INF4230 – Intelligence Artificielle

Réseaux bayésiens

Hiver 2017

Sommaire

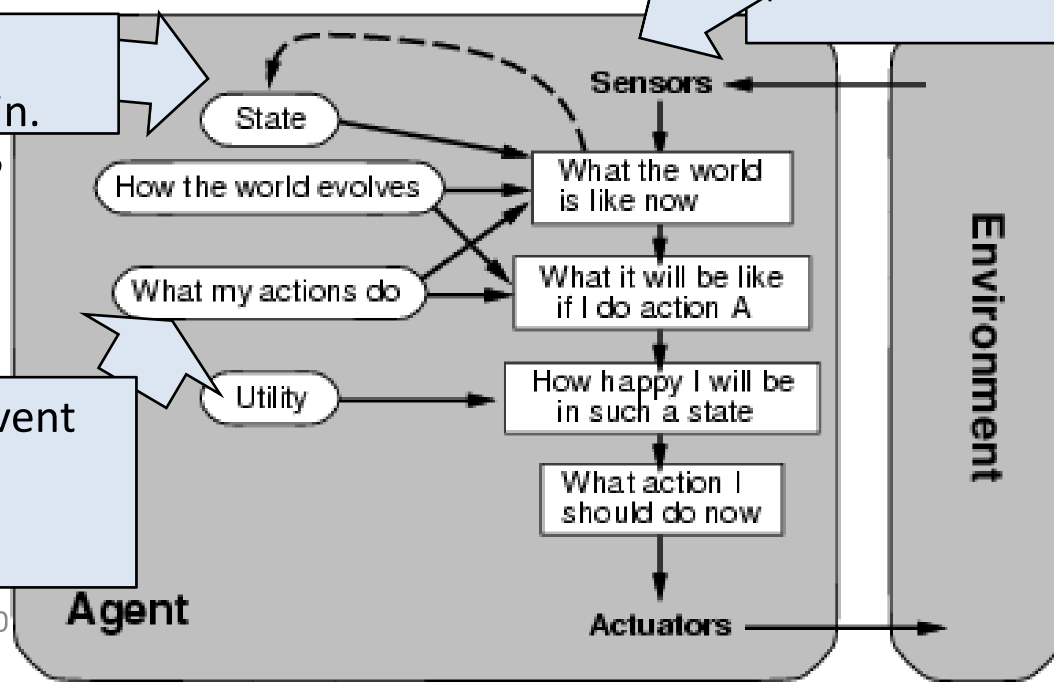
- Qu'est-ce qu'un **réseau bayésien** (RB) ?
  - Structure et formalisme
  - Signification
- Indépendance conditionnelle dans un RB.
- Inférence dans un réseau bayésien
  - Inférence exacte
  - Inférence approximative
- Diagrammes d'influence

Contexte

- Jusqu'à présent, nous avons notamment étudié des techniques pour des environnements déterministes
- Que faire lorsque l'environnement est non déterministe?
- Quel type d'agent? Un qui doit pouvoir:
  - Gérer l'incertitude lié à l'environnement
    - ⇨ probabilité
  - Gérer la qualité de ses
    - ⇨ utilité

État incertain.

(1) Les capteurs peuvent être bruités…
(2) Observabilité partielle.



Les actions peuvent avoir des effets incertains!

Exemple de décision sous incertitude

- Soit l'action A_t d'aller à l'aéroport t minutes avant le départ de l'avion.
- A_t me permettra-t-il d'arriver à temps?
- Caractéristique (de l'environnement) du problème:
  - Observabilité partielle (conditions routières, etc.)
  - Senseurs bruités (annonces du trafic, etc.)
  - Incertitude dans l'effet des actions (crevaisons, pannes, etc.)
  - Immense complexité pour modéliser les actions et le trafic.

Exemple de décision sous incertitude

- Un raisonnement purement logique
  - Risque de tirer des conclusions erronées: « A_{25} me permettra d'arriver à temps », ou
  - Risque de tirer des conclusions peu exploitable du point de vue « prise de décision »:
    - « A_{25} me permettra d'arriver à temps, s'il ne pleut pas, s'il n'y a pas d'accident, si mes pneus ne crèvent pas, etc. »
    - « A_{1440} me permettra raisonnable d'arriver à temps, mais je devrai passer une nuit à l'aéroport. »

Prise de décision sous incertitude

- Supposons que je crois ceci :

  $P(A_{25}$ me permet d'arriver à temps | …) = 0.04
  $P(A_{90}$ me permet d'arriver à temps | …) = 0.70
  $P(A_{120}$ me permet d'arriver à temps | …) = 0.95
  $P(A_{240}$ me permet d'arriver à temps | …) = 0.999
  $P(A_{1440}$ me permet d'arriver à temps | …) = 0.9999

- Quelle action devrai-je choisir?
  - Cela dépend de mes ***préférences***: manquer l'avion *vs.* trop d'attente.

- ***La théorie de l'utilité*** est utilisée pour modéliser et inférer sur les préférences.
  - Une préférence exprime le degré d'utilité d'une action/situation.

- ***Théorie de la décision*** = **théorie des probabilités + théorie de l'utilité**.

Discussion : utilité de l’argent

- Lors d’un quiz, vous gagnez 1 000 000 $!
- On vous propose un dernier jeu facultatif à pile ou face:
  - On lance une pièce de monnaie équilibrée :
    - Vous ne jouez pas : vous partez avec 1 000 000 $.
    - Face : vous tripler votre gain 3 000 000 $.
    - Pile : vous repartez les mains vide : 0 $
  - Acceptez-vous ce dernier jeu?
- Même décision si gain initial était de 10$?

Réseaux bayésiens: Préambule

Notions de bases: **Probabilité conditionnelle & Variables indépendantes**

- Variables aléatoires X et Y avec les domaines de valeurs possibles suivants:

  Exemple: X= {1,2,3,4,5,6}, Y = {oui, non, peut-être}

- Règle de produit: $\mathbf{P(X,Y)=P(X|Y)*P(Y) = P(Y|X)*P(X)}$
- Règle de Bayes: $\mathbf{P(X|Y) = P(Y|X)\ P(X)\ /\ P(Y) = \alpha P(Y|X)P(X)}$
- Variables aléatoires booléennes Heureux et Riche avec les domaines de valeurs = {vrai, faux}

- P(Heureux=vrai, Riche=faux) == P(Heureux, ¬ Riche)

  = P(Heureux| ¬ Riche) * P(¬ Riche)

Réseaux bayésiens: Préambule

- Avec trois variables aléatoires…
- Soit une variable aléatoire Z = {a,b,c}

P(X=3, Y=non, Z=a)

= P(X=3| Y=non,Z=a)*P(Y=Non,Z=a)

= P(X=3| Y=non,Z=a)***P(Y=Non,Z=a)**

= P(X=3| Y=non,Z=a)***P(Y=Non|Z=a)*P(Z=a)**

Réseaux bayésiens: Préambule

- **Variables indépendantes**

Si X et Y sont indépendantes,

$P(X,Y) = P(X)*P(Y)$

Ex: $P(Pluie,Riche) = P(Pluie)*P(Riche)$

Par ailleurs:

$P(X|Y) = P(X)$

Ex: $P(Pluie|Riche) = P(Pluie)$

Réseaux bayésiens: Préambule

- 4 variables aléatoires booléennes:
  - Détecte (vrai si le dentiste détecte une anomalie)
  - Douleur (vrai si le patient ressent une douleur)
  - Carie (vrai si le patient a une carie)
  - Compétent (vrai si le dentiste est compétent)

P(Detecte,Douleur,Carie,Compétent)
= P(Détecte|Douleur,Carie,Compétent)*P(Douleut,Carie,Compétent)
= P(Détecte|Douleur,Carie,Compétent)*P(Douleur|Carie,Compétent)
 * P(Carie,Compétent)
= **P(Détecte|Douleur,Carie,Compétent)*P(Douleur|Carie,Compétent)**
 *** P(Carie|Compétent)*P(Compétent)**

Réseaux bayésiens: Préambule

- Détection et élimination des liens d'indépendance

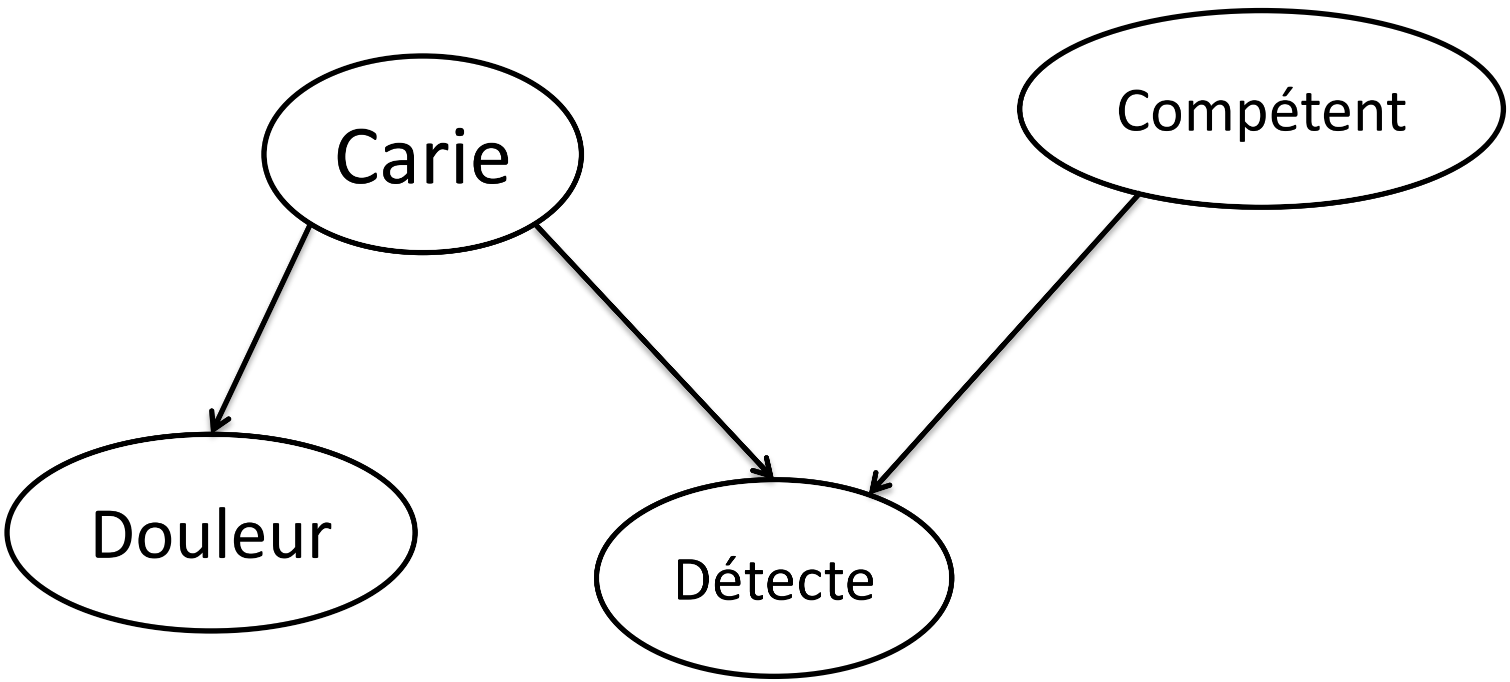
P(Detecte,Douleur,Carie,Compétent)
= P(Détecte|Douleur,Carie,Compétent)*P(Douleur,Carie,Compétent)
= P(Détecte|Douleur,Carie,Compétent)*P(Douleur|Carie,Compétent)
* P(Carie,Compétent)
= **P(Détecte|~~Douleur~~,Carie,Compétent)*P(Douleur|Carie~~,Compétent~~)**
*** P(Carie~~|Compétent~~)*P(Compétent)**
=P(Détecte|Carie,Compétent)*P(Douleur|Carie)*P(Carie)*P(Compétent)

Ces simplifications dérivent des faits de l'indépendance conditionnelle entre les variables concernées.
Par exemple, puisque Compétent est conditionnellement indépendant de Douleur , P(Douleur|Carie,Compétent) = P(Douleur|Carie).

