Décision dans l'incertain

Cours 4: Réseaux Bayésiens

Stéphane Airiau

Université Paris-Dauphine

Introduction

- la table des probabilités peut répondre à toutes les questions sur le domaine
- Si le nombre de variables est grand, la table devient rapidement trop grosse
  - difficile d'obtenir tous les nombres
  - va nécissiter beaucoup d'opérations pour calculer le résultat !
- spécifier les probabilités de chaque monde possible est difficile et pas forcemment naturel.
- l'indépendance et l'indépendance conditionnelle peuvent grandement aider à faire baisser le nombre de probabilités à connaître pour calculer la table complète.

➩ structure de données Réseaux Bayésiens : elle représente la dépendance entre chaque variable pour représenter de façon concise toute distribution complète de probabilité.

Rappel indépendance

Deux variables sont indépendantes ssi

$$\forall (x,y) \in Dom(X) \times Dom(Y) : \mathbb{P}(x,y) = \mathbb{P}(x)\mathbb{P}(y)$$

⇨ la distribution jointe est le produit de deux distributions plus simples.

on peut aussi écrire
$\forall (x,y) \in Dom(X) \times Dom(Y) : \mathbb{P}(x \mid y) = \mathbb{P}(x)$

On va noter $X \perp\!\!\!\perp Y$

Indépendance est parfois une simplification d'une hypothèse sur la modélisation.

- on observe une distribution empirique "proche" de l'indépendance.
- Que peut on attendre pour $\{Météo, Traffic, Carie, MalDeDents\}$?
  $\{Météo, Traffic\}$ sont surement indépendants de $\{Carie, MalDeDents\}$
  ⇨pas complète indépendance, mais des indépendances.

⇨ simplement vérifier les produits.

Rappel indépendance

Supposons qu'on a n dés parfaitement équilibrés

➯ si on veut représenter la distribution jointe : 6^n nombres à spécifier

On peut suppose que les dés sont indépendants !

➯ avec l'indépendance, on n'a besoin que de $6n$ nombres !

On est intéressé à des domaines où il y a des dépendances, sinon, on n'a pas grand chose à faire !

Rappel indépendance conditionnelle

- Si j'ai une carie, la détection d'une carie par la sonde est indépendante de si j'ai mal aux dents ou non !

$$\mathbb{P}(testSonde \mid malAuDent, carie) = \mathbb{P}(testSonde \mid carie)$$

- la même chose fonctionne si je n'ai pas de carie !

$$\mathbb{P}(testSonde \mid malAuDent, \neg carie) = \mathbb{P}(testSonde \mid \neg carie)$$

- La variable *TestSonde* est conditionnellement indépendante de la variable *MalauDents* étant donnée la présence de carie.

$$\mathbb{P}(TestSonde \mid MalAuDent, Carie) = \mathbb{P}(TestSonde \mid Carie)$$

d'autres

- $\mathbb{P}(MalAuDent \mid TestSonde, Carie) = \mathbb{P}(MalAuDent \mid Carie)$
- $\mathbb{P}(MalAuDent, TestSonde \mid Carie) = \mathbb{P}(MalAuDent \mid Carie)\mathbb{P}(TestSonde \mid Carie)$

Rappel indépendance conditionnelle

X est indépendante de Y sachant Z si $\forall(x,y,z) \in Dom(X) \times Dom(Y) \times Dom(Z) : \mathbb{P}(x,y \mid z) = \mathbb{P}(x \mid z)\mathbb{P}(y \mid z)$

On va noter
$X \perp\!\!\!\perp Y \mid Z$

$$(1) \quad \forall(x,y,z) \in Dom(X) \times Dom(Y) \times Dom(Z) : \mathbb{P}(x,y \mid z) = \mathbb{P}(x \mid z)\mathbb{P}(y \mid z)$$
$$(2) \quad \forall(x,y,z) \in Dom(X) \times Dom(Y) \times Dom(Z) : \mathbb{P}(x \mid y,z) = \mathbb{P}(x \mid z)$$

