Raisonnement Incertain

Boi Faltings

Laboratoire d'Intelligence Artificielle
boi.faltings@epfl.ch
http://moodle.epfl.ch/

Limites de la logique

Le monde ne suit pas exactement la logique:

- informations insuffisantes.
- imprécision de la modélisation.
- conclusions disjonctives non modélisable par des clauses de Horn.

$\Rightarrow$ le raisonnement est *incertain*.

Représentation de l'Incertitude

Principe:

associer à chaque proposition une représentation numérique de l'incertitude: plausibilité.

Raisonnement *Bayesien*:

- la plausibilité tient compte de l'ensemble des évidences.

⇒ mise à jour pendant le raisonnement:

plausibilité à-priori ⇒ *plausibilité* à-posteriori.

Desiderata

- Associer une plausibilité à chaque prémisse.
- Pour chaque déduction, calculer la plausibilité de la conclusion sur la base des plausibilités des prémisses et des règles.
- Nécessite que la plausibilité doit être indépendent se sa dérivation.

Formalismes pour l'incertitude

- Logique floue:
  facile à appliquer mais sans base théorique solide.
- Raisonnement probabiliste:
  très bien fondée, mais difficile à appliquer.
- Réseaux Bayesiens: facile à appliquer et théoriquement bien fondées, mais applicables uniquement dans une interprétation *causale*.

Exemple: système d'alarme

- 3 détecteurs: M(ouvement), E(ntrée), V(itre).
- Déclenchées soit par le propriétaire soit par un cambrioleur.
- On aimerait trouver une fonction d'alarme qui détecte les cambrioleurs, mais pas les propriétaires.

Modélisation avec des facteurs de certitude

$$A \Rightarrow B : CF(B) = CF(A) * CF(\text{Règle})$$

M = *"Détecteur de mouvement", CF = 1.0*
R1: $M \Rightarrow C$ = *"Cambrioleur présent", CF=0.1*
R2: $M \Rightarrow P$ = *"Proprietaire présent", CF=0.9*

$\Rightarrow$

C, *CF = 0.1*
P, *CF = 0.9*

Lacune 1: influence du context

- Suppons que le propriétaire est enregistré à un autre endroit.

$\Rightarrow$ $\neg P$ avec certitude.

$\Rightarrow$ seul un cambrioleur peut déclencher le détecteur:
$M \Rightarrow C$ avec CF $>> 0.1$ (par exemple, 0.99).

- Il faut modéliser l'interdépendance de P et