- Notes: Dans des cas idéals, l'exploitation de l'indépendance conditionnelle réduit la complexité de représentation de la distribution conjointe d'une exponentielle ***(O(2^n))*** en linéaire ***(O(n)****), pour n variables aléatoires.*

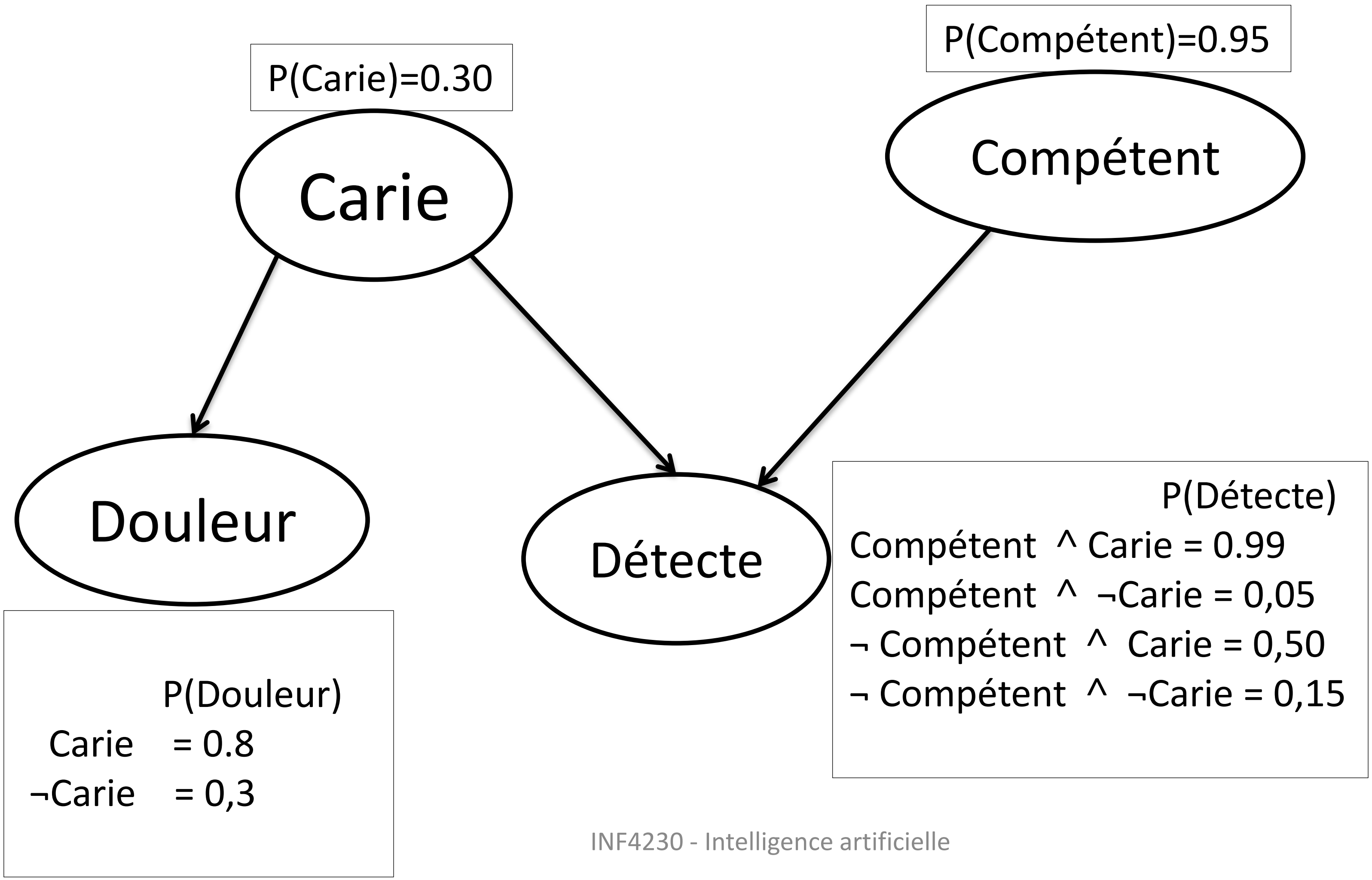
Réseaux bayésiens: Préambule

- Représentation graphique: modélisation du graphe



Réseaux bayésiens: Préambule

- Représentation graphique: Probabilités



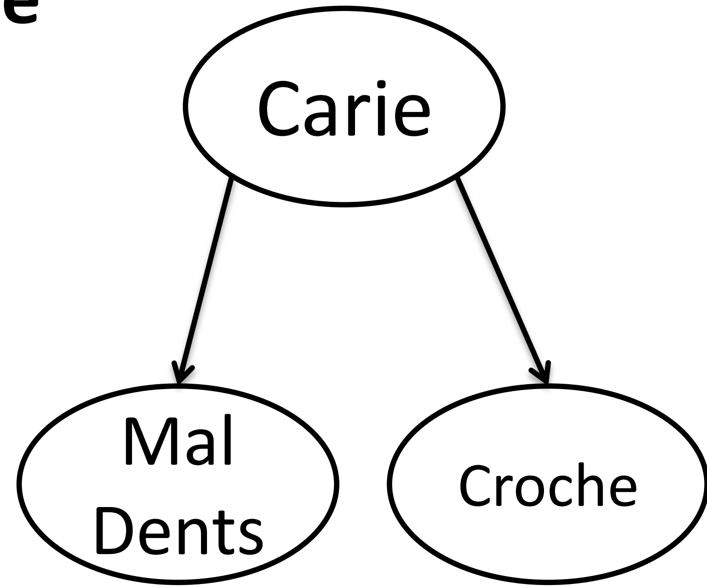
Réseaux bayésiens

En bref:

- Les RB sont un mariage entre la théorie des graphes et la théorie des probabilités.
- Un RB représente les connaissances probabilistes à partir d'une situation 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 causent d'autres.
  - Notre compréhension de la causalité des événements est incomplète : on doit recourir aux probabilités.

Formalisme

- Un **réseau bayésien** est un **graphe orienté** et **acyclique** où :
  - les **nœuds** sont des **variables** (généralement aléatoires);
  - les **arcs** représentent des relations de **dépendance** (causalités) entres les variables;
  - des distributions de probabilités conditionnelles («tables de probabilités») sont attachées aux variables.

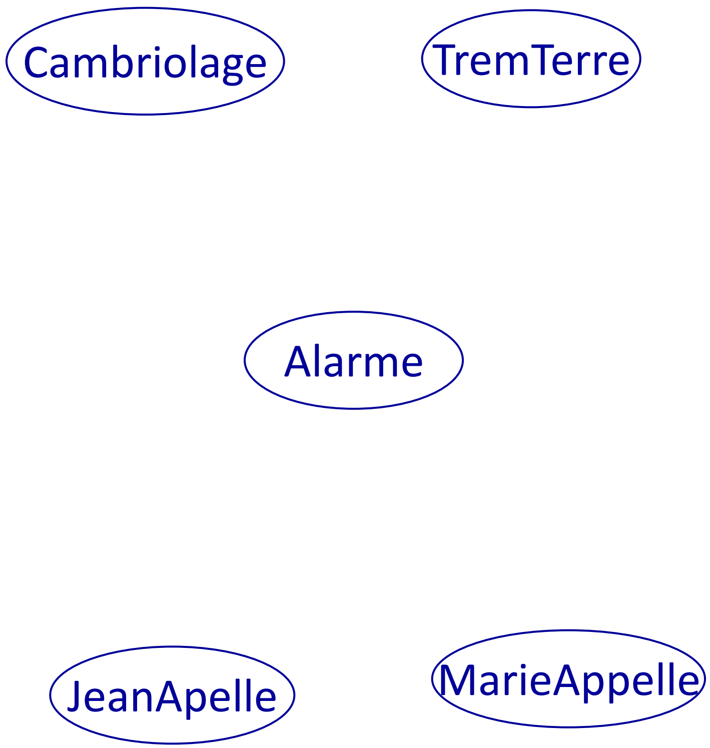


Exemple : Détection de cambriolage

- Je suis souvent à l’extérieur de chez moi.
- Je m’inquiète d’un cambriolage durant mon absence.
- J’ai fait installé un système d’alarme chez moi.
- Mes voisins, Marie et Jean, m’ont promis de m’appeler à chaque fois qu’ils entendent mon système d’alarme.
- Jean confond parfois l’alarme avec le téléphone.
- Marie n’entend pas toujours l’alarme, car elle met la musique trop forte chez elle.
- Parfois mon système d’alarme déclenche lors de légers tremblements de terre.
- Comment conclure qu’il y a un cambriolage chez moi ?

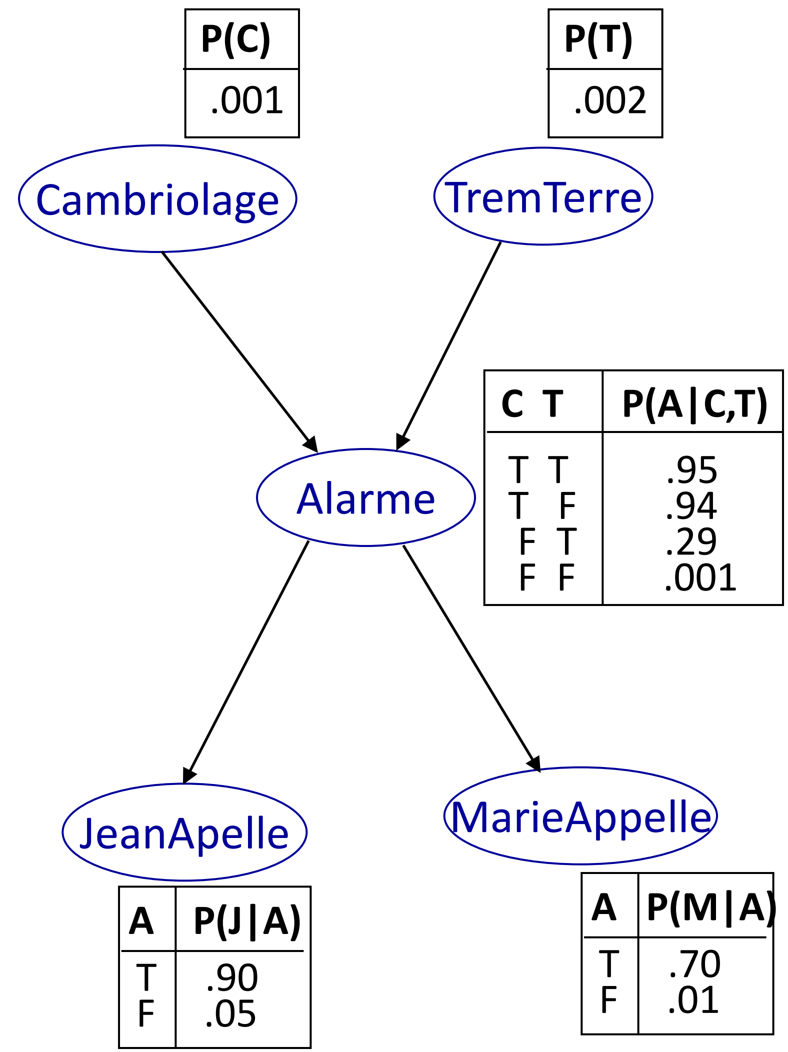
Exemple

- Variables aléatoires :
  - *Cambriolage*
  - *TremblementDeTerre*
  - *Alarme*
  - *Jean appelle*
  - *Marie appelle*.



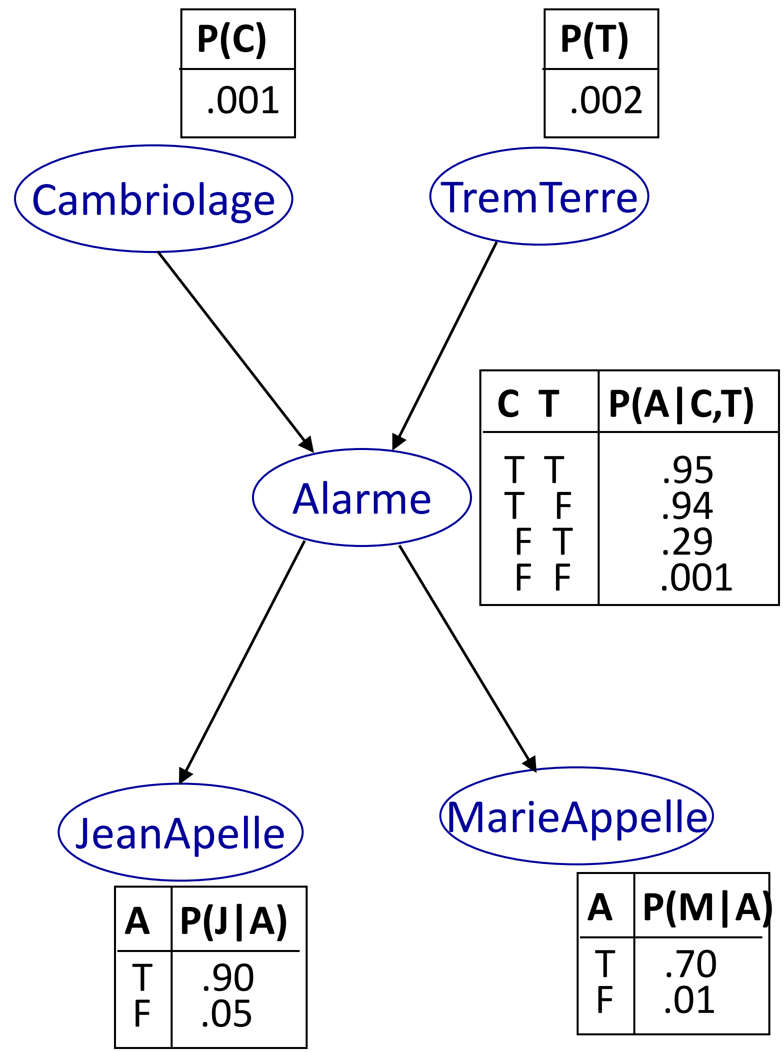
Exemple

- La topologie du RB modélise la connaissance causale.
- 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 tremblement de terre aussi.
  - L'alarme peut inciter Jean à appeler.
  - Idem pour Marie à appeler.
- Pour chaque nœud, une table de probabilité conditionnelle (TPC) donne la probabilité pour chaque valeur du nœud étant donné les combinaisons des valeurs des parents du nœud.



