Décision dans l'incertain

Cours 6: Inférence dans les Réseaux Bayésiens

Stéphane Airiau

Université Paris-Dauphine

Inférence avec les réseaux bayésiens

Calculer des proabilités utiles à partir de la distribution jointe.

exemples :

- calculer $\mathbb{P}(Q \mid e_1, \ldots, e_k)$
- calculer l'explication la plus plausible

$$\max_q \mathbb{P}(q \mid e_1, \ldots, e_k)$$

Technique 1 : énumération

Déjà rencontrée :

1- On utilise la définition des probabilités conditionnelles

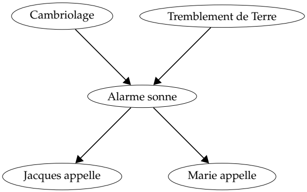
$$\mathbb{P}(Q \mid e_1, \ldots, e_k) \propto \mathbb{P}(Q, e_1, \ldots, e_k)$$

2- On somme sur toutes les variables non présentes

$$\mathbb{P}(Q \mid e_1, \ldots, e_k) \propto \sum_{z_1, \ldots z_l} \mathbb{P}(\underbrace{Q, z_1, \ldots, z_l, e_1, \ldots, e_k}_{x_1, \ldots, x_n})$$

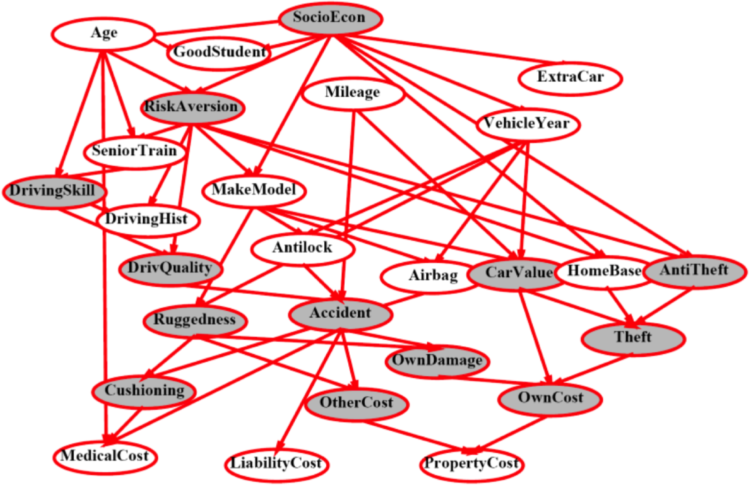
3- on normalise pour obtenir une probabilité

Technique 1 : énumération – exemple



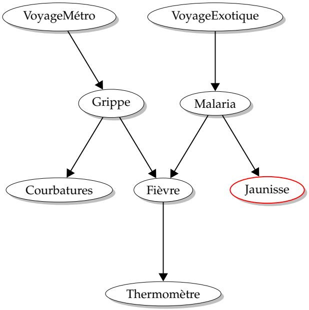
$$
\begin{aligned}
\mathbb{P}(C \mid j, m) &\propto \mathbb{P}(C, j, m) \\
&= \sum_{t,a} \mathbb{P}(C, t, a, j, m) \\
&= \sum_{t,a} \mathbb{P}(C)\mathbb{P}(t)\mathbb{P}(a \mid C, t)\mathbb{P}(j \mid a)\mathbb{P}(m \mid a) \\
&= \mathbb{P}(C)\mathbb{P}(t)\mathbb{P}(a \mid C, t)\mathbb{P}(j \mid a)\mathbb{P}(m \mid a) \\
&\quad + \mathbb{P}(C)\mathbb{P}(t)\mathbb{P}(\bar{a} \mid C, t)\mathbb{P}(j \mid \bar{a})\mathbb{P}(m \mid \bar{a}) \\
&\quad + \mathbb{P}(C)\mathbb{P}(\bar{t})\mathbb{P}(a \mid C, \bar{t})\mathbb{P}(j \mid a)\mathbb{P}(m \mid a) \\
&\quad + \mathbb{P}(C)\mathbb{P}(\bar{t})\mathbb{P}(\bar{a} \mid C, \bar{t})\mathbb{P}(j \mid \bar{a})\mathbb{P}(m \mid \bar{a})
\end{aligned}
$$

