INF7370

Réseaux bayésiens

Mohamed Bouguessa

Rappels de probabilités

❑ Probabilité conditionnelle

- A et M deux événements
- Information a priori sur A : P(A)
- M s'est produit : $P(M) \neq 0$
- Information a posteriori : $$P(A|M) = \frac{P(A,M)}{P(M)}$$

Rappels de probabilités

❑ Indépendance

A et B sont indépendants ssi :

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

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

❑ Indépendance conditionnelle

A et B sont indépendants conditionnellement à C ssi :

$$P(A|B, C) = P(A|C)$$

Deux variables indépendantes

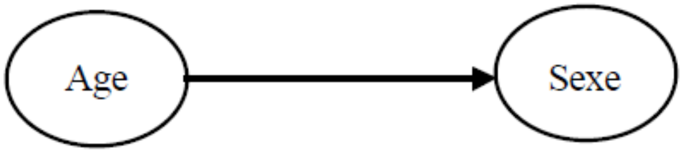
- ***Age*** et ***Sexe*** sont indépendants

$$P(A,S) = P(A).P(S)$$

$$P(A|S) = P(A) \quad \text{car} \quad A \perp S$$
$$P(S|A) = P(S) \quad \text{car} \quad A \perp S$$

$$P(A,S) = P(S|A).P(A) = P(S).P(A)$$
$$P(A,S) = P(A|S).P(S) = P(A).P(S)$$

Deux variables dépendantes



- Dans certaines espèces, ***Age*** et ***Sexe*** sont **<u>dépendants</u>**

$$P(A,S) = P(A)\ .\ P(S|A)$$

$$P(S = s) = \sum_{\hat{a}ge} P(S \mid \hat{a}ge)\ .\ P(\hat{a}ge)$$

Plan

1 – Les réseaux bayésiens

2 – Apprentissage des réseaux bayésiens

3 – Les inférences dans les réseaux bayésiens

Notations et définition

- On suppose que notre problème (étude de cas) peut être décrit par une collection de variables aléatoires $V_1, V_2, \ldots, V_k$
- Chaque variable V_i prend une valeur dans un domaine de définition (booléen, réel …) que nous noterons v_i

Définition :

Un réseau bayésien fournit une représentation graphique des relations probabilistes qui existent entre les variables aléatoires.