Définitions

- S'il y a un arc d'un nœud *X* vers un nœud *Y,* cela signifie que la variable *X* influence la variable *Y*.
  - X est appelé le parent de Y.
  - *Parents(X)* est l'ensemble des parents de X.
- Si X n'a pas de parents, sa distribution de probabilités est dite *inconditionnelle* ou *à priori*.
- Si X a des parents, sa distribution de probabilités est dite *conditionnelle* (par rapport aux parents) ou *à postériori*.
- Si X est une variable observée, ont dit que c'est une *évidence*.



Sémantique

- Un RB est une façon compacte de représenter des probabilités conjointes.
- Par définition, la probabilité conjointe de X_1 et X_2 est la probabilité d'avoir X_1 et X_2 : $P(X_1,X_2)$.
- La probabilité 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 :

  $$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$
- En d'autres mots, la distribution conjointe des variables d'un RB est définie comme étant le produit des distributions conditionnelles (locales).
  - *Seules comptent les probabilités de chaque variable connaissant ses parents, pour calculer la distribution conjointe.*

Sémantique

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

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

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

- Ceci est cohérent avec l'assertion précédente pour autant que $Parents(X_i)$ soit un sous-ensemble de $\{X_{i-1}, \ldots, X_1\}$.

- Cette condition est satisfaite en étiquetant les nœuds du RB dans un ordre cohérent avec la relation d'ordre partielle implicite dans la topologie du RB.

Exemple

- $P(X_1, \ldots, X_n) =$

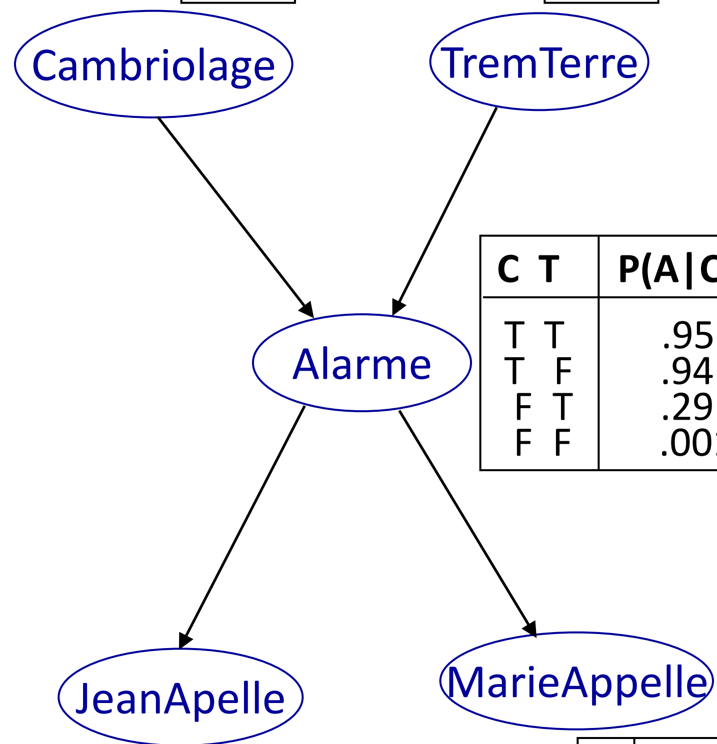
  $$\prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

- $P(j, m, a, \neg c, \neg t)$

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

$= .90 \times .70 \times .001 \times 0.999 \times .998$

$= .00062$



P(C)
.001

P(T)
.002

C	T	P(A|C,T)
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

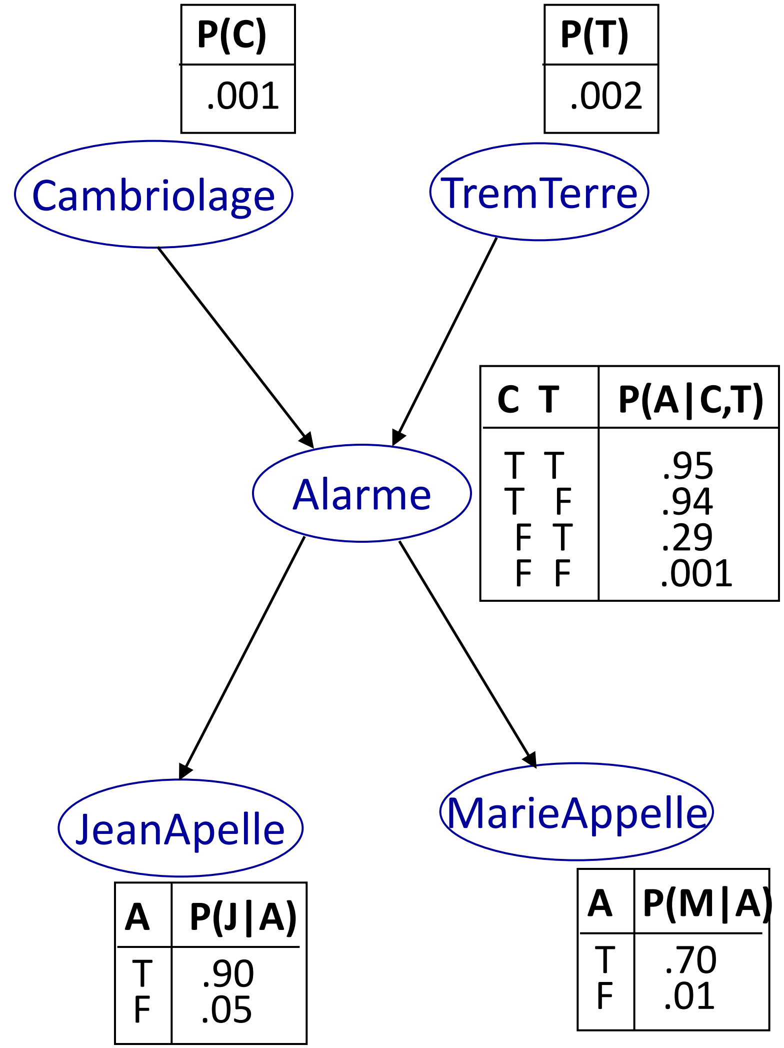
Construction d'un RB

- Il y a un arc de X vers Y si X influence ***directement*** Y.

- S'il y a un arc d'un nœud X vers un nœud Y, on dit que :

  – X donne le support causal à Y.

  – Y donne le support diagnostique à X.



P(C)
.001

P(T)
.002

C	**T**	**P(A|C,T)**
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 correcte, on s'assure que chaque nœud est indépendant de tous ses prédécesseurs, étant donné 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/cause 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 $X_1, \ldots, X_n$

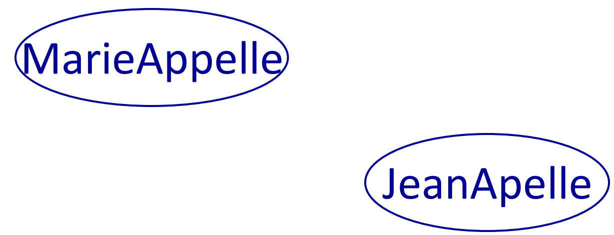
- 2. Pour i = 1 to n:

  - 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:
      - $\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})$ (chain rule)
        $= \prod_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$ (par construction)

Exemple

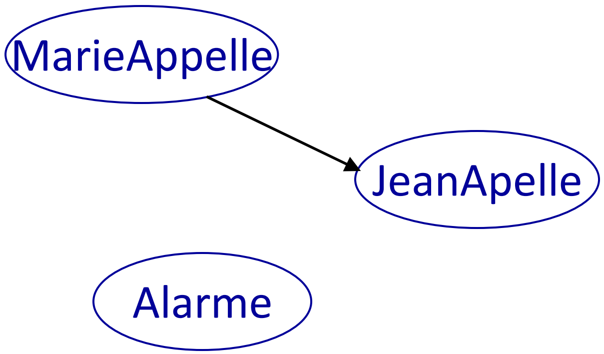
- Supposons qu'on ordonne les variables comme suit: *M, J, A, C, T.*



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

Exemple

- Supposons qu'on ordonne les variables comme suit: *M, J, A, C, T*



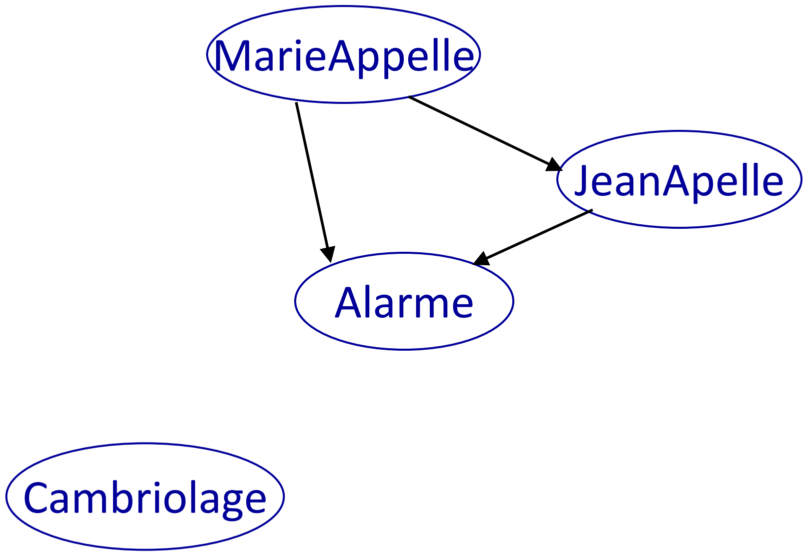
***P**(J | M) =* ***P**(J)?*

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, T.*



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)*?

Example

- Supposons qu'on ordonne les variables comme suit: *M, J, A, C, T.*

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

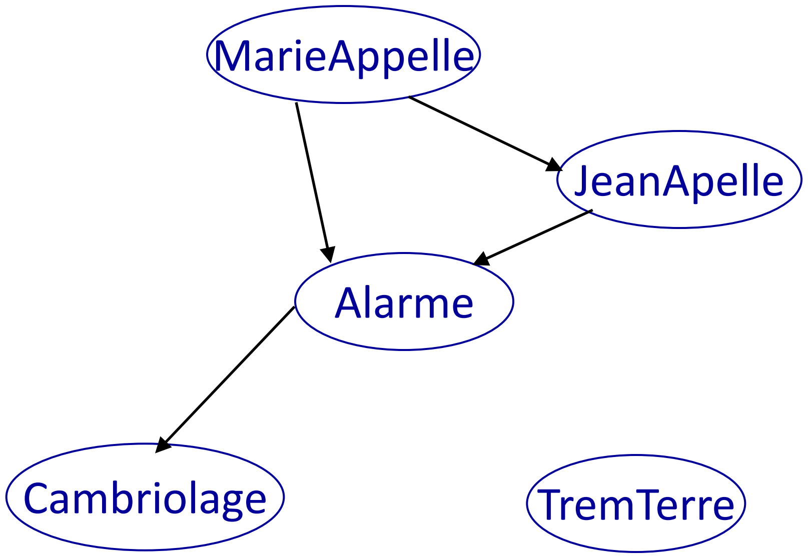
Non.

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(*T | C, A ,J, M) =* ***P***(*T | A)*?



Exemple

- Supposons qu'on ordonne les variables comme suit: *M, J, A, C, T.*

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

Non.

