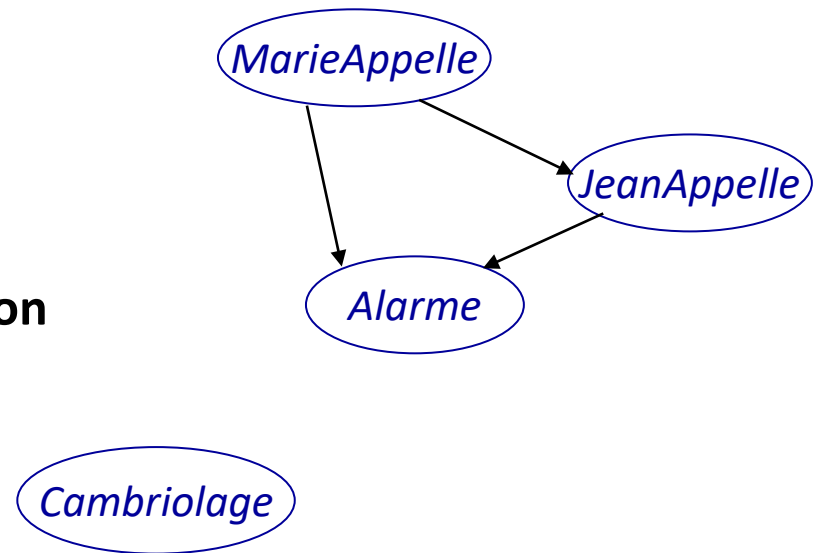
- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$?



Exemple

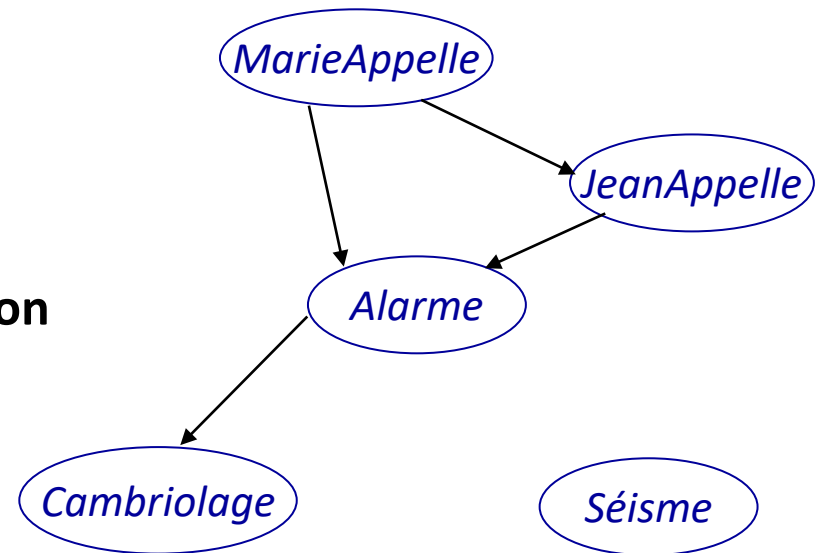
- Supposons qu'on ordonne les variables comme suit : M, J, A, C, S

- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? **Non**
- $P(C|A,J,M) = P(C|A)$?
- $P(C|A,J,M) = P(C)$?



Exemple

- Supposons qu'on ordonne les variables comme suit : M, J, A, C, S

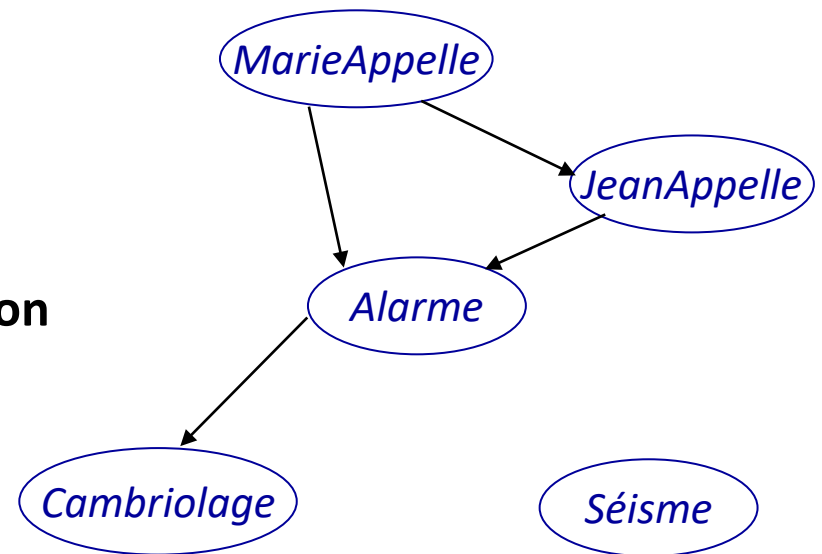


- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? **Non**
- $P(C|A,J,M) = P(C|A)$? **Oui**
- $P(C|A,J,M) = P(C)$?
- $P(S|C,A,J,M) = P(S|A)$?
- $P(S|C,A,J,M) = P(S|A,C)$?