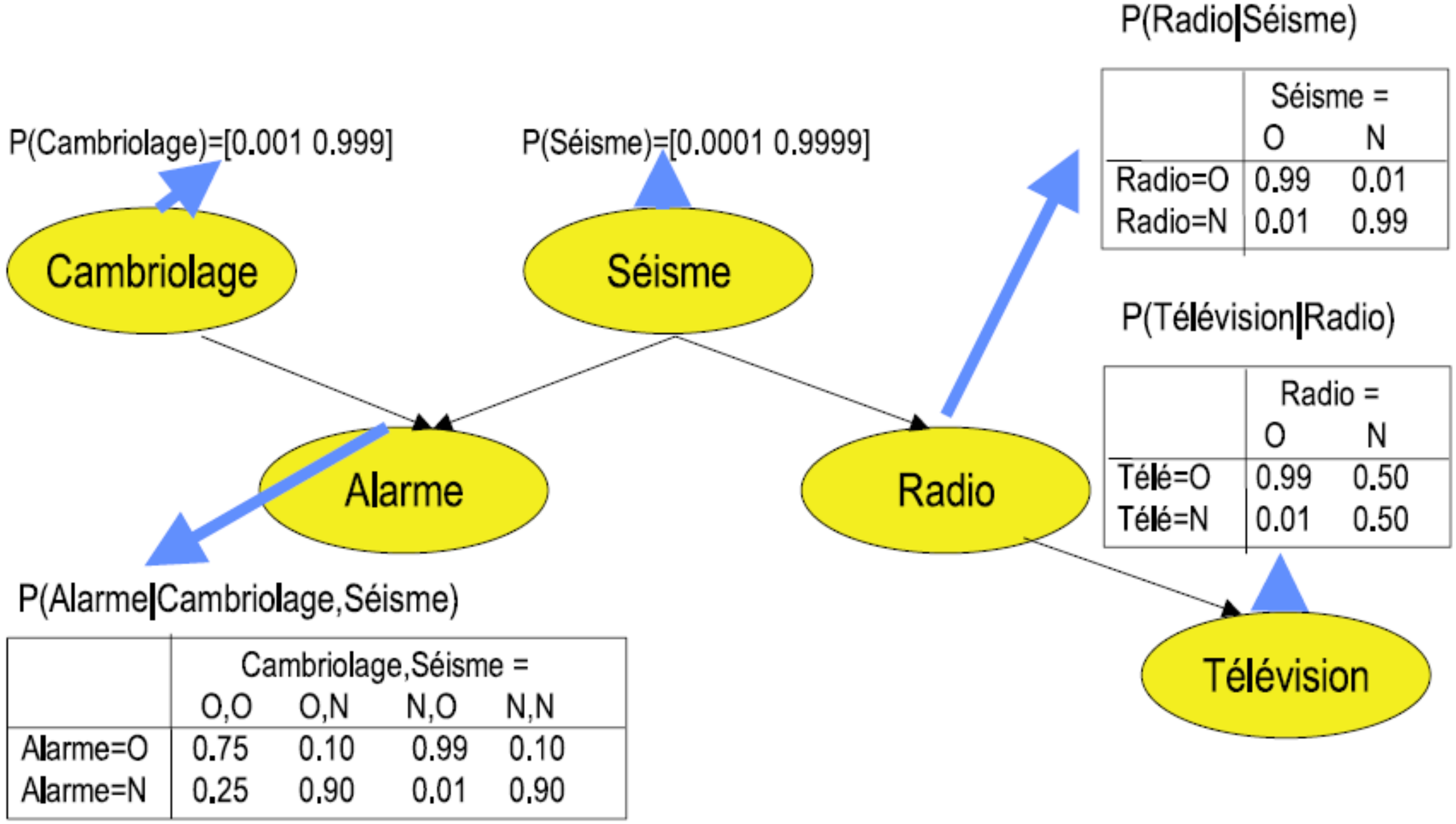
Caractéristiques

- Un réseau bayésien est un graphe orienté sans cycle (DAG : Directed Acyclic Graph).
- Chaque nœud du graphe correspond à une variable aléatoire.
- Les connexions (arcs) du graphe représentent les relations de dépendance entre les variables aléatoires.
- Une table de probabilités est associée à chaque nœud qui relate les probabilités conditionnelles du nœud courant sachant ses parents immédiats.

Exemple

Les réseaux bayésiens (RB) représentent graphiquement les indépendances conditionnelles

ordre topologique : C, S, A, R, T (non unique)



P(Radio|Séisme)

	Séisme = O	Séisme = N
Radio=O	0.99	0.01
Radio=N	0.01	0.99

P(Télévision|Radio)

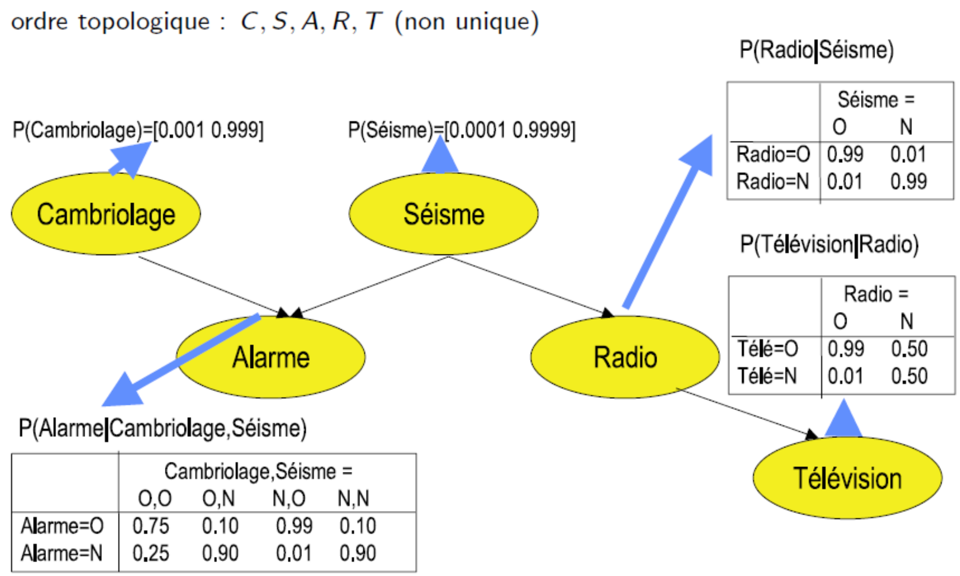
	Radio = O	Radio = N
Télé=O	0.99	0.50
Télé=N	0.01	0.50

P(Alarme|Cambriolage,Séisme)

	Cambriolage,Séisme = O,O	O,N	N,O	N,N
Alarme=O	0.75	0.10	0.99	0.10
Alarme=N	0.25	0.90	0.01	0.90

Source de la figure : Tutoriel - Philipe Leray.

Exemple



Les réseaux bayésiens conjuguent deux aspects:

- Une partie exprimant des indépendances conditionnelles entre variables et des liens de causalités.
- Une partie constituée des tables de probabilités conditionnelles de chaque variable étant donné ses parents dans le graphe.

Source de la figure : Tutoriel - Philipe Leray.

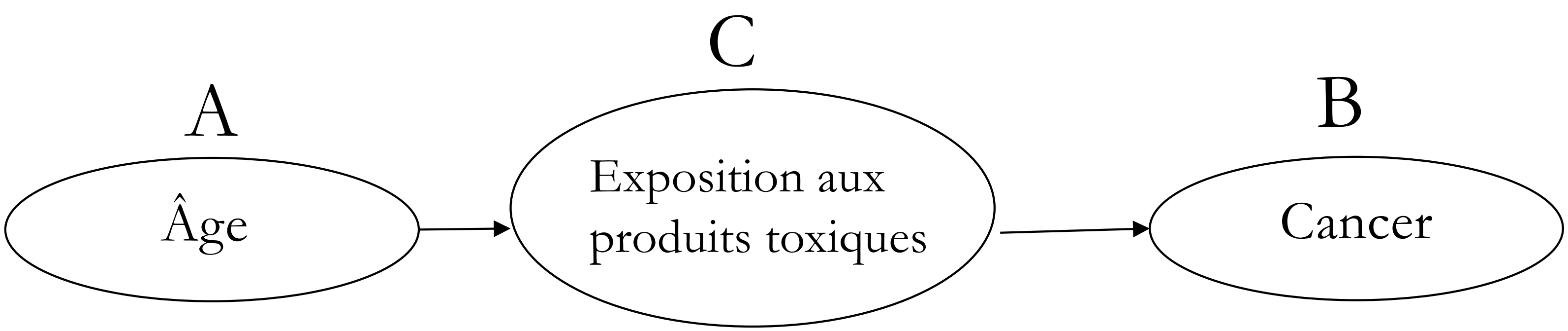
Relations possibles

Exemple sur trois noeuds :

Trois types de relations entre A, B et C :

- A → C → B : connexion en série
- A ← C → B : connexion divergente
- A → C ← B : connexion convergente (V-structure)