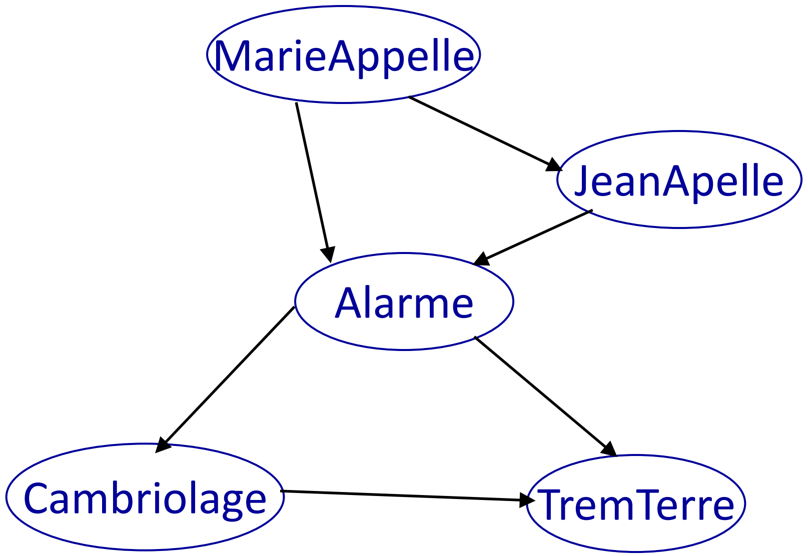
P*(A | J, M) =* ***P****(A | J)*? ***P****(A | J, M) =* ***P****(A)*? **Non.**

P*(C | A, J, M) =* ***P****(C | A)*? **Oui.**

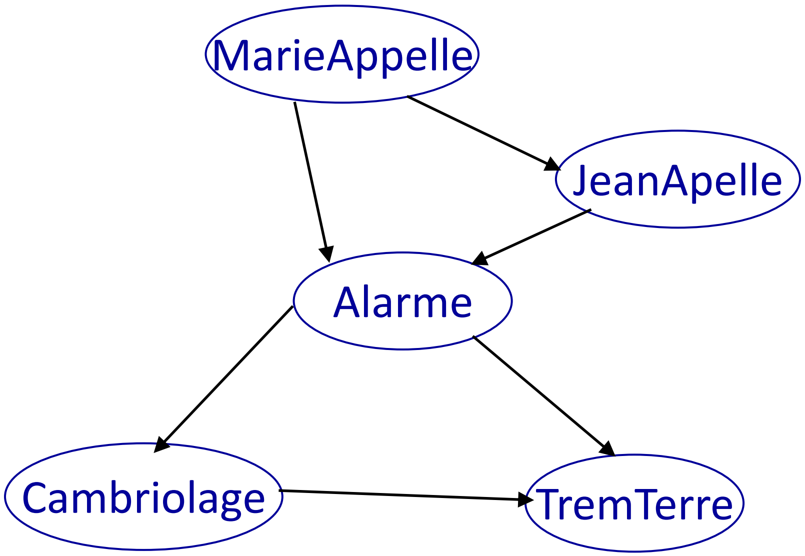
P*(C | A, J, M) =* ***P****(C)*? **Non.**

P*(T | C, A ,J, M) =* ***P****(T | A)*? **Non.**

P*(T | C, A, J, M) =* ***P****(T | A, C)*? **Oui.**



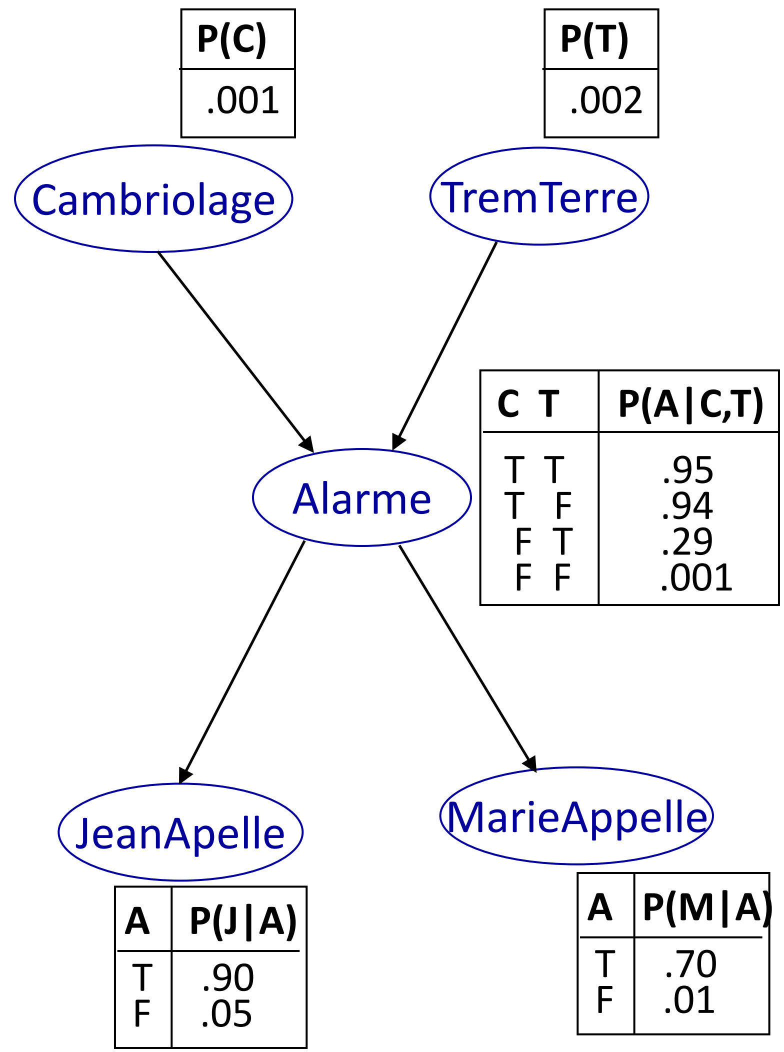
Exemple



- Déterminer l'indépendance conditionnelle est très difficile dans le sens non causal.
  - Par exemple, en médecine, des études ont montré que les experts préfèrent donner des probabilités dans le sens causal (pathologie → symptôme) plutôt que dans le sens diagnostique.
- Un réseau avec des dépendance diagnostique (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

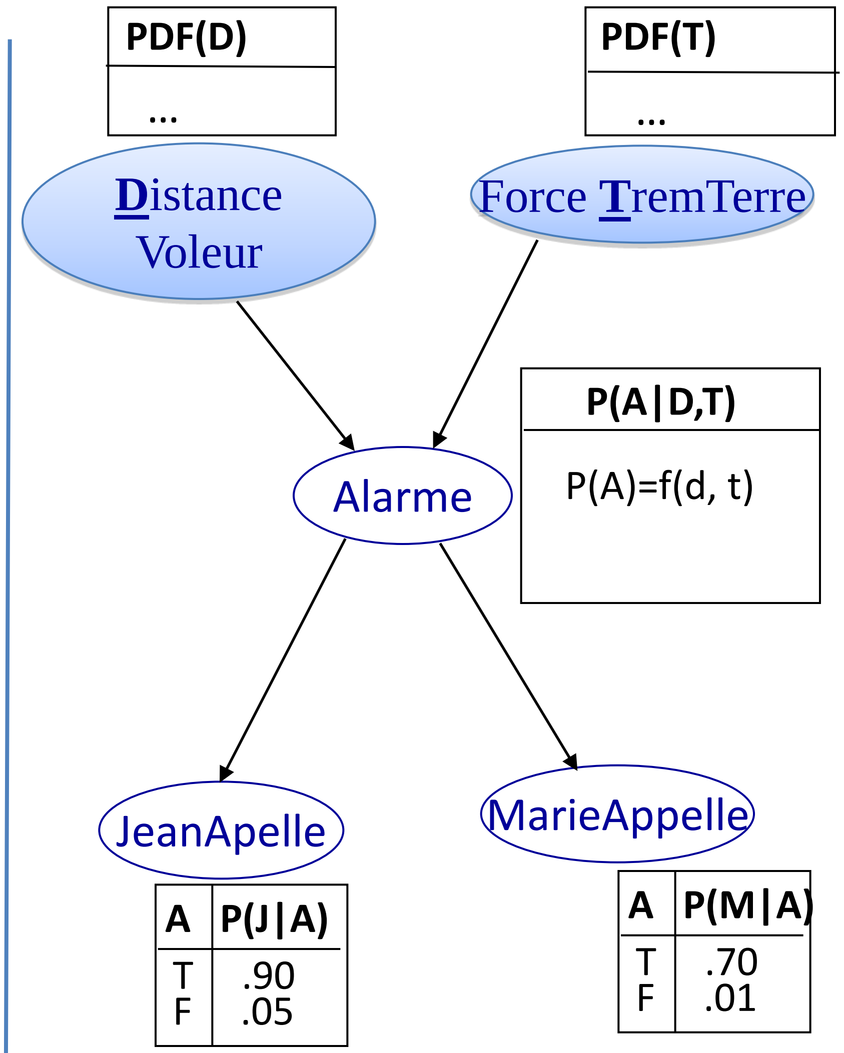
- Il y a d'autres appellations pour les RB :
  - Réseaux de croyances (*beleif networks*).
  - Réseaux probabilistes.
  - Réseaux causals.
- Les nœuds d'un réseau bayésien sont aussi appelés des « nœuds chances ».



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.

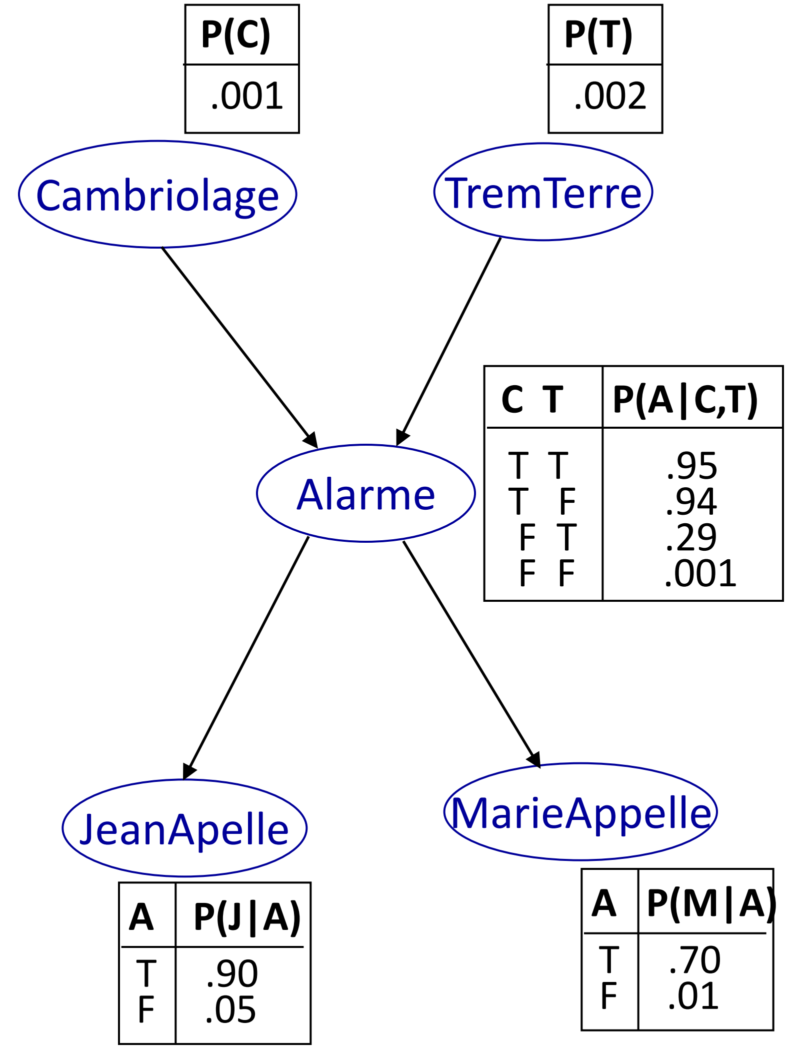
- Mais les RB fonctionnent aussi avec des variables continues :
  - Les probabilités conditionnelles sont spécifiées par des fonctions de densité de probabilités (PDF).
  - Ex:
    - Distance entre voleur et le capteur de mouvement.
    - Force du séisme sur l'échelle de Richter.



