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

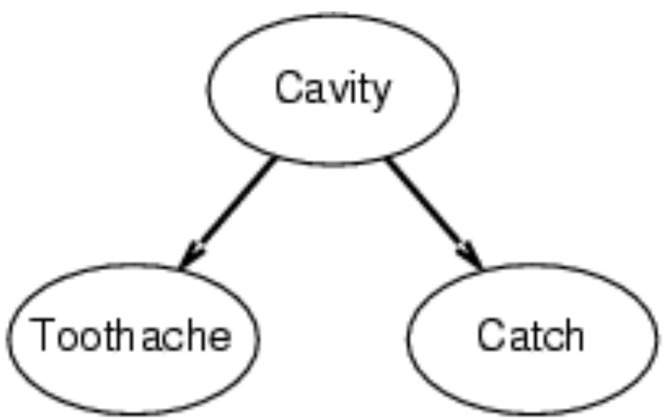
- C’est quoi un réseau bayésien (RB)?
  - structure d’un RB
  - signification
- Indépendance conditionnelle dans un RB
- Inférence dans un réseau bayésien
  - inférence exacte
  - inférence approximative
- Diagrammes d’influence

Réseaux bayésiens

- Les 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)
  - 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



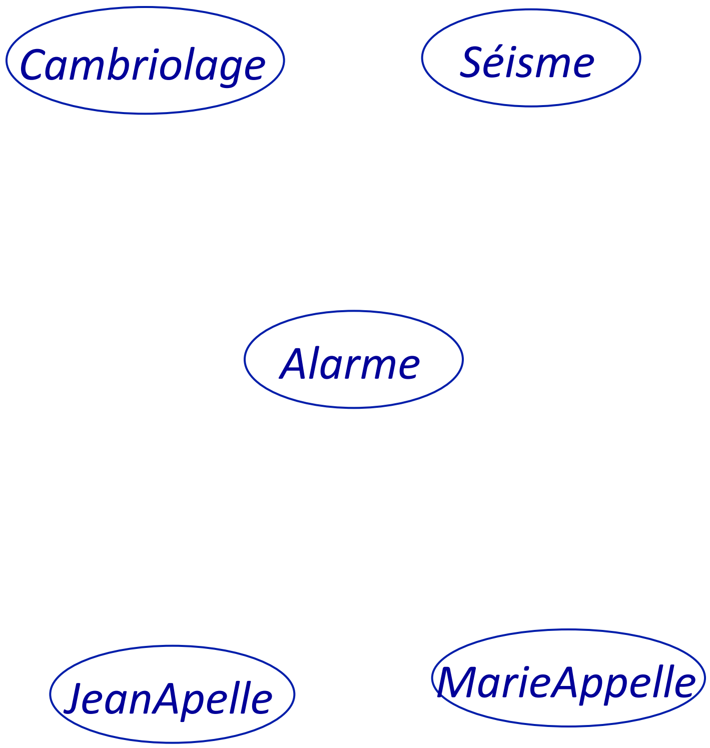
Exemple

- 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
    - 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

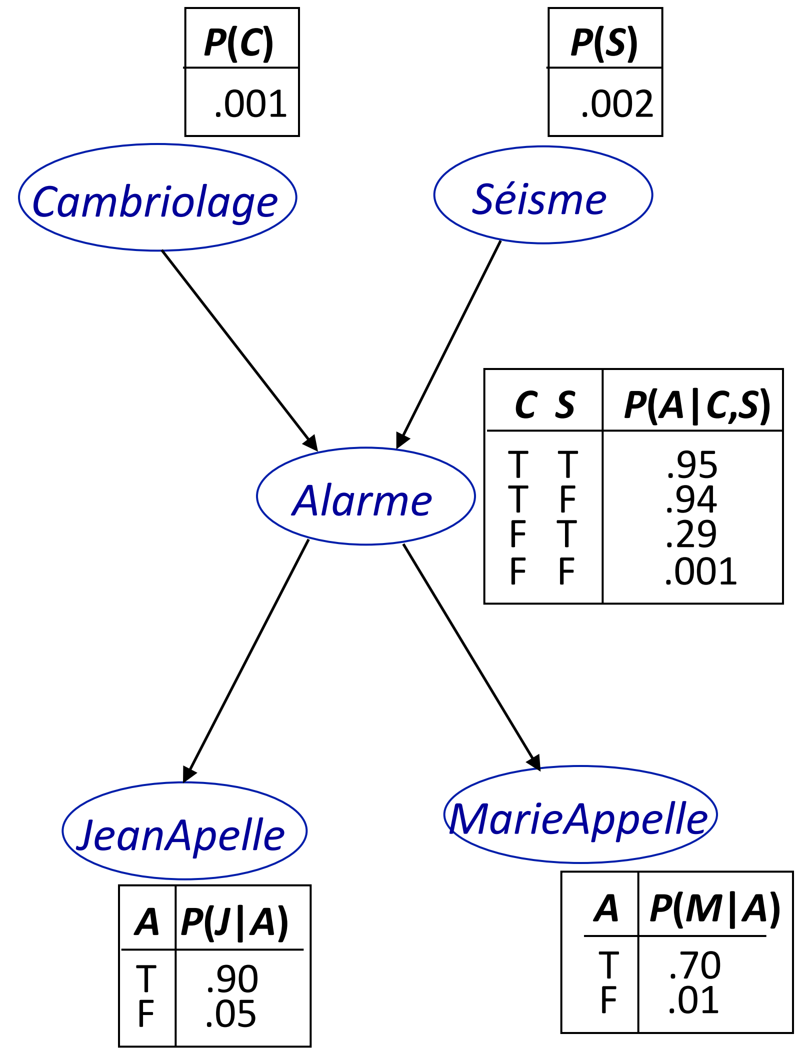
Exemple

- Variables aléatoires:
  - *Cambriolage*
  - *Séisme*
  - *Alarme*
  - *JeanAppelle*
  - *MarieAppelle*



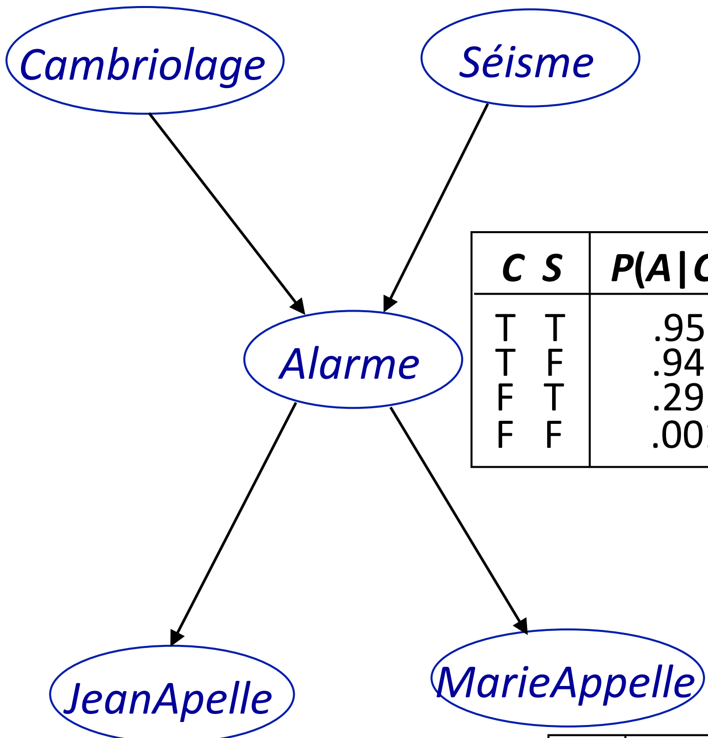
Exemple

- 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**)



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**
- Si *X* est une variable observée, ont dit que c’est une observation (*evidence*)



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

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 $\mathbf{P}(X_1 | X_2)$
  - $\mathbf{P}(X_1,X_2) = \mathbf{P}(X_1 \mid X_2)\, \mathbf{P}(X_2)$
- Soit $X = \{X_1, \ldots, X_n\}$, l'ensemble des variables d'un RB:
  - $\mathbf{P}(X_1, \ldots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$
- En d'autres mots, la distribution conjointe des variables d'un RB est définie comme étant le produit des distributions conditionnelles (locales)