Connexion série

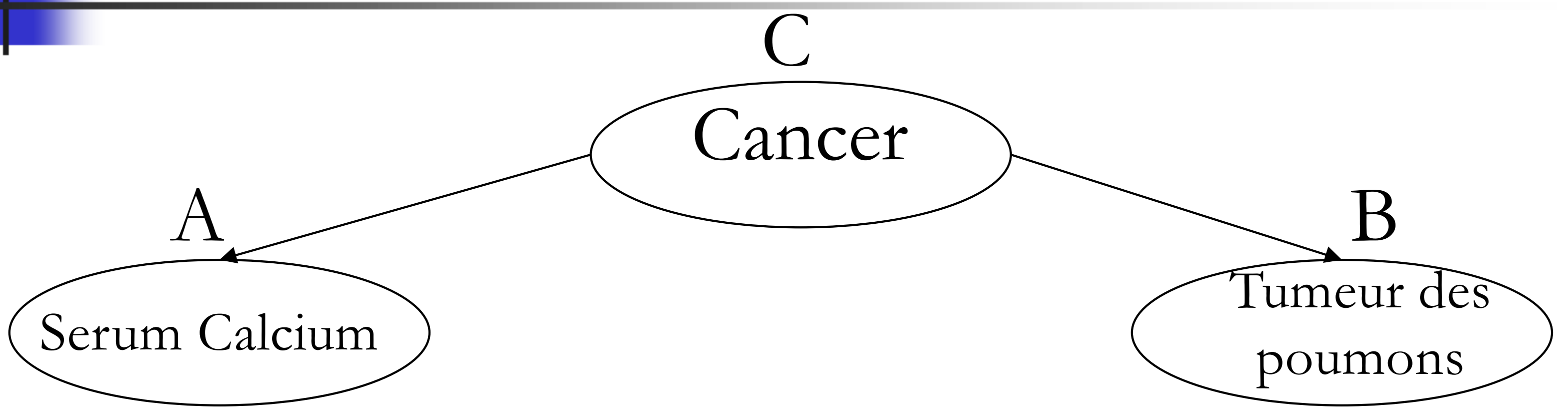


- B est conditionnellement indépendant de A sachant C

$$P(B \mid C, A) = P(B \mid C) = P(B \mid \mathrm{Parent}(B))$$

La connaissance de C rend A et B indépendants (cause intermédiaire).
Une fois on connait qu'il "Exposition aux produits toxiques" les valeurs de l'âge ne changent pas notre croyance concernant le fait d'avoir un cancer.

Connexion divergente

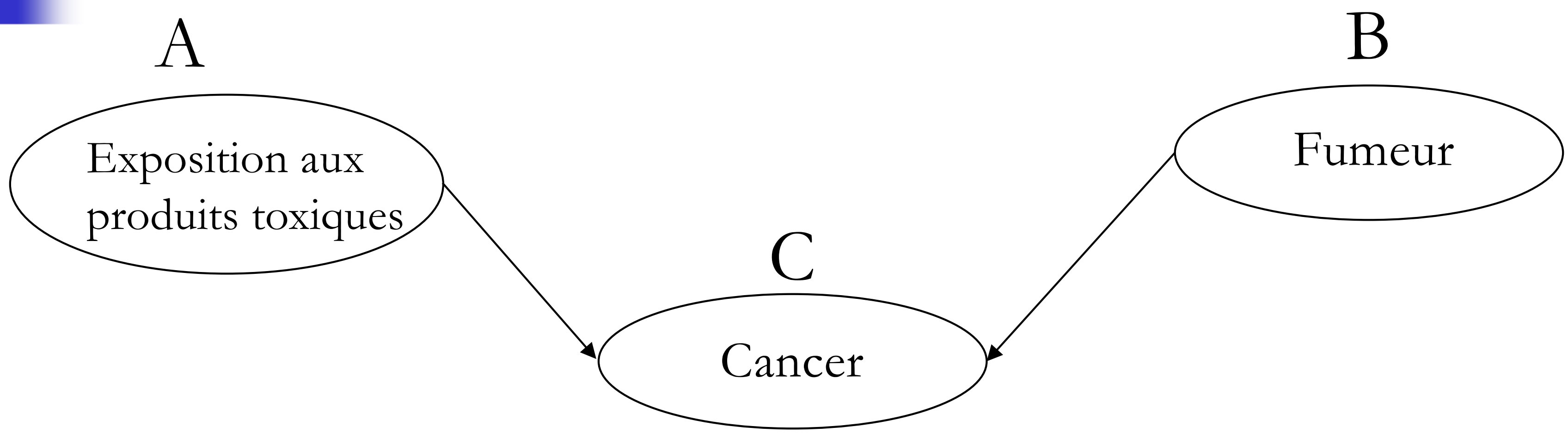


- A et B sont conditionnellement indépendants à C

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

- Il s'agit d'un scénario de « la variable cachée » :
  - Si C est inconnue, alors A et B peuvent apparaître comme dépendants l'un de l'autre.
  - A et B deviennent indépendants si l'on connaît la valeur de la variable C.