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*** :

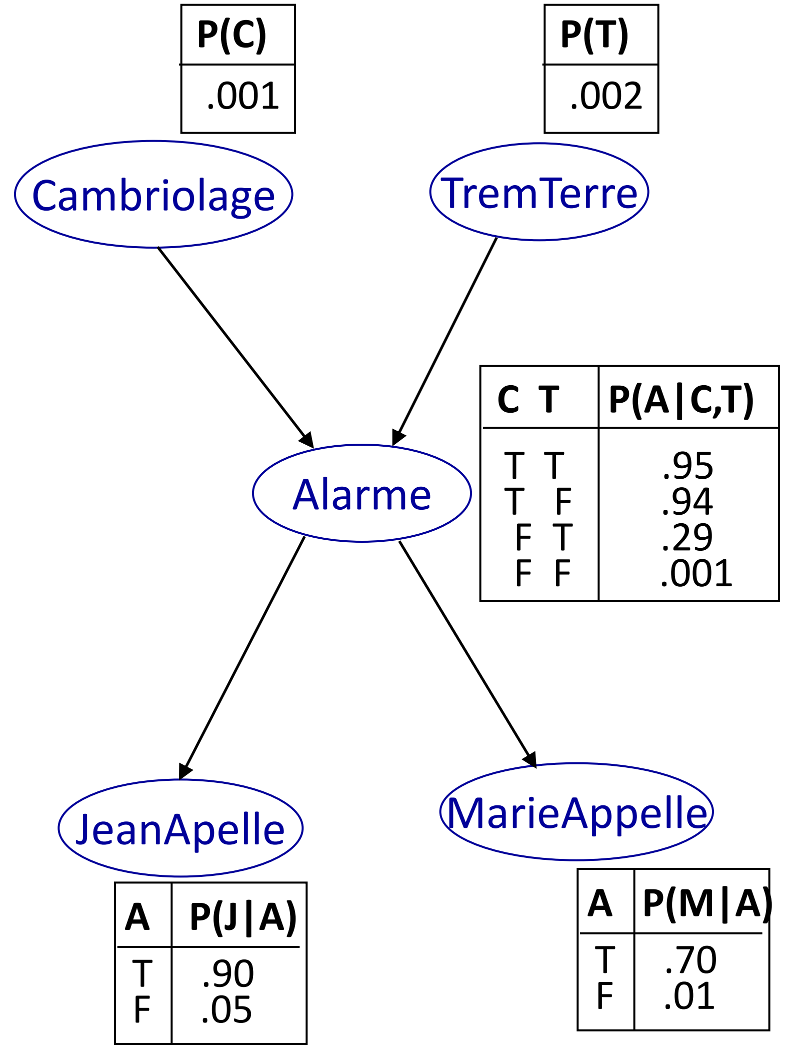
      $\mathbf{P(\mathit{M} \mid \mathit{A}, \mathit{C}) = P(\mathit{M} \mid \mathit{A})}$.
    - Si ***A*** est connu, alors ***C*** n'intervient pas dans le calcul.



Indépendance conditionnelle dans un RB (2)

- 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*** :

      $$P(J \mid A, M) = P(J \mid A).$$
    - Si ***A*** est connu, alors ***M*** n'intervient pas dans le calcul.



P(C)
.001

P(T)
.002

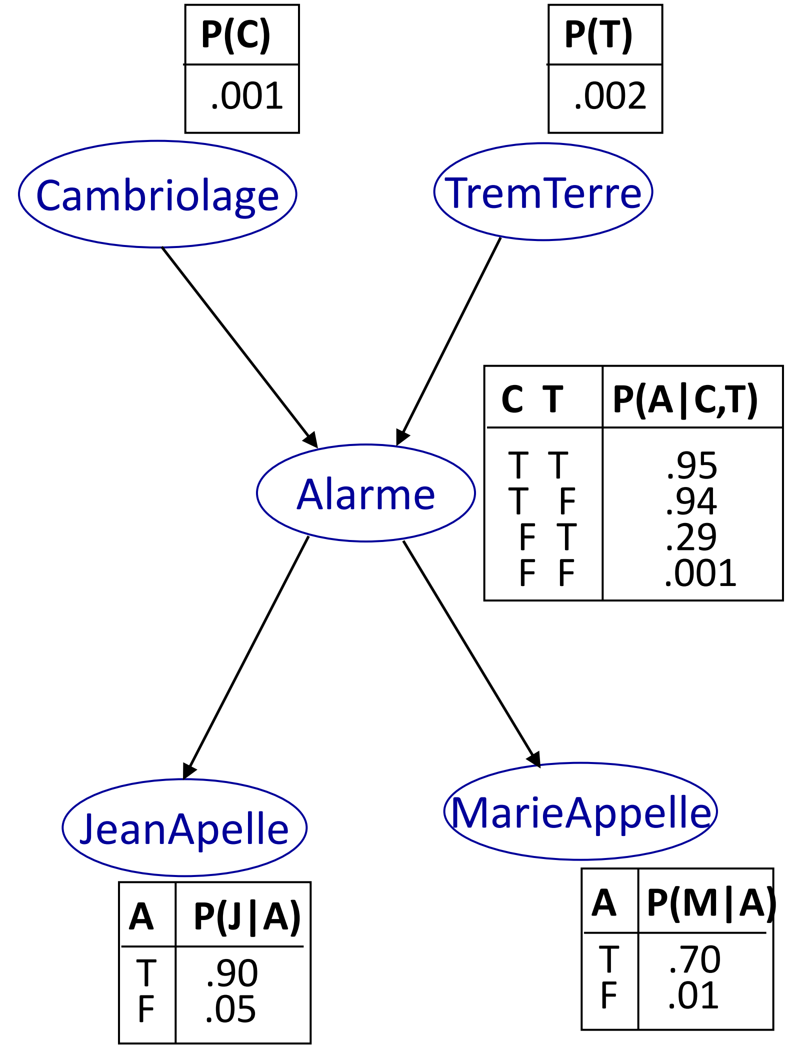
C	T	P(A|C,T)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

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

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

Dépendance conditionnelle dans un RB (1)

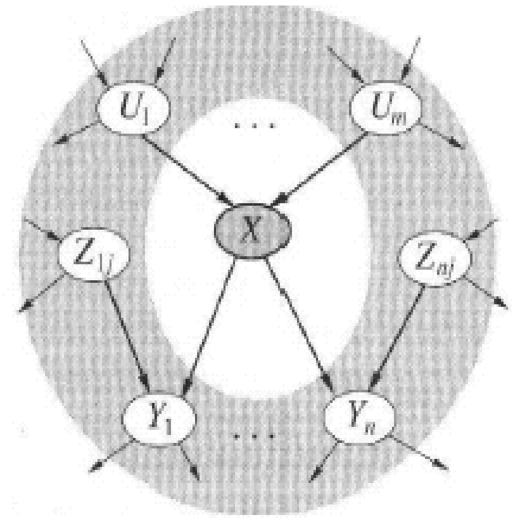
- Un RB modélise les relations d'indépendance conditionnelle suivantes
  - Un nœud est indépendante de ses non-descendants, étant donné ses parents.
  - Exemples :
    - ***Cambriolage*** et ***TremTerre*** sont **indépendants**.
    - Mais ils sont **dépendants** étant donné ***Alarme***.
      - **P(*A* | *C*, *S*)** n'est pas simplifiable.



Indépendance conditionnelle dans un RB (2)

– Un nœud est conditionnellement indépendant étant donné tous ses parents, enfants et parents de ses enfants (autrement dit étant donné sa **couverture de Markov)**.

Couverture de Markov de *X*



Inférence / Interrogation d'un RB

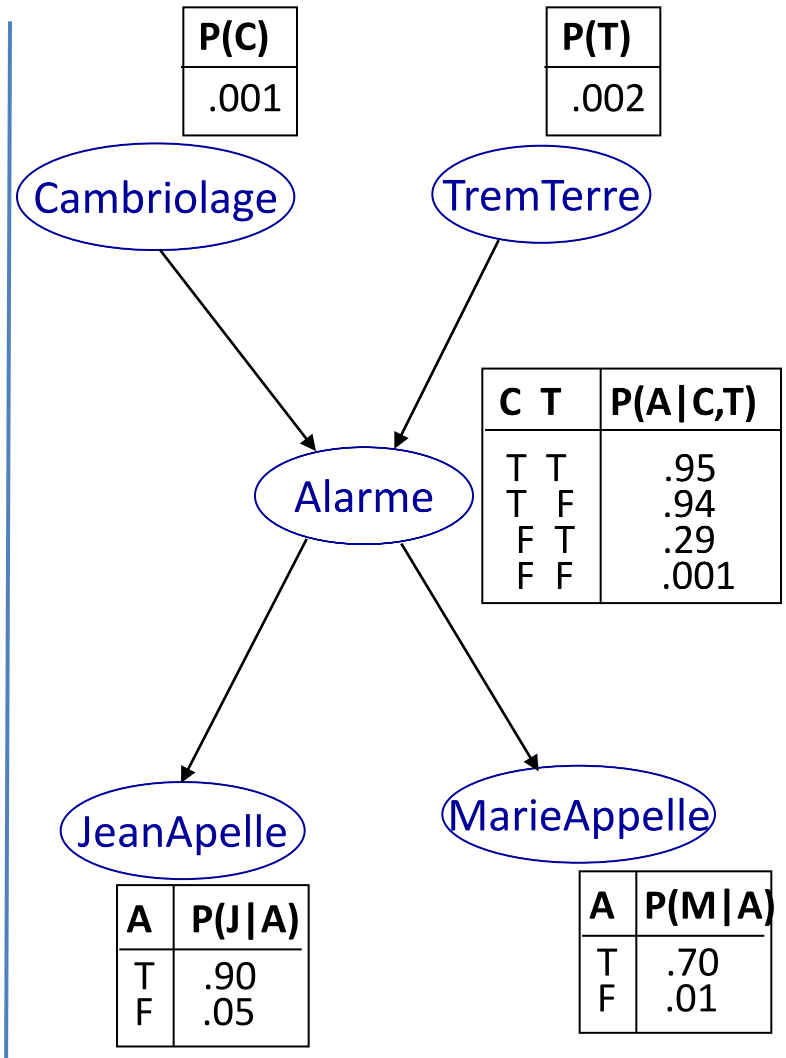
- On exploite un réseau bayésien (RB) pour calculer **la distribution de probabilités à postériori** pour un ensemble de variables, étant donné un **événement observé**.
- Un **événement observé** est un ensemble d'assignations de valeur à un ensemble de variables observées (évidences).
- Plus formellement, on a :
  - ***X*** l'ensemble de variables pour laquelle on fait l'interrogation;
  - ***E*** les variables d'évidences (qu'on peut observer);
  - ***H*** les variables cachées (qu'on ne peut pas observer).
- Une interrogation à un RB a la forme $\mathbf{P}(\boldsymbol{X}|\boldsymbol{e})$, où ***e*** est une assignation de valeurs aux variables dans ***E***.
- Une interrogation s'appelle aussi faire une **inférence**.

Exemple d'inférence/interrogation

- Situation **: Jean et Marie appellent**.
- Interrogation **: quelle est la probabilité d'un cambriolage**?
- **P(*Cambriolage*|*JeanApelle=True*, *MarieAppelle=True*)**

  = [0.284,0.716]

- Comment faire ce calcul ?
  - **Inférence exacte (prohibitif)**
    - Par énumération
    - Élimination de variables
  - **Inférence approximative** par échantillonnage avec les méthode de Monte-Carlo
    - Échantillonnage direct
    - Vraisemblance pondérée
    - Chaînes de Markov simulées



Rappel de notions de base en probabilités

- Ceci nous donne
  - $P(X|Z) = \alpha P(X,Z)$
  - α est une constante de normalisation pour s'assurer que la somme des probabilités de la distribution P(X,Z) soit égale à 1.

- De manière générale, soit
  - *X*, l'ensemble de variables pour laquelle on fait l'interrogation,
  - *E*, les variables d'évidences (qu'on peut observer) et
  - *Y*, les variables cachées (qu'on ne peut pas observer).
  - *e*, les valeurs observées pour les variables dans E.