Technique 1 : énumération



$$\mathbb{P}(\textit{Antilock} \mid \text{variables observées}) = ?$$

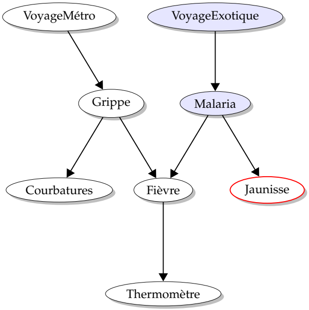
Simple inférence



$$
\begin{aligned}
\mathbb{P}(J) &= \sum_{M,Vexo} \mathbb{P}(J \mid M, Vexo)\mathbb{P}(M, Vexo) && (1)\\
&= \sum_{M,Vexo} \mathbb{P}(J \mid M)\mathbb{P}(M \mid Vexo)\mathbb{P}(Vexo) && (2)\\
&= \sum_{M} \mathbb{P}(J \mid M) \sum_{Vexo} \mathbb{P}(M \mid Vexo)\mathbb{P}(Vexo) && (3)
\end{aligned}
$$

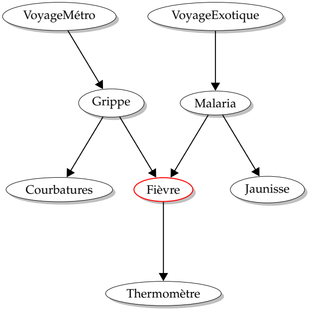
Seuls les ancêtres de J sont utilisés.
Je n'ai plus besoin que de regarder des TPC.

Simple inférence



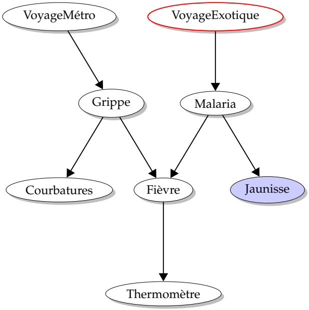
$$\begin{aligned}\mathbb{P}(J \mid vexo) &= \sum_M \mathbb{P}(J \mid M, vexo)\mathbb{P}(M \mid vexo) \\ &= \sum_M \mathbb{P}(J \mid M)\mathbb{P}(M \mid vexo)\end{aligned}$$

Simple inférence – parents multiples



$$
\begin{aligned}
\mathbb{P}(F) &= \sum_{G,M} \mathbb{P}(F \mid G,M)\mathbb{P}(G,M) \\
&= \sum_{G,M} \mathbb{P}(F \mid G,M)\mathbb{P}(G)\mathbb{P}(M) \\
&= \sum_{G,M} \mathbb{P}(F \mid G,M) \sum_{Vm\acute{e}t} \mathbb{P}(G|VM\acute{e}t)\mathbb{P}(Vm\acute{e}t) \sum_{Vexo,T} \mathbb{P}(M \mid Vexo)\mathbb{P}(Vexo)
\end{aligned}
$$

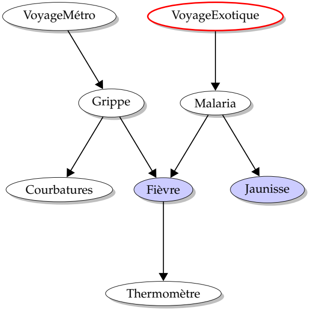
Simple inférence – à l'envers



observation est plus basse que la variable demandée, on utilise Bayes.