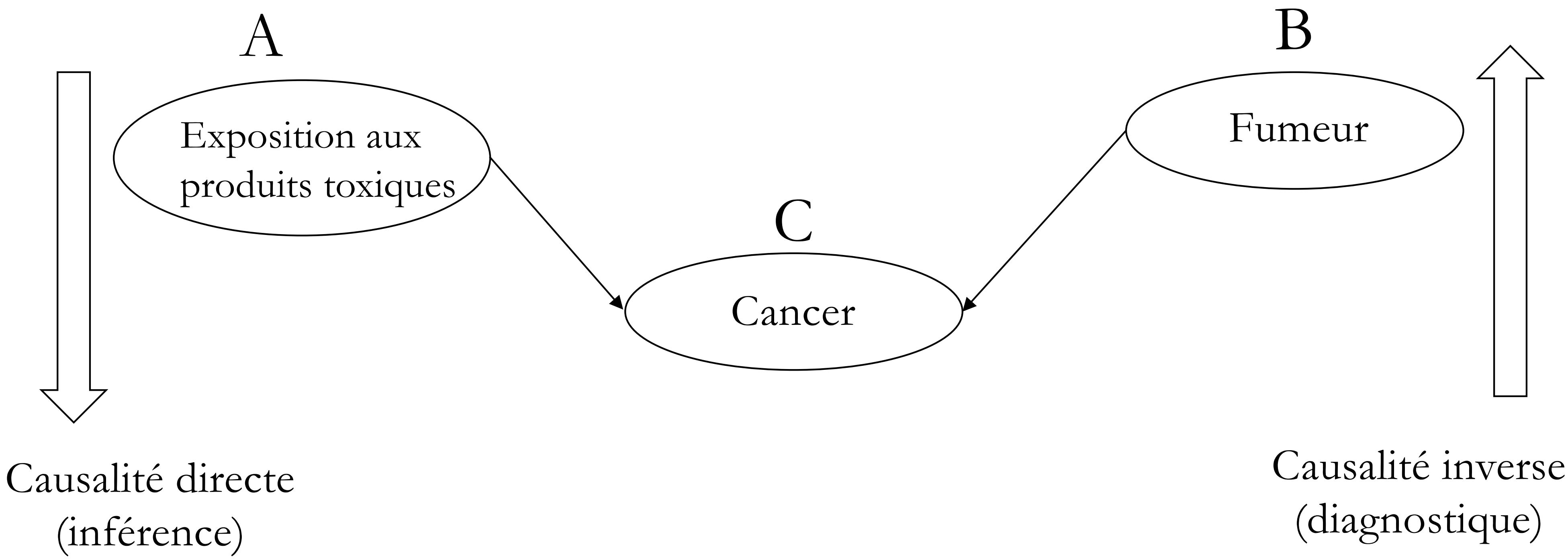
Connexion convergente : V-structure



- A et B sont a priori indépendants, mais peuvent devenir dépendants étant donné C.

- A et B sont dépendants conditionnellement à C
  - $P(C|A,B) = P(C|Parent(C))$

Connexion convergente : V-structure



Peut aussi aider à l'explication/justification par analyse de causalité inverse, en révélant tous les facteurs en amont.

Indépendance conditionnelle

- Dans un réseau bayésien, tout nœud est conditionnellement indépendant de ses non-descendants, sachant ses parents.
- En termes plus formels, notons *A(v)* n'importe quel ensemble de nœuds qui ne sont pas descendants du nœud *v* et Pr(v) l'ensemble des parents de *v*. Ceci s'écrit :

$$P(v|A(v), \mathrm{Pr}(v)) = P(v|\mathrm{Pr}(v))$$

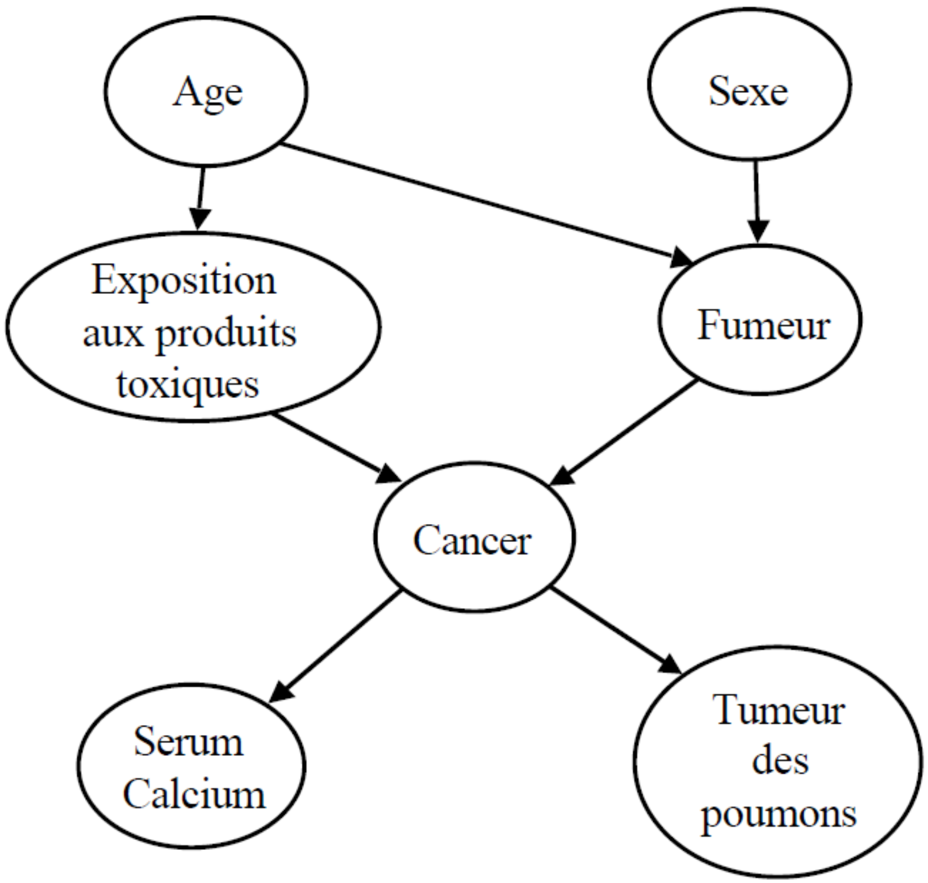
Indépendance conditionnelle

- La condition $P(v|A(v),\Pr(v)) = P(v|\Pr(v))$ peut se récrire sous la forme suivante :