Rappel indépendance conditionnelle

X est indépendante de Y sachant Z si $\forall (x,y,z) \in Dom(X) \times Dom(Y) \times Dom(Z) : \mathbb{P}(x,y \mid z) = \mathbb{P}(x \mid z)\mathbb{P}(y \mid z)$

On va noter
$X \perp\!\!\!\perp Y \mid Z$

$$\begin{array}{ll} (1) & \forall (x,y,z) \in Dom(X) \times Dom(Y) \times Dom(Z) : \mathbb{P}(x,y \mid z) = \mathbb{P}(x \mid z)\mathbb{P}(y \mid z) \\ (2) & \forall (x,y,z) \in Dom(X) \times Dom(Y) \times Dom(Z) : \mathbb{P}(x \mid y,z) = \mathbb{P}(x \mid z) \end{array}$$

$(1) \Rightarrow (2)$

$$\begin{aligned} \mathbb{P}(x \mid y,z) &= \frac{\mathbb{P}(x,y,z)}{\mathbb{P}(y,z)} \\ &= \frac{\mathbb{P}(z) \cdot \mathbb{P}(x,y \mid z)}{\mathbb{P}(y,z)} \\ &= \frac{\mathbb{P}(z)\mathbb{P}(x \mid z)\mathbb{P}(y \mid z)}{\mathbb{P}(z)\mathbb{P}(y \mid z)} \\ &= \mathbb{P}(x \mid z) \end{aligned}$$

Rappel indépendance conditionnelle

- Traffic
- Parapluie
- Pluie

Rappel indépendance conditionnelle

- Traffic
- Parapluie
- Pluie

$Traffic \perp\!\!\!\perp Parapluie \mid Pluie$

Rappel indépendance conditionnelle

- Traffic
- Parapluie
- Pluie

$Traffic \perp\!\!\!\perp Parapluie \mid Pluie$

- Incendie
- Fumée
- DetecteurFumée

Rappel indépendance conditionnelle

- Traffic
- Parapluie
- Pluie

$Traffic \perp\!\!\!\perp Parapluie \mid Pluie$

- Incendie
- Fumée
- DetecteurFumée

$Incendie \perp\!\!\!\perp DetecteurFumée \mid Fumée$

Rappel indépendance conditionnelle & Dérivation

Dérivation : $\mathbb{P}(X_1, X_2, \ldots, X_n) = \mathbb{P}(X_1)\mathbb{P}(X_2 \mid X_1)\mathbb{P}(X_3 \mid X_1, X_2)\ldots$

$\mathbb{P}(Traffic, Pluie, Parapluie) = \mathbb{P}(Pluie) \cdot \mathbb{P}(Traffic \mid Pluie) \cdot \mathbb{P}(Parapluie \mid Traffic, Pluie)$

Supposons que $Traffic \perp\!\!\!\perp Parapluie \mid Pluie$.

$\mathbb{P}(Traffic, Pluie, Parapluie) = \mathbb{P}(Pluie) \cdot \mathbb{P}(Traffic \mid Pluie) \cdot \mathbb{P}(Parapluie \mid Pluie)$

⇨ moins de nombres à fournir !

plus on conditionne sur des variables ⇨ plus on a à fournir des nombres

Il faut bien choisir l'ordre avec laquelle on utilise la dérivation.