Exemple

- Supposons qu'on ordonne les variables comme suit : M, J, A, C, S

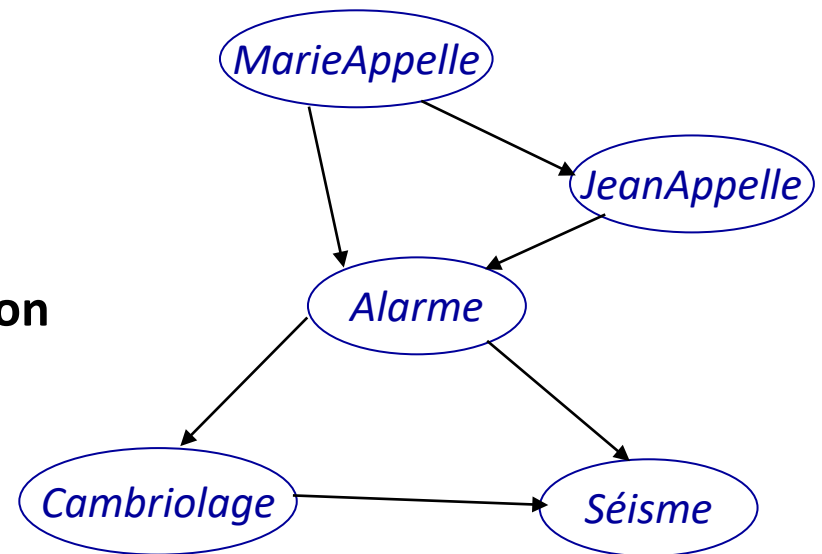
- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? **Non**
- $P(C|A,J,M) = P(C|A)$? **Oui**
- $P(C|A,J,M) = P(C)$? **Non**
- $P(S|C,A,J,M) = P(S|A)$?
- $P(S|C,A,J,M) = P(S|A,C)$?



Exemple

- Supposons qu'on ordonne les variables comme suit : M, J, A, C, S

- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? **Non**
- $P(C|A,J,M) = P(C|A)$? **Oui**
- $P(C|A,J,M) = P(C)$? **Non**
- $P(S|C,A,J,M) = P(S|A)$? **Non**
- $P(S|C,A,J,M) = P(S|A,C)$? **Oui**