- Soit $V=\{v_1, v_2, \ldots, v_d\}$ l'ensemble des nœuds du graphe.

$$P(v_1, v_2, \cdots, v_d) = \prod_{i=1}^{d} P(v_i|\Pr(v_i))$$

Exemple



$$P(A,S,E,F,C,T,SC) =$$

$$P(A)\ .\ P(S)\ .$$

$$P(E \mid A)\ .\ P(F \mid A,S)\ .$$

$$P(C \mid E,F)\ .$$

$$P(SC \mid C)\ .\ P(T \mid C)$$

Exemple de réseau bayésien

➢ Construire un réseau qui modélise les patients qui souffrent des brulures d'estomac de ceux qui ont un problème cardiaque.

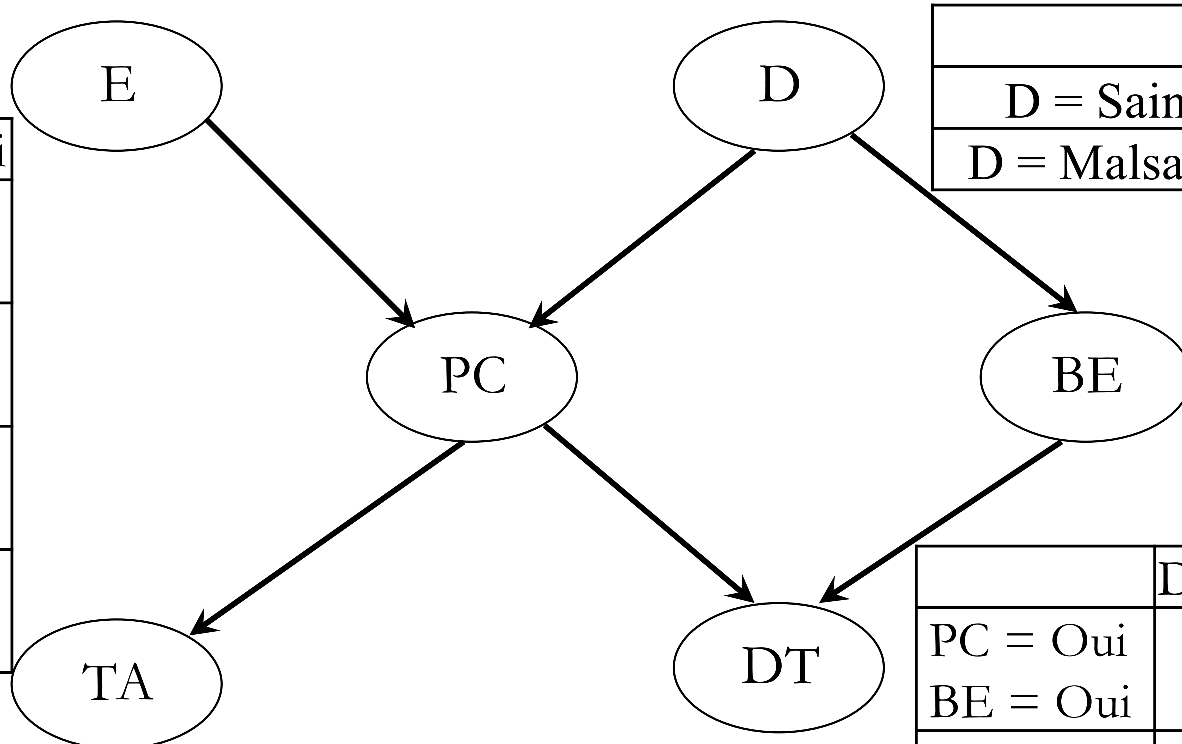
- Les variables :
  - Exercice (E)
  - Diète (D)
  - Problème cardiaque (PC)
  - Brulures d'estomac (BE)
  - Tension artérielle (TA)
  - Douleur thoracique (DT)

➢ Toutes ces variables sont booléennes (0 ou 1)

Exemple de réseau bayésien (suite)

E = Oui
0.7

D = Sain
0.25



	BE = Oui
D = Sain	0.2
D = Malsain	0.85

	PC = Oui
E = Oui D = Sain	0.25
E = Oui D = Malsain	0.45
E = Non D = Sain	0.55
E = Non D = Malsain	0.75

	TA = Élevée
PC = Oui	0.85
PC = Non	0.2

	DT = Oui
PC = Oui BE = Oui	0.8
PC = Oui BE = Non	0.6
PC = Non BE = Oui	0.4
PC = Non BE = Non	0.1

Exemple de réseau bayésien (suite et fin)

- Il convient de noter que certaines probabilités ne sont pas mentionnées dans les tableaux (contraintes d'espace). Ces probabilités peuvent être facilement obtenues comme suit :

$$P(X = \bar{x}) = 1 - P(X = x)\, et$$

$$P(X = \bar{x} | Y) = P(X = x | Y)\, avec\, \bar{x}\, est\, l'\, opposé\, de\, x$$

- Exemple

P(PC = Non | E = Non, D = Sain)

= 1 - P(PC = Oui | E = Non, D = Sain)