- $P(X|E=e) = \alpha P(X,E=e) = \Sigma_y P(X, E=e, Y = y)$

- *Noté aussi* $P(X|e) = \alpha \Sigma_y P(X, e, y)$

Inférence par énumération dans un RB

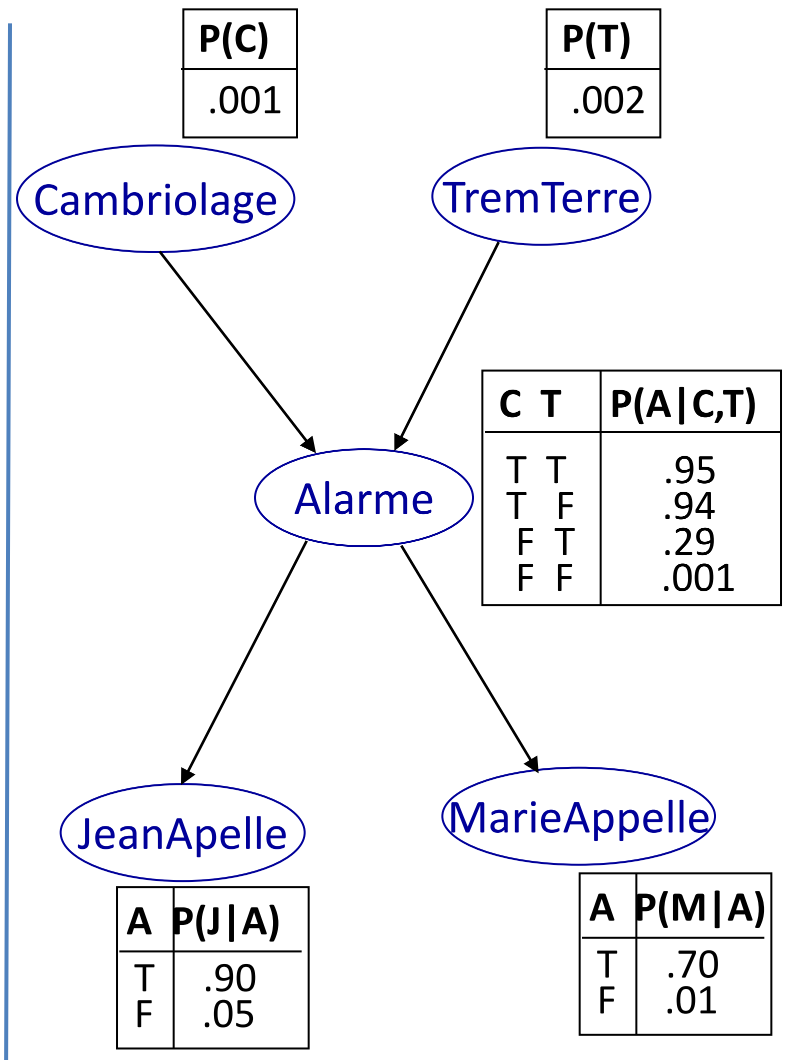
- On a vu que : $P(X|e) = \alpha \, \Sigma_y \, P(X, e, y)$
- On a vu aussi que selon la sémantique d'un RB

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

- Les termes 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 interrogation *P(X|e)* sur un RB, simplement en *calculant les sommes des produits des probabilités conditionnelles du RB*.
  - Algorithme Figure 14.9, Page 506 du livre.

Exemple

- Requête : **P(*Cambriolage* | *JeanApelle*=T, *MarieAppelle*=T)**
- Aussi noté P(*C* | *j*, *m*)
- Les variables cachées : H={**TremTerre**, ***Alarme***}.
- $P(C \mid j, m) = \alpha\, \Sigma_{s,a}\, P(C, s, a, j, m)$
- *Note : s* et *a* veulent dire, 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*).



P(C)
.001

P(T)
.002

C	T	P(A|C,T)
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

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

- On calcule pour $C = true$

  $P(c \mid j, m)$

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

  $= 0.284$

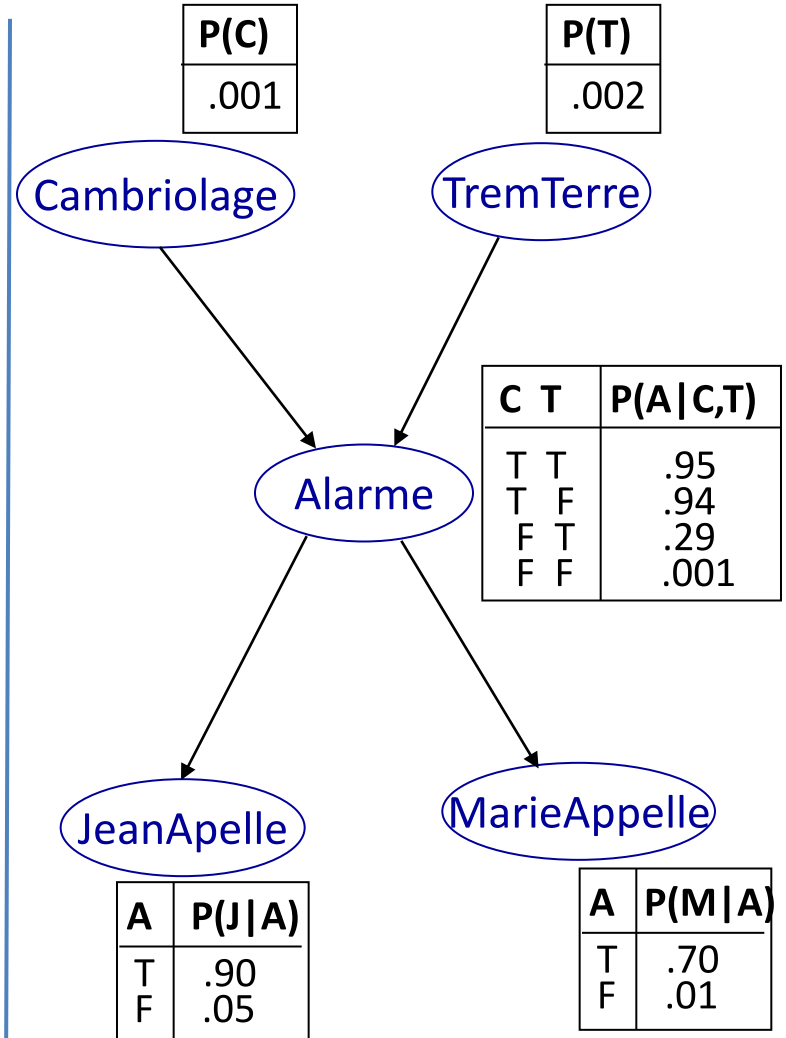
- Et $C = false$

  $P(c \mid j, m)$

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

  $= 0.716$

- Ce qui donne $P(C \mid j, m) = [0.284, 0.716]$



P(C)
.001

P(T)
.002

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

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

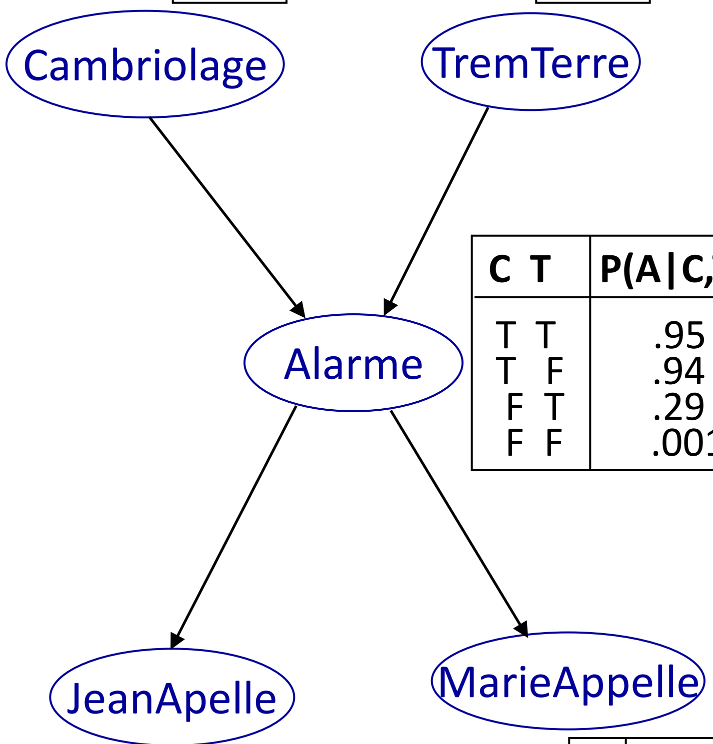
A	P(M|A)
T	.70
F	.01

Inférence par élimination des variables

- Même principe que l'inférence par énumération, mais on évite les répétions de calculs déjà faits.
  - Voir section 14.4.2 du livre (3e édition).

Types d’interrogations d’un RB

- **Diagnostique** (on connaît les effets, on cherche les causes)
  - P(Cambriolage|JeanAppelle=T)
  - Garder à l’esprit qu’on a des arcs « causes $\rightarrow$ effets ».
- **Prédiction** (étant donné les causes, quels sont les effets)
  - P(JeanAppelle|Cambriolage=T).
- **Probabilité conjointe**
  - P(Alarme)



P(C)
.001

P(T)
.002

C	T	P(A|C,T)
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

Exercices (sens prédictif)

Exercice 1 : Inférence par énumérations

<u>Requête</u>:

Calculer P(T=true)

<u>Variables connues:</u>

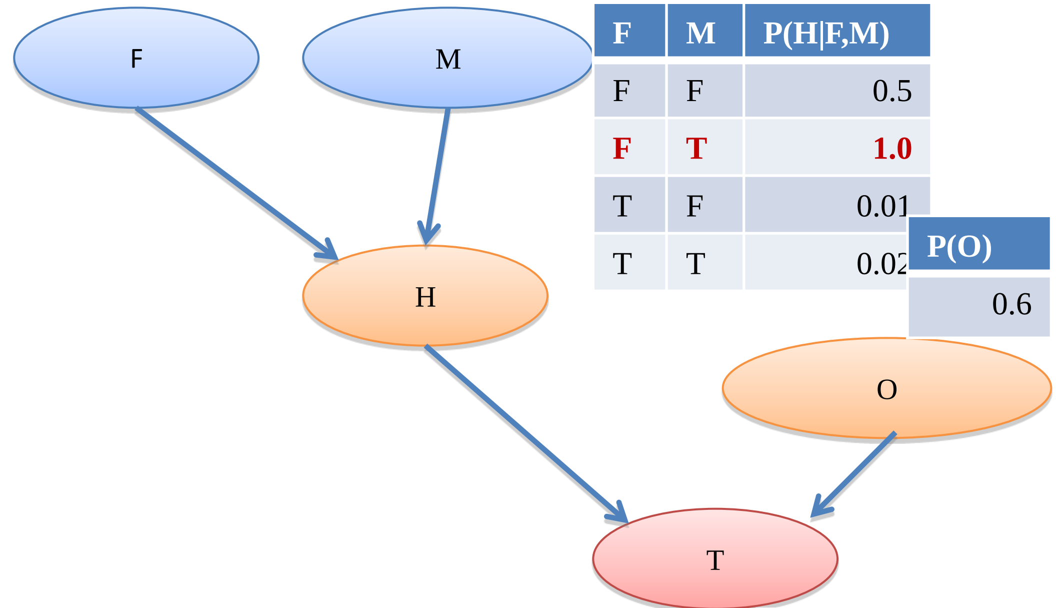
F = False

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

Exercice 1 : Inférence par énumérations

<u>**Requête**:</u>

Calculer P(T=true)

<u>Variables connues:</u>

F = False

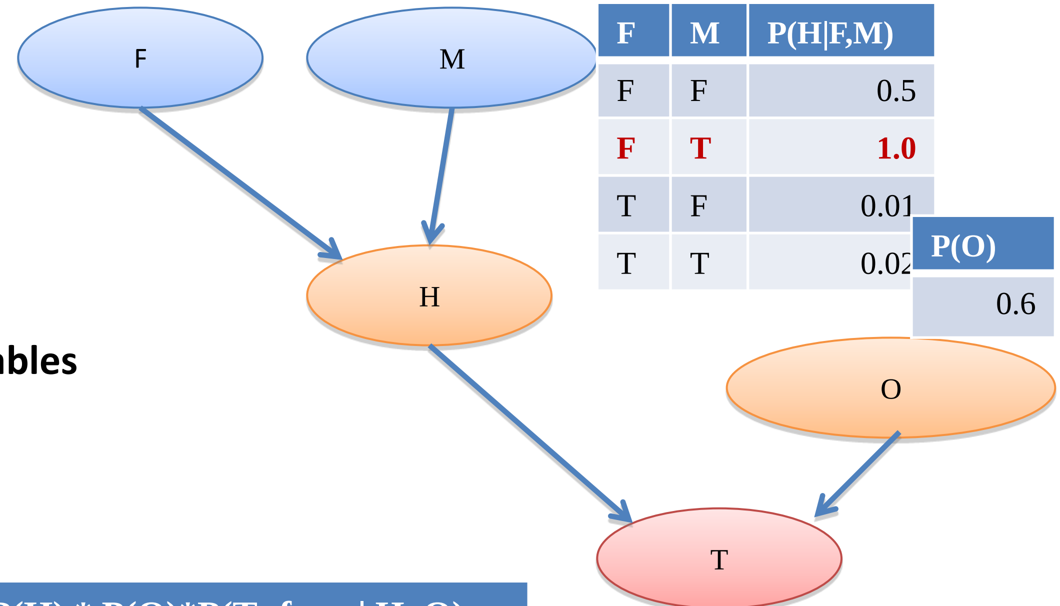
M = true

<u>Variables inconnues:</u>

H

O

Énumération des valeurs possible des variables cachées (2*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