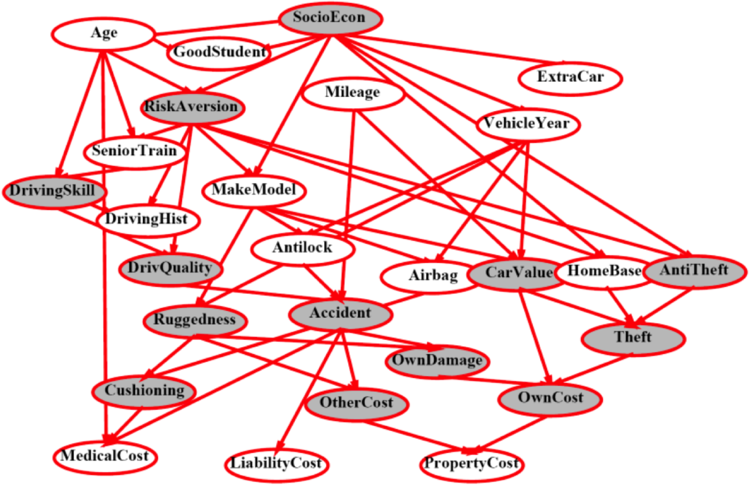
Souvent, utiliser un processus causal est une bonne solution !

Réseaux Bayésiens ou modèles graphiques

nouvelle manière de représenter une distribution jointe

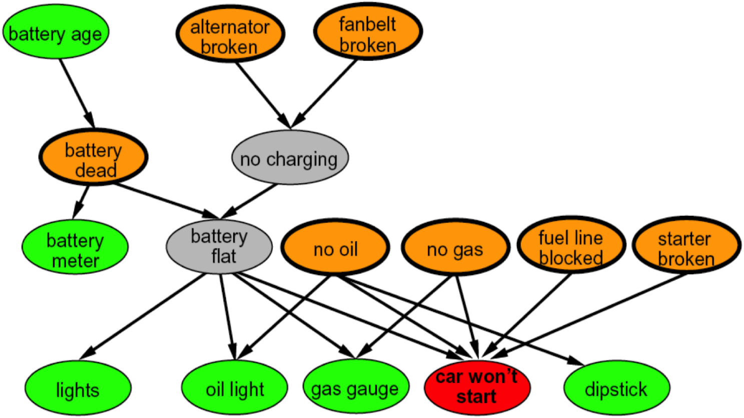
- plus compacte : moins de nombres
  - la distribution jointe est trop grande
  - difficile à estimer
- représente directement certaines hypothèses sans avoir à lire les nombres
- on va représenter des interactions locales entre variables : plus facile à obtenir et raisonner.

Examples de réseau bayésien : domaine de l'assurance



27 variables binaires ≈ 134 millions
a-t-on même assez de données pour estimer chacune ?

Examples de réseau bayésien : diagnostique



en bas, on a souvent les données que l'on peut observer.
16 variable ⇒ $2^{16} \approx 65000$ valeurs.
La plus grande table qu'on va avoir $2^5 = 120$ valeurs.

Réseaux Bayésiens ou modèles graphiques

- Les noeuds du graphe sont les variables
  - on peut connaître ou non la valeur
- On a un graphe orienté acyclique
- un arc indique une influence "directe" entre variables.

➩ encode une indépendance conditionnelle.

exemple :

- n dés équilibrés
- pluie, traffic. Onpeut considérer plusieurs hypothèses :

Réseaux Bayésiens ou modèles graphiques

- Les noeuds du graphe sont les variables
  - on peut connaître ou non la valeur
- On a un graphe orienté acyclique
- un arc indique une influence "directe" entre variables.
- ➩ encode une indépendance conditionnelle.

exemple :

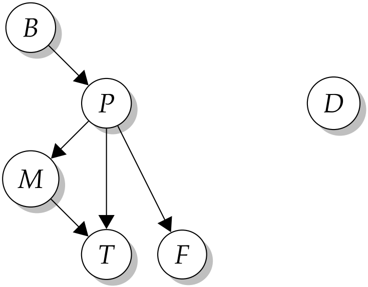
- n dés équilibrés
- pluie, traffic. Onpeut considérer plusieurs hypothèses :
  - indépendance
  - la pluie cause le traffic

Réseaux Bayésiens ou modèles graphiques

- traffic (T)
- pluie (P)
- basses pressions (B)
- match de foot (M)
- mal de dents (D)
- toit qui fuit (F)