= 1 – 0.55 = 0.45

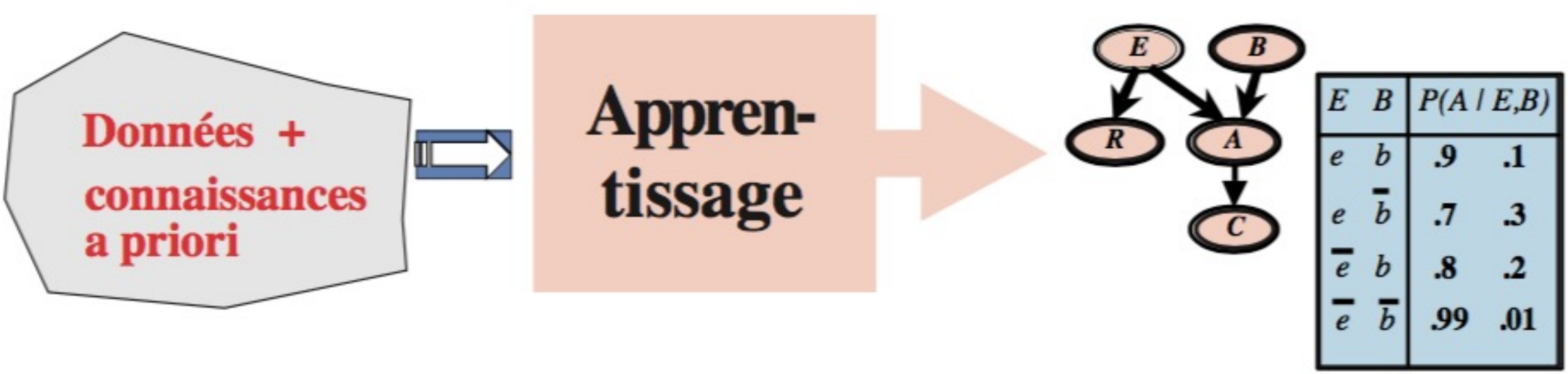
Plan

1 – Les réseaux bayésiens

2 – Apprentissage des réseaux bayésiens

3 – Les inférences dans les réseaux bayésiens

Construction d'un réseau bayésien



- Un réseau bayésien comprend à la fois une **structure** et des **paramètres** associés aux nœuds de cette structure (estimer les probabilités conditionnelles dans les tables associées à chaque nœud).

- L'apprentissage consiste à estimer les paramètres et parfois aussi la structure à partir de données et, éventuellement, des connaissances préalables.

Apprentissage de la structure

1. Établir un ordre des variables : $T = \{V_1, V_2, \ldots, V_d\}$

2. pour $j = 1, \ldots, d$ **faire**

3. Soit V_j la première variable dans l'ensemble T

4. Soit $Z_j = \{V_1, V_2, \ldots, V_{(j-1)}\}$ l'ensemble des variables qui précède V_j

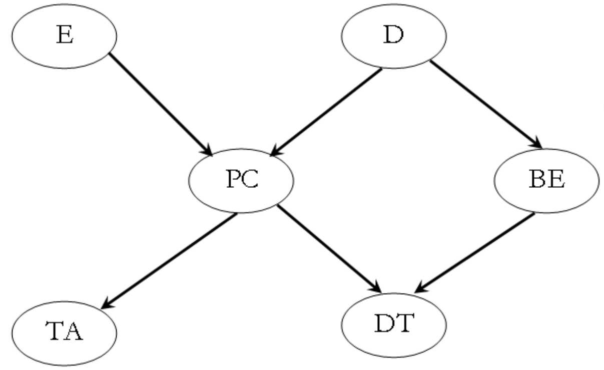
5. Éliminer les variables de l'ensemble Z_j qui n'affectent pas V_j (**en utilisant des connaissances a priori**)

6. Créer un arc entre V_j et les variables restantes dans Z_j

7. fin pour

*Source : Tan et. Al "Introduction to Data Mining", 2nd edition, Pearson, 2018.

Exemple



On considère l'exemple →

Étape 1 : établir un ordre T = {E, D, PC, BE, DT, TA}

Étape 2 – 7 : on peut estimer les probabilités suivantes:

- P(E)
- P(D|E) : simplifiée à P(D)
- P(PC|E, D)
- P(BE|E, D, PC) : simplifiée à P(BE|D)
- P(DT|E, D, PC, BE) : simplifiée à P(DT|PC, BE)
- P(TA|E, D, PC, BE, DT) : simplifiée à P(TA|PC)

➢ Maintenant on peut créer des arcs entre les nœuds: (E, PC), (D, PC), (D, BE), (PC, DT), (BE, DT) et (PC, TA)

Remarques

- Il est clair que la topologie du réseau change si on choisit un ordre différent des variables.
- Dans ce cas, il est possible d'avoir des structures de RB non significative.

Remarques

Une solution simple consiste à :

1. Diviser l'ensemble des variables en deux sous-ensembles : (1) cause et (2) effet.

2. Établir les connexions (arcs) entre chaque variable « causale » vers la variable « effet » correspondante.

- ➢ Une solution qui nécessite toujours des connaissances a priori du domaine pour simplifier les probabilités conditionnelles

- ➢ **Un RB est un système expert à base de connaissance**

Apprentissage par recherche directe des dépendances conditionnelles

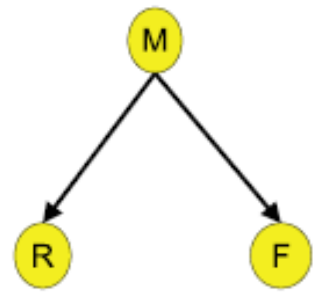
- L'idée générale de ces approches est de déterminer dans un premier temps un graphe non dirigé exprimant les dépendances entre variables détectées par des tests statistiques.
- Il est alors possible de reconstruire la structure du réseau bayésien à partir de l'ensemble des relations d'indépendances conditionnelles découvertes.
- En pratique, un graphe complètement connecté sert de point de départ. Lorsqu'une indépendance conditionnelle est détectée, l'arc correspondant est retiré.