Sémantique

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

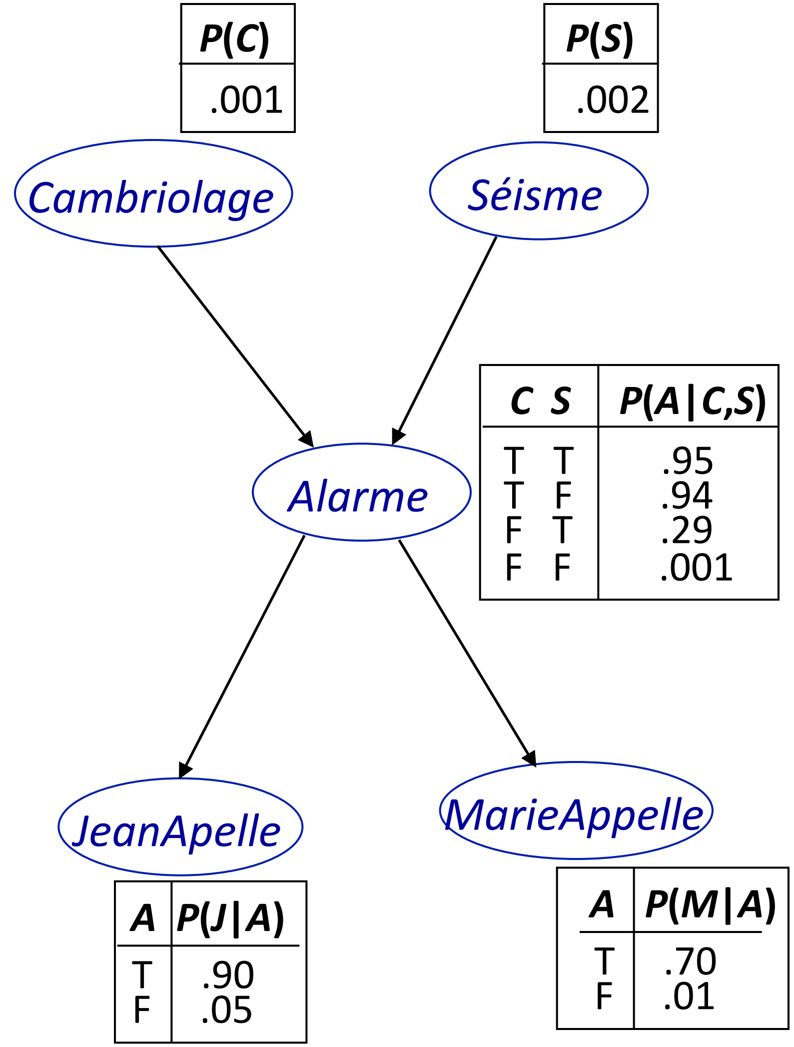
$$\begin{aligned} \mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_n \mid X_{n-1}, \ldots, X_1)\, \mathbf{P}(X_{n-1}, \ldots, X_1) \\ &= \mathbf{P}(X_n \mid X_{n-1}, \ldots, X_1)\, \mathbf{P}(X_{n-1} \mid X_{n-2}, \ldots, X_1) \ldots \mathbf{P}(X_2 \mid X_1)\, \mathbf{P}(X_1) \\ &= \Pi_{i=1}^{n}\, \mathbf{P}(X_i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

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

Exemple

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

- $P(j, m, a, \neg c, \neg s)$
  $= P(j \mid a)\, P(m \mid a)\, P(a \mid \neg c, \neg s)$
  $P(\neg c)\, P(\neg s)$
  $= .90 \times .70 \times .001 \times$
  $0.999 \times .998$
  $= .00062$



P(C)
.001

P(S)
.002

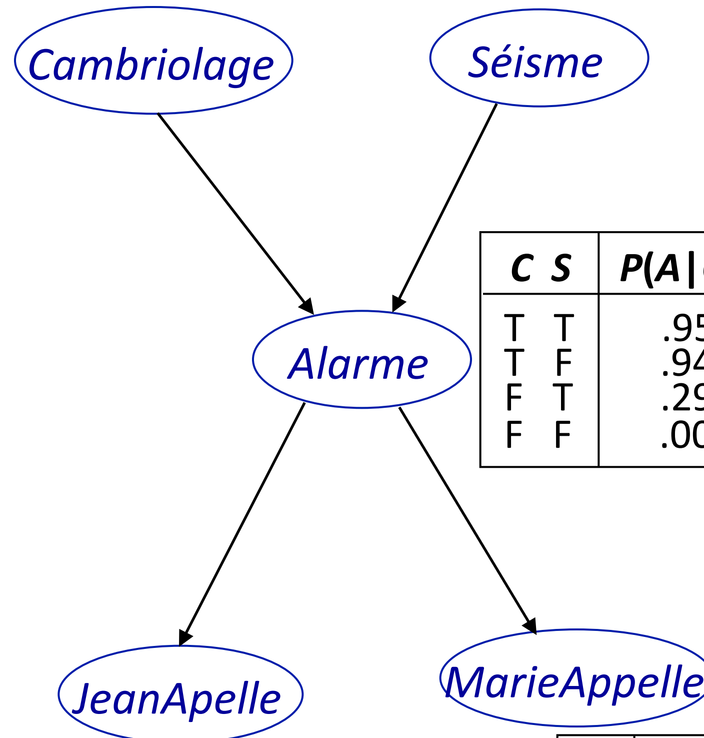
C	*S*	*P(A|C,S)*
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

Construction d'un RB

- Il y a un arc de *Y* vers *X* si *Y* influence directement *X*
- S'il y a un arc d'un nœud *Y* vers un nœud *X*, on dit que:
  - *Y* donne le support causal à *X*
  - *X* donne le support diagnostique à *Y*



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

Construction d’un RB

- Pour construire un RB correct, on s’assure que chaque nœud est indépendant de tous ses prédécesseurs, étant donnés ses parents
  - en d’autres mots, les parents du nœud X_i devraient être tous les nœuds dans $\{X_1, \ldots, X_{i-1}\}$ qui influencent/causent directement X_i

- Dans quel ordre ajouter les nœuds au réseau?
  - mettre les « causes racines » d’abord, ensuite les nœuds qu’ils influencent directement

Construction d'un RB

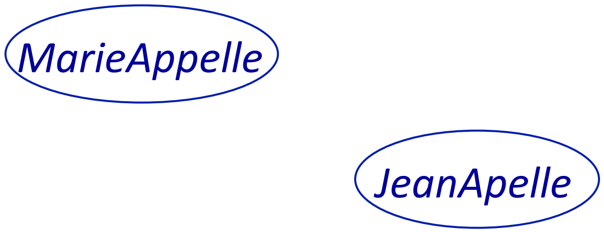
1. Choisir un ordre des variables $X_1, \ldots, X_n$
2. Pour i = 1 to n:
   - ajouter X_i au réseau
   - choisir les parents $X_1, \ldots, X_{i-1}$ tel que $\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots X_{i-1})$
   - Ce choix garantit que:

$$P(X_1, \ldots, X_n) = \Pi_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1}) \quad (\textit{chain rule})$$

$$= \Pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i)) \quad (\text{par construction})$$

Exemple

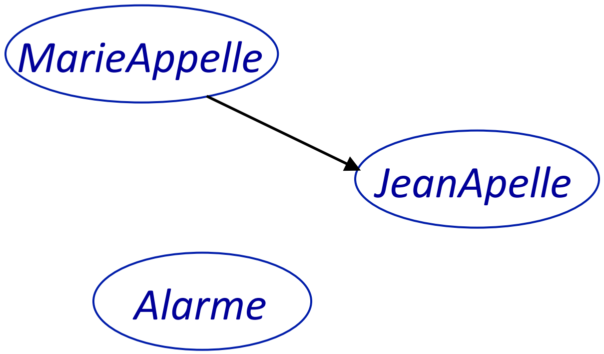
- Supposons qu'on ordonne les variables comme suit: M, J, A, C, S



- $\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

Exemple

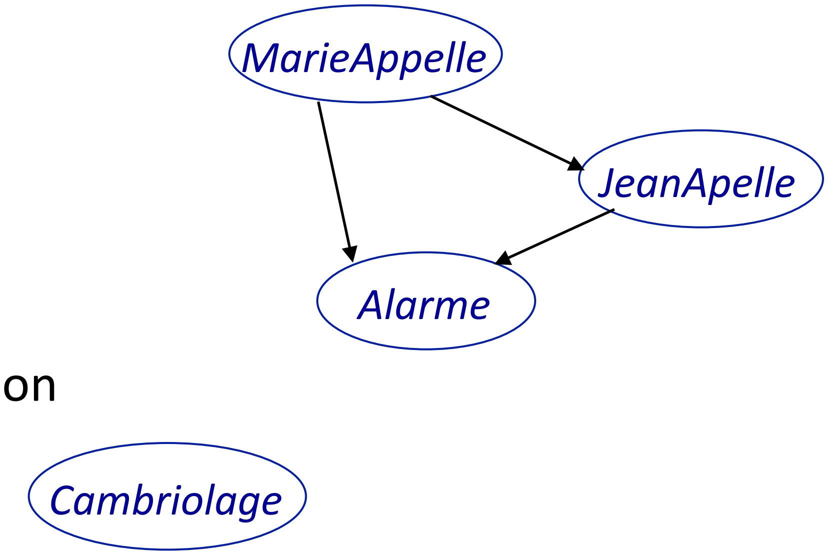
- Supposons qu'on ordonne les variables comme suit: *M*, *J*, *A*, *C*, *S*



- $\mathbf{P}(J|M) = \mathbf{P}(J)$?
- Non
- $\mathbf{P}(A|J,M) = \mathbf{P}(A|J)$? $\mathbf{P}(A|J,M) = \mathbf{P}(A)$?