H	O	P(H) * P(O)*P(T=vrai | H, O)	P(H) * P(O)*P(T=faux | H, O)
F	F	0.0 * 0.4 * 0.1 = 0.0	0.0 * 0.4 * 0.9 = 0.0
F	T	0.0 * 0.6 * 0.5 = 0.0	0.0 * 0.6 * 0.5 = 0.0
T	F	1.0 * 0.4 * 0.5 = **0.2**	1.0 * 0.4 * 0.5 = **0.2**
T	T	1.0 * 0.6 * 1.0 = 0.6	1.0 * 0.6 * 0.0 = 0.0
TOTAL		**0.8**	0.2

Exercice 2 : Inférence par énumérations

<u>Requête</u>:

Calculer P(T=true)

<u>Variables connues:</u>

M = true

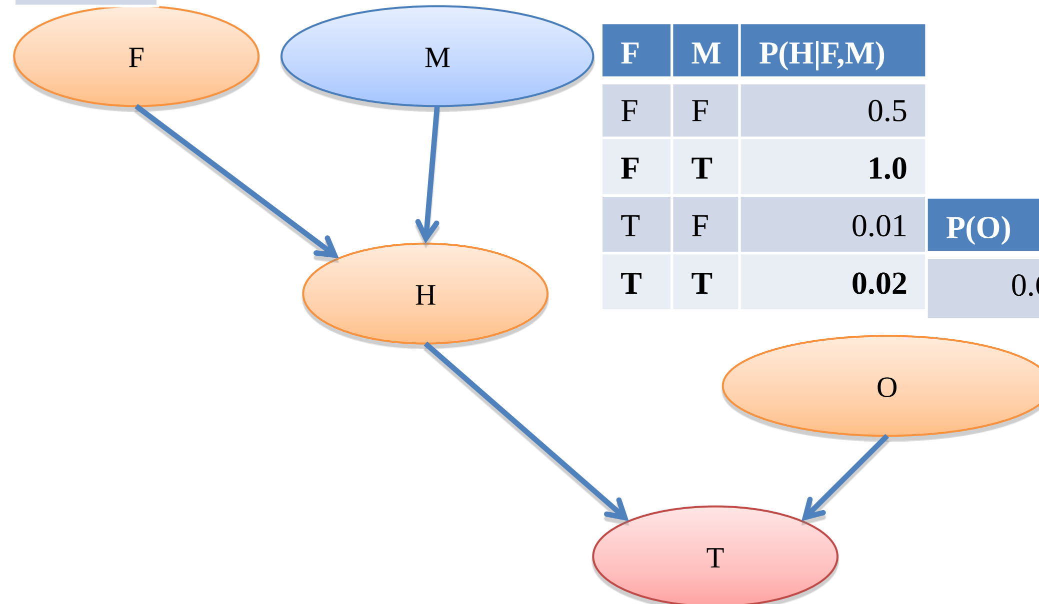
<u>Variables inconnues:</u>

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

Exercice 2 : Inférence par énumérations

<u>Requête</u>:

Calculer P(T=true)

<u>Variables connues:</u>

M = true

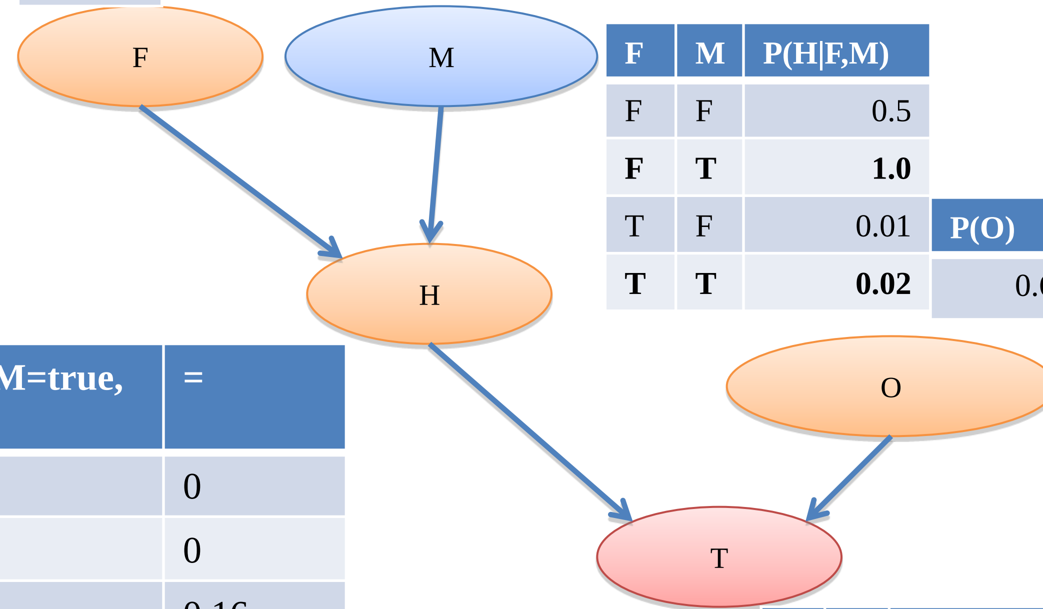
<u>Variables inconnues:</u>

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)*P(O|F,H)* P(T | F, M=true, 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**

Exercice 3: Évaluation avec décomposition

Requête:

Calculer P(T=true)

Variables connues:

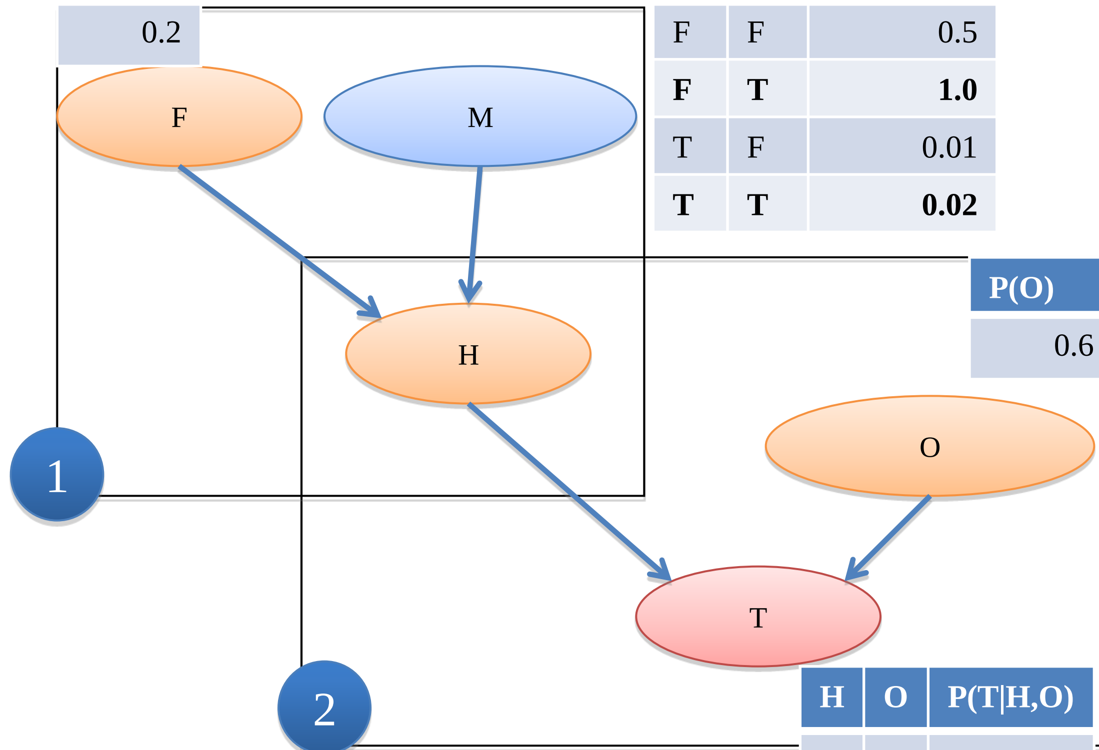
M = true

Variables inconnues:

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

Exercice 3: Évaluation avec décomposition

<u>Requête</u>:

Calculer P(T=true)

<u>Variables connues:</u>

M = true

<u>Variables inconnues:</u>

H

O

F

1

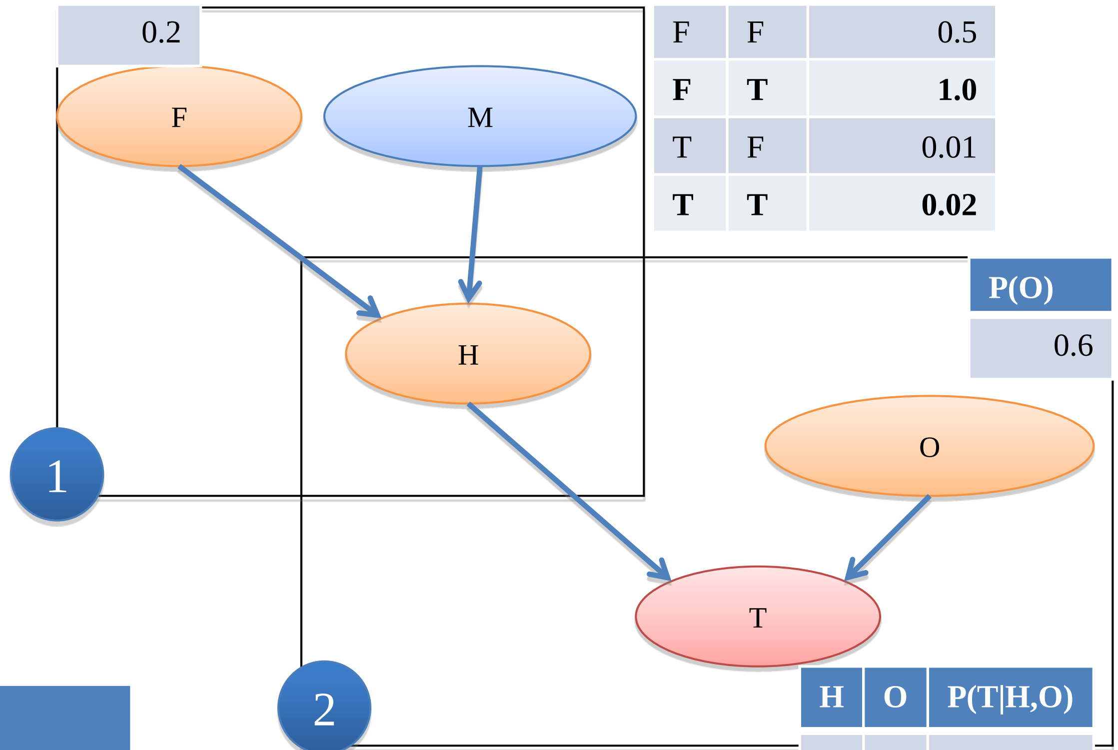
F	P(F)*P(H | F)	=
F	0.8*1.0	0.8
T	0.2*0.02	0.004
	Total	**0.804**

2

H	O	P(H)*P(O)*P(T|H,O)	=
F	F	(1-0.804)*0.4 * 0.1	0.00784
F	T	(1-0.804)*0.6 * 0.5	0.0588
T	F	0.804*0.4 * 0.5	0.1068
T	T	0.804*0.6 * 1.0	0.4824
		TOTAL	**0.71**

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**

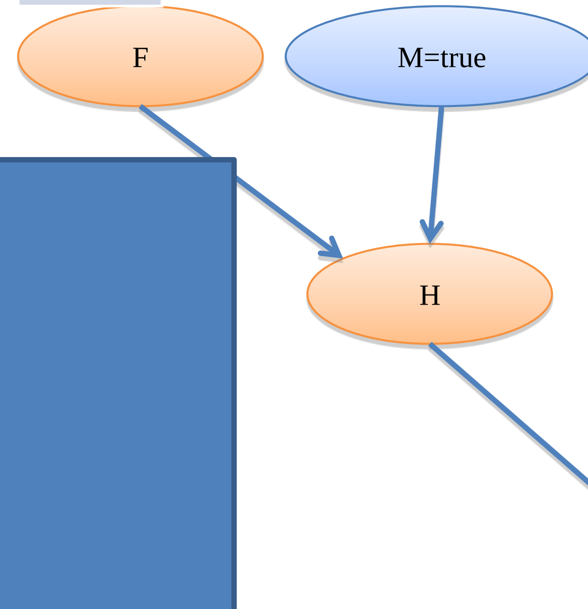


P(O)
0.6

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

Exercice 4 : par échantillonnage direct (approximation)

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

```
Random random = new Random();
    int n=10000000;
    int nbT = 0;
    for(int i=0;i<n;i++){
      boolean F, M, H, O, T;
      M=true;
      double pF=0.2;
      F = random.nextDouble() < pF;
      double pH=0;
      if(!F && !M) pH = 0.5;
      if(!F && M) pH = 1.0;
      if(F && !M) pH = 0.01;
      if(F && M) pH = 0.02;
      H = random.nextDouble() < pH;
      double pO = 0.6;
      O = random.nextDouble() < pO;
      double pT=0;
      if(!H && !O) pT = 0.1;
      if(!H && O) pT = 0.5;
      if(H && !O) pT = 0.5;
      if(H && O) pT = 1.0;
      T = random.nextDouble() < pT;
      if(T) nbT++;
    }
    System.out.println("P(T|M=true)=" + ((double)nbT)/n);
```

Requête:

Calculer P(T=true)

Variables connues:

M = true

Variables inconnues:

H

O

F

us l'erreur d'estimation est faible.