Exemple

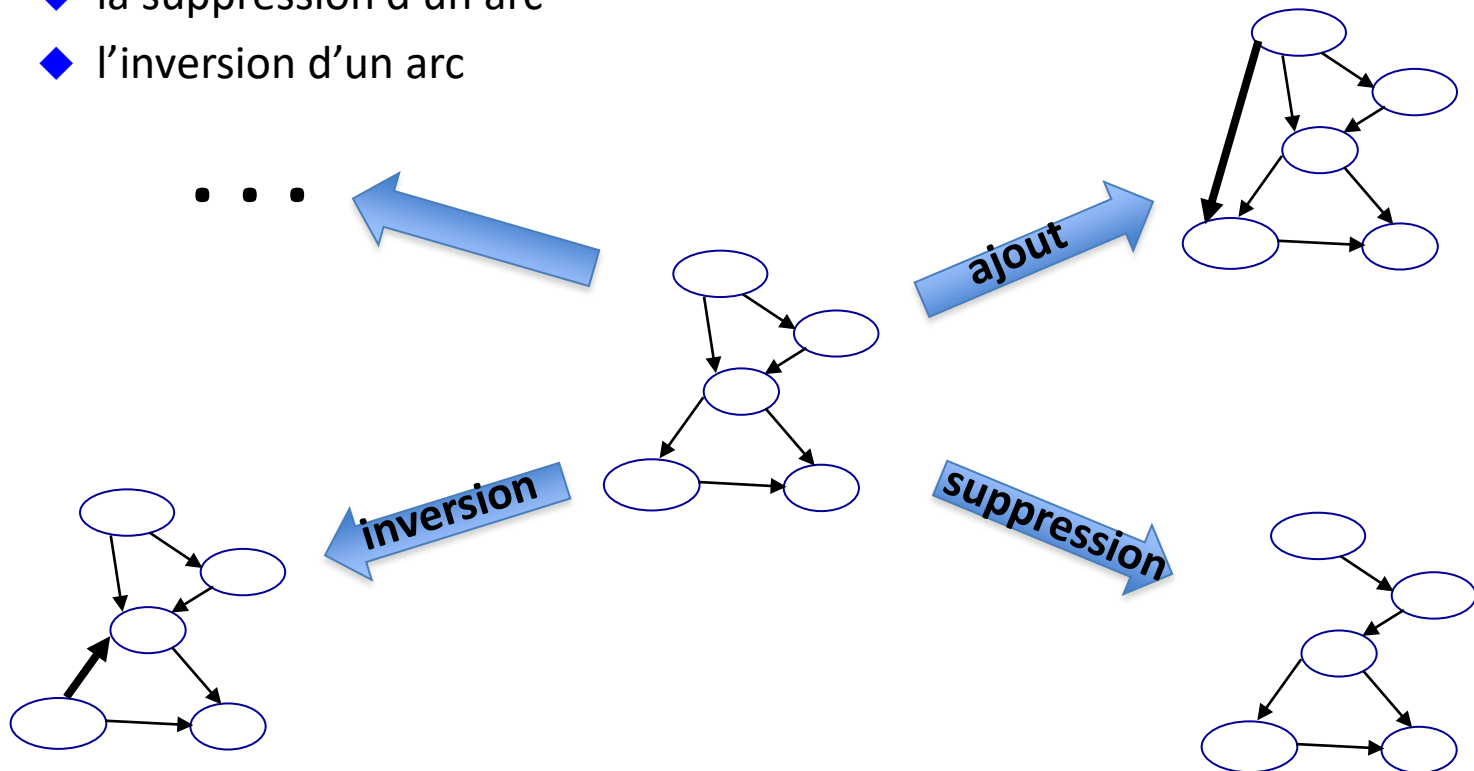
- Déterminer l'indépendance conditionnelle est très difficile dans le sens non causal
 - ◆ par exemple, en médecine, souvent les experts préfèrent donner des probabilités dans le sens causal (pathologie → symptôme) plutôt que dans le sens diagnostique
- Un réseau avec des dépendances diagnostiques (effet → cause) est généralement moins compacte
 - ◆ dans le cas présent : $1 + 2 + 4 + 2 + 4 = 13$ nombres pour représenter les tables de probabilité conditionnelle du réseau au lieu de 10 pour la première version

Génération automatique de la structure du RB

- Quoi faire si on n'a pas accès à un expert pour nous donner un bon graphe de RB ?
- On peut aussi tenter d'obtenir la structure du RB à partir de données, à l'aide de la recherche locale (par exemple *Hill Climbing* que nous verrons plus tard dans le cours) :
 1. on débute avec un **graphe acyclique aléatoire** comme graphe courant
 2. on **obtient ses tables de probabilités** à partir des fréquences d'observation du graphe courant
 3. on utilise la recherche locale pour **générer des graphes successeurs** du graphe courant
 4. on obtient les tables de probabilités du graphe successeur
 5. on remplace le graphe courant par le successeur s'il est « meilleur »
 3. on retourne à 2. jusqu'à un certain critère d'arrêt

Génération automatique de la structure du RB

- On génère des successeurs à partir des modifications au graphe suivantes
 - ◆ l'ajout d'un arc
 - ◆ la suppression d'un arc
 - ◆ l'inversion d'un arc



Génération automatique de la structure du RB

- La fonction objectif à maximiser par la recherche locale est :

$$\underbrace{\sum_t \log P(X_1 = x_1^t, \dots, X_n = x_n^t)}_{\text{log probabilité des données}} - \underbrace{M (\log T) / 2}_{\text{complexité du graphe}}$$

- ◆ $\{x_1^t, \dots, x_n^t\}$ est la $t^{\text{ième}}$ donnée de mon ensemble de T données
- ◆ M est le nombre de paramètres requis par les tables de probabilités conditionnelles du réseau bayésien
- On cherche donc un graphe
 - ◆ qui **explique bien les données** (leur donne une haute probabilité)
 - ◆ qui est **compacte** (qui a peu de paramètres)