Exemple

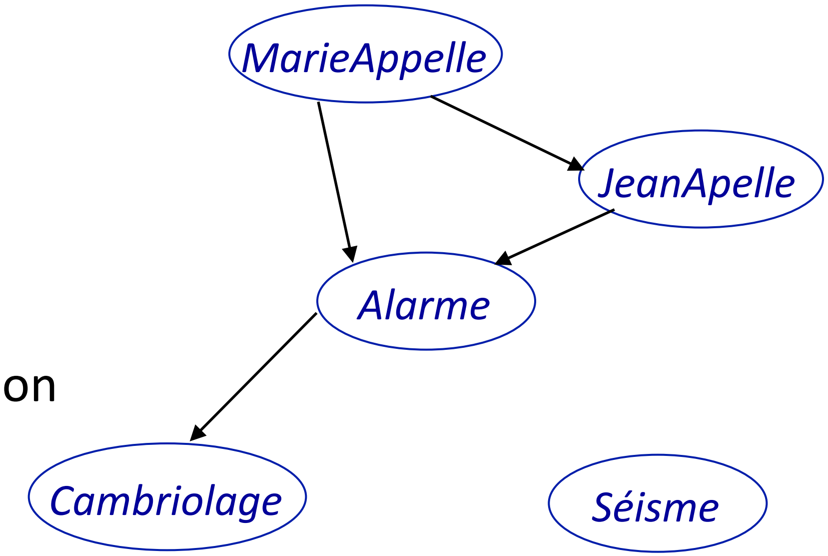
- Supposons qu'on ordonne les variables comme suit: *M*, *J*, *A*, *C*, *S*



- $\mathbf{P}(J \mid M) = \mathbf{P}(J)$?
- Non
- $\mathbf{P}(A \mid J,M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J,M) = \mathbf{P}(A)$? Non
- $\mathbf{P}(C \mid A,J,M) = \mathbf{P}(C \mid A)$?
- $\mathbf{P}(C \mid A,J,M) = \mathbf{P}(C)$?

Exemple

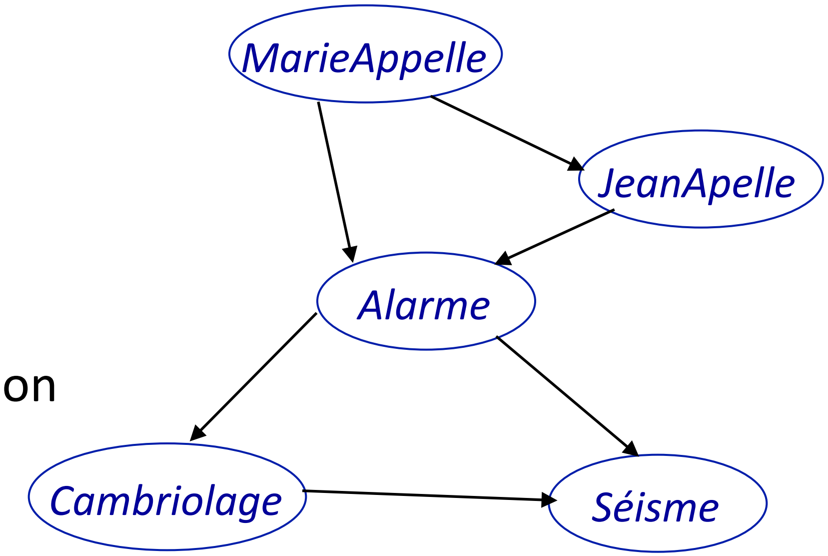
- Supposons qu'on ordonne les variables comme suit: *M*, *J*, *A*, *C*, *S*



- $\mathbf{P}(J|M) = \mathbf{P}(J)$?
- Non
- $\mathbf{P}(A|J,M) = \mathbf{P}(A|J)$? $\mathbf{P}(A|J,M) = \mathbf{P}(A)$? Non
- $\mathbf{P}(C|A,J,M) = \mathbf{P}(C|A)$? Oui
- $\mathbf{P}(C|A,J,M) = \mathbf{P}(C)$? Non
- $\mathbf{P}(S|C,A,J,M) = \mathbf{P}(S|A)$?
- $\mathbf{P}(S|C,A,J,M) = \mathbf{P}(S|A,C)$?

Exemple

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



- $\mathbf{P}(J\,|\,M) = \mathbf{P}(J)$?
- Non
- $\mathbf{P}(A\,|\,J,M) = \mathbf{P}(A\,|\,J)$? $\mathbf{P}(A\,|\,J,M) = \mathbf{P}(A)$? Non
- $\mathbf{P}(C\,|\,A,J,M) = \mathbf{P}(C\,|\,A)$? Oui
- $\mathbf{P}(C\,|\,A,J,M) = \mathbf{P}(C)$? Non
- $\mathbf{P}(S\,|\,C,A,J,M) = \mathbf{P}(S\,|\,A)$? Non
- $\mathbf{P}(S\,|\,C,A,J,M) = \mathbf{P}(S\,|\,A,C)$? Oui

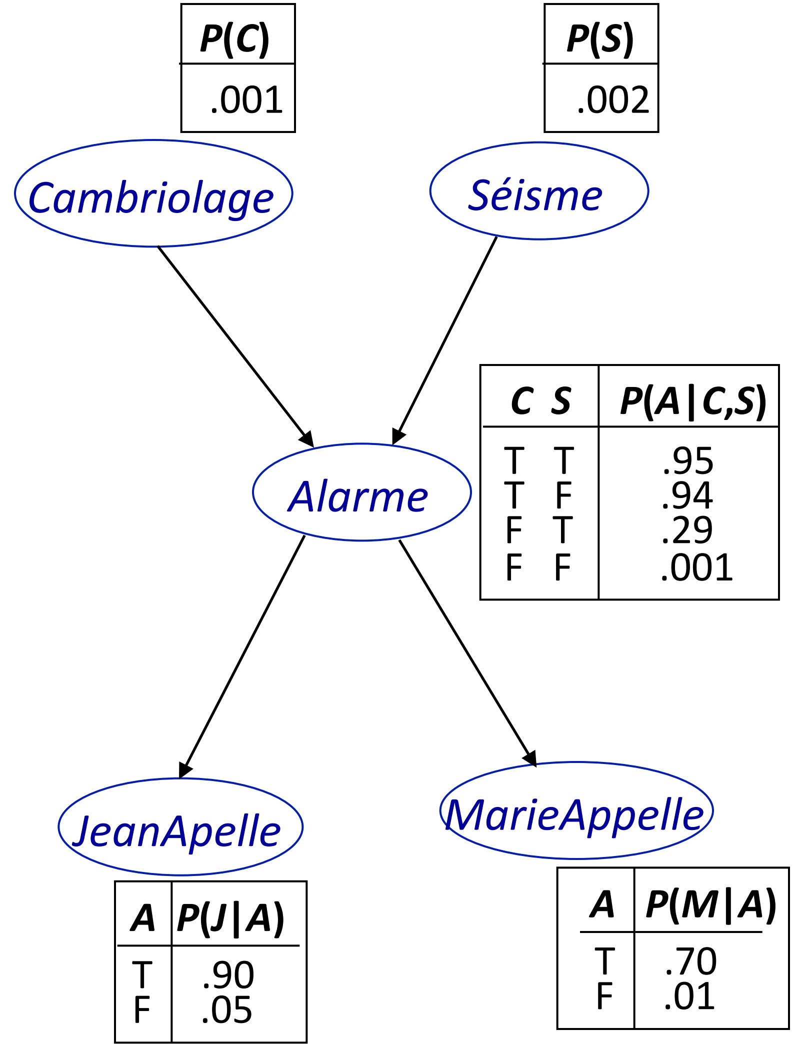
Exemple

- 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

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**



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

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

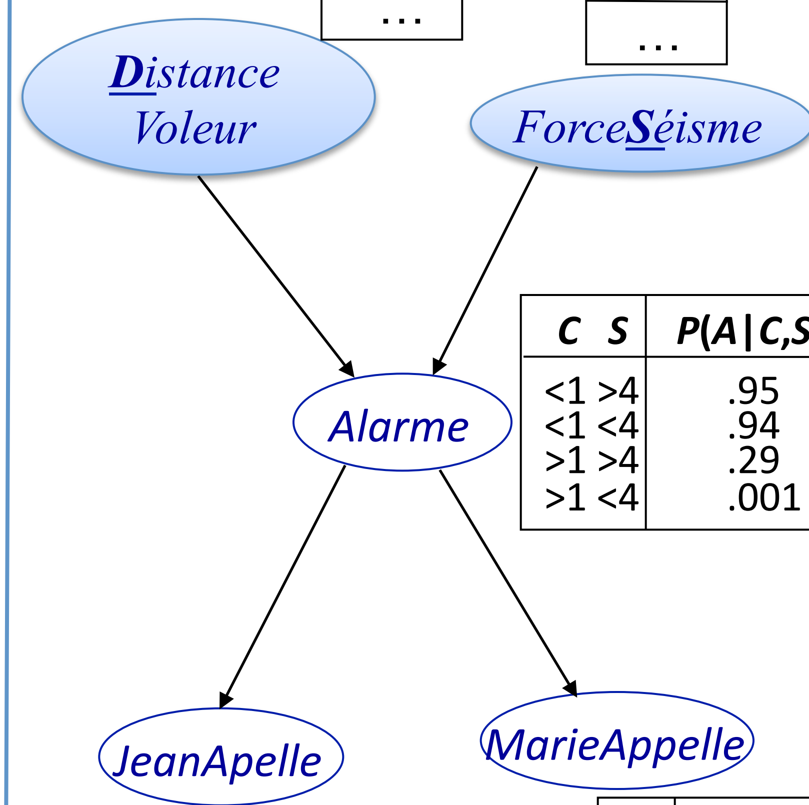
C	***S***	***P*(*A*|*C*,*S*)**
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