Apprentissage des paramètres d'un RB

- L'apprentissage des paramètres d'un réseau bayésien consiste à estimer les probabilités conditionnelles associées à chaque nœud.
- L'estimation de ces probabilités est identique à l'estimation des probabilités conditionnelles dans le cas d'un classifieur bayésien naïf (estimation par fréquence).

Exemple

$$\hat{P}(M = m_0) = 6/15 = 0.4$$
$$\hat{P}(M = m_1) = 8/15 = 0.53$$
$$\hat{P}(M = m_2) = 1/15 = 0.07$$



$$\hat{P}(F = OK | M = m_0) = 1/6 = 0.17$$
$$\hat{P}(F = BAD | M = m_0) = 5/6 = 0.83$$

etc ...

M	F	R
m_0	BAD	O
m_0	BAD	O
m_0	BAD	O
m_0	BAD	O
m_0	BAD	N
m_0	OK	O
m_1	BAD	O
m_1	BAD	N
m_1	OK	O
m_1	OK	N
m_1	OK	O
m_1	OK	N
m_1	OK	O
m_1	OK	N
m_2	OK	N

Source : Tutoriel - Philipe Leray.

Plan

1 – Les réseaux bayésiens

2 – Apprentissage des réseaux bayésiens

3 – Les inférences dans les réseaux bayésiens

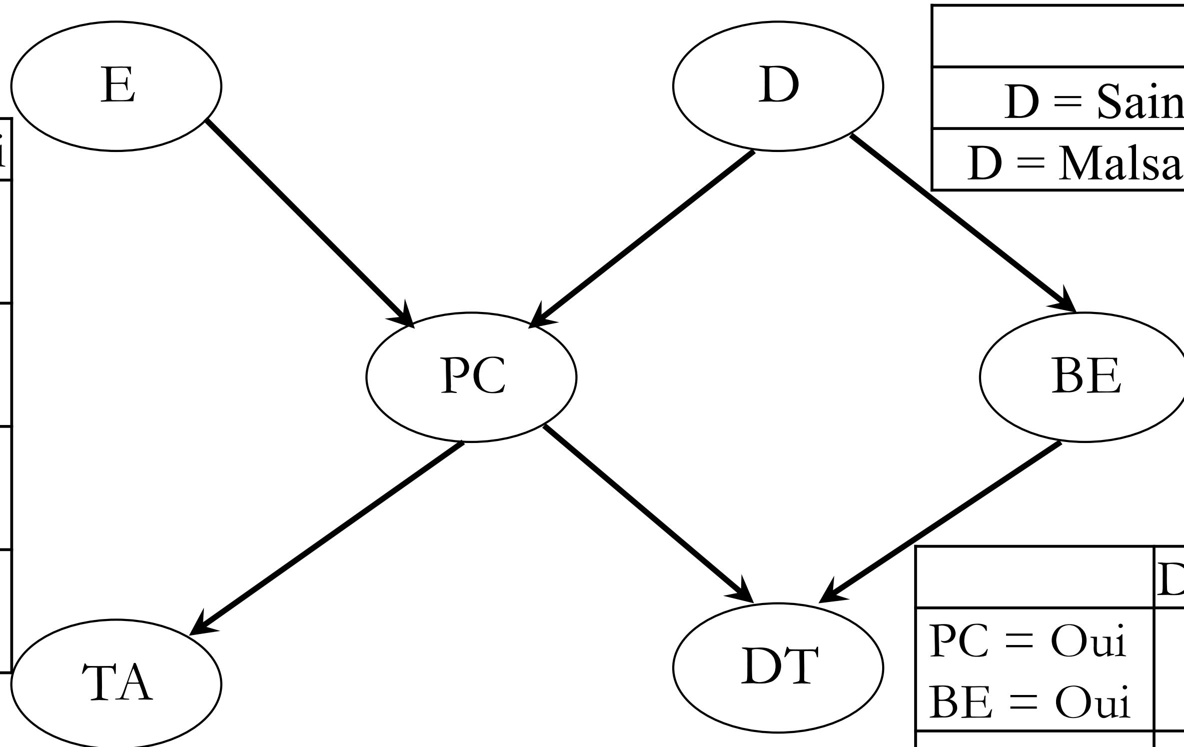
Exemple d’inférence dans les RB

- On suppose que nous sommes intéressés à utiliser le réseau bayésien de l’exemple illustré dans l’acétate suivant pour établir un diagnostic si une personne est susceptible d’avoir un problème cardiaque ou non.
- Les trois prochains cas illustrent comment le diagnostic peut se faire sous différents scénarios.

Exemple

E = Oui
0.7

D = Sain
0.25



	BE = Oui
D = Sain	0.2
D = Malsain	0.85

	PC = Oui
E = Oui D = Sain	0.25
E = Oui D = Malsain	0.45
E = Non D = Sain	0.55
E = Non D = Malsain	0.75

	DT = Oui
PC = Oui BE = Oui	0.8
PC = Oui BE = Non	0.6
PC = Non BE = Oui	0.4
PC = Non BE = Non	0.1

	TA = Élevée
PC = Oui	0.85
PC = Non	0.2

Exemple tiré de : Tan et. Al “Introduction to Data Mining”, 2nd edition, Pearson, 2018.

Cas 1 : Aucune information a priori

- Avec aucune information a priori, on peut déterminer si une personne est susceptible d'avoir un problème cardiaque ou non en calculant les probabilités :
  P(PC = Oui) et P(PC = Non).

- Notations:
  - $\alpha \in \{Oui, Non\}$ désigne l'ensemble des valeurs que la variable **Exercice** peut prendre.
  - $\beta \in \{Sain, Malsain\}$ désigne l'ensemble des valeurs que la variable **Diète** peut prendre.