$$\begin{aligned}
\mathbb{P}(Exo|j) &\propto \mathbb{P}(j \mid Vexo)\mathbb{P}(Vexo) \\
&\propto \sum_M \mathbb{P}(j \mid M, Vexo)\mathbb{P}(M \mid Vexo)\mathbb{P}(Vexo) \\
&\propto \sum_M \mathbb{P}(j \mid M)\mathbb{P}(M \mid Vexo)\mathbb{P}(Vexo)
\end{aligned}$$

Simple inférence – à l'envers avec plusieurs observations



$$
\begin{aligned}
\mathbb{P}(Exo|j,f) &\propto \mathbb{P}(j,f \mid Vexo)\mathbb{P}(Vexo) \\
&\propto \sum_M \mathbb{P}(j,f \mid M, Vexo)\mathbb{P}(M \mid Vexo)\mathbb{P}(Vexo) \\
&\propto \sum_M \mathbb{P}(j,f \mid M)\mathbb{P}(M \mid Vexo)\mathbb{P}(Vexo) \\
&\propto \sum_M \mathbb{P}(j \mid M)\mathbb{P}(f \mid M)\mathbb{P}(M \mid Vexo)\mathbb{P}(Vexo)
\end{aligned}
$$

Technique 1 : énumération – problème

Avec un approche naïve on passe par la table complète avant de sommer sur les variables non présentes... explosion.

on va essayer de ne pas faire cela d'un coup !

Dans les exemples précédents, on a reussi à utiliser l'indépendance et à mettre la somme "à l'intérieur" pour faciliter le calcul et utiliser les TPC.

mais avant de voir l'algorithme d'élimination des variables, on a besoin d'outils.

Factor – 1

- Distribution jointe $\mathbb{P}(X,Y)$
  - entrées $\mathbb{P}(x,y)$ pour tout x, y
  - somme de la table : 1

$\mathbb{P}(T,M)$

T	M	$\mathbb{P}$
chaud	soleil	0.4
chaud	pluie	0.1
froid	soleil	0.2
froid	pluie	0.3

- Selection $\mathbb{P}(x,Y)$
  - on veut une tranche de la table complète
  - entrées $\mathbb{P}(x,y)$ pour x **fixé** et tout y
  - somme de la table : $\mathbb{P}(x)$

$\mathbb{P}(\mathit{froid},M)$

T	M	$\mathbb{P}$
froid	soleil	0.2
froid	pluie	0.3

Factor – 2

- Conditionnelle simple $\mathbb{P}(Y \mid x)$
  - entrées $\mathbb{P}(y \mid x)$ pour tout x **fixé** et tout y
  - somme de la table : 1

$\mathbb{P}(M \mid froid)$

T	M	$\mathbb{P}$
froid	soleil	0.4
froid	pluie	0.6

- famille de conditionnelles $\mathbb{P}(X \mid Y)$
  - de multiples conditionnelles
  - entrées $\mathbb{P}(x \mid y)$ pour x et tout y
  - somme de la table : ?

$\mathbb{P}(M \mid T)$

T	M	$\mathbb{P}$
chaud	soleil	0.8
chaud	pluie	0.2
froid	soleil	0.4
froid	pluie	0.6

Factor – 3

famille spécifique de conditionnelles $\mathbb{P}(y \mid X)$

- entrées $\mathbb{P}(y \mid x)$ pour y **fixé** et tout x
- somme de la table : ?

$\mathbb{P}(pluie \mid T)$

T	M	$\mathbb{P}$
froid	soleil	0.2
froid	pluie	0.6

Factors

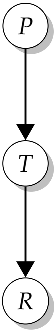
$$\mathbb{P}(Y_1, Y_2, \ldots, Y_k \mid X_1, \ldots, X_m)$$

- c'est un tableau a plusieurs dimensions
- les valeurs sont $\mathbb{P}(y_1, y_2, \ldots, y_k \mid x_1, \ldots, x_m)$
- toute valeur affectée (i.e. en minuscule) est une dimension manquante (tranche de la table générale)
- chaque TCP du réseau bayésien est un factor.