Réseaux Bayésiens ou modèles graphiques

- traffic (T)
- pluie (P)
- basses pressions (B)
- match de foot (M)
- mal de dents (D)
- toit qui fuit (F)

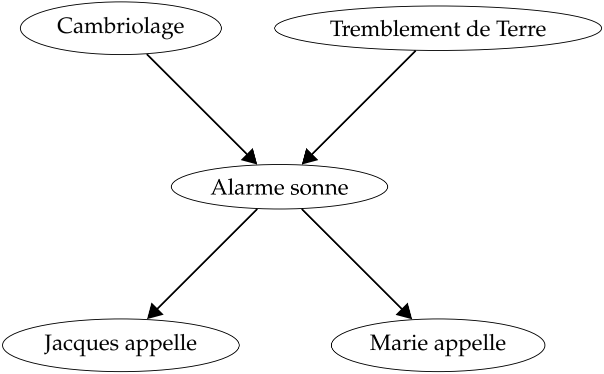


Réseaux Bayésiens : exemple classique!

- Cambriolage
- L'alarme sonne
- il y a un tremblement de terre
- le voisin (Jacques) appelle pour prévenir que quelque chose se passe
- la voisine (Marie) appelle pour prévenir que quelque chose se passe

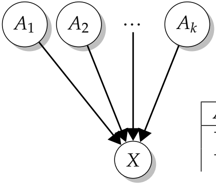
Réseaux Bayésiens : exemple classique!

- Cambriolage
- L'alarme sonne
- il y a un tremblement de terre
- le voisin (Jacques) appelle pour prévenir que quelque chose se passe
- la voisine (Marie) appelle pour prévenir que quelque chose se passe



Réseaux Bayésiens ou modèles graphiques

- Les noeuds du graphe sont les variables
- On a un graphe orienté acyclique
- Une table de probabilités conditionnelles (TPC)
  pour le noeud de la variable X dont les parents sont les noeuds $A_1, \ldots, A_k$, on a une table qui représente $\mathbb{P}(X \mid a_1, \ldots, a_k)$



A_1	A_2	$\ldots$	A_k	$\mathbb{P}(A)$
$\top$	$\top$	$\ldots$	$\top$	0.95
$\top$	$\bot$	$\ldots$	$\top$	0.94
		$\ldots$		
$\bot$	$\bot$	$\ldots$	$\bot$	0.001

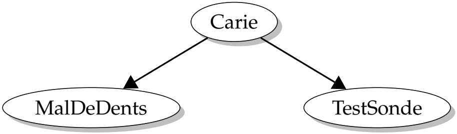
réseau bayésien = graphe + TPC

Réseaux Bayésiens & probabilités

Un réseau bayésien encode la distribution jointe complète en tant que produit de distributions conditionnelles locales.

$$\mathbb{P}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \mathbb{P}(x_i \mid parents(X_i))$$

exemple :



$$\mathbb{P}(carie, testSonde, \neg malDents) = \mathbb{P}(carie) \cdot \mathbb{P}(testSonde \mid carie) \cdot \mathbb{P}(\neg malDents \mid carie)$$

A-t-on la garantie que le produit est bien une probabilité jointe correcte ?
$\mathbb{P}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \mathbb{P}(x_i \mid parents(X_i))$

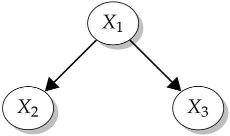
règle de dérivation :
$\mathbb{P}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \mathbb{P}(x_i \mid x_1 \ldots x_{i-1})$

hypothèse : probabilités conditionnelles
$\mathbb{P}(x_1, x_2, \ldots, x_{i-1}) = \mathbb{P}(x_i \mid parents(X_i))$

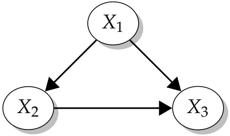
(on suppose qu'on a ordonné les noeuds de telle sorte que
$parents(X_i) \subseteq \{x_1, \ldots, x_{i-1}\}$)