Cas 1 : suite et fin

$$P(PC=Oui)=\sum_{\alpha}\sum_{\beta}P(PC=Oui\,|\,E=\alpha,D=\beta)\,P(E=\alpha,D=\beta)$$

$$=\sum_{\alpha}\sum_{\beta}P(PC=Oui\,|\,E=\alpha,D=\beta)\,P(E=\alpha)\,P(D=\beta)$$

$$=(0.25\times0.7\times0.25)+(0.45\times0.7\times0.75)+(0.55\times0.3\times0.25)$$

$$+(0.75\times0.3\times0.75)$$

$$=0.49$$

$$P(PC=Non)=1-P(PC=Oui)=0.51$$

- La personne en question à peu de chance de ne plus avoir un risque cardiaque.

Cas 2 : La tension artérielle est haute

- Si la tension artérielle d'une personne est haute, on peut faire un diagnostic au sujet de son risque cardiaque, et ce en comparant les probabilités a posteriori suivantes:

P(PC = Oui | TA = Élevée) et P(PC = Non | TA = Élevée)

- On ne peut pas estimer P(PC | TA) directement du réseau → on applique donc le théorème de Bayes

$$P(PC = Oui \mid TA = \acute{E}lev\acute{e}e) = \frac{P(TA = \acute{E}lev\acute{e}e \mid PC = Oui)\, P(PC = Oui)}{P(TA = \acute{E}lev\acute{e}e)}$$

➢ On doit donc calculer P(TA=Élevée)

Cas 2 : suite et fin

- $\lambda \in \{Oui, Non\}$ désigne l'ensemble des valeurs que la variable Problème Cardiaque (PC) peut prendre.

$$P(TA = \acute{E}lev\acute{e}e) = \sum_{\lambda} P(TA = \acute{E}lev\acute{e}e \mid PC = \lambda) P(PC = \lambda)$$

$$= (0.85 \times 0.49) + (0.2 \times 0.51) = 0.5185$$

$$P(PC = Oui \mid TA = \acute{E}lev\acute{e}e) = \frac{P(TA = \acute{E}lev\acute{e}e \mid PC = Oui)\, P(PC = Oui)}{P(TA = \acute{E}lev\acute{e}e)} = \frac{0.85 \times 0.49}{0.5185} = 0.8033$$

$$P(PC = Non \mid TA = \acute{E}lev\acute{e}e) = 1 - 0.8033 = 0.1967$$

- ➢ Donc, si la tension artérielle est élevée, la personne en question court un grand risque d'avoir un problème cardiaque.

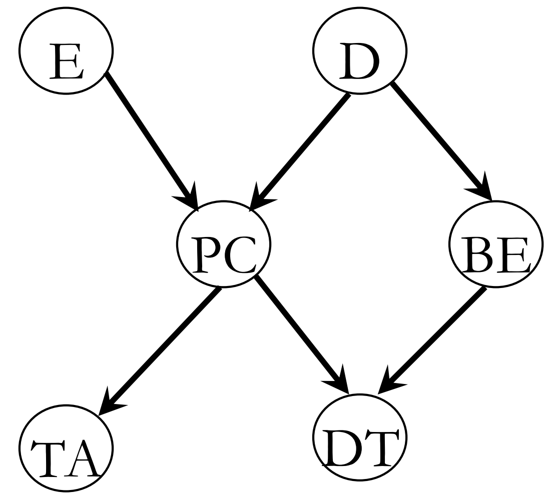
Cas 3 : TA = Élevée, D = Sain, E = Oui

- Supposons maintenant que la personne en question fait régulièrement de l'exercice et mange sain. Comment ces nouvelles informations peuvent affecter le diagnostic?
- Avec les nouvelles informations (D = Sain et E = Oui), la probabilité a posteriori que la personne en question risque d'avoir un problème cardiaque peut s'écrire comme suit :

P(PC = Oui | TA = Élevée, D = Sain, E = Oui) = ?

Cas 3 : TA = Élevée, D = Sain, E = Oui

P(PC = Oui | TA = Élevée, D = Sain, E = Oui) = ?



$$P(PC = Oui \mid TA = \acute{E}lev\acute{e}e, D = Sain, E = Oui)$$

$$= \left[\frac{P(TA = \acute{E}lev\acute{e}e \mid PC = Oui, D = Sain, E = Oui)}{P(TA = \acute{E}lev\acute{e}e \mid D = Sain, E = Oui)} \right]$$

$$\times P(PC = Oui \mid D = Sain, E = Oui)$$

$$= \frac{P(TA = \acute{E}lev\acute{e}e \mid PC = Oui) \times P(PC = Oui \mid D = Sain, E = Oui)}{\sum_{\lambda} P(TA = \acute{E}lev\acute{e}e \mid PC = \lambda) P(PC = \lambda \mid D = Sain, E = Oui)}$$

$$= \frac{0.85 \times 0.25}{(0.85 \times 0.25) + (0.2 \times 0.75)}$$

$$= 0.5862$$

Cas 3 : suite et fin

La probabilité qu'une personne puisse ne pas avoir un risque cardiaque

$$P(PC = \text{Non} \mid TA = \text{Élevée}, D = \text{Sain}, E = \text{Oui}) = 1 - 0.5862$$

$$= 0.4138$$

- Le modèle suggère donc que manger santé et faire des exercices peut réduire le risque d'avoir un problème cardiaque.

Exercice

En considérant le réseau bayésien illustré dans ce qui suit, calculer les probabilités suivantes :

(a) P(B = good, F = empty, G = empty, S = yes)

(b) P(B = bad, F = empty, G = not empty, S = no)

(c) Calculer la probabilité : la voiture démarre sachant que la batterie est mauvaise.