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

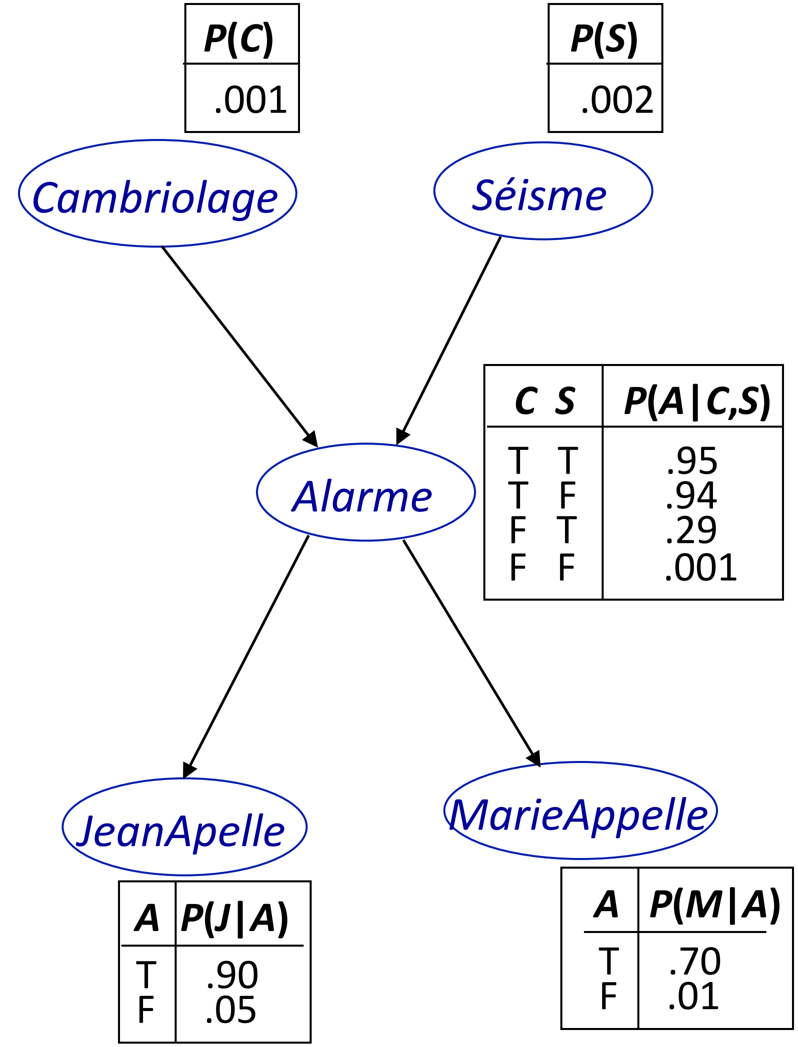
A	*P(J|A)*
T	.90
F	.05

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

Indépendance conditionnelle dans un RB (1)

- 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**
    - » mais ils sont **indépendants** étant donné *Alarme*:

      $\mathbf{P}(M|A,C) = \mathbf{P}(M|A)$
    - » si *A* est connu, *C* n'intervient pas dans le calcul



P(*C*)
.001

P(*S*)
.002

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

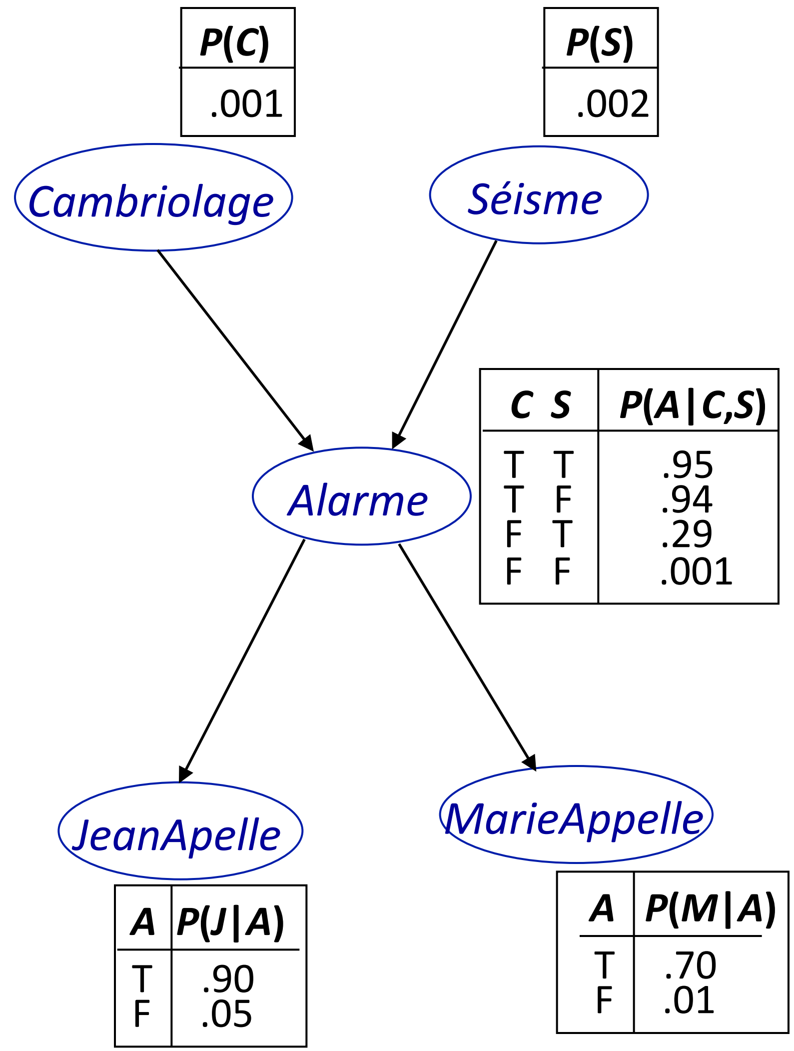
A	*P*(*J*|*A*)
T	.90
F	.05

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

Indépendance conditionnelle dans un RB (1)

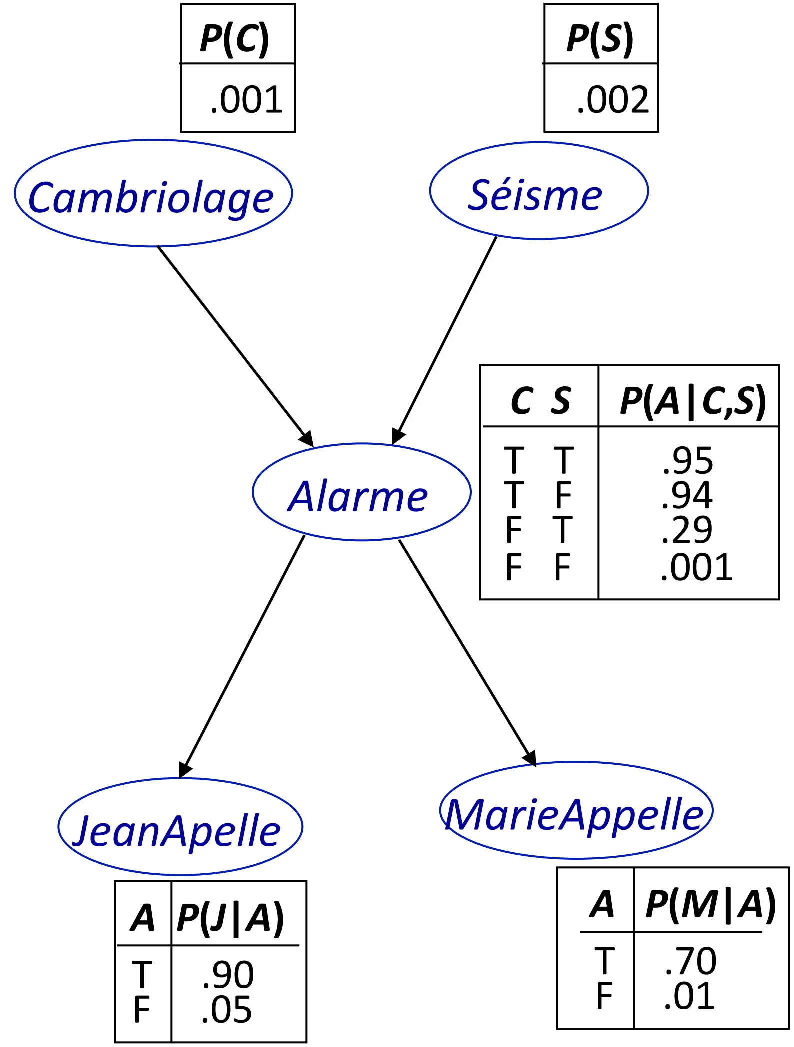
- Un RB modélise les relations d'indépendance conditionnelle suivantes
  - un nœud est indépendant de ses non-descendants, étant donné ses parents
  - exemples:
    - » *JeanAppelle* et *MarieAppelle* sont **dépendants**.
    - » mais ils sont **indépendant** étant donné *Alarme*:

      $\mathbf{P}(J|A,M) = \mathbf{P}(J|A)$
    - » si *A* est connu, *M* n'intervient pas dans le calcul



Indépendance conditionnelle dans un RB (1)

- 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**
    - » 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