➩ $\mathbb{P}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \mathbb{P}(x_i \mid parents(X_i))$

exemple avec trois variables X_1, X_2 et X_3



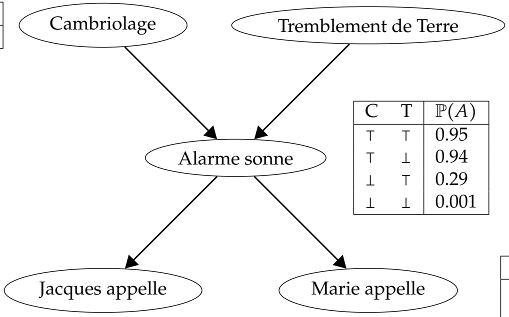
encode seulement
$X_2 \perp\!\!\!\perp X_3 \mid X_1$



encode toutes distributions !

La topologie du réseau impose des probabilités conditionnelles : étant donné un réseau bayésien, on peut représenter certaines distributions jointes, mais pas toutes !

Réseaux Bayésiens & probabilités



$\mathbb{P}(C)$
0.001

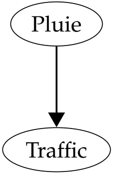
$\mathbb{P}(T)$
0.002

C	T	$\mathbb{P}(A)$
$\top$	$\top$	0.95
$\top$	$\bot$	0.94
$\bot$	$\top$	0.29
$\bot$	$\bot$	0.001

A	$\mathbb{P}(J)$
$\top$	0.90
$\bot$	0.05

A	$\mathbb{P}(M)$
$\top$	0.70
$\bot$	0.001

Réseaux Bayésiens & causalité



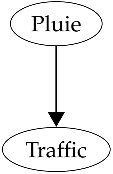
$\mathbb{P}(Pluie)$
$1/4$

Pluie	$\mathbb{P}(Traffic \mid Pluie)$
$\top$	$3/4$
$\bot$	$1/4$

on en déduit

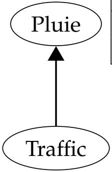
Pluie	Traffic	$\mathbb{P}(Pluie, Traffic)$
$\top$	$\top$	$3/16$
$\top$	$\bot$	$1/16$

Réseaux Bayésiens & causalité



$\mathbb{P}(Pluie)$
1/4

Pluie	$\mathbb{P}(Traffic \mid Pluie)$
$\top$	3/4
$\perp$	1/4



Traffic	$\mathbb{P}(Pluie \mid Traffic)$
$\top$	1/3
$\perp$	2/3

$\mathbb{P}(Traffic)$
9/16

on en déduit

Pluie	Traffic	$\mathbb{P}(Pluie, Traffic)$
$\top$	$\top$	3/16
$\top$	$\perp$	1/16

on en déduit

Pluie	Traffic	$\mathbb{P}(Pluie, Traffic)$
$\top$	$\top$	3/16
$\top$	$\perp$	1/16

Réseaux Bayésiens & causalité

- quand le réseau reflète la vraie structure causale
  - souvent plus petit (les noeuds ont moins de parents)
  - souvent plus facile pour le raisonnement (plus naturel)
  - plus facile pour demander aux experts (ont plus d'intuition dans ce sens)
- mais pas besoin de respecter la structure causale !
  - s'il manque des variables, la structure causale n'est pas claire. ex on veut faire un RB avec toit qui fuit et traffic, mais sans la pluie.
- la topologie encode l'indépendance conditionnelle.

Réseaux Bayésiens & taille des tables

- n variables booléénnes : 2^n
- Supposons qu'on est un réseau bayésien avec n variables, et chaque noeud possède au plus k parents.
  Il faut une table de $O(n \cdot 2^k \cdot 2)$.

On passe linéaire en n !

Si on arrive à maintenir k relativement petit !

- moins de nombres à utiliser pour les calculs !
- les nombres sont aussi plus facile à trouver !