Exemple

Variables

- P : il pleut
- T : il y a du traffic
- R : je suis en retard pour le cours



$\mathbb{P}(P)$

p	0.1
$\bar{p}$	0.9

$\mathbb{P}(T \mid P)$

p	t	0.8
p	$\bar{t}$	0.2
$\bar{p}$	t	0.1
$\bar{p}$	$\bar{t}$	0.9

$\mathbb{P}(R \mid T)$

t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

$$\begin{aligned}
\mathbb{P}(R) &= ? \\
&= \sum_{p,t} \mathbb{P}(p,t,R) \\
&= \sum_{p,t} \mathbb{P}(p)\mathbb{P}(t \mid p)\mathbb{P}(R \mid t)
\end{aligned}$$

Elimination de variables

- initialement, les factors sont les tables locales TCP

$\mathbb{P}(P)$

p	0.1
$\bar{p}$	0.9

$\mathbb{P}(T \mid P)$

p	t	0.8
p	$\bar{t}$	0.2
$\bar{p}$	t	0.1
$\bar{p}$	$\bar{t}$	0.9

$\mathbb{P}(R \mid T)$

t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

- chacune des valeurs connues sont selectionnées
  par exemple si je suis en retard r

$\mathbb{P}(P)$

p	0.1
$\bar{p}$	0.9

$\mathbb{P}(T \mid P)$

p	t	0.8
p	$\bar{t}$	0.2
$\bar{p}$	t	0.1
$\bar{p}$	$\bar{t}$	0.9

$\mathbb{P}(r \mid T)$

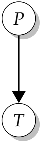
t	r	0.3
$\bar{t}$	r	0.1

- on joint les factors, puis on élimine les variables cachées.

Opération1 : Jointure de factors

- comme jointure en base de données

➯ construire un nouveau factor sur l'union des variables impliquées



$\mathbb{P}(P)$

p	0.1
$\bar{p}$	0.9

$\bowtie$

$\mathbb{P}(T \mid P)$

p	t	0.8
p	$\bar{t}$	0.2
$\bar{p}$	t	0.1
$\bar{p}$	$\bar{t}$	0.9

$\Longrightarrow$

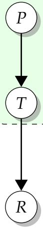
$\mathbb{P}(P,T)$

p	t	0.08
p	$\bar{t}$	0.02
$\bar{p}$	t	0.09
$\bar{p}$	$\bar{t}$	0.81

Pour se faire, on fait juste des produits **termes à termes**

$$\forall p,t : \mathbb{P}(p,t) = \mathbb{P}(p) \cdot \mathbb{P}(t \mid p)$$

Exemple : jointures multiples



$\mathbb{P}(P)$

p	0.1
$\bar{p}$	0.9

$\bowtie$

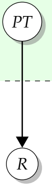
$\mathbb{P}(T \mid P)$

p	t	0.8
p	$\bar{t}$	0.2
$\bar{p}$	t	0.1
$\bar{p}$	$\bar{t}$	0.9

$\Longrightarrow$

$\mathbb{P}(P,T)$

p	t	0.08
p	$\bar{t}$	0.02
$\bar{p}$	t	0.09
$\bar{p}$	$\bar{t}$	0.81



$\mathbb{P}(R \mid T)$

t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

$\bowtie$

$\mathbb{P}(R \mid T)$

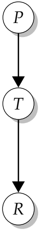
t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

$\Longrightarrow$

$\mathbb{P}(P,T,R)$

p	t	r	0.024
p	t	$\bar{r}$	0.056
p	$\bar{t}$	r	0.002
p	$\bar{t}$	$\bar{r}$	0.018
$\bar{p}$	t	r	0.027
$\bar{p}$	t	$\bar{r}$	0.063
$\bar{p}$	$\bar{t}$	r	0.081
$\bar{p}$	$\bar{t}$	$\bar{r}$	0.729