***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

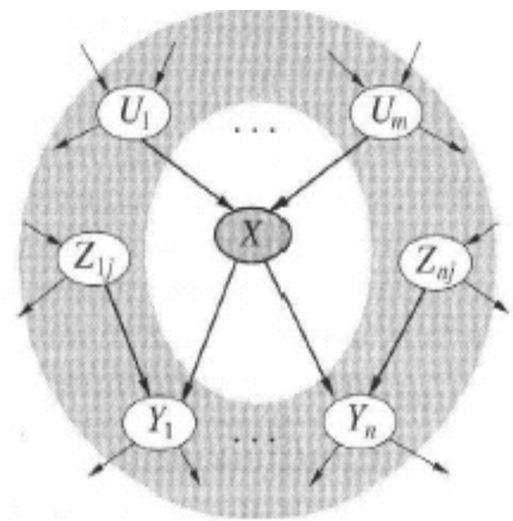
A	***P*(*M*|*A*)**
T	.70
F	.01

Indépendance conditionnelle dans un RB (2)

- Soit la **couverture de Markov** (***Markov blanket***) *MB*(*X*) d'un noeud *X*, c'est à dire:
  - les parents de *X*
  - les enfants de *X*
  - et les parents des enfants de *X*
- Le noeud *X* est conditionnellement indépendant des autres noeuds (hors de la couverture de Markov), étant donné les noeuds de la couverture de Markov:

  $\mathbf{P}(X|MB(X),Others) = \mathbf{P}(X|MB(X))$

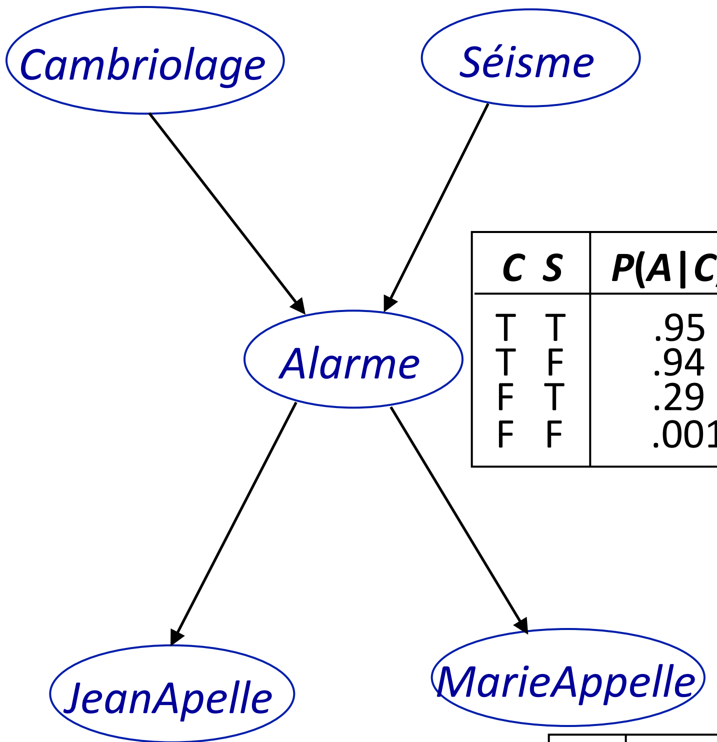
Couverture de Markov de *X*



Indépendance conditionnelle dans un RB (3)

- **D-séparation**: critère pour décider si un nœud *X* est indépendant d'un nœud *Y*, étant donné un autre nœud *Z*

- Ce cas est non traité ici



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

Requête dans un RB

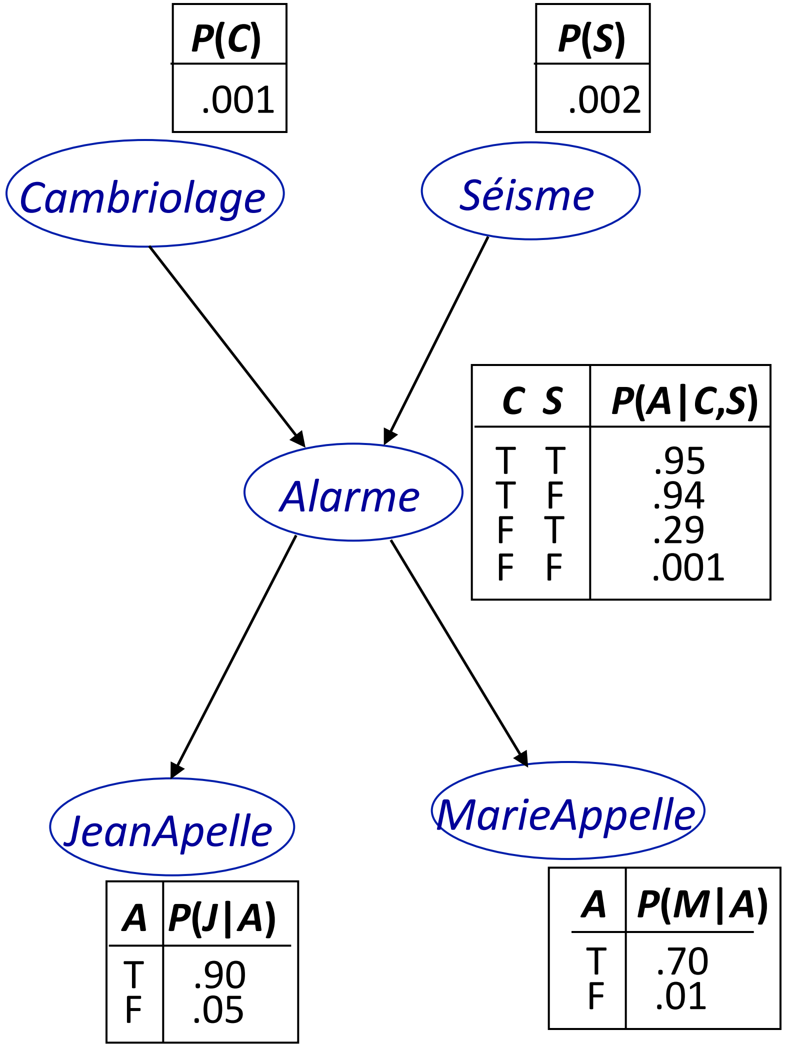
- L'usage principal d'un RB est de calculer la distribution de probabilités a posteriori pour un ensemble de variables, étant donné un événement observé
- Un événement est une assignation de valeurs à certaines variables d'observation
- Soit
  - *X* l'ensemble de variables pour lesquelles on fait une requête
  - *E* les variables d'observation et
  - *Y* les variables cachées (qui ne sont pas observées)
- 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* | *JeanApelle = true*,
  MarieAppelle = true)
  = <0.287,0.716>

- Comment fait-on un tel calcul?
  - **inférence exacte** (prohibitif)
    - » **par énumération**
    - » **élimination de variables**
  - **inférence approximative par échantillonnage** avec les méthodes Monte-Carlo (plus efficace)
    - » **échantillonnage direct**
    - » **échantillonnage par chaînes de Markov**



Rappel de notions de base en probabilités

- $\mathbf{P}(X,Z) = \mathbf{P}(X|Z)\ \mathbf{P}(Z)$
- $\mathbf{P}(Z,X) = \mathbf{P}(Z|X)\ \mathbf{P}(X)$

- On en déduit
  - $\mathbf{P}(X|Z) = \mathbf{P}(X,Z)\ /\ \mathbf{P}(Z)$
  - $\mathbf{P}(X|Z) = \mathbf{P}(Z|X)\ \mathbf{P}(X)\ /\ \mathbf{P}(Z)$ (règle de Bayes)
- **Marginalisation:** si on a une distribution conjointe $\mathbf{P}(Z,Y)$ on peut calculer la distribution $\mathbf{P}(Z)$ par la somme des probabilités pour toutes les valeurs possibles de Y: $\mathbf{P}(Z) = \Sigma_y\ \mathbf{P}(Z, Y = y)$
- Si on a une distribution conditionnelle $\mathbf{P}(Z|Y)$: $\mathbf{P}(Z) = \Sigma_y\ \mathbf{P}(Z \mid Y = y)\ \mathbf{P}(Y = y)$
- $\mathbf{P}(Z)$ peut donc être considéré comme un facteur α

Rappel de notions de base en probabilités

- Ceci nous donne
  - $\mathbf{P}(X|Z) = \alpha\ \mathbf{P}(X,Z)$
  - α est une constante de normalisation pour s'assurer que la somme des probabilités de la distribution $\mathbf{P}(X,Z)$ soit égale à 1
- De manière générale, soit
  - X l'ensemble de variables pour laquelle on fait la requête
  - E les variables d'observation
  - Y les variables cachées (qui ne sont pas observées)
  - e les valeurs observées pour les variables dans E
- $\mathbf{P}(X|E{=}e) = \alpha\ \mathbf{P}(X,E{=}e) = \alpha\ \Sigma_y\ \mathbf{P}(X, E{=}e, Y = y)$
- Noté aussi $\mathbf{P}(X|e) = \alpha\ \Sigma_y\ \mathbf{P}(X, e, y)$

Inférence par énumération

- On a vu que: $\mathbf{P}(X|e) = \alpha \Sigma_y \mathbf{P}(X, e, y)$
- On a vu aussi que selon la sémantique d'un RB