Exercice 4 : par échantillonnage direct (approximation)

Requête:

Calculer P(T=true)

Variables connues:

M = true

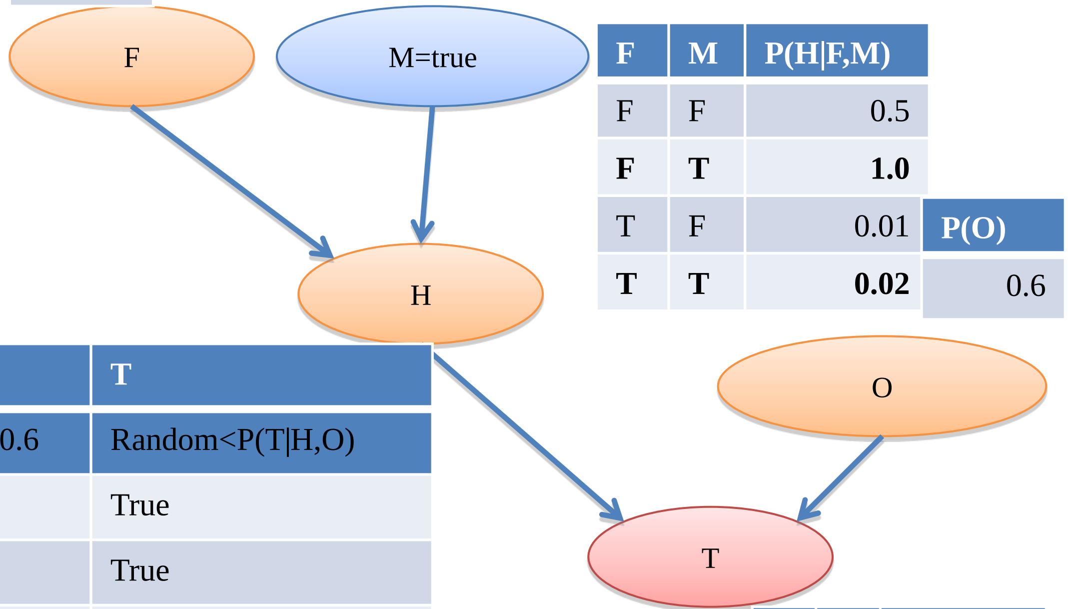
Variables inconnues:

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

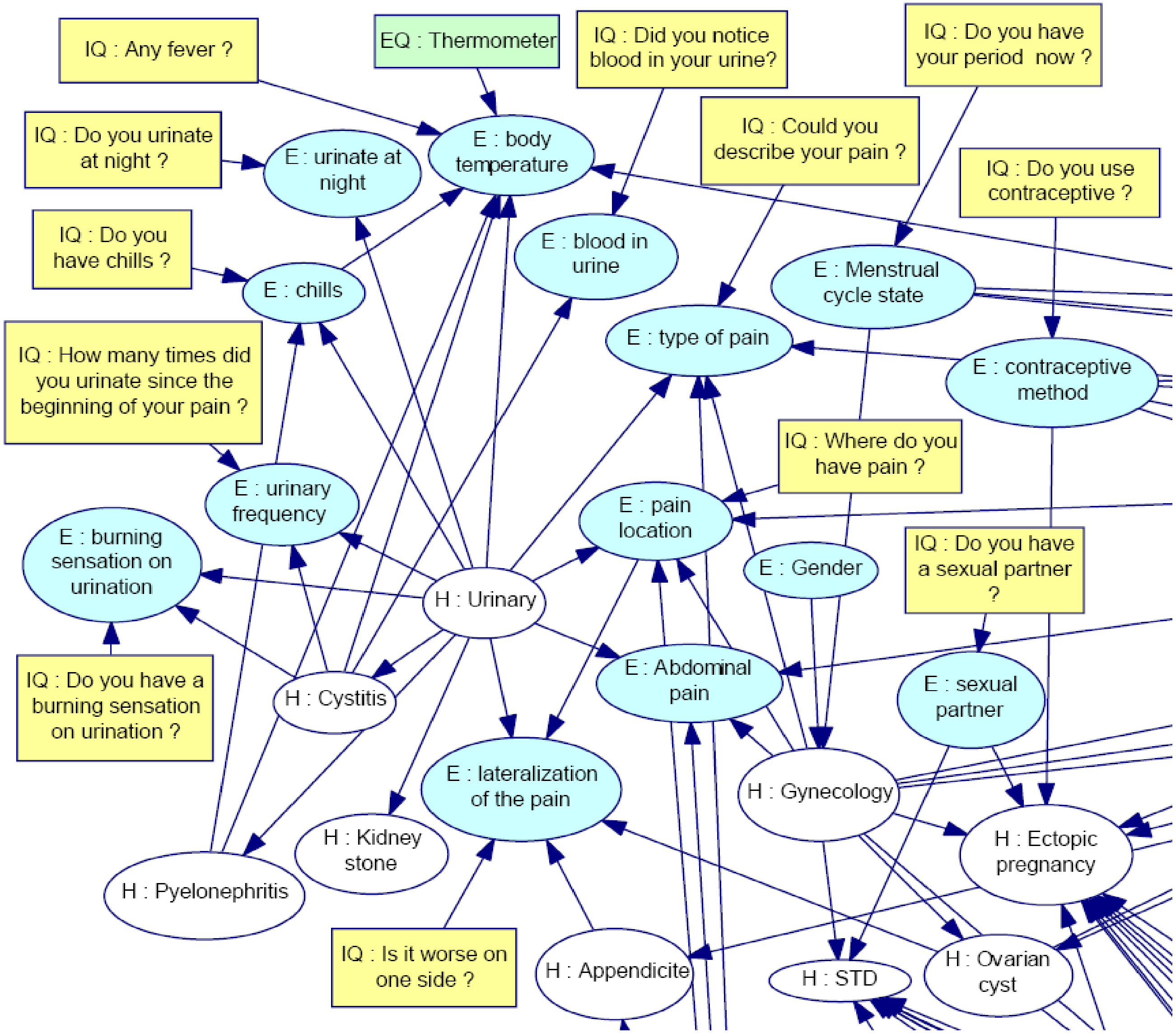
	F	H	O	T
#	Random<0.2	Random<P(H|F,M)	Random<0.6	Random<P(T|H,O)
1	False	True	True	True
2	False	True	True	True
3	False	True	False	False
4	True	False	False	False
5	False	True	True	True
6	False	True	True	True
7	False	True	True	True
8	False	True	False	True
			Average of T=True	**6/8 = 0.75**

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

Apprentissage d’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 machine.*
- 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 machine.*

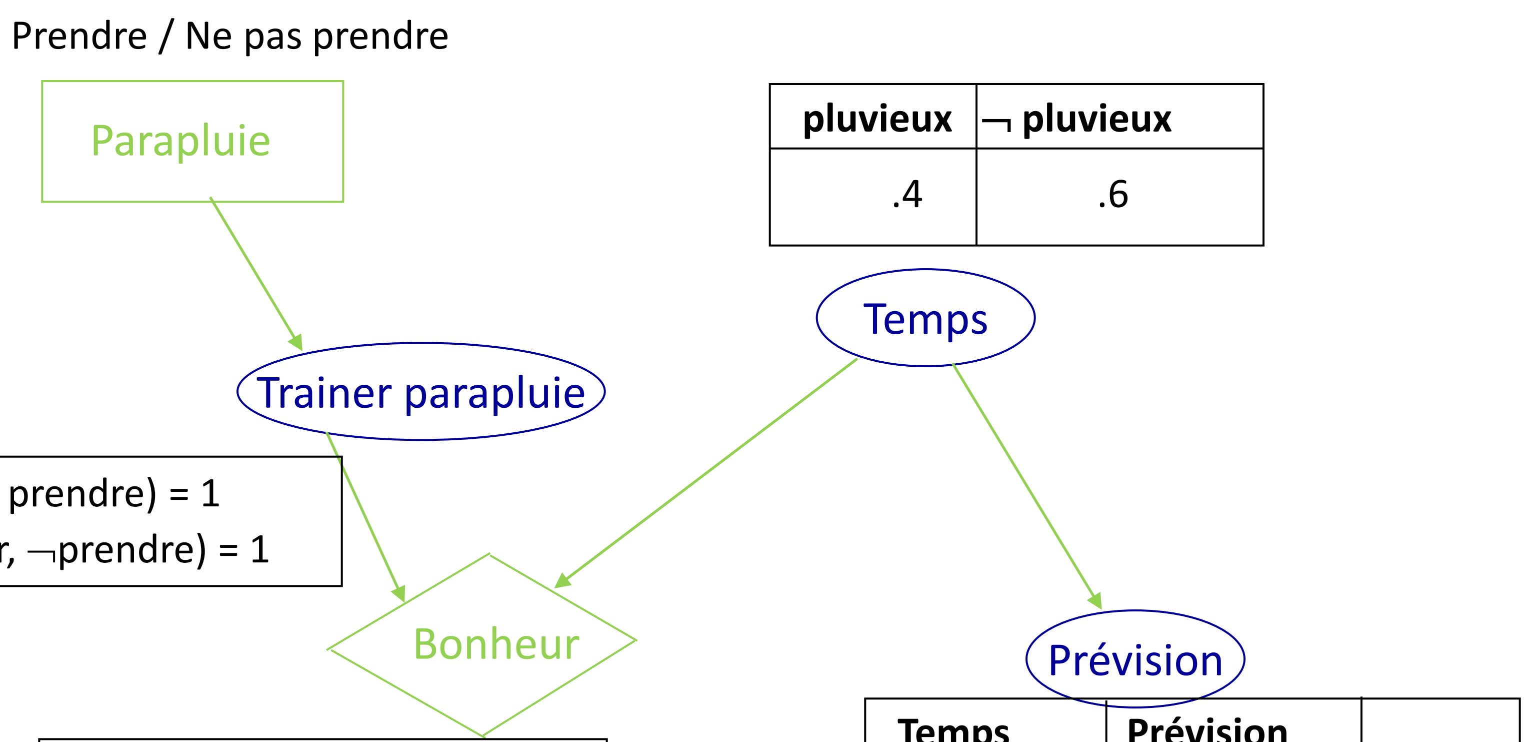
Exemple de RB complexe



Diagrammes d’influence

- Un diagramme d’influence (DI) est une extension d’un RB avec des **nœuds de décision** et des **nœuds d’utilité**.
  - Les nœuds habituels d’un RB sont appelés des **nœuds chances**.
  - 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 chances influencés par les actions.
- Ainsi on peut modéliser des prises des décisions simples
  - Pour des décisions complexes (séquentielles), les processus de décision de Markov sont généralement préférables.

Exemple



pluvieux	¬ pluvieux
.4	.6

P(trainer | prendre) = 1
P(¬ trainer, ¬prendre) = 1

U(trainer, pluvieux) = 25
U(trainer, ¬pluvieux) = -20
U(¬trainer, pluvieux) = -100
U(¬ trainer, ¬pluvieux) = 100

Temps	Prévision	
pluvieux	ensoleillé	.3
pluvieux	pluvieux	.7
¬ pluvieux	ensoleillé	.8
¬ pluvieux	pluvieux	.2

Évaluation des diagrammes d'influence

1. Mettre à jour les variables d'évidence.

2. Pour chaque valeur possible d'un nœud décision

    a. Change le nœud décision pour lui donner cette valeur.
    b. Calcule les probabilités à postérioris des parents des nœuds utilités (en utilisant un algorithme d'inférence standard pour les RB).
    c. Calcule l'utilité résultante pour l'action.

3. Retourne l'action avec la plus grande utilité

Autres appellations pour DI

- Diagrammes d’influence (*influence diagrams*).
- Réseaux de décision (*decision networks*)
- Diagrammes de pertinence (*relevance diagrams*)

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.

Logiciels

- Hugin (commercial).
- **Génie/Smile** (domaine publique)

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.