Exemple : jointures multiples



$\mathbb{P}(P)$

p	0.1
$\bar{p}$	0.9

$\bowtie$

$\mathbb{P}(T \mid P)$

p	t	0.8
p	$\bar{t}$	0.2
$\bar{p}$	t	0.1
$\bar{p}$	$\bar{t}$	0.9

$\mathbb{P}(R \mid T)$

t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

$\Longrightarrow$

$\mathbb{P}(P,T)$

p	t	0.08
p	$\bar{t}$	0.02
$\bar{p}$	t	0.09
$\bar{p}$	$\bar{t}$	0.81

$\bowtie$

$\mathbb{P}(R \mid T)$

t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

$\Longrightarrow$

$\mathbb{P}(P,T,R)$

p	t	r	0.024
p	t	$\bar{r}$	0.056
p	$\bar{t}$	r	0.002
p	$\bar{t}$	$\bar{r}$	0.018
$\bar{p}$	t	r	0.027
$\bar{p}$	t	$\bar{r}$	0.063
$\bar{p}$	$\bar{t}$	r	0.081
$\bar{p}$	$\bar{t}$	$\bar{r}$	0.729

Opération 2 : marginalisation

- prendre un factor et sommer sur une variable
- ➩ comme un opérateur de projection

$\mathbb{P}(P,T)$

p	t	0.08
p	$\bar{t}$	0.02
$\bar{p}$	t	0.09
$\bar{p}$	$\bar{t}$	0.81

$\Longrightarrow$

$\mathbb{P}(T)$

t	0.17
$\bar{t}$	0.83

marginalisations multiples

$\mathbb{P}(P,T,R)$

p	t	r	0.024
p	t	$\bar{r}$	0.056
p	$\bar{t}$	r	0.002
p	$\bar{t}$	$\bar{r}$	0.018
$\bar{p}$	t	r	0.027
$\bar{p}$	t	$\bar{r}$	0.063
$\bar{p}$	$\bar{t}$	r	0.081
$\bar{p}$	$\bar{t}$	$\bar{r}$	0.729

$\Longrightarrow$

$\mathbb{P}(T,R)$

t	r	0.051
t	$\bar{r}$	0.119
$\bar{t}$	r	0.083
$\bar{t}$	$\bar{r}$	0.747

$\Longrightarrow$

$\mathbb{P}(R)$

r	0.134
$\bar{r}$	0.886

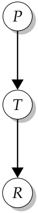
Eliminations de variables

marginaliser tôt ⇒ éliminer une variable !

$$\mathbb{P}(R) = \underbrace{\sum_t \underbrace{\sum_p \underbrace{\mathbb{P}(R \mid T) \underbrace{\mathbb{P}(T \mid P)\mathbb{P}(P)}_{\text{jointure avec } P}}_{\text{Jointure avec } T}}_{\text{élimination } p}}_{\text{élimination } t}$$

$$\mathbb{P}(R) = \underbrace{\sum_t \underbrace{\mathbb{P}(R \mid T) \underbrace{\sum_p \underbrace{\mathbb{P}(P)\mathbb{P}(T \mid P)}_{\text{jointure avec } P}}_{\text{élimination } p}}_{\text{jointure avec } T}}_{\text{élimination } t}$$