  $\mathbf{P}(X_1, \ldots, X_n) = \Pi^{n}_{i=1} \mathbf{P}(X_i \mid Parents(X_i))$
- Les termes $\mathbf{P}(X, e, y)$ peuvent donc s'écrire comme le produit des probabilités conditionnelles du réseau
- En d'autre termes, on peut calculer la réponse à une requête $\mathbf{P}(X|e)$ dans un RB, simplement en calculant les sommes des produits des probabilités conditionnelles du RB

Inférence par énumération

function ENUMERATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* **Y** = *hidden variables* */

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{x_i}$)
      where $\mathbf{e}_{x_i}$ is **e** extended with $X = x_i$
  return NORMALIZE($\mathbf{Q}(X)$)

---

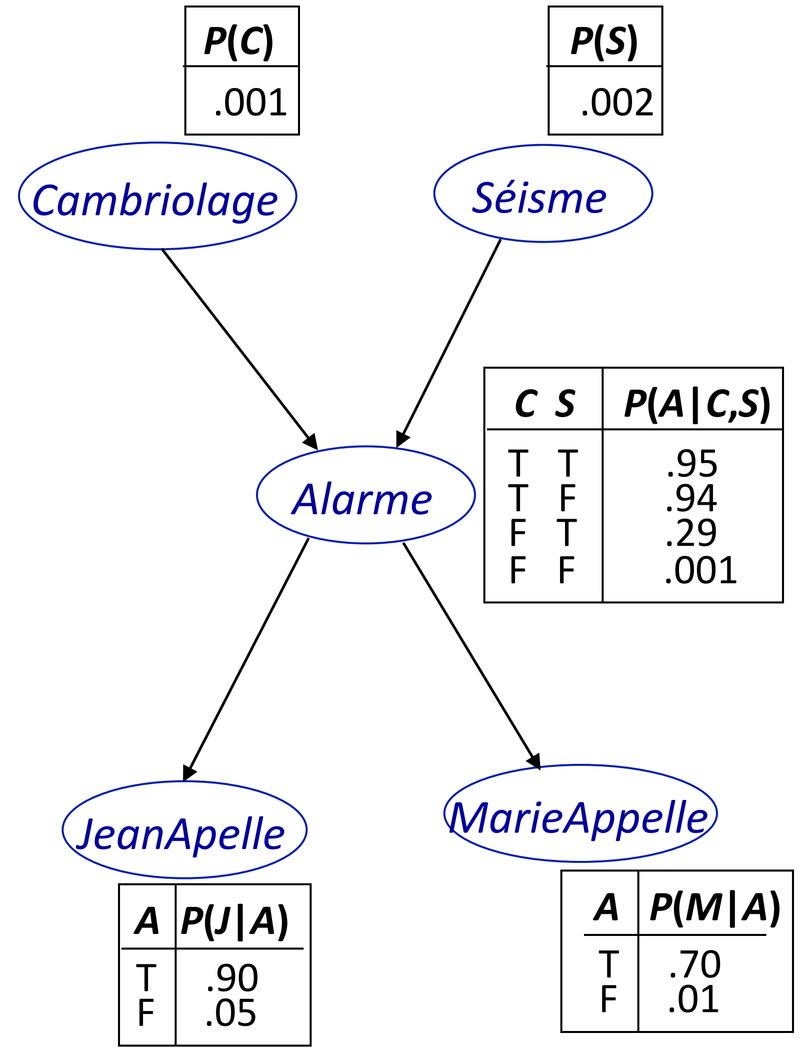
function ENUMERATE-ALL($vars$, **e**) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $Y \leftarrow$ FIRST($vars$)
  if Y has value y in **e**
    then return $P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), **e**)
    else return $\sum_y P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)
      where $\mathbf{e}_y$ is **e** extended with $Y = y$

Exemple

- **P**(*Cambriolage* | *JeanApelle = true*, *MarieAppelle = true*)
- Noté **P**(C | j, m)
- Les variables cachées sont *Séisme* et *Alarme*

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

- Note:
  - s et a prennent toutes les valeurs possibles de $S=s$ et $A=a$ variables
  - ne pas confondre avec j et m qui sont des évidences fixes ($J=j$ et $M=m$)



Exemple

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

$P(c \mid j, m)$

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

$= \alpha$ (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)

- Et $C = false$

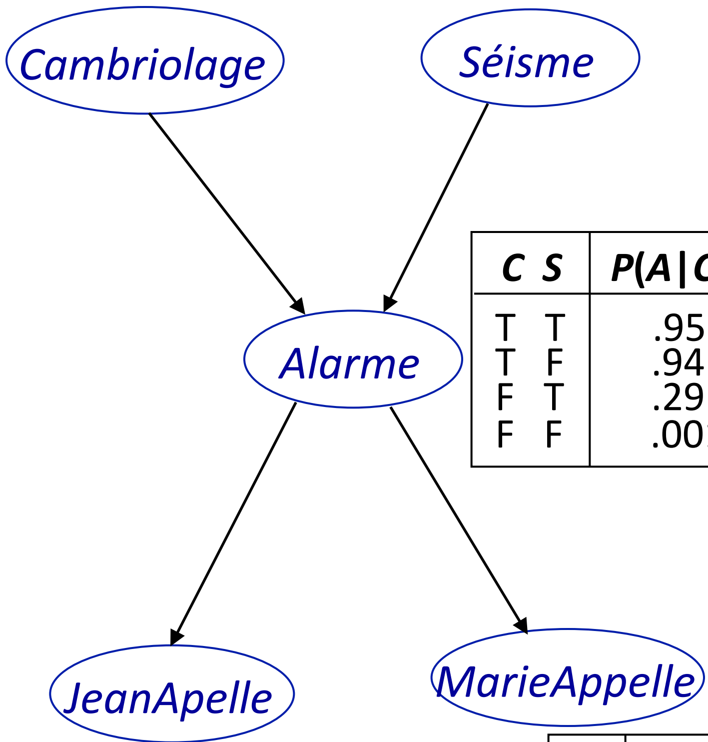
$P(\neg c \mid j, m)$

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

$= \alpha$ (0.0014919)

$\alpha = 1/(0.00059224 + 0.0014919)$

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



$P(C)$
.001

$P(S)$
.002

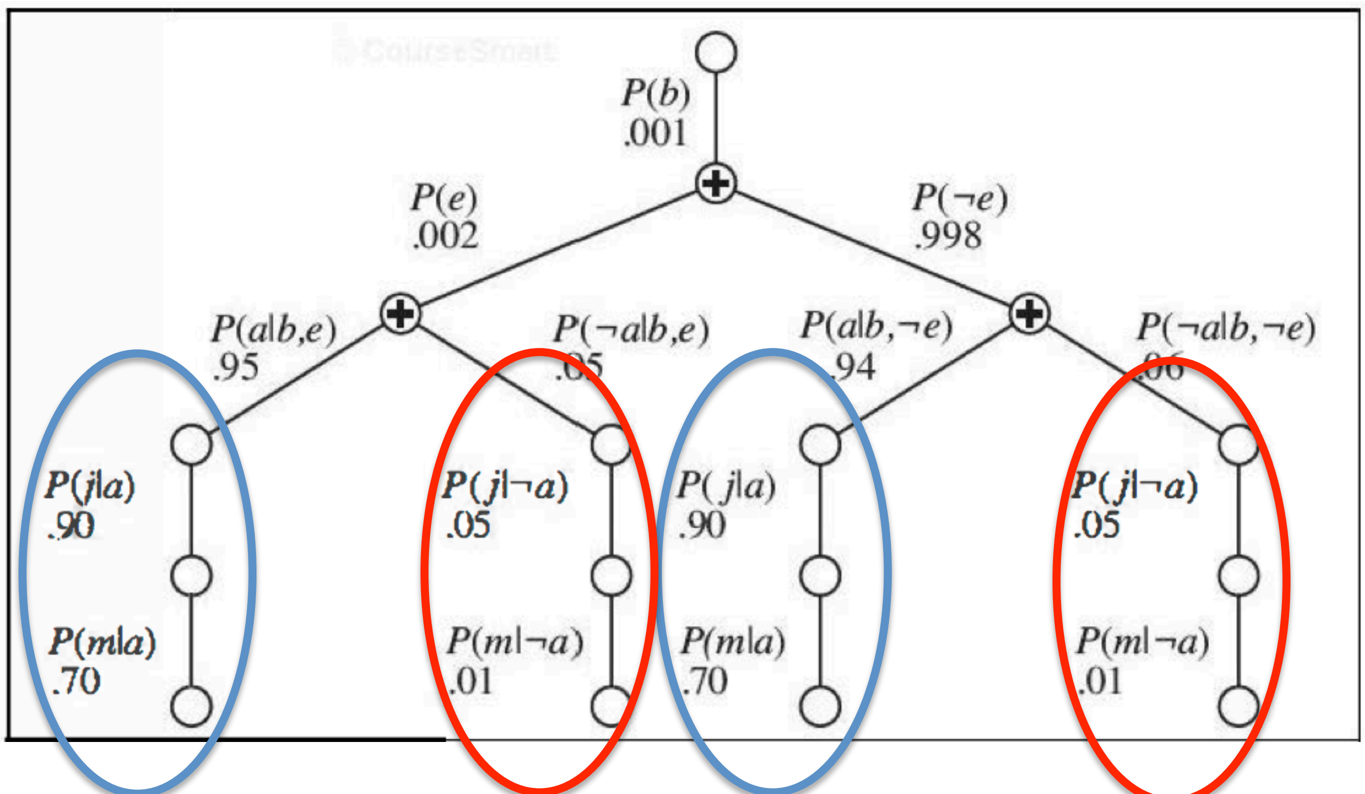
C	S	$P(A\|C,S)$
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

