IFT 615 – Intelligence artificielle

Réseaux bayésiens

Hugo Larochelle
Département d'informatique
Université de Sherbrooke
http ://www.dmi.usherb.ca/~larocheh/cours/ift615.html

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
- Sujet plus poussés
  - apprentissage automatique de réseaux bayésiens

Réseaux bayésiens

- On a vu les bases du raisonnement probabiliste et de la théorie des probabilité
  - à partir d'une table des probabilités conjointes, comment calculer toute autre probabilité

- On a utilisé un exemple simple (*MalDeDent*, *Croche*, *Carie*)
  - on aurait pu considérer le cas de la détection de « pourriels »
    - » *AuteurInconnu* : l'auteur du courriel est inconnu
    - » *ContientMotListeNoire* : le courriel contient un mot d'une « liste noire »
    - » *Pourriel* : le courriel est un pourriel
  - souvent, on aura besoin de centaines de variables aléatoires
    - » la table des probabilités ne pourra pas être stockée en mémoire

- Dans ce cours, on va voir une façon plus efficace de construire un modèle de raisonnement probabiliste

Réseaux bayésiens

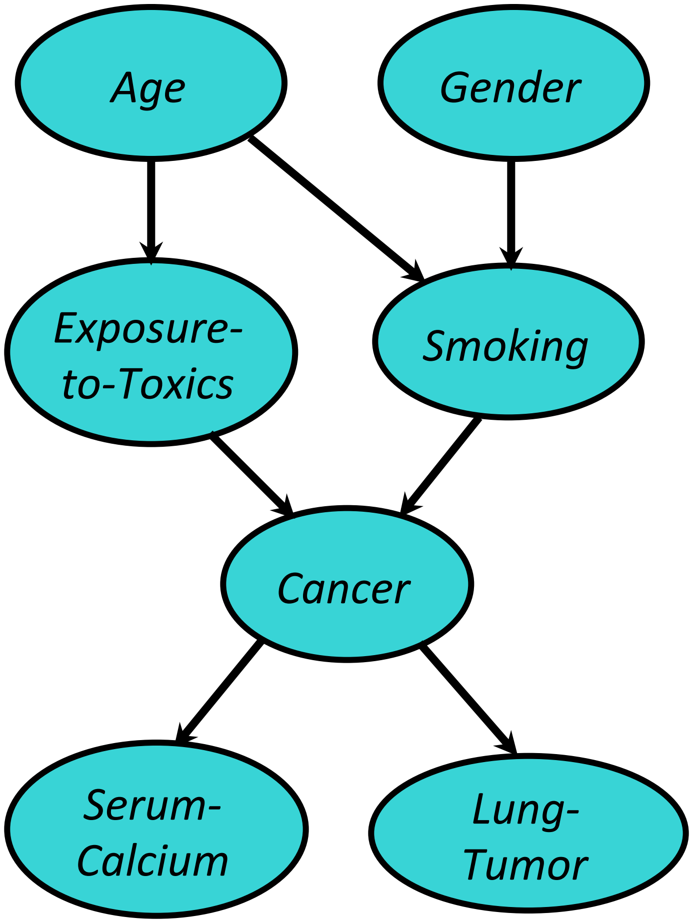
- Les **réseaux bayésiens** (RB) sont un mariage entre la théorie des graphes et la théorie des probabilités

- 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)

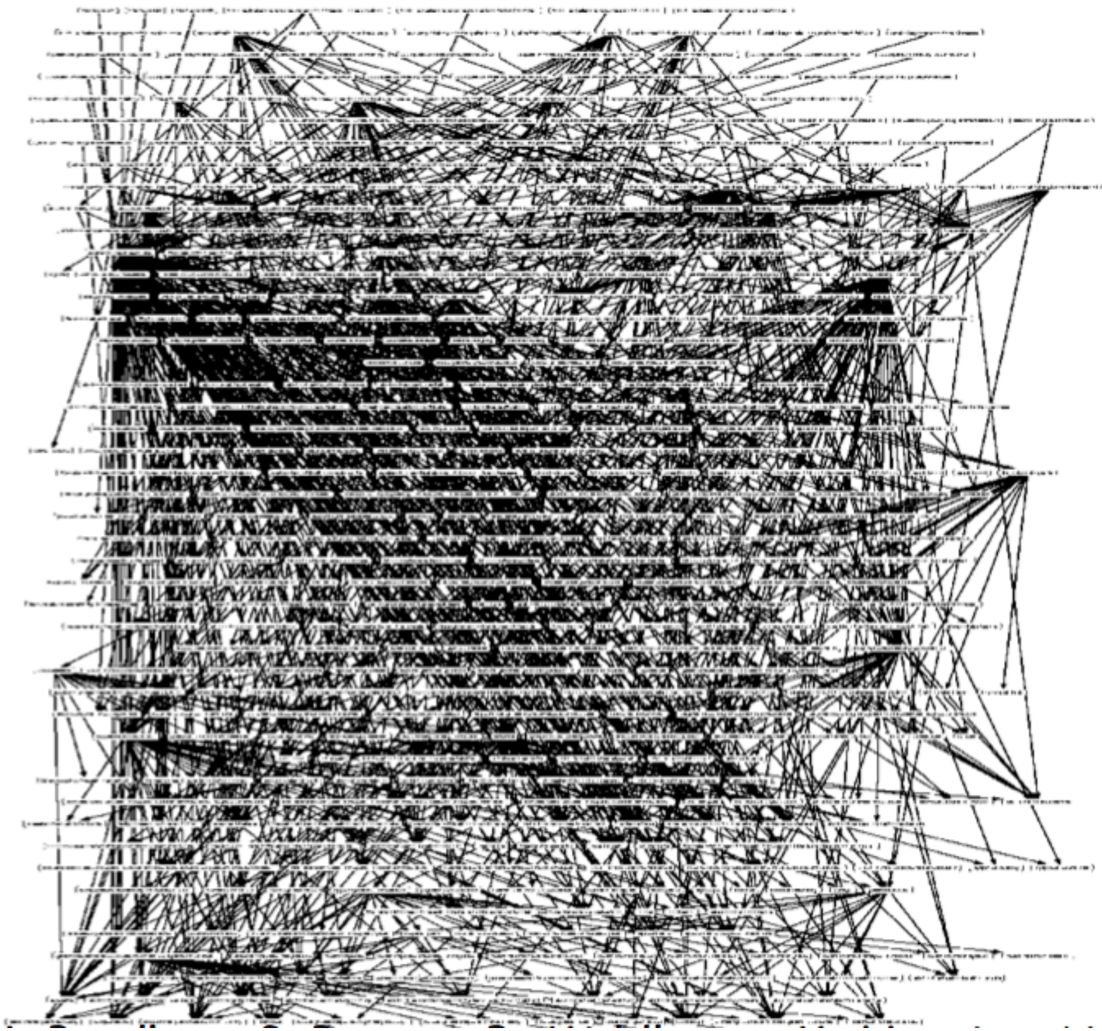
Application : diagnostique médical

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



Application : diagnostique médical

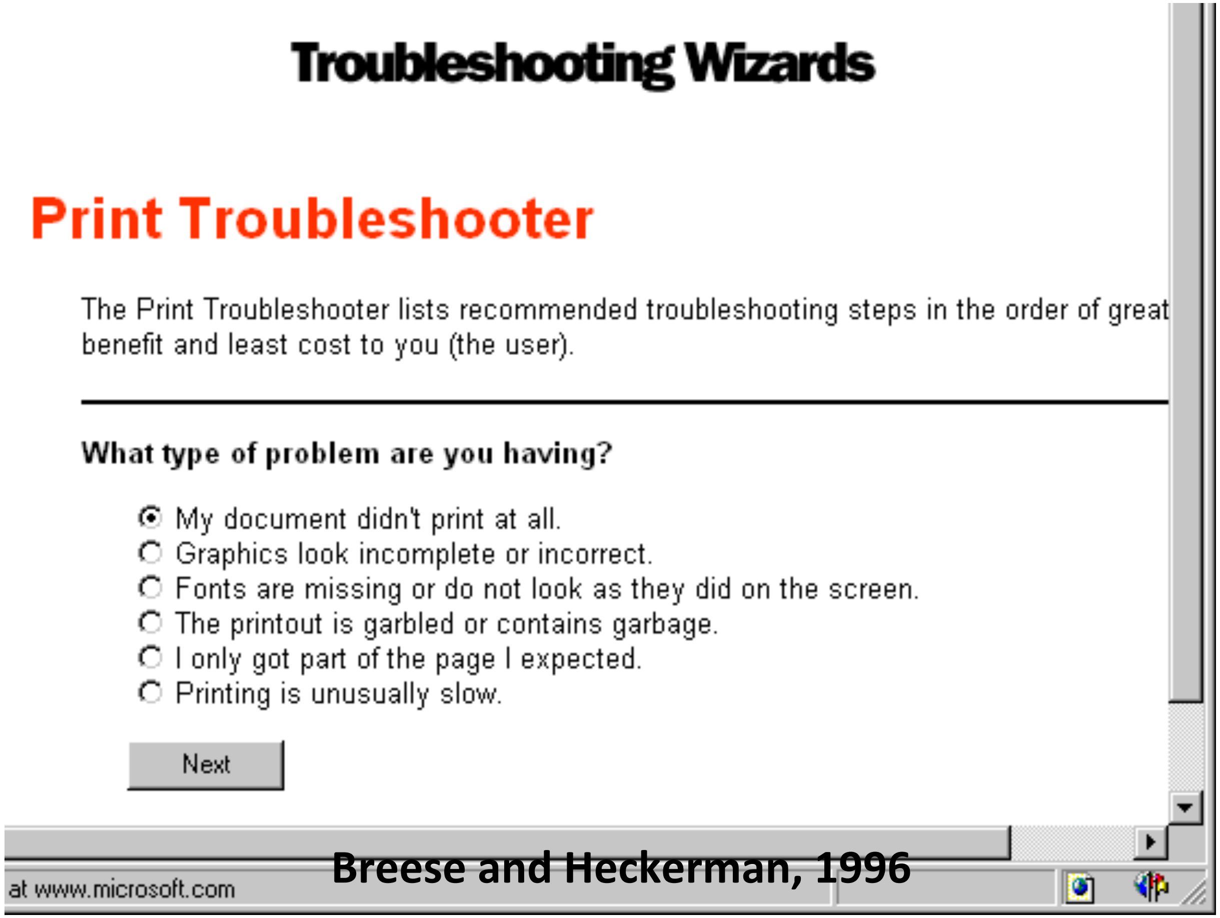
- *Pathfinder* (version 4)
  - en accord avec un panel d'experts 50 fois sur 53
  - prédictions aussi bonnes que celles des experts qui ont développé le système



M. Pradhan, G. Provan, B. Middleton, M. Henrion, UAI 1994

Application : diagnostique informatique

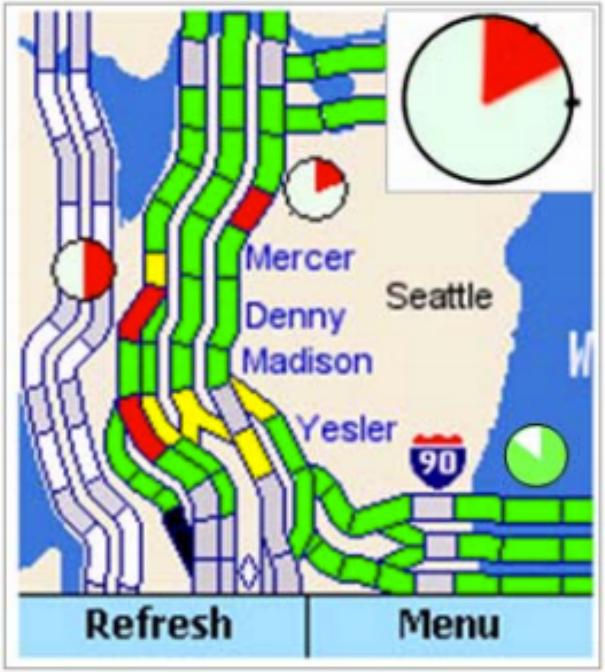
- *Troubleshooting Wizard* : détection et analyse de problèmes informatiques



Breese and Heckerman, 1996

Application : diagnostique informatique

- *JamBayes* : système de prédiction du trafic routier à Seattle



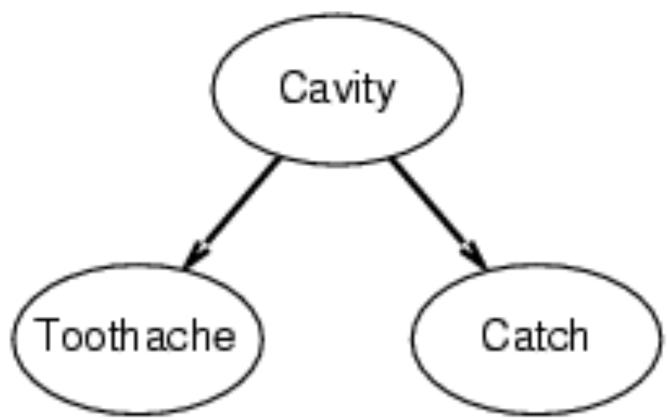
Horvitz, Apacible, Sarin and Liao 2005

Autres applications

- NASA
  - support au diagnostique en temps réel des pannes du système de propulsion des navettes spatiales
- AT&T
  - détections des fraudes et des mauvais payeurs pour les factures de téléphone
- Classification de documents (à venir)
  - détection de pourriels
- Localisation d'un robot dans une carte (à venir)

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



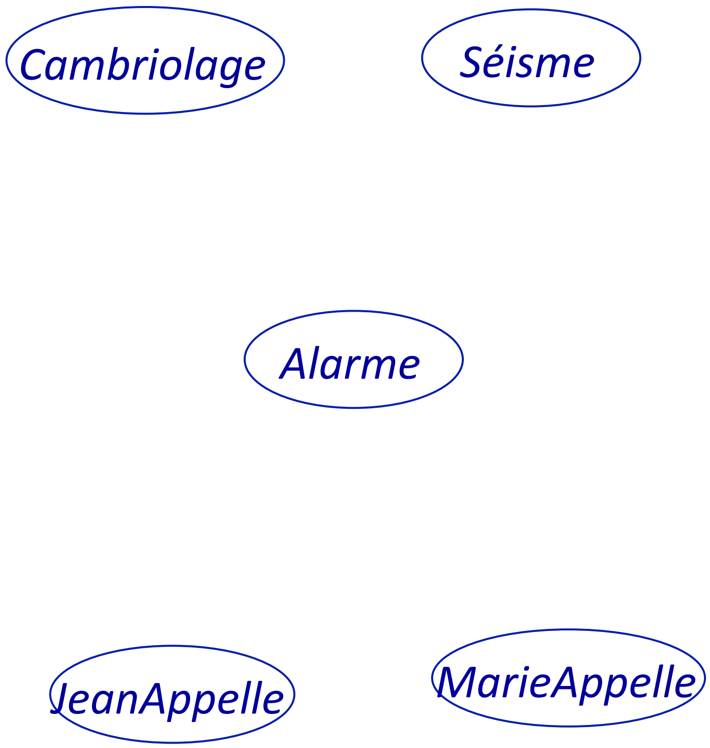
Exemple

- 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 un cambriolage chez moi?

- On peut représenter ce problème par un RB

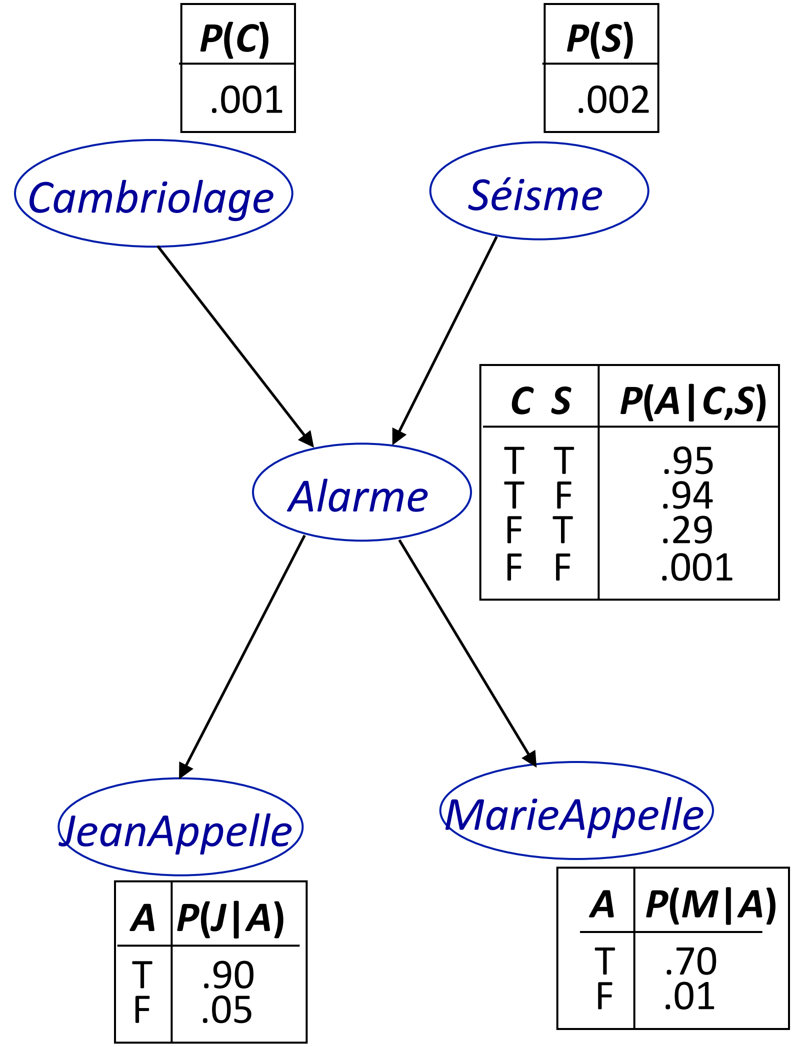
Exemple

- 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**)



***P*(*C*)**
.001

***P*(*S*)**
.002

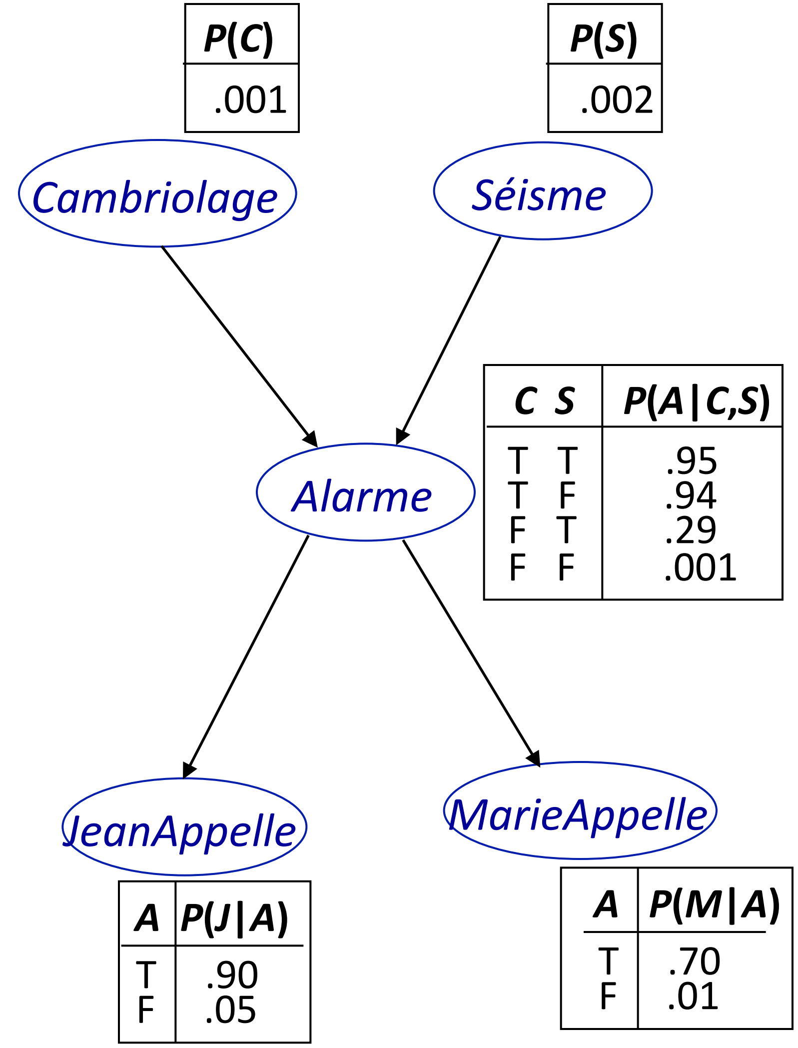
C	***S***	***P*(*A*|*C*,*S*)**
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	***P*(*J*|*A*)**
T	.90
F	.05

A	***P*(*M*|*A*)**
T	.70
F	.01

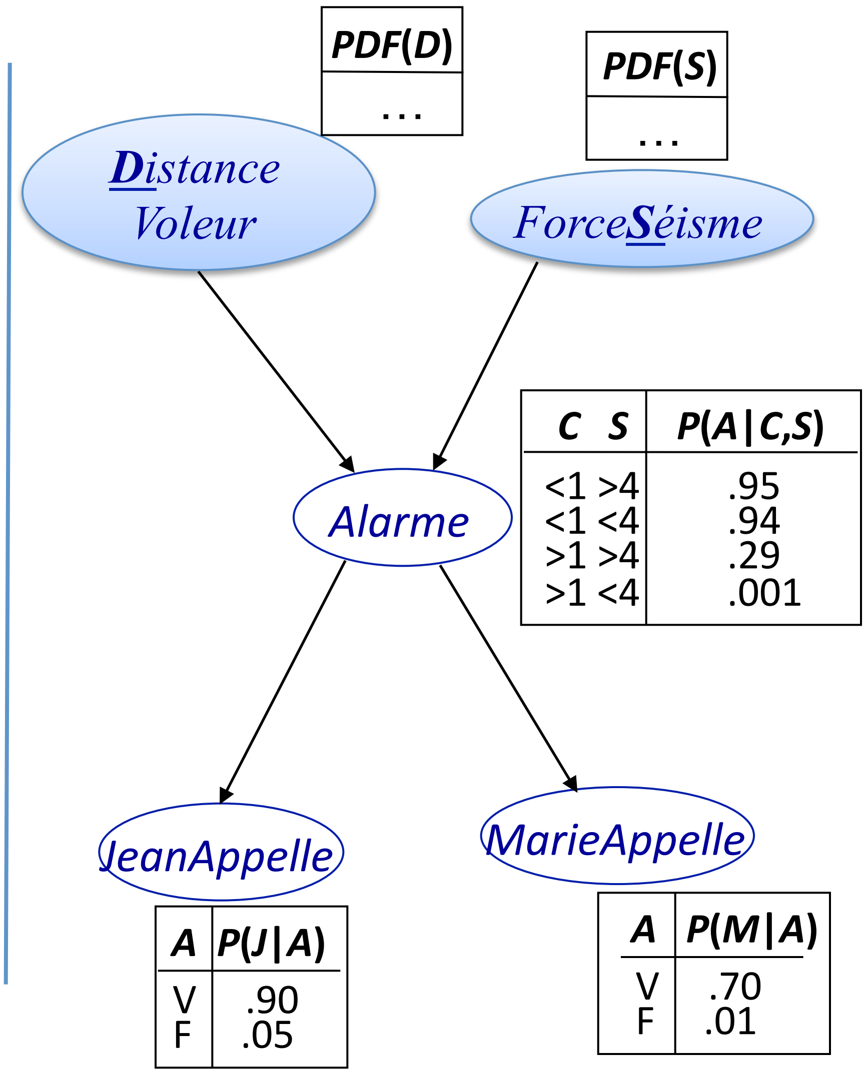
Définitions

- 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



PDF(D)
...

PDF(S)
...

C *S*	*P(A|C,S)*
<1 >4	.95
<1 <4	.94
>1 >4	.29
>1 <4	.001

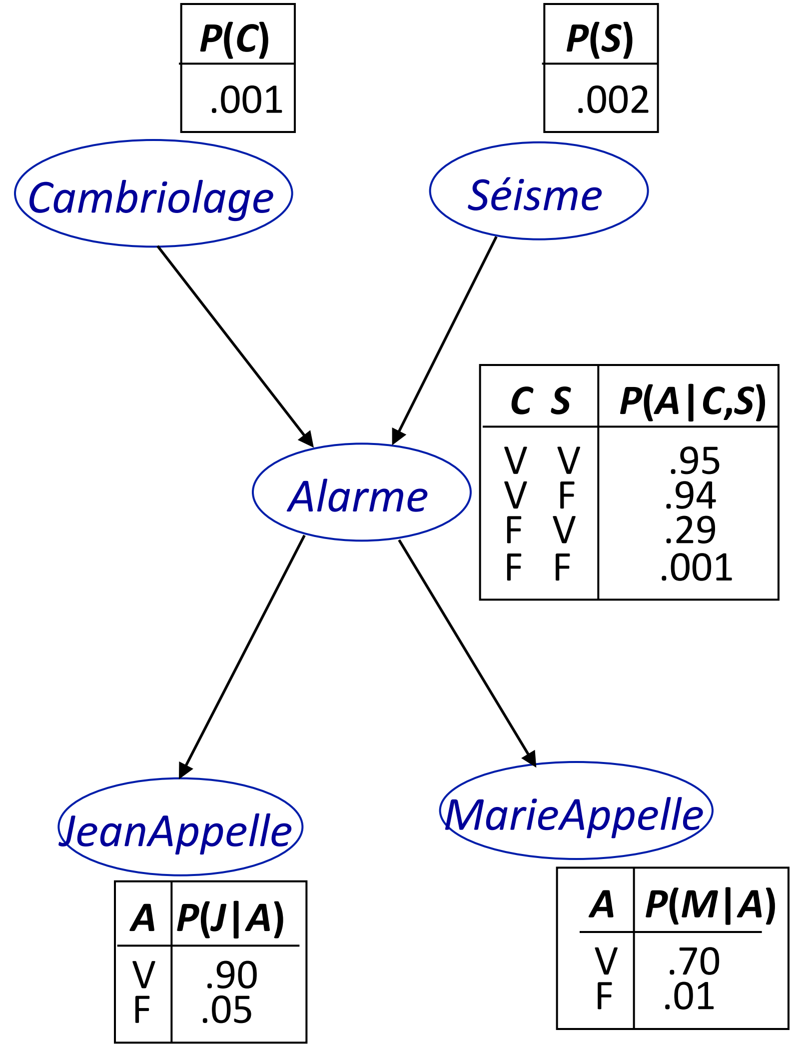
A	*P(J|A)*
V	.90
F	.05

A	*P(M|A)*
V	.70
F	.01

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 générale 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 $P(X_1,X_2)$, pour une valeur donnée de X_1 et X_2

- La distribution conditionnelle de X_1 sachant X_2 est notée $P(X_1 | X_2)$
  - $P(X_1,X_2) = P(X_1 \mid X_2)\, P(X_2)$

- Soit $X = \{X_1, \ldots, X_n\}$, l'ensemble des variables d'un RB :

$$P(X_1, \ldots, X_n) = \Pi_{i=1}^{n} P(X_i \mid Parents(X_i))$$

- 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

- En fait, quelque soit l'ensemble de variables $X = \{X_1, \ldots, X_n\}$, par définition :

$$P(X_1, \ldots, X_n) = P(X_n \mid X_{n-1}, \ldots, X_1)\, P(X_{n-1}, \ldots, X_1)$$
$$= P(X_n \mid X_{n-1}, \ldots, X_1)\, P(X_{n-1} \mid X_{n-2}, \ldots, X_1) \ldots P(X_2 \mid X_1)\, P(X_1)$$
$$= \Pi^{n}_{i=1} P(X_i \mid X_{i-1}, \ldots, X_1)$$

- Pour un RB : $P(X_1, \ldots, X_n) = \Pi^{n}_{i=1} P(X_i \mid Parents(X_i))$
  - ceci est cohérent avec l'assertion précédente pour autant que $Parents(X_i)$ soit l'ensemble de $\{X_{i-1}, \ldots, X_1\}$
  - sinon, un RB est alors une façon de **représenter les indépendances conditionnelles**

Exemple : probabilité conjointe

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

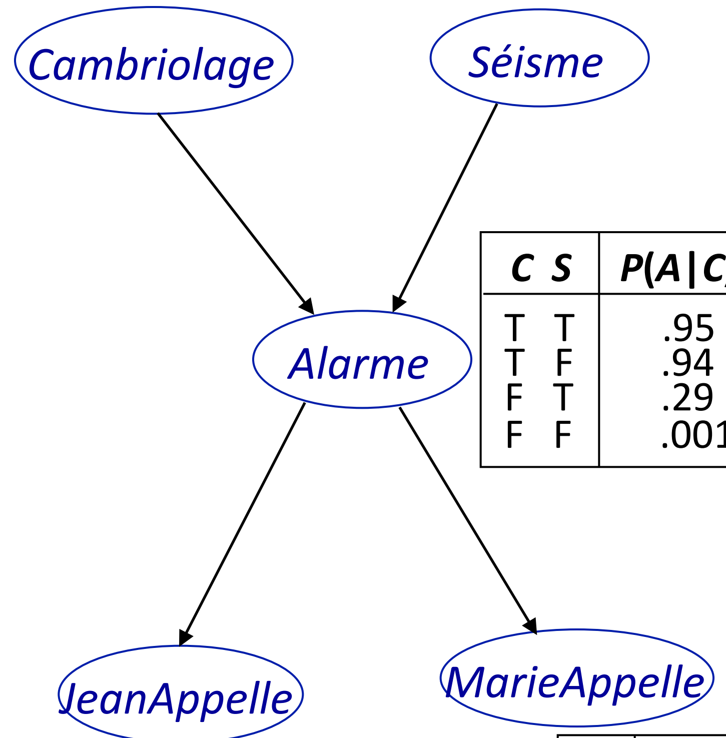
$P(j, m, a, \neg c, \neg s)$

$= P(j|a)\, P(m|a)\, P(a|\neg c, \neg s)\, P(\neg c)\, P(\neg s)$

$= .90 * .70 * .001 * .999 * .998$

$\approx .00062$

P(*J*=*j*, *M*=*m*, *A*=*a*, *C*=¬*c*, *S*=¬*s*) est aussi noté *P*(*j*, *m*, *a*, ¬*c*, ¬*s*)



P(C)
.001

P(S)
.002

C	**S**	**P(A|C,S)**
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	**P(J|A)**
T	.90
F	.05

A	**P(M|A)**
T	.70
F	.01

Exemple : probabilité marginale

$P(\neg c, a) = \Sigma_m \, \Sigma_j \, \Sigma_s \, P(J{=}j, M{=}m, a, \neg c, S{=}s)$

$= \Sigma_m \, \Sigma_j \, \Sigma_s \, P(j|a) \, P(m|a) \, P(a| \neg c, s) \, P(\neg c) \, P(s)$

$= \Sigma_s \, \Sigma_j \, \Sigma_m \, P(j|a) \, P(m|a) \, P(a| \neg c, s) \, P(\neg c) \, P(s)$

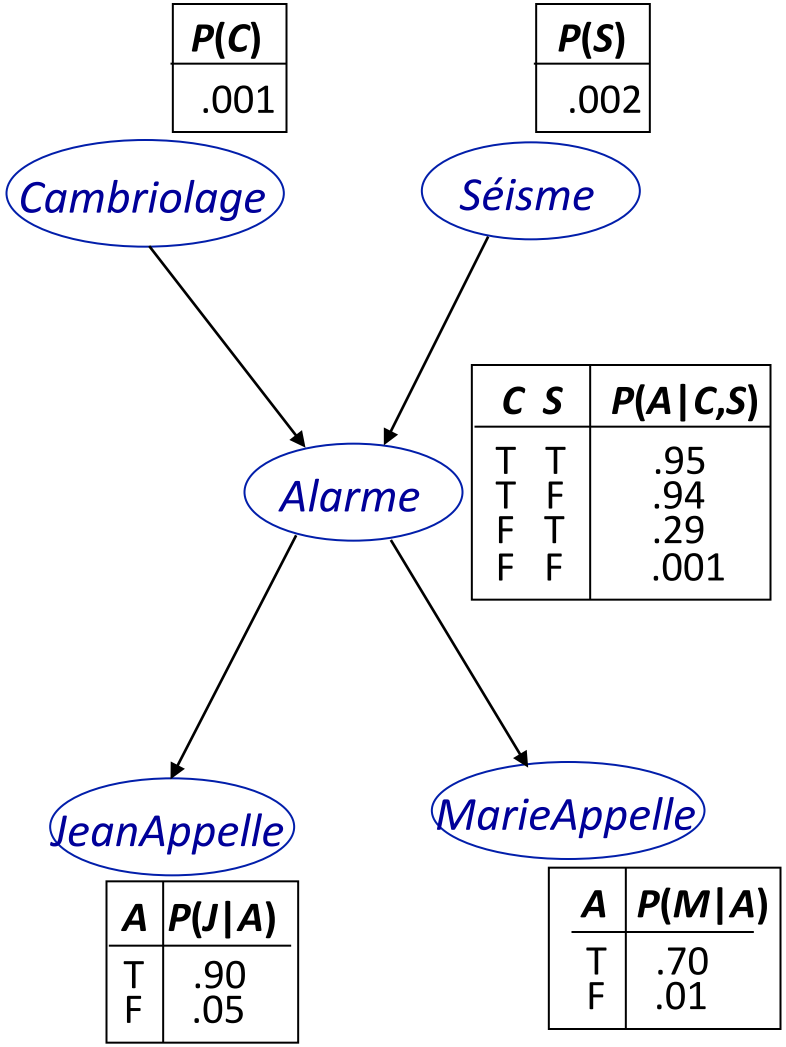
$= \Sigma_s \, \Sigma_j \, P(j|a) \, P(a| \neg c, s) \, P(\neg c) \, P(s) \, \underbrace{\cancel{\Sigma_m \, P(m|a)}}_{= 1}$

$= \Sigma_s \, P(a| \neg c, s) \, P(\neg c) \, P(s) \, \underbrace{\cancel{\Sigma_j \, P(j|a)}}_{= 1}$

$= P(a|\neg c,s) \, P(\neg c) \, P(s) + P(a|\neg c,\neg s) \, P(\neg c) \, P(\neg s)$

$= .29 * .999 * .002 + .001 * .999 * .998$

≈ 0.0016



P(C)
.001

P(S)
.002

C	*S*	*P(A|C,S)*
T	T	.95
T	F	.94
F	T	.29
F	F	.001

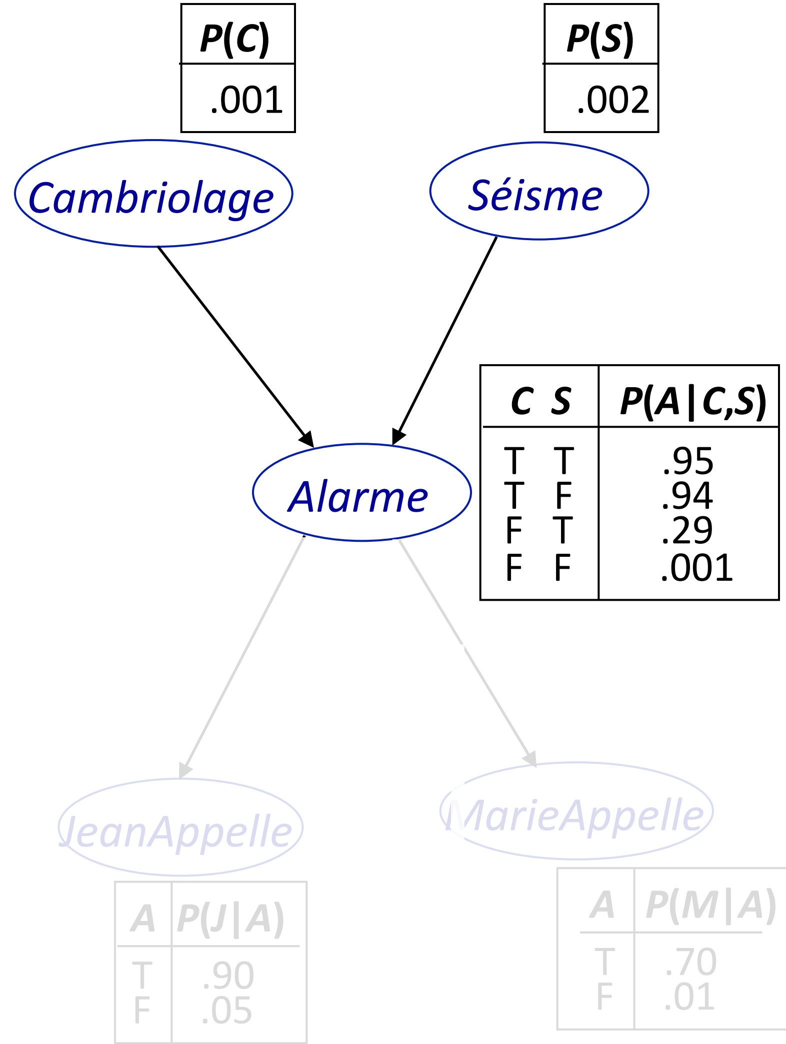
A	*P(J|A)*
T	.90
F	.05

A	*P(M|A)*
T	.70
F	.01

Probabilité marginale

$$P(c, \neg a) = \Sigma_m\, \Sigma_j\, \Sigma_s\, P(J{=}j, M{=}m, a, \neg c, S{=}s)$$
$$= \Sigma_s\, P(a \mid \neg\, c, S{=}s)\, P(\neg c)\, P(S{=}s)$$

- Pour les probabilités marginales, on peut ignorer les noeuds **dont les descendants ne sont pas les noeuds observés**
  - *Cambriolage* et *Alarme* ne sont pas des ancêtres de *JeanAppelle* ou *MarieAppelle*, alors on peut les ignore
  - *Alarme* est un descendant de *Séisme*, alors on doit le marginaliser explicitement



$P(C)$
.001

$P(S)$
.002

C	S	$P(A \mid C,S)$
T	T	.95
T	F	.94
F	T	.29
F	F	.001

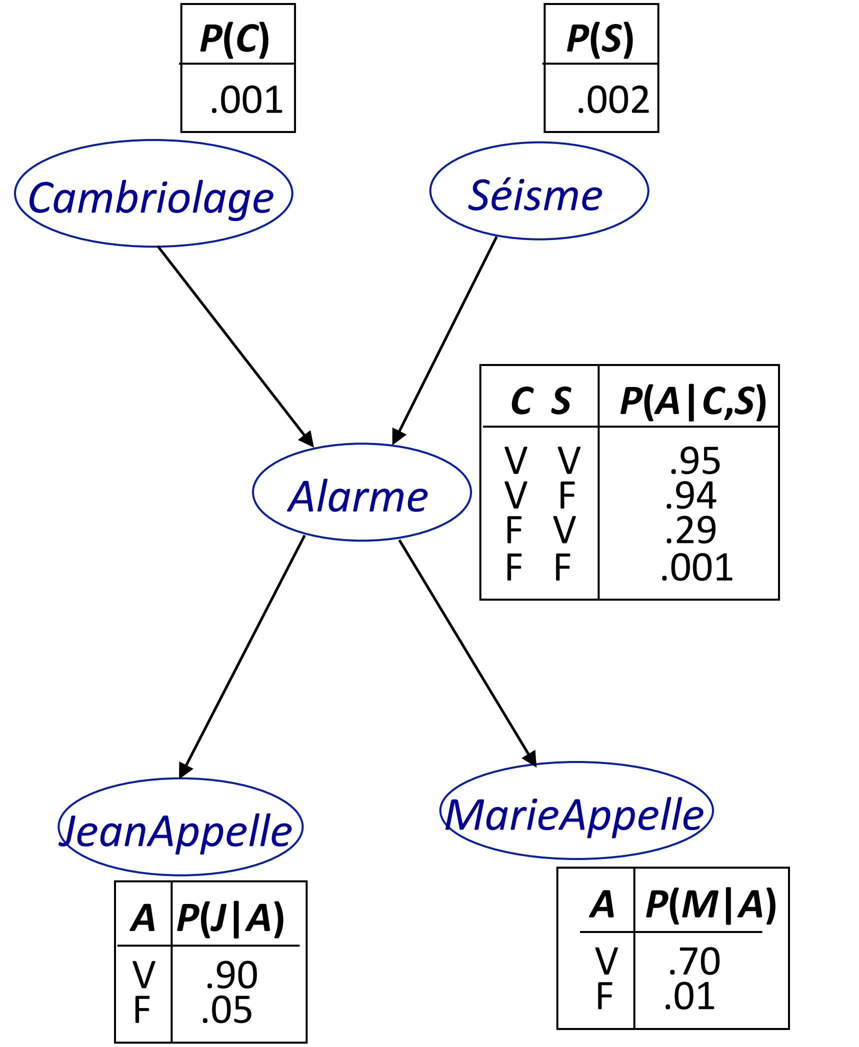
A	$P(J \mid A)$
T	.90
F	.05

A	$P(M \mid A)$
T	.70
F	.01

Indépendance conditionnelle dans un RB

- Un RB modélise les relations d'indépendance conditionnelle suivantes
  - un nœud est indépendant de ses non-descendants, étant donné ses parents
  - 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*



***P*(*C*)**
.001

***P*(*S*)**
.002

C	***S***	***P*(*A*|*C*,*S*)**
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	***P*(*J*|*A*)**
V	.90
F	.05

A	***P*(*M*|*A*)**
V	.70
F	.01

Indépendance conditionnelle dans un RB

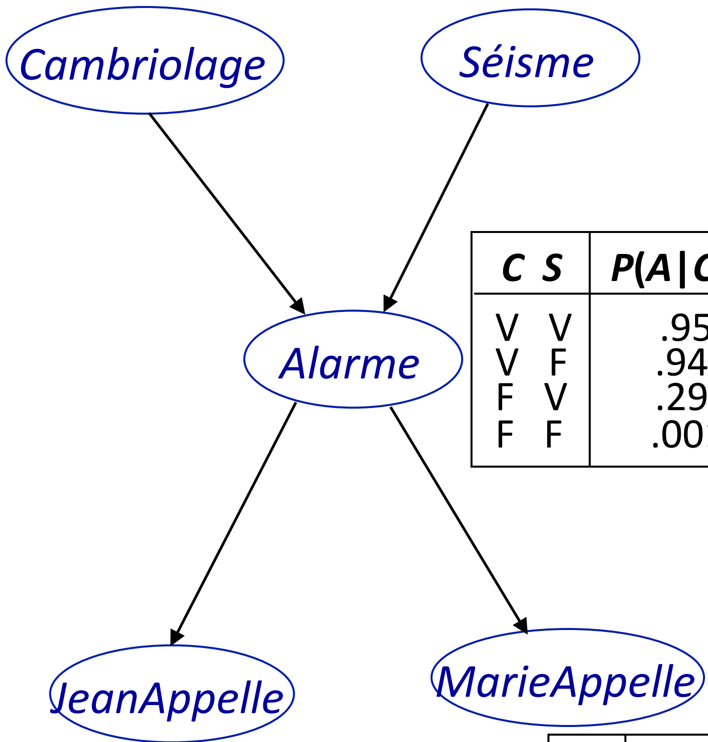
$$P(M|A,C) = P(M,A,C) \,/\, P(A,C)$$

$$= \frac{\Sigma_s P(M,A,C,S=s)}{\Sigma_s P(A,C,S=s)}$$

$$= \frac{\Sigma_s P(M|A)\, P(A|C,S=s)\, P(S=s)\, P(C)}{\Sigma_s P(A|C,S=s)\, P(S=s)\, P(C)}$$

$$= \frac{P(M|A)\, \cancel{\Sigma_s P(A|C,S=s)\, P(S=s)\, P(C)}}{\cancel{\Sigma_s P(A|C,S=s)\, P(S=s)\, P(C)}}$$

$$= P(M|A)$$



$P(C)$
.001

$P(S)$
.002

C	S	$P(A\|C,S)$
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	$P(J\|A)$
V	.90
F	.05

A	$P(M\|A)$
V	.70
F	.01

Indépendance conditionnelle dans un RB

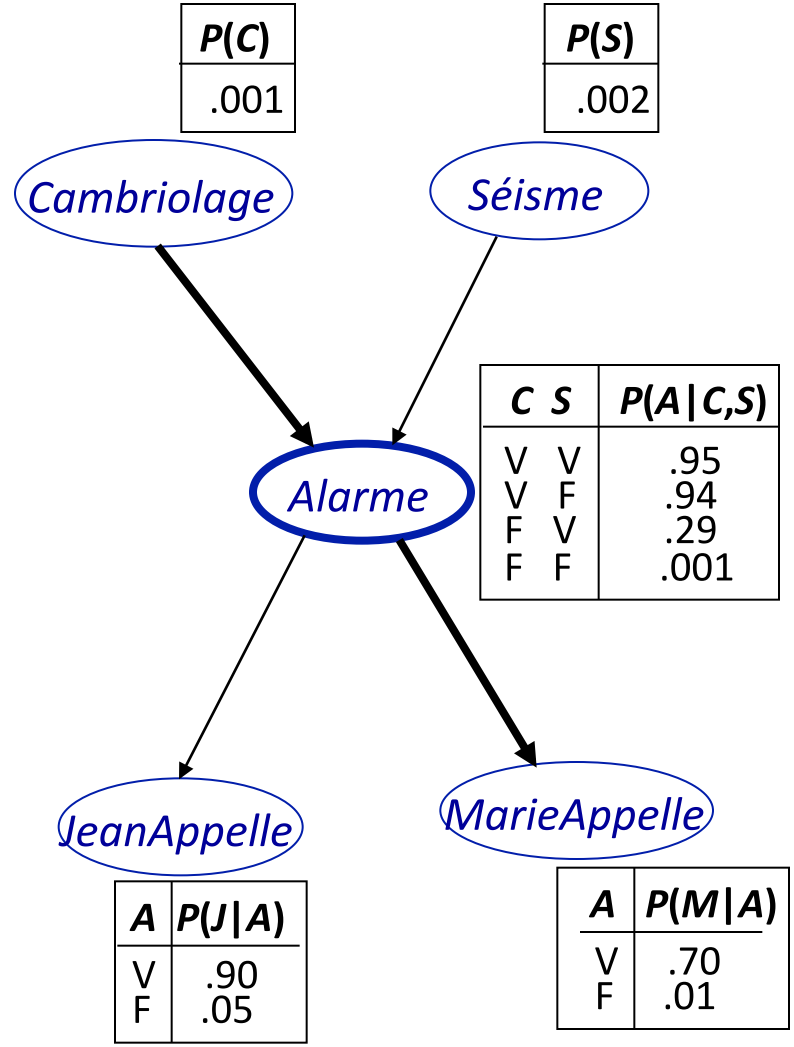
$$P(M|A,C) = P(M,A,C) \,/\, P(A,C)$$

$$= \frac{\Sigma_s\, P(M,A,C,S{=}s)}{\Sigma_s\, P(A,C,S{=}s)}$$

$$= \frac{\Sigma_s\, \mathbf{P(M|A)\; P(A|C,S{=}s)}\; P(S{=}s)\; P(C)}{\Sigma_s\, P(A|C,S{=}s)\; P(S{=}s)\; P(C)}$$

$$= P(M|A)\, \frac{\cancel{\Sigma_s\, P(A|C,S{=}s)\; P(S{=}s)\; P(C)}}{\cancel{\Sigma_s\, P(A|C,S{=}s)\; P(S{=}s)\; P(C)}}$$

$$= P(M|A)$$



***P*(*C*)**
.001

***P*(*S*)**
.002

C	***S***	***P*(*A*|*C*,*S*)**
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	***P*(*J*|*A*)**
V	.90
F	.05

A	***P*(*M*|*A*)**
V	.70
F	.01

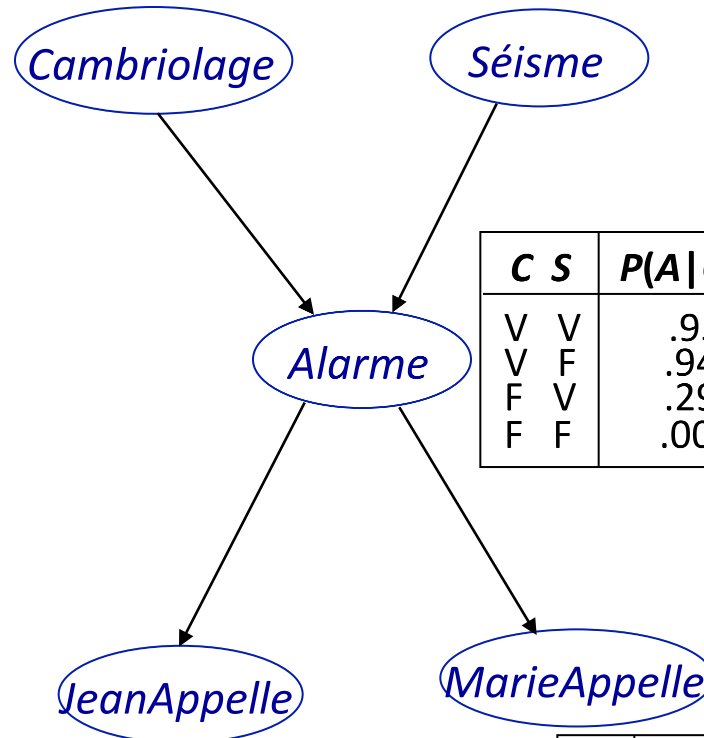
Indépendance conditionnelle dans un RB

- Un RB modélise les relations d'indépendance conditionnelle suivantes
  - un nœud est indépendant de ses non-descendants, étant donné ses parents
  - exemples :
    - » *JeanAppelle* et *MarieAppelle* sont **dépendants** a 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*

***P*(*C*)**
.001

***P*(*S*)**
.002



C	***S***	***P*(*A*|*C*,*S*)**
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	***P*(*J*|*A*)**
V	.90
F	.05

A	***P*(*M*|*A*)**
V	.70
F	.01

Indépendance conditionnelle dans un RB

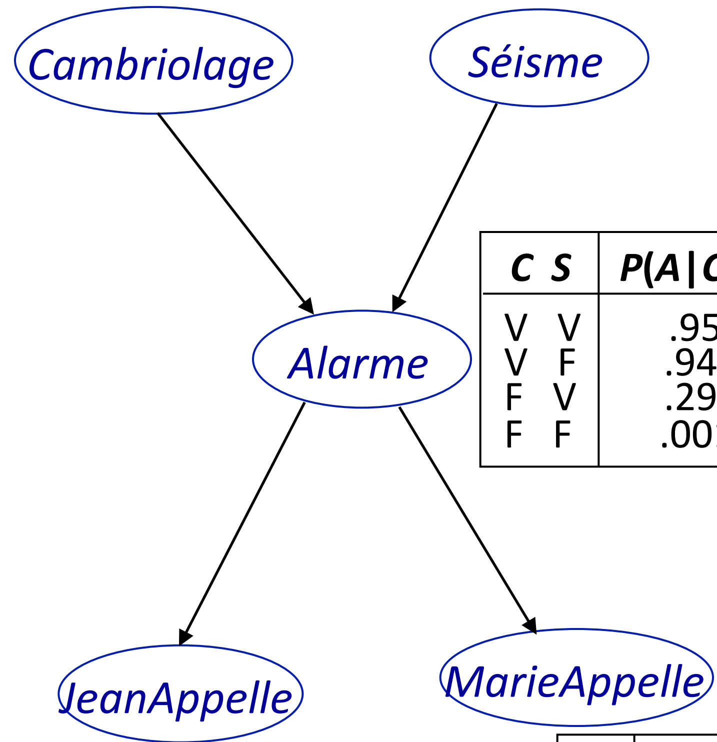
$P(M|A,J) = P(M,A,J) \,/\, P(A,J)$

$$= \frac{\Sigma_s \Sigma_c P(M,A,J,S{=}s,C{=}c)}{\Sigma_s \Sigma_c P(A,J,S{=}s,C{=}c)}$$

$$= \frac{\Sigma_s \Sigma_c P(J|A)\, P(M|A)\, P(A,S{=}s,C{=}c)}{\Sigma_s \Sigma_c P(J|A)\, P(A,S{=}s,C{=}c)}$$

$$= P(M|A) \frac{\cancel{\Sigma_s \Sigma_c P(J|A)\, P(A,S{=}s,C{=}c)}}{\cancel{\Sigma_s \Sigma_c P(J|A)\, P(A,S{=}s,C{=}c)}}$$

$$= P(M|A)$$



P(C)
.001

P(S)
.002

C	***S***	***P(A|C,S)***
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	***P(J|A)***
V	.90
F	.05

A	***P(M|A)***
V	.70
F	.01

Indépendance conditionnelle dans un RB

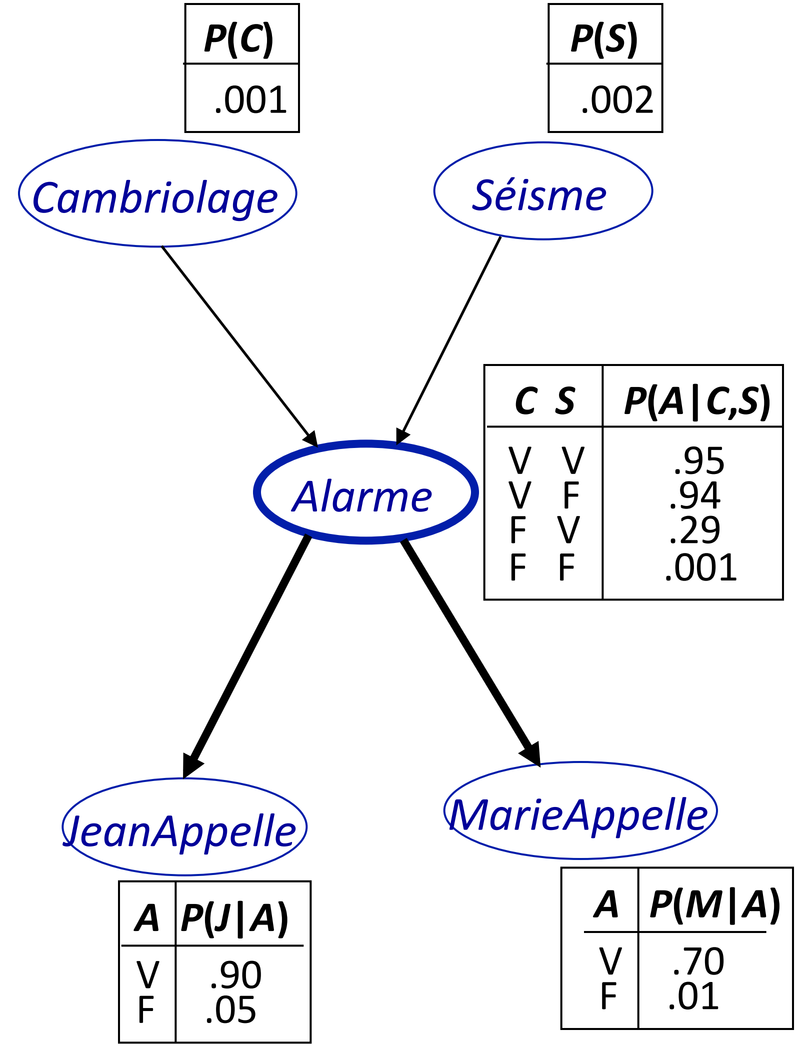
$$P(M|A,J) = P(M,A,J) \,/\, P(A,J)$$

$$= \frac{\Sigma_s \Sigma_c P(M,A,J,S=s,C=c)}{\Sigma_s \Sigma_c P(A,J,S=s,C=c)}$$

$$= \frac{\Sigma_s \Sigma_c \mathbf{P(J|A)\ P(M|A)}\ P(A,S=s,C=c)}{\Sigma_s \Sigma_c P(J|A)\ P(A,S=s,C=c)}$$

$$= \frac{P(M|A)\ \cancel{\Sigma_s \Sigma_c P(J|A)\ P(A,S=s,C=c)}}{\cancel{\Sigma_s \Sigma_c P(J|A)\ P(A,S=s,C=c)}}$$

$$= P(M|A)$$



P(C)
.001

P(S)
.002

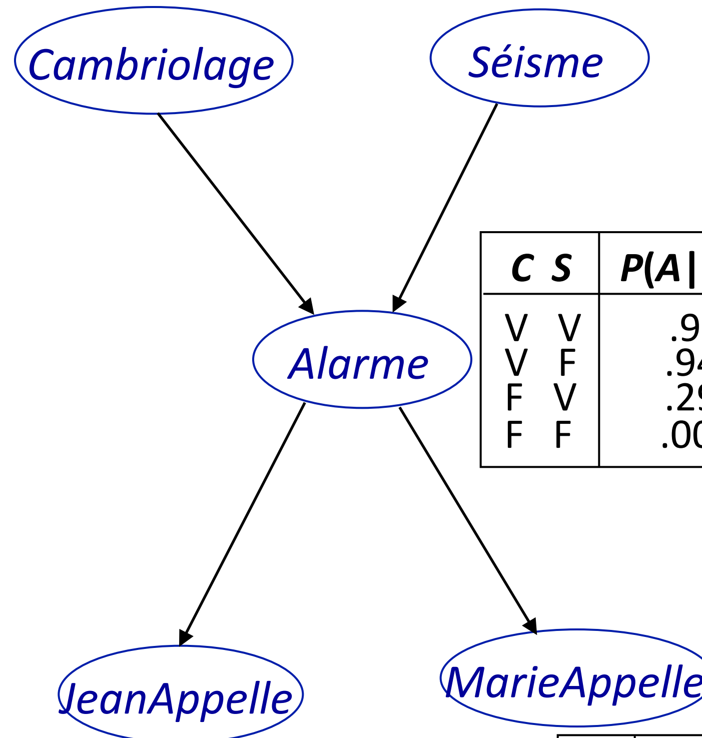
C	***S***	***P(A|C,S)***
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	***P(J|A)***
V	.90
F	.05

A	***P(M|A)***
V	.70
F	.01

Indépendance conditionnelle dans un RB

- Un RB modélise les relations d'indépendance conditionnelle suivantes
  - 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



P(C)
.001

P(S)
.002

C	***S***	***P(A|C,S)***
V	V	.95
V	F	.94
F	V	.29
F	F	.001

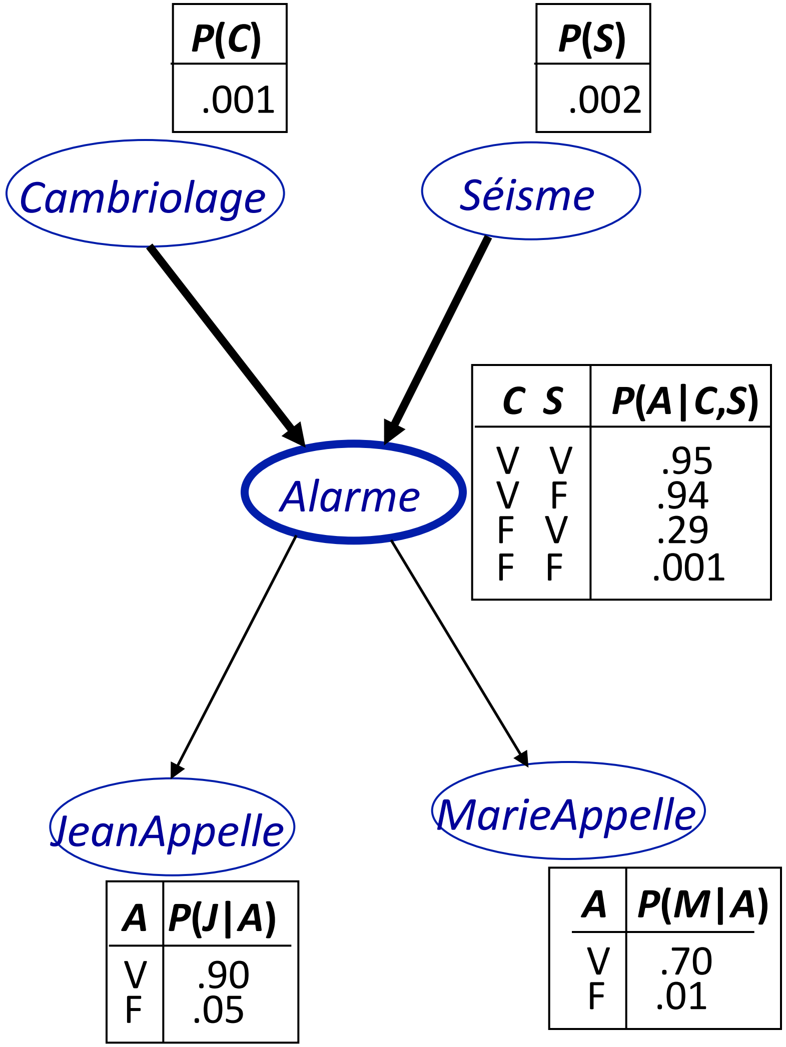
A	***P(J|A)***
V	.90
F	.05

A	***P(M|A)***
V	.70
F	.01

Indépendance conditionnelle dans un RB

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

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



P(C)
.001

P(S)
.002

C	***S***	***P(A|C,S)***
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	***P(J|A)***
V	.90
F	.05

A	***P(M|A)***
V	.70
F	.01

Indépendance conditionnelle dans un RB : D-séparation

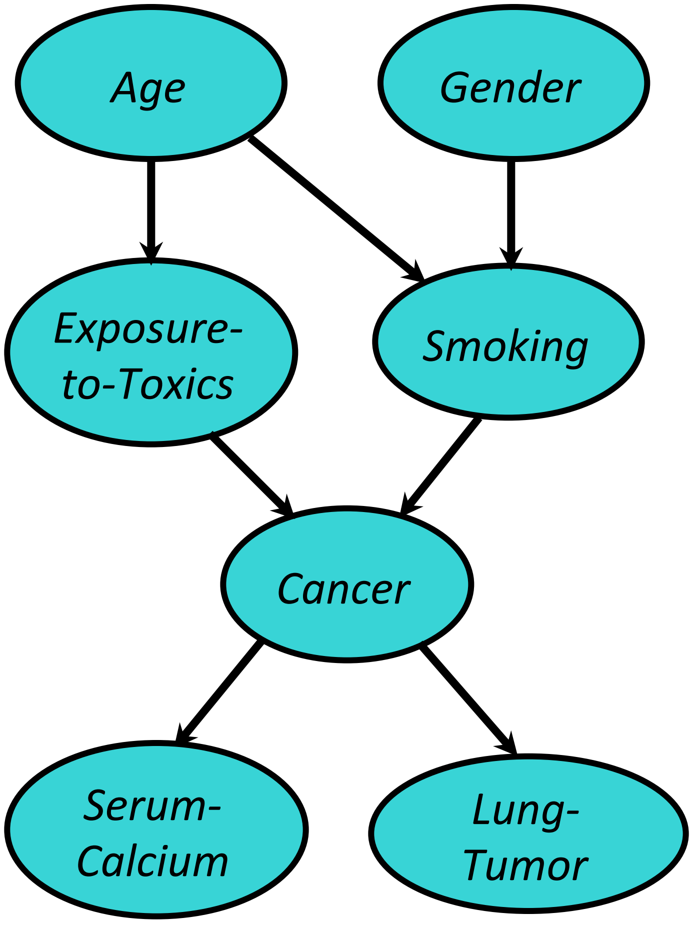
- **D-séparation** : critère général pour décider si un nœud X est indépendant d'un nœud Y, étant donnés d'autres noeuds $Z = \{Z_1, \ldots, Z_m\}$

- X est indépendant de Y sachant Z si tous **les chemins non-dirigés** entre X et Y sont **bloqués** par Z

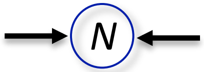
- Un **chemin est bloqué** s'il contient au moins un noeud N qui satisfait une ou l'autre des conditions suivantes :
  1. il inclue un noeud $\rightarrow N \rightarrow$ ou $\leftarrow N \rightarrow$, où $N \in \{Z_1, \ldots, Z_m\}$
  2. il inclue un noeud $\rightarrow N \leftarrow$ et $N \notin \{Z_1, \ldots, Z_m\}$, ni **aucun des descendants** de N

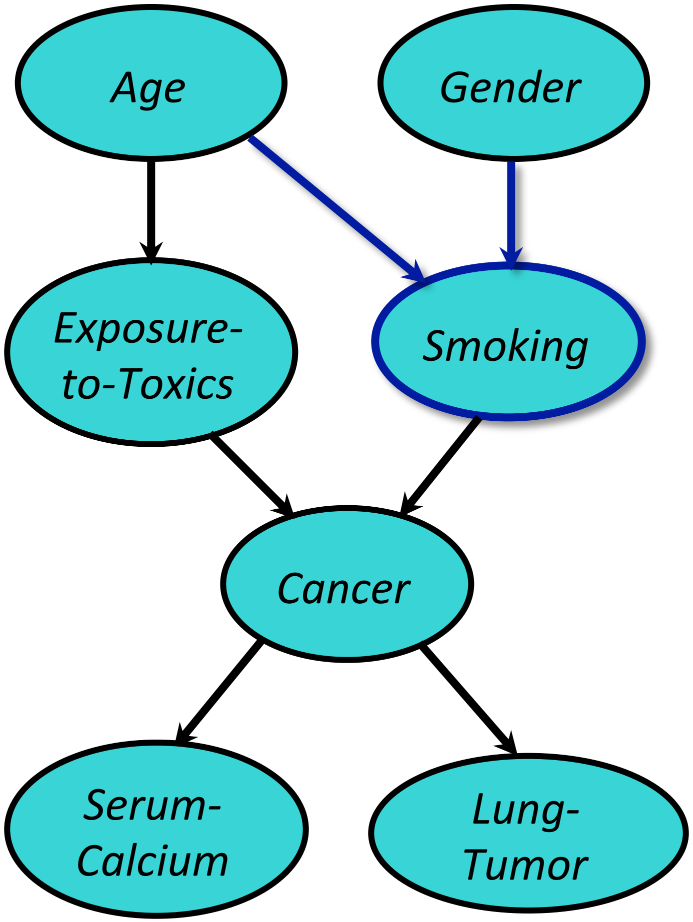
Exemple

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



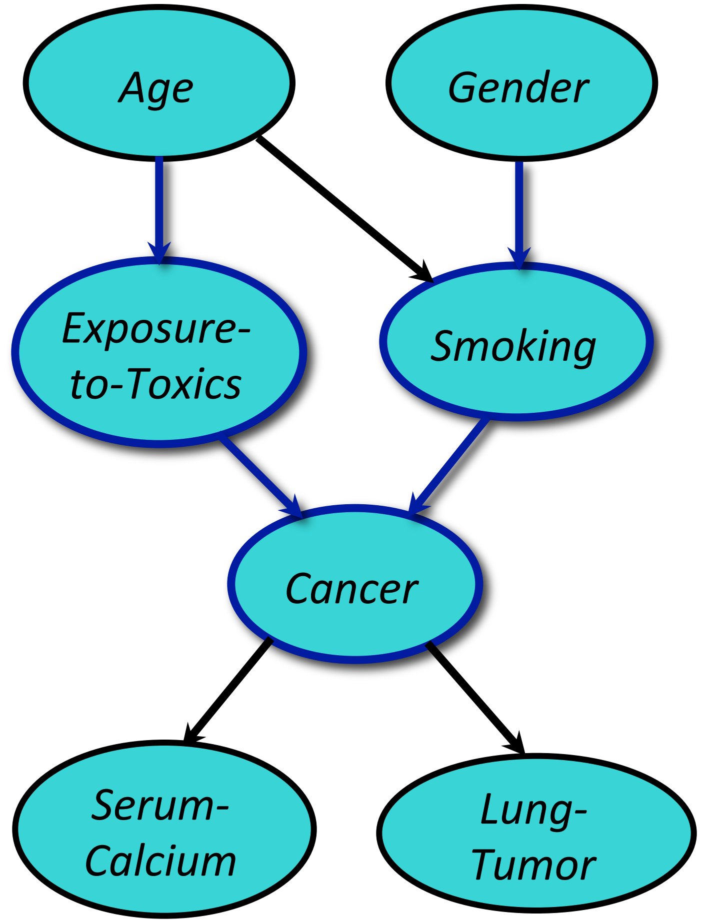
Exemple

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



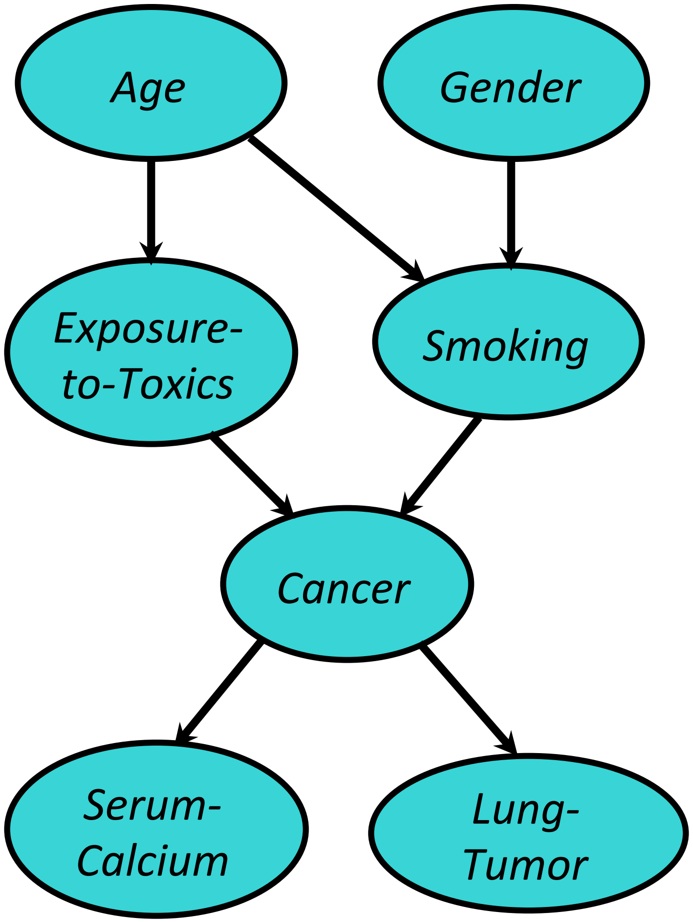
Exemple

- 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 bloqué au niveau de *Smoking* → *N* ←
    - » même raisons
- Réponse : **oui**



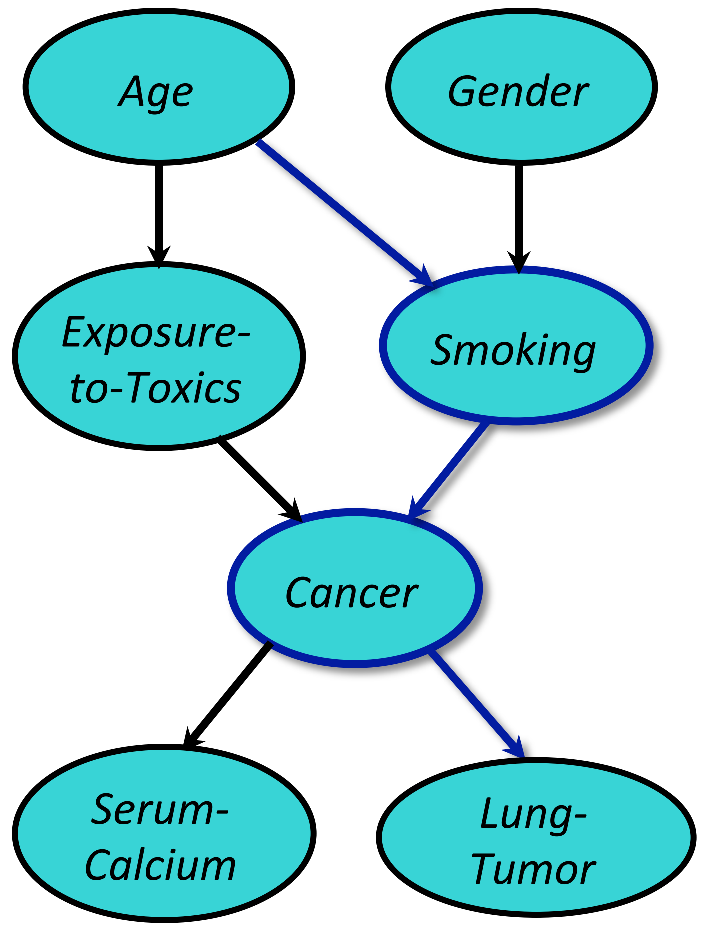
Exemple

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



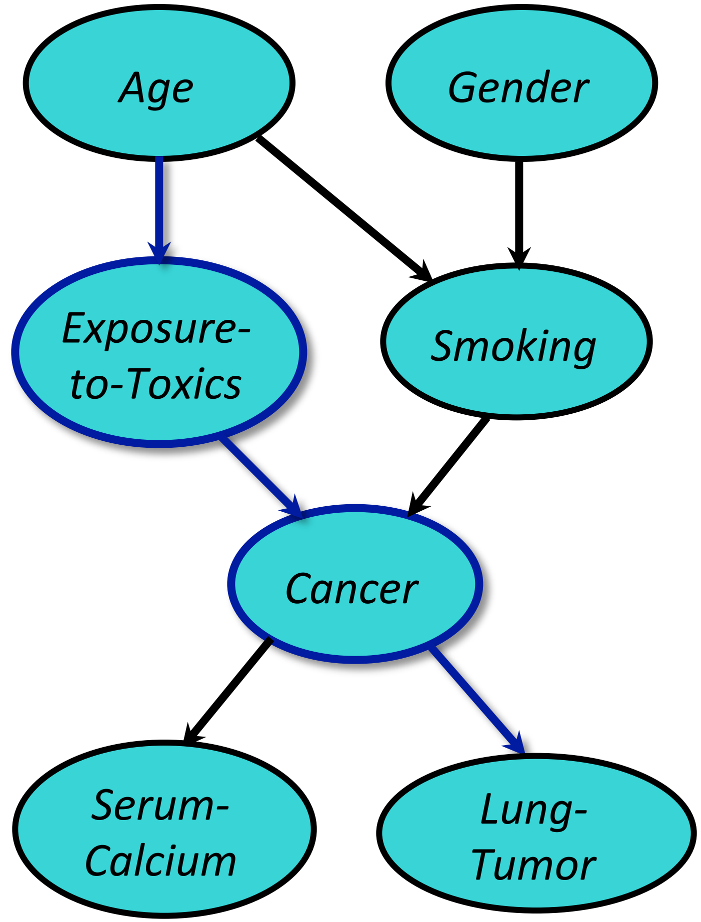
Exemple

- Est-ce que *Age* et *Lung-Tumor* sont indépendants sachant *Smoking* ?
  - chemin 1 est bloqué au niveau de *Smoking* $\rightarrow N \rightarrow$
    - » *Smoking* est observé



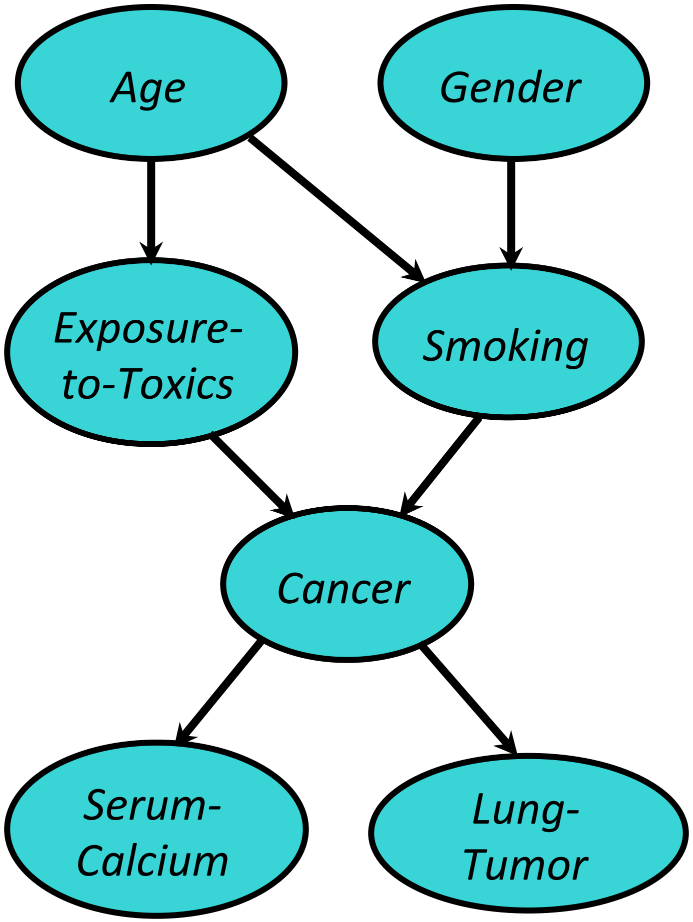
Exemple

- Est-ce que *Age* et *Lung-Tumor* sont indépendants sachant *Smoking* ?
  - chemin 1 est bloqué au niveau de *Smoking* $\rightarrow N \rightarrow$
    - » *Smoking* est observé
  - chemin 2 n’est pas bloqué
    - » *Exposure-to-Toxics* $\rightarrow N \rightarrow$ n’est pas observé
    - » *Cancer* $\rightarrow N \rightarrow$ n’est pas observé
- Réponse : **non**



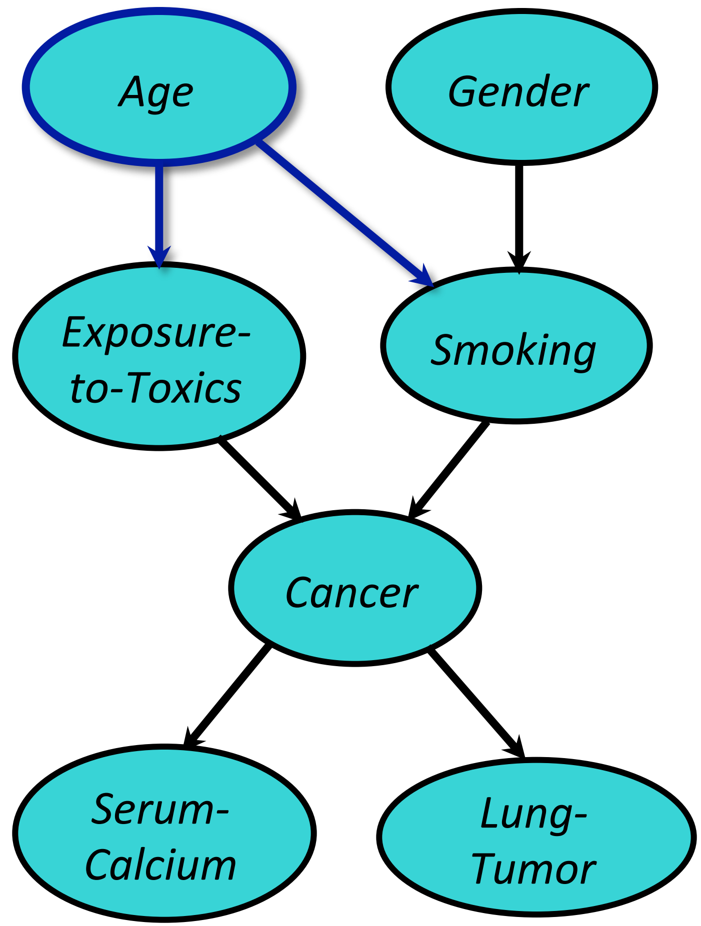
Exemple

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



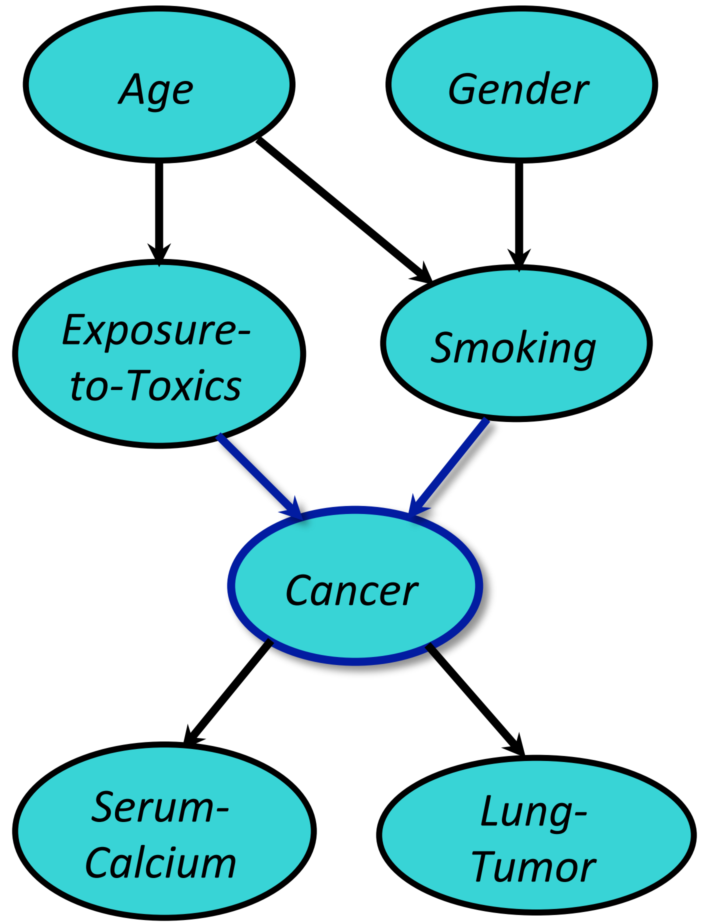
Exemple

- Est-ce que *Exposure-to-Toxics* et *Smoking* sont indépendants sachant *Age* et *Lung-Tumor*?
    - chemin 1 est bloqué au niveau de *Age* $\leftarrow$ *N* $\rightarrow$
        - *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**



Requête dans un RB

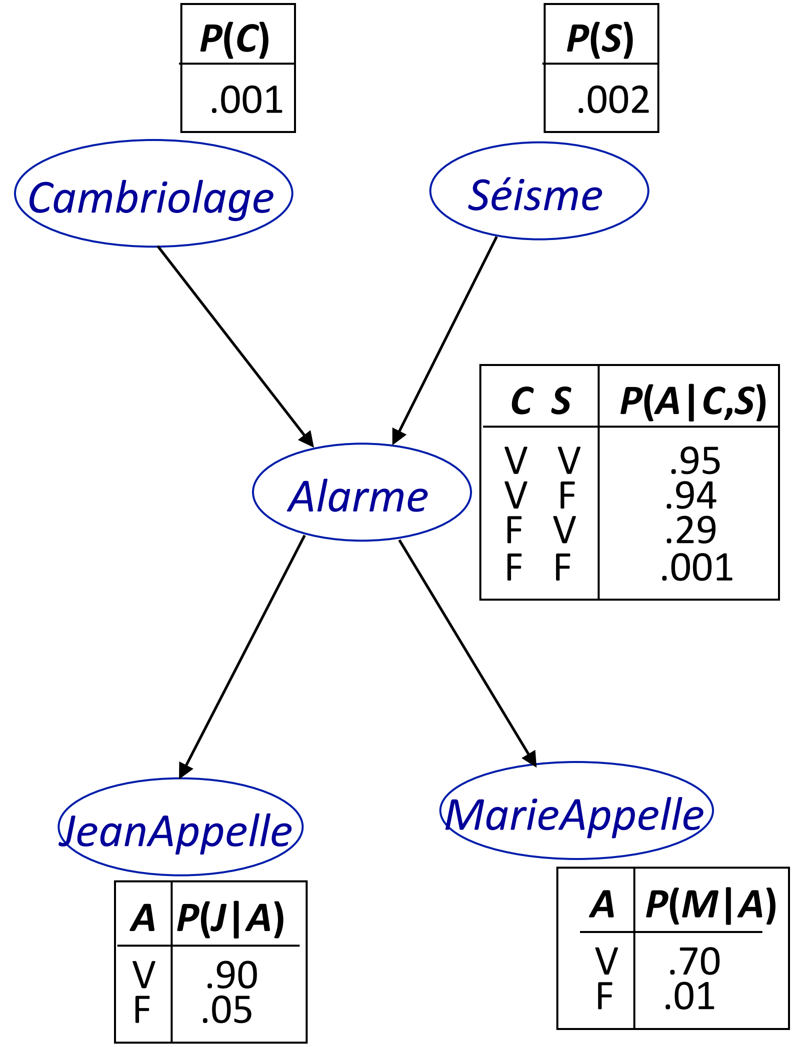
- L'usage principal d'un RB est de calculer les probabilités a posteriori, étant donné un événement observé
  - un événement est une assignation de valeurs à certaines variables d'observation
  - ex. : sachant le résultat d'une batterie de test, quelle est maintenant la probabilité qu'un patient ait 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

Requête dans un RB

- Exemple :

  P(*Cambriolage* | *JeanAppelle = vrai,*
  MarieAppelle = 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**



***P*(*C*)**
.001

***P*(*S*)**
.002

C	***S***	***P*(*A*|*C*,*S*)**
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	***P*(*J*|*A*)**
V	.90
F	.05

A	***P*(*M*|*A*)**
V	.70
F	.01

Inférence par énumération

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

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

- 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

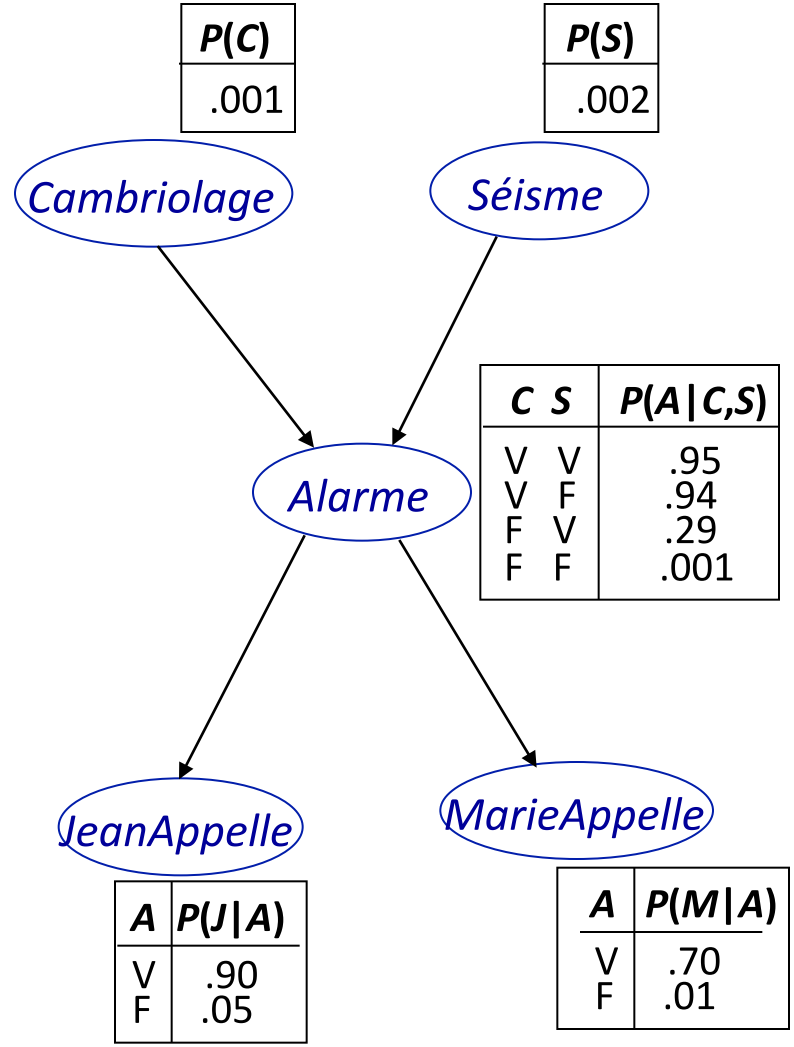
Exemple

- $\mathbf{P}(Cambriolage \mid JeanAppelle = vrai,$
  $MarieAppelle = vrai)$
  - noté $\mathbf{P}(C \mid j, m)$

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

$$\mathbf{P}(C \mid j, m) = \Sigma_{s,a}\, \mathbf{P}(C, s, a, j, m) \,/\, \alpha$$
$$= \Sigma_s\, \Sigma_a\, \mathbf{P}(C, s, a, j, m) \,/\, \alpha$$

- 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*=*vrai* et *M*=*vrai*)



***P*(*C*)**
.001

***P*(*S*)**
.002

C	***S***	***P*(*A*|*C*,*S*)**
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	***P*(*J*|*A*)**
V	.90
F	.05

A	***P*(*M*|*A*)**
V	.70
F	.01

Exemple

- $\mathbf{P}(C \mid j, m) = \Sigma_{s,a}\, \mathbf{P}(C, s, a, j, m) / \alpha$
- On calcule pour $C = vrai$

  $P(c \mid j, m)$

  $= \Sigma_{s,a}\, P(c)\, P(s)\, P(a|c,s)\, P(j|a)\, P(m|a) / \alpha$

  $= (\ 0.001*0.002*0.95*0.90*0.70+$

  $0.001*0.998*0.94*0.90*0.70+$

  $0.001*0.002*0.05*0.05*0.01+$

  $0.001*0.998*0.06*0.05*0.01) / \alpha$

  $= 0.00059224 / \alpha$

- Et $C = faux$

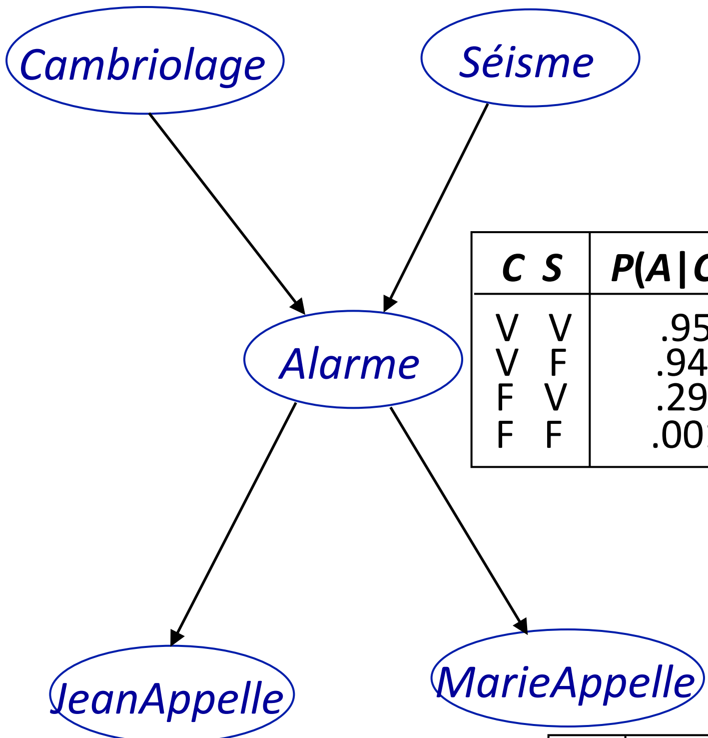
  $P(\neg c \mid j, m)$

  $= \Sigma_{s,a}\, P(\neg c)\, P(s)\, P(a| \neg c,s)\, P(j|a)\, P(m|a) / \alpha$

  $= 0.0014919 / \alpha$

  $\alpha = 0.00059224 + 0.0014919$

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



P(*C*)
.001

P(*S*)
.002

C	*S*	*P*(*A*|*C*,*S*)
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	*P*(*J*|*A*)
V	.90
F	.05

A	*P*(*M*|*A*)
V	.70
F	.01

Exercice

- Soit le RB ayant les tables de probabilités conditionnelles suivantes
  - Dessinez le graphe du RB.
  - Calculez *P*(*A*=faux | *E*=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=vrai*
faux	*faux*	0.1
faux	*vrai*	0.2
vrai	*faux*	0.8
vrai	*vrai*	0.7

E	*C=vrai*
faux	0.2
vrai	0.4

D=vrai
0.2

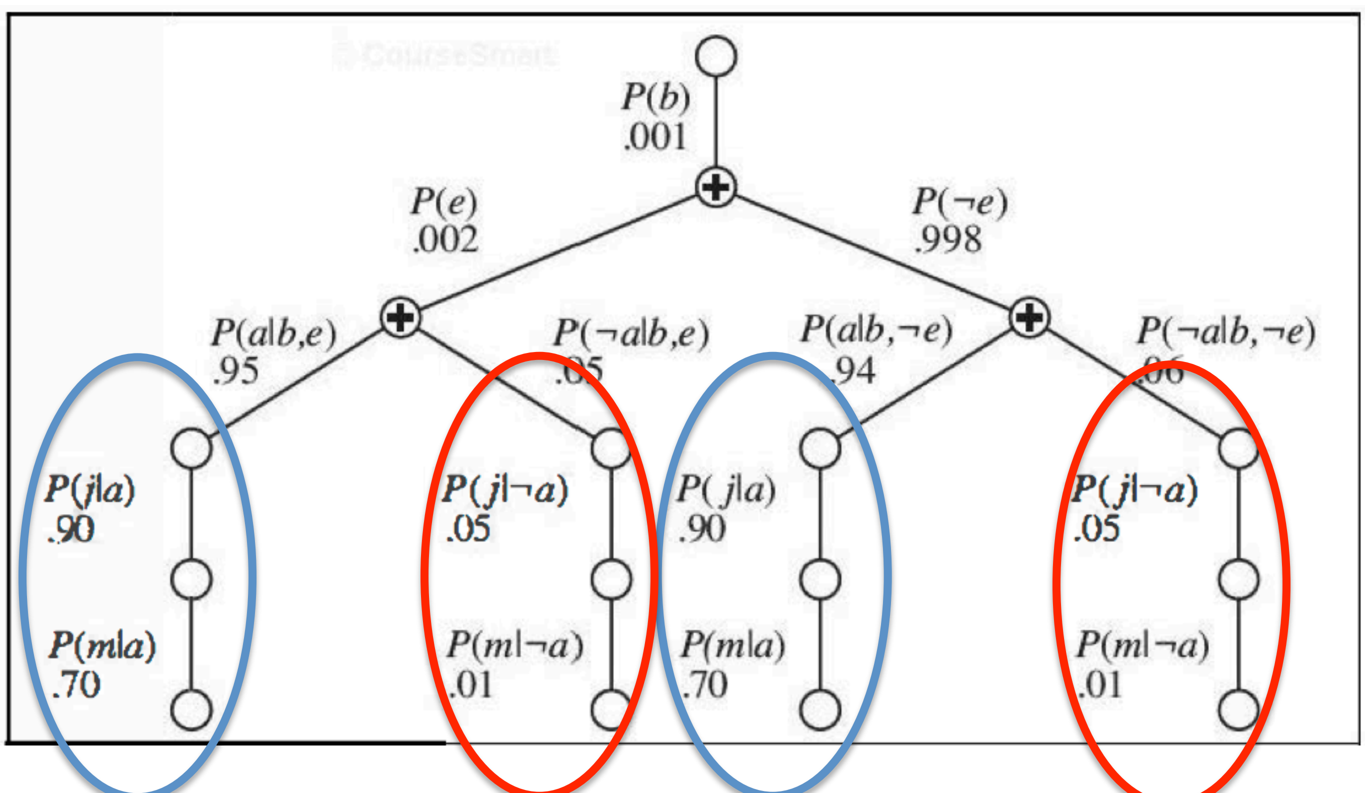
B	*D*	*E*	*A=vrai*
faux	*faux*	*faux*	0.7
faux	*faux*	*vrai*	0.2
faux	*vrai*	*faux*	0.5
faux	*vrai*	*vrai*	0.1
vrai	*faux*	*faux*	0.2
vrai	*faux*	*vrai*	0.9
vrai	*vrai*	*faux*	0.8
vrai	*vrai*	*vrai*	0.6

E=vrai
0.9

B=vrai
0.7

Inférence par élimination des variables

- Même principe que l'inférence par énumération, mais on évite les répétions de calculs déjà faits (comme en programmation dynamique)
- Voir section 14.4.2 du livre



Inférence approximative

- Les méthodes d'inférence exactes sont inefficaces
  - le problème d'inférence 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*)

- **Idée :** simuler des observations complètes du RB et estimer les probabilités à partir de la fréquence des observations échantillonnées

  $P(X=x|e) = \Sigma_y P(X=x, e, y) / \alpha \approx \mathrm{freq}(x,e) / \Sigma_{x'} \mathrm{freq}(x',e) = \mathrm{freq}(x,e) / \mathrm{freq}(e)$

  où freq(x,e) est le nombre de fois que $X=x$ et $E=e$ a été échantillonné
  et freq(e) = $\Sigma_{x'}$ freq(x',e) est le nombre de fois que $E=e$

- Cette technique est appelée **méthode de rejet** (***rejection sampling***)
  - le problème avec cette méthode est que si 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 1 : Évaluation par énumérations

<u>**Requête**</u> :

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

<u>Variables connues :</u>

$F = faux$

$M = vrai$

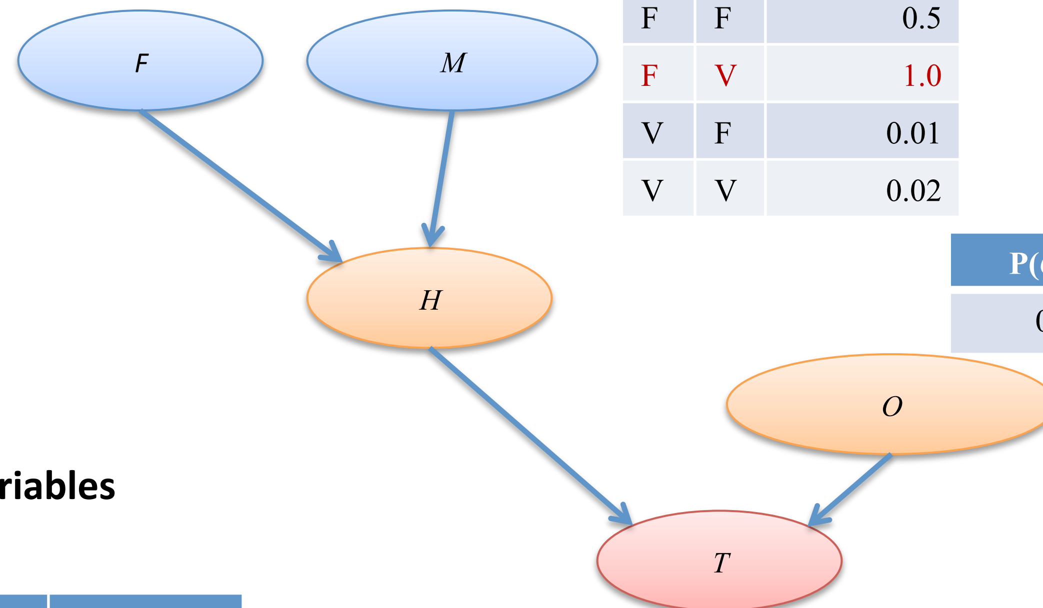
<u>Variables inconnues :</u>

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 possible 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 2 : Évaluation par énumérations

<u>Requête :</u>

Calculer $P(T{=}vrai \mid M{=}vrai)$

<u>Variables connues :</u>

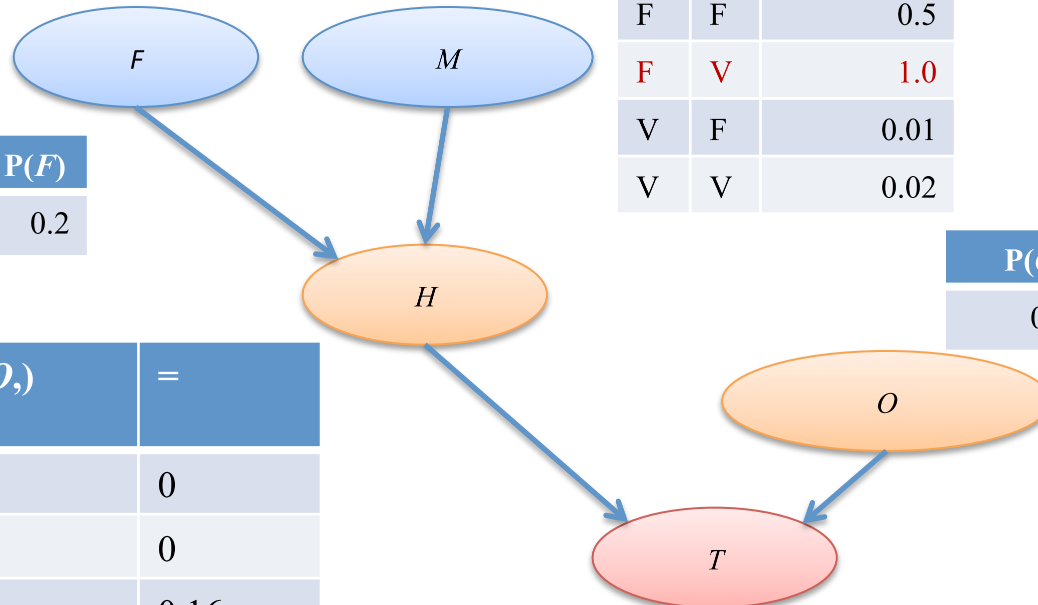
$M = vrai$

<u>Variables inconnues :</u>

H

O

F



P(*F*)
0.2

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

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

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**

Exemple 2 : Échantillonage direct

<u>Requête</u> :

Calculer $P(T{=}vrai \mid M{=}vrai)$

<u>Variables connues :</u>

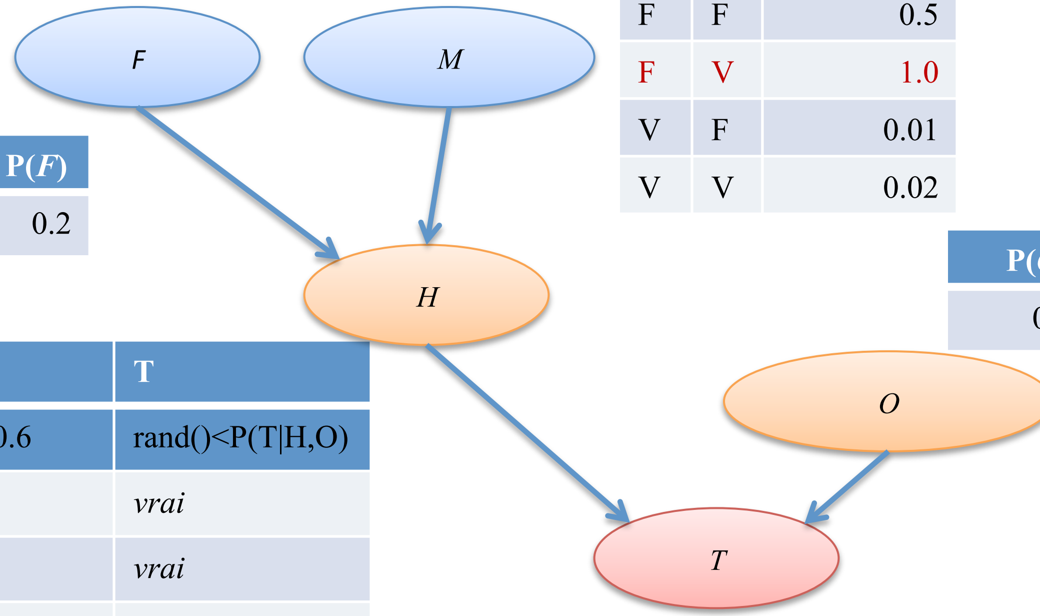
$M = vrai$

<u>Variables inconnues :</u>

H

O

F



P(*F*)
0.2

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

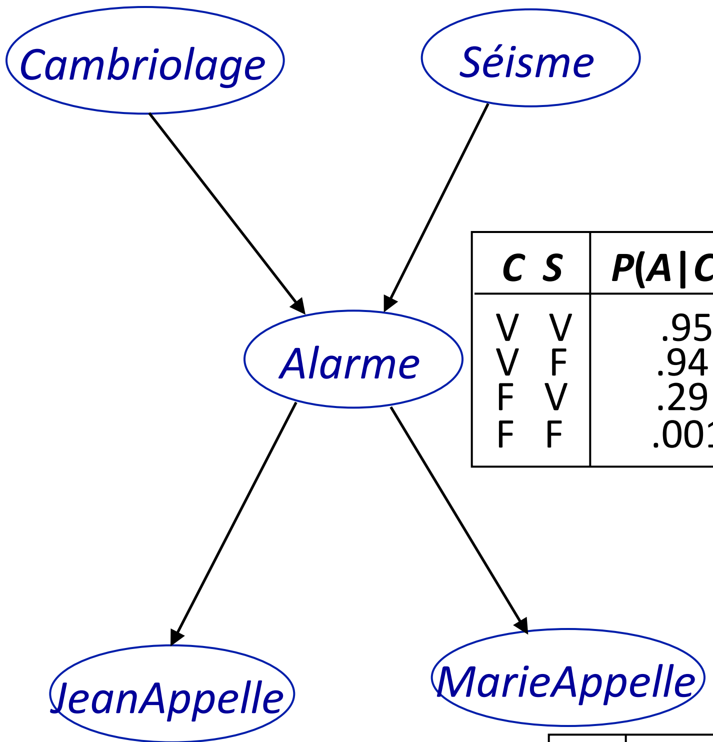
H	*O*	P(*T*|*H*,*O*)
F	F	0.1
F	V	0.5
V	F	0.5
V	V	1.0

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

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

Types d'interrogations d'un RB

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



P(*C*)
.001

P(*S*)
.002

C	*S*	*P*(*A*|*C*,*S*)
V	V	.95
V	F	.94
F	V	.29
F	F	.001

A	*P*(*J*|*A*)
V	.90
F	.05

A	*P*(*M*|*A*)
V	.70
F	.01

Construction d'un RB

1. Choisir un ordre des variables $X_1, \ldots, X_n$
2. Pour i = 1 to n :
   - ajouter X_i au réseau
   - choisir les parents $X_1, \ldots, X_{i-1}$ tels que $P(X_i \mid Parents(X_i)) = P(X_i \mid X_1, \ldots X_{i-1})$
   - ce choix garantit que :

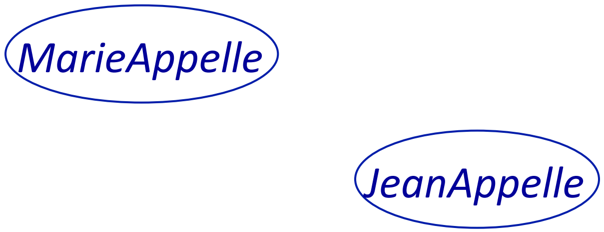
$$
\begin{aligned}
P(X_1, \ldots, X_n) &= \Pi_{i=1}^{n} P(X_i \mid X_1, \ldots, X_{i-1}) \quad (\textit{chain rule}) \\
&= \Pi_{i=1}^{n} P(X_i \mid Parents(X_i)) \quad \text{(par construction)}
\end{aligned}
$$

- 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

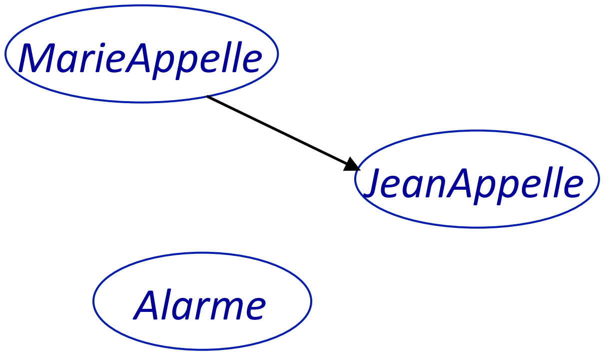
- Supposons qu'on ordonne les variables comme suit : *M*, *J*, *A*, *C*, *S*



- $P(J \mid M) = P(J)$?

Exemple

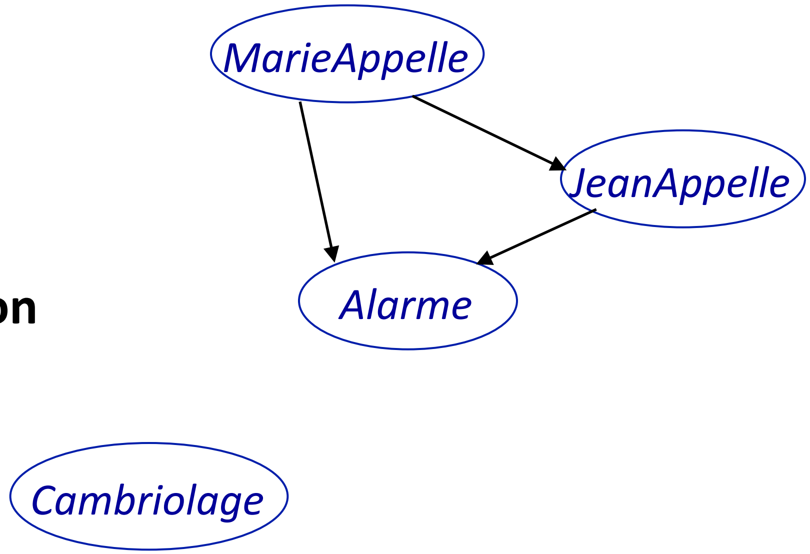
- Supposons qu'on ordonne les variables comme suit : *M*, *J*, *A*, *C*, *S*



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

Exemple

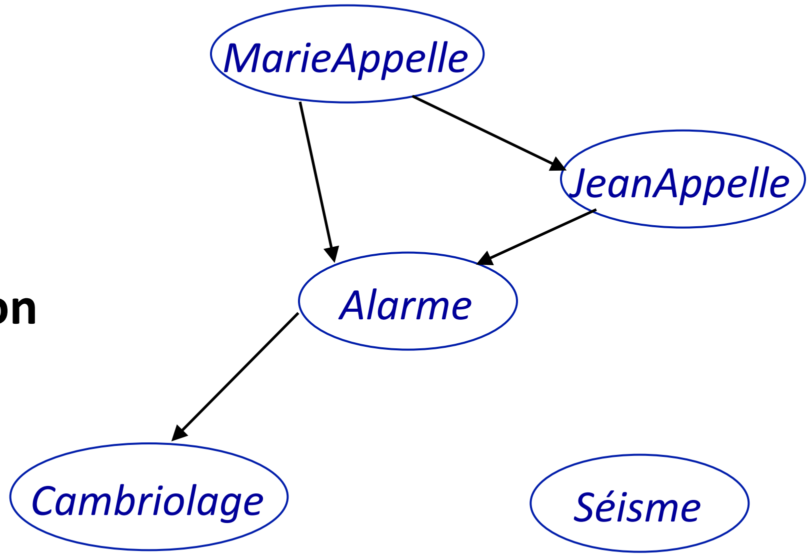
- Supposons qu'on ordonne les variables comme suit : *M*, *J*, *A*, *C*, *S*



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

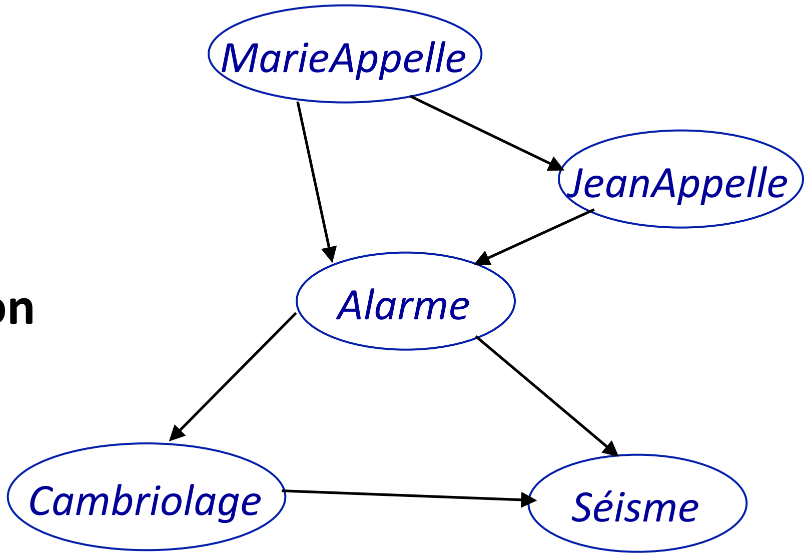
- Supposons qu'on ordonne les variables comme suit : *M, J, A, C, S*



- $P(J|M) = P(J)$? **Non**
- $P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? **Non**
- $P(C|A,J,M) = P(C|A)$? **Oui**
- $P(C|A,J,M) = P(C)$? **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, des études ont démontré que les experts préfèrent donner des probabilités dans le sens causal (pathologie → symptôme) plutôt que dans le sens diagnostique

- Un réseau avec des dépendance 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

Apprentissage dans un RB : paramètres

- Supposons que le graphe d'un RB ait été spécifiée par un expert
- Comment estimer les tables de probabilités $P(X_i \mid Parents(X_i))$?
- Si on a un ensemble de données **où tous les noeuds $\boldsymbol{X_i}$ sont observés**, c'est facile :

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

- 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 freq(x,e) = δ à une petite constante δ (ex. : δ=1)

Apprentissage dans un RB : paramètres

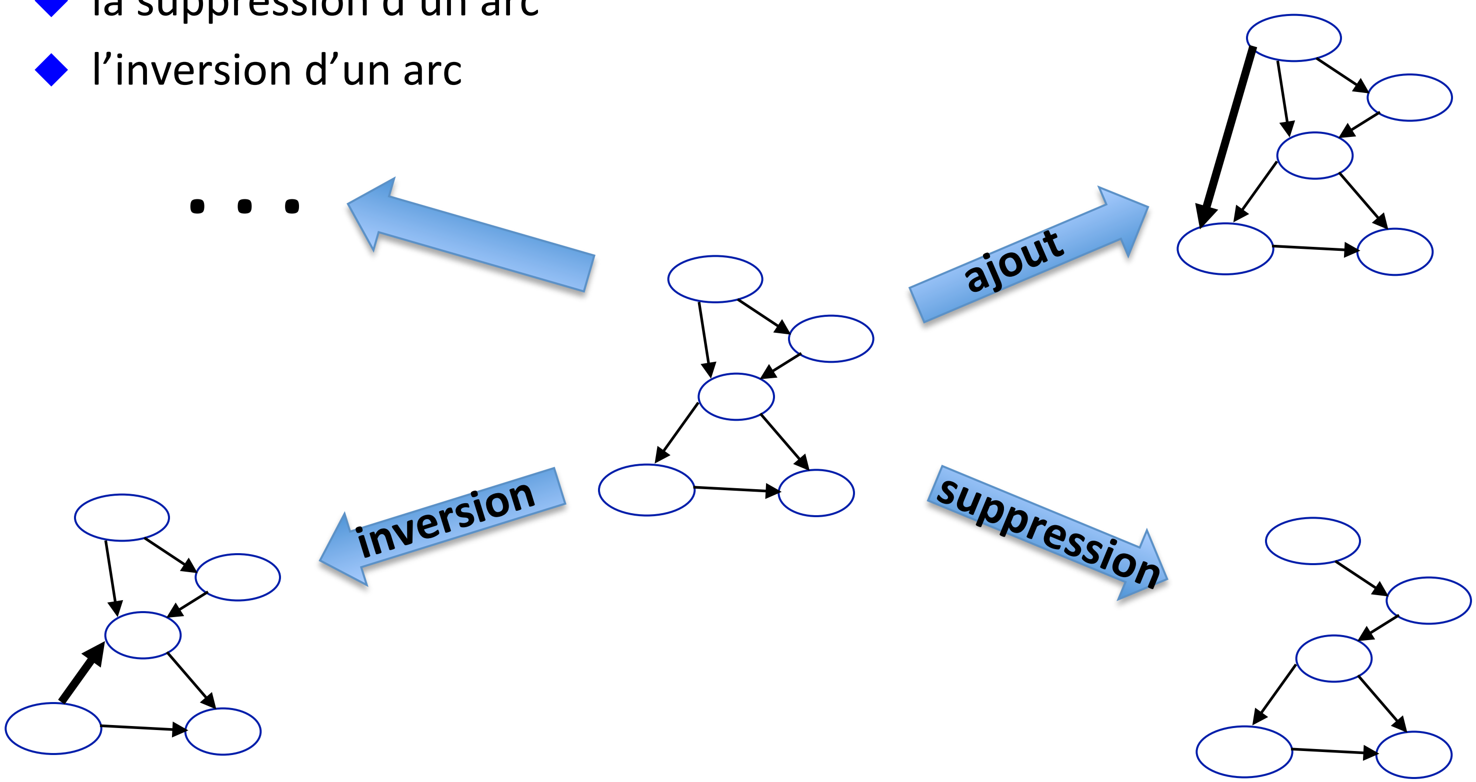
- Supposons que le graphe d'un RB ait été spécifiée par un expert
- Comment estimer les tables de probabilités $P(X_i \mid Parents(X_i))$?
- Si on a un ensemble de données **où certains des noeuds ne sont pas observés**, on doit utiliser des méthodes plus sophistiquées
    - Algorithme EM (voir section 20.3.2)

Apprentissage dans un RB : graphe

- 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*) :
  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 »
  6. on retourne à 2. jusqu'à un certain critère d'arrêt

Apprentissage dans un RB : graphe

- 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



Apprentissage dans un RB : graphe

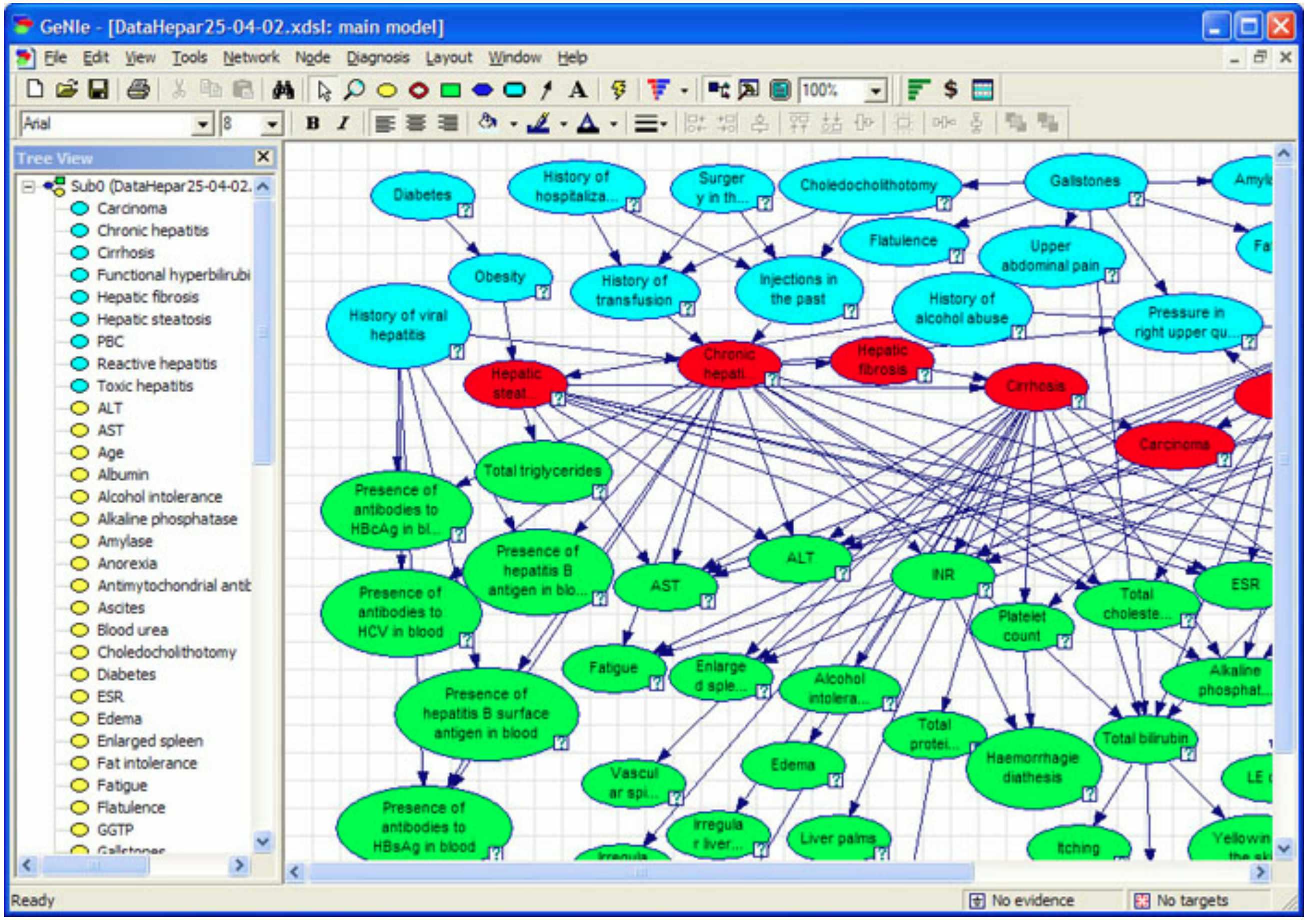
- La fonction objectif à maximiser par la recherche locale est :

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

  - $\{x_1^t, \ldots, x_n^t\}$ 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)
- Pour en savoir plus : voir section 20.2.5

Logiciels

- SMILE / GeNIe (http://genie.sis.pitt.edu/)



Résumé

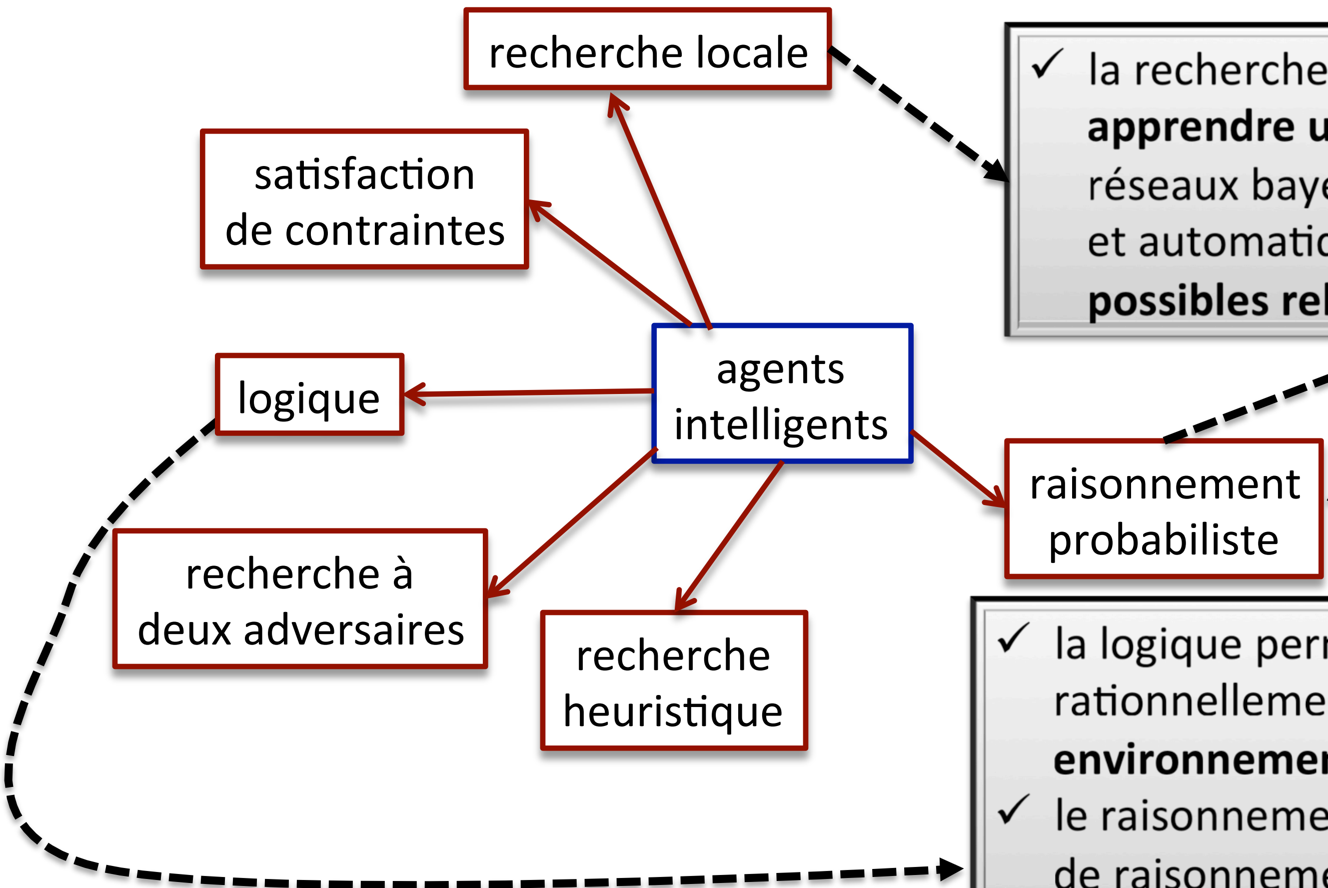
- Un RB est un graphe orienté, acyclique, représentant des connaissances causales, et reflétant les dépendances conditionnelles entre des variables

- La topologie du réseau (arcs entres 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

Objectifs du cours

Algorithmes et concepts



- ✓ la recherche locale peut servir à **apprendre une bonne structure** de réseaux bayésien à partir de données et automatiquement **découvrir de possibles relations causes/effets**

- ✓ la logique permet de raisonner rationnellement dans un **environnement déterministe**
- ✓ le raisonnement probabiliste permet de raisonner rationnellement dans un **environnement stochastic**

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 comment on
  - apprend les paramètres d'un réseau bayésien à partir de données
  - apprend la structure d'un réseau bayésien à partir de données