Eliminations de variables



$\mathbb{P}(P)$

p	0.1
$\bar{p}$	0.9

$\mathbb{P}(T \mid P)$

p	t	0.8
p	$\bar{t}$	0.2
$\bar{p}$	t	0.1
$\bar{p}$	$\bar{t}$	0.9

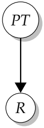
$\mathbb{P}(R \mid T)$

t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

$\Rightarrow$ Jointure

$\mathbb{P}(P,T)$

p	t	0.08
p	$\bar{t}$	0.02
$\bar{p}$	t	0.09
$\bar{p}$	$\bar{t}$	0.81



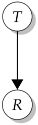
$\mathbb{P}(R \mid T)$

t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

$\Rightarrow$ Elimination

$\mathbb{P}(T)$

t	0.17
$\bar{t}$	0.83



$\mathbb{P}(R \mid T)$

t	r	0.3
t	$\bar{r}$	0.7
$\bar{t}$	r	0.1
$\bar{t}$	$\bar{r}$	0.9

$\Rightarrow$ Jointure

$\mathbb{P}(T,R)$

t	r	0.051
t	$\bar{r}$	0.119
$\bar{t}$	r	0.083
$\bar{t}$	$\bar{r}$	0.747

$\Rightarrow$ Elimination

$\mathbb{P}(R)$

r	0.134
$\bar{r}$	0.866

Observations

S'il y a des observations, on commence avec les factors qui selectionnent cette observation.
Si on veut calculer $\mathbb{P}(R \mid p)$, on commence par les factors

$\mathbb{P}(P)$

p	0.1
$\bar{p}$	0.9

$\mathbb{P}(T \mid P)$

p	t	0.8
p	$\bar{t}$	0.2
$\bar{p}$	t	0.1
$\bar{p}$	$\bar{t}$	0.9

$\mathbb{P}(r \mid T)$

t	r	0.3
$\bar{t}$	r	0.1

Observations – 2

On obtiendra une jointure entre une variable demandée et une observation.

Il suffira alors de normaliser !

$\mathbb{P}(p, R)$

p	r	0.026
p	$\bar{r}$	0.074

$\Longrightarrow$

$\mathbb{P}(R \mid p)$

r	0.26
$\bar{r}$	0.74

Algorithme Général

- Requête $\mathbb{P}(Q \mid e_1, \ldots, e_k)$
- On prend les TPC locales, et on selectionne les tranches en fonction des observations $e_1, \ldots, e_k$
- Tant qu'il reste des variables cachées (i.e. des variables qui ne sont ni des variables de Q ni des variables dans $\{e_1, \ldots, e_k\}$)
  - on choisit une variable cachée H
  - on joint tous les factors qui mentionnent H
  - on élimine H
- on joint tous les factors restant et on normalise.

Exemple

1 $\mathbb{P}(C \mid j,m) \propto \mathbb{P}(C,j,m)$

2 Initialisation des tables

$\mathbb{P}(C) \quad \mathbb{P}(T) \quad \mathbb{P}(A \mid C,T) \quad \mathbb{P}(j \mid A) \quad \mathbb{P}(m \mid A)$

3 On choisit comme variable cachée A

- jointure : $\mathbb{P}(A \mid C,T) \bowtie \mathbb{P}(j \mid A) \Rightarrow \mathbb{P}(j,A \mid C,T)$
- jointure : $\mathbb{P}(j,A \mid C,T) \bowtie \mathbb{P}(m \mid A) \Rightarrow \mathbb{P}(j,m,A \mid C,T)$
- élimination : $\mathbb{P}(j,m,A \mid C,T) \Rightarrow \mathbb{P}(j,m \mid C,T)$

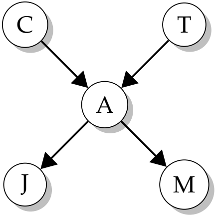
$\mathbb{P}(C) \quad \mathbb{P}(T) \quad \mathbb{P}(j,m \mid C,T)$

4 On choisit comme variable cachée restante T

- jointure : $\mathbb{P}(T) \bowtie \mathbb{P}(j,m \mid C,T) \Rightarrow \mathbb{P}(j,m,T \mid C)$
- élimination : $\mathbb{P}(j,m,E \mid C) \Rightarrow \mathbb{P}(j,m \mid C)$

$\mathbb{P}(C) \quad \mathbb{P}(j,m \mid C)$

- Il ne reste plus que C
  - jointure : $\mathbb{P}(C) \bowtie \mathbb{P}(j,m \mid C) \Rightarrow \mathbb{P}(j,m,C)$
  - normalisation $\mathbb{P}(C \mid j,m)$