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

Inférence par échantillonnage direct

- Simuler des observations complètes du RB
- Estimer les distributions de probabilités à partir de la fréquence des observations échantillonnées

  $P(X{=}x|e) = \alpha\, \Sigma_y\, P(X{=}x, e, y) \approx \beta\, \text{freq}(x,e,y)$

  où freq(x,e,y) est le nombre de fois que $X{=}x$, $E{=}e$ et $Y{=}y$ a été échantillonné
- 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

Méthode de rejet

function PRIOR-SAMPLE(bn) **returns** an event sampled from the prior specified by bn
inputs: bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

x ← an event with n elements
foreach variable X_i **in** $X_1, \ldots, X_n$ **do**
  x$[i]$ ← a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
return x

function REJECTION-SAMPLING(X, **e**, bn, N) **returns** an estimate of $\mathbf{P}(X|\mathbf{e})$
inputs: X, the query variable
  e, observed values for variables **E**
  bn, a Bayesian network
  N, the total number of samples to be generated
local variables: **N**, a vector of counts for each value of X, initially zero

for $j = 1$ **to** N **do**
  x ← PRIOR-SAMPLE(bn)
  if x is consistent with **e then**
    N$[x]$ ← **N**$[x]$+1 where x is the value of X in **x**
return NORMALIZE(**N**)

Exemple 1: Évaluation par énumérations

<u>**Requête**</u>:

Calculer <u>P</u>(T=*true*|F=*false*, M=*true*)

<u>Variables connues:</u>

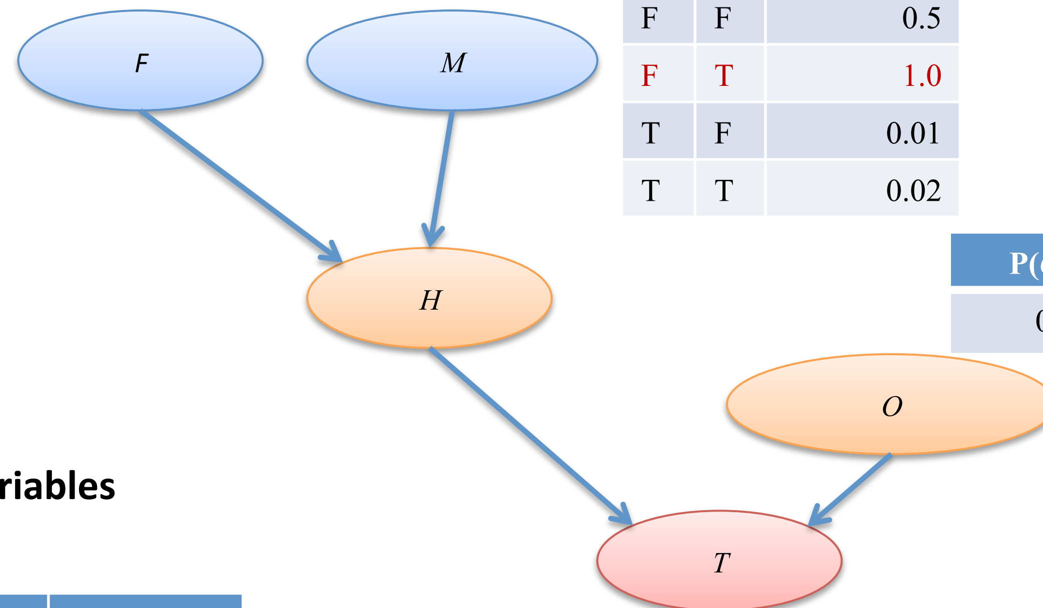
F = false

M = true

<u>Variables inconnues:</u>

H

O



F	*M*	P(*H*|*F*,*M*)
F	F	0.5
F	T	1.0
T	F	0.01
T	T	0.02

P(*O*)
0.6

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

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	T	0.0 * 0.6 * 0.5	0
T	F	1.0 * 0.4 * 0.5	0.20
T	T	1.0 * 0.6 * 1.0	0.60
		TOTAL	**0.80**

Exemple 2: Évaluation par énumérations

__Requête__:

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

<u>Variables connues:</u>

$M = true$

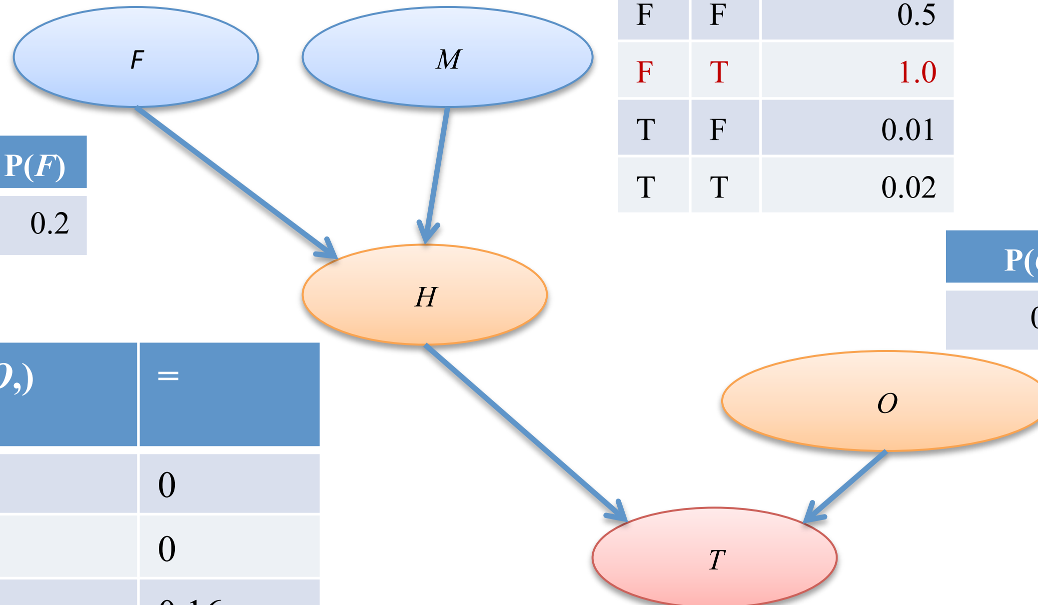
<u>Variables inconnues:</u>

H

O

F

P(*F*)
0.2



F	*M*	P(*H*|*F*,*M*)
F	F	0.5
F	T	1.0
T	F	0.01
T	T	0.02

P(*O*)
0.6

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

Exemple 2: Échantillonage direct

<u>Requête</u>:

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

<u>Variables connues:</u>

$M = true$

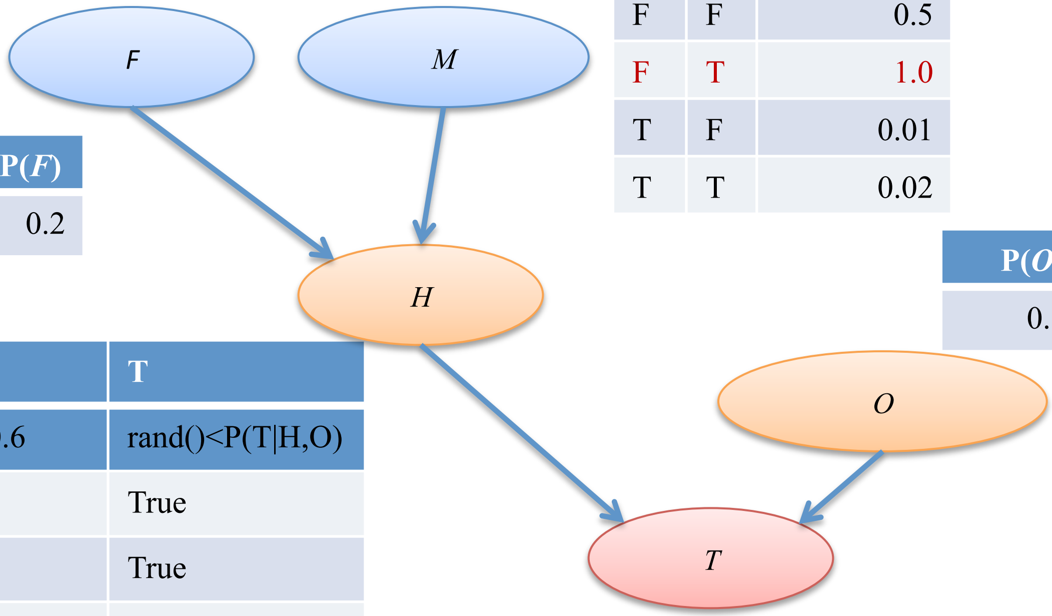
<u>Variables inconnues:</u>

H

O

F

P(*F*)
0.2



F	*M*	P(*H*|*F*,*M*)
F	F	0.5
F	T	1.0
T	F	0.01
T	T	0.02

P(*O*)
0.6

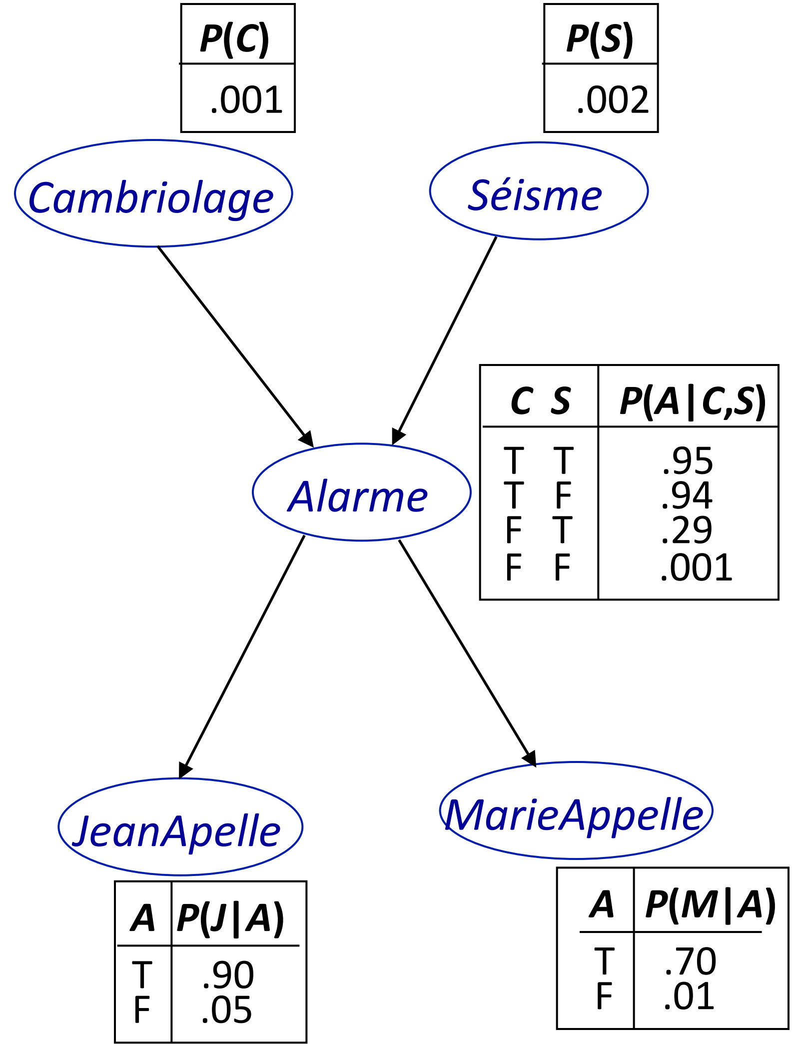
H	*O*	P(*T*|*H*,*O*)
F	F	0.1
F	T	0.5
T	F	0.5
T	T	1.0

	F	H	O	T
#	rand()<0.2	rand()<P(H|F,M)	rand()<0.6	rand()<P(T|H,O)
1	False	True	True	True
2	False	True	True	True
3	False	True	False	False
4	True	False	False	False
5	False	True	True	True
6	False	True	True	True
7	False	True	True	True
8	False	True	False	True
			Average of T=True	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=true*)
  - garder à l'esprit qu'on a des arcs « causes / effets ».
- **Prédiction** (étant données les causes, quels sont les effets)
  - **P**(*JeanAppelle*|*Cambriolage=true*)
- **Probabilité conjointe ou marginale**
  - **P**(*Alarme*)



Apprentissage dans un RB

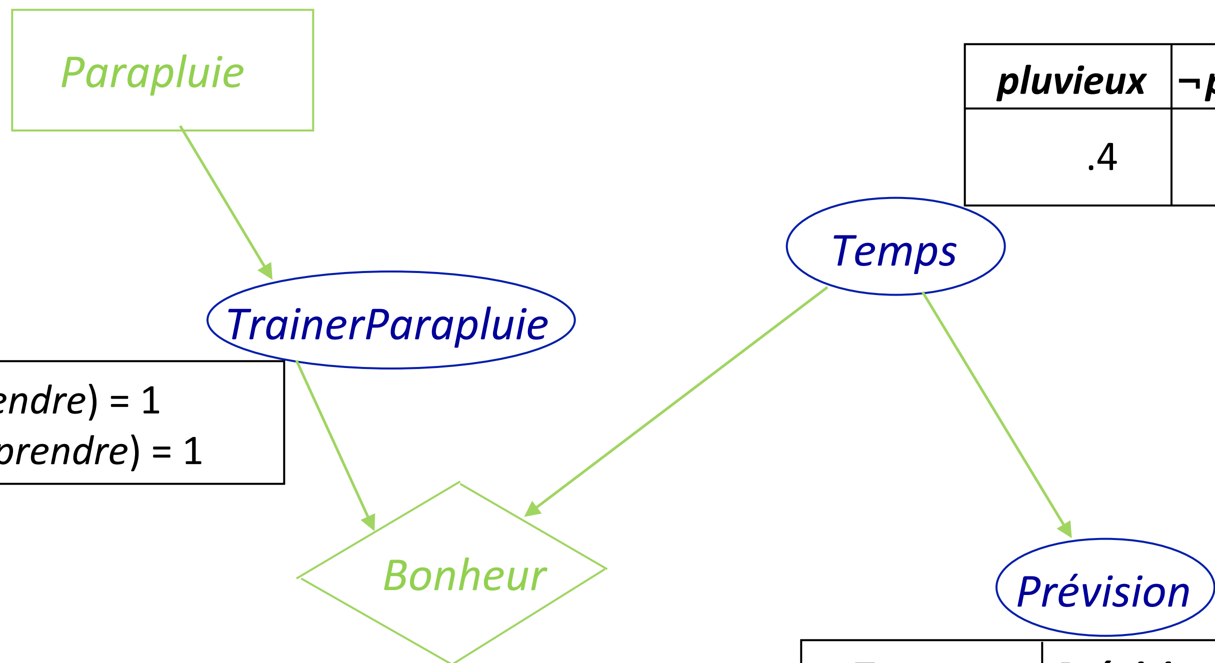
- La structure d'un RB (le graphe) est le plus souvent spécifiée à l'aide d'un expert
- Dans d'autres applications, la structure est générée automatiquement à partir des données statistiques
  - c'est un des problèmes d'apprentissage automatique
- Dans d'autres problèmes, on connaît la structure du RB, mais on ne connaît pas les TPC
  - là aussi, on peut les apprendre à partir des données statistiques
  - c'est un autre problème d'apprentissage automatique.

Diagrammes d'influence

- Un **diagramme d'influence** (DI) est une extension d'un RB avec en plus des **nœuds de décision** et des **nœuds d'utilité**
  - les nœuds habituels d'un RB sont appelés des **nœuds chance**
  - on ajoute:
    - des nœuds de décision représentant une prise de décision
    - des nœuds d'utilité représentant l'utilité (coût ou degré de désirabilité) des nœuds chance influencés par les actions
- Ainsi on peut modéliser des prises de décision simples
  - pour des décisions complexes (séquentielles), les processus de décision markoviens sont généralement préférables
- Les diagrammes d'influence sont aussi appelés réseaux de décision (*decision networks*)

Exemple

Prendre / Ne pas prendre



pluvieux	***¬pluvieux***
.4	.6

$P(trainer \mid prendre) = 1$
$P(\neg trainer, \neg prendre) = 1$

$U(trainer, pluvieux) = 25$
$U(trainer, \neg pluvieux) = -20$
$U(\neg trainer, pluvieux) = -100$
$U(\neg trainer, \neg pluvieux) = 100$

Temps	***Prévision***	**P(*P*|*T*)**
pluvieux	*ensoleillé*	.3
pluvieux	*pluvieux*	.7
¬ pluvieux	*ensoleillé*	.8
¬ pluvieux	*pluvieux*	.2

Évaluation des diagrammes d'influence

- Mettre à jour les variables d'observation
- Pour chaque valeur possible d'un nœud décision
  - change le nœud décision pour lui donner cette valeur
  - calcule les probabilités a posteriori des parents des nœuds utilités (en utilisant un algorithme d'inférence standard pour les RB)
  - calcule l'utilité espérée résultante pour l'action
- Retourne l'action avec la plus grande utilité

- Pour en savoir plus: section 16.5

Valeur de l'information

- Parfois un agent est amené à prendre des décisions sans posséder toute l'information
- Un aspect important de la prise de décision est de déterminer les questions à poser pour chercher de l'information (pour obtenir des observations *E=e*)
- La valeur d'une information pour une action dépend de deux choses:
  - est-ce que l'obtention d'une observation particulière *E=e* augmenterait grandement notre utilité espérée
  - est-ce que cette observation est vraisemblable
- Les inférences sur les DI permettent de déterminer les questions qui apportent le plus d'information, pour chaque action
- Pour en savoir plus: section 16.6

Exemples d'applications

- Microsoft
  - Windows: identification des problèmes d'impression
  - Office: Microsoft Agent
- NASA
  - Support au diagnostique en temps réel des pannes du système de propulsion des navettes spatiales
- Médecine
  - Intellipath: aide au diagnostique des maladies (proposer le diagnostique le plus probable à partir des symptômes; recommander les tests de laboratoires les plus pertinents; recommander les traitements)
- AT&T
  - Détections des fraudes et des mauvais payeurs pour les factures de téléphone

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
- Un diagramme d'influence est un réseau bayésien avec des nœuds de décision et des nœuds d'utilité