Exemple : revisité

$\mathbb{P}(C \mid j,m) \propto \mathbb{P}(C,j,m)$

$\mathbb{P}(C) \quad \mathbb{P}(T) \quad \mathbb{P}(A \mid C,T) \quad \mathbb{P}(j \mid A) \quad \mathbb{P}(m \mid A)$

$$\begin{aligned}
\mathbb{P}(C \mid j,m) &\propto \mathbb{P}(C,j,m) \\
&= \sum_t \sum_a \mathbb{P}(C,e,a,j,m) \\
&= \sum_t \sum_a \mathbb{P}(C)\mathbb{P}(t)\mathbb{P}(a \mid c,t)\mathbb{P}(j \mid a)\mathbb{P}(m \mid a) \\
&= \sum_t \mathbb{P}(C)\mathbb{P}(t) \sum_a \mathbb{P}(a \mid c,t)\mathbb{P}(j \mid a)\mathbb{P}(m \mid a) \\
&= \sum_t \mathbb{P}(C)\mathbb{P}(t) f_1(C,t,j,m) \\
&= \mathbb{P}(C) \sum_t \mathbb{P}(t) f_1(C,t,j,m) \\
&= \mathbb{P}(C) f_2(C,j,m) \\
&= f_3(C,j,m) \\
&= \mathbb{P}(C \mid j,m)
\end{aligned}$$

Avant dernier exemple

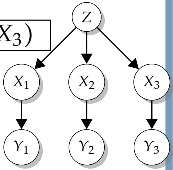
Requête : $\mathbb{P}(X_3 \mid y_1, y_2, y_3)$

1

$\mathbb{P}(Z)$	$\mathbb{P}(X_1 \mid Z)$	$\mathbb{P}(X_2 \mid Z)$	$\mathbb{P}(X_3 \mid Z)$	$\mathbb{P}(y_1 \mid X_1)$	$\mathbb{P}(y_2 \mid X_2)$	$\mathbb{P}(y_3 \mid X_3)$

2 élimination de X_1

$\mathbb{P}(Z)$	$f_1(y_1, Z)$	$\mathbb{P}(X_2 \mid Z)$	$\mathbb{P}(X_3 \mid Z)$	$\mathbb{P}(y_2 \mid X_2)$	$\mathbb{P}(y_3 \mid X_3)$



3 élimination de X_2

$\mathbb{P}(Z)$	$f_1(y_1, Z)$	$f_2(y_2, Z)$	$\mathbb{P}(X_3 \mid Z)$	$\mathbb{P}(y_3 \mid X_3)$

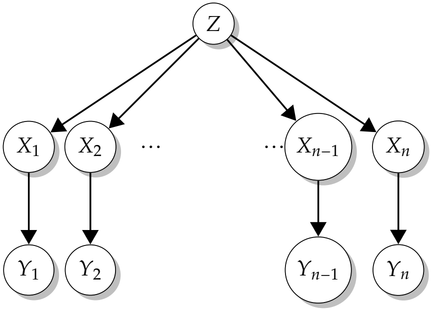
4 élimination de Z

$f_3(y_1, y_2, X_3)$	$\mathbb{P}(y_3 \mid X_3)$

5 Dernière jointure et normalisation $\boxed{f_4(y_1, y_2, y_3, X_3)}$

➩ $\mathbb{P}(X_3 \mid y_1, y_2, y_3)$

Dernier exemple et ordre des variables



Requête $\mathbb{P}(X_n \mid y_1, \ldots, y_n)$

- Avec ordre $Z, X_1, \ldots X_{n-1}$ comme précédemment
- Avec ordre $X_1, \ldots X_{n-1}$ puis Z

⇨ quelle est la taille maximale des factors ?

⇨ l'ordre a une **grande** influence sur l'efficacité !

L'inférence dans un réseau Bayésien est est fait NP difficile !
(réduction par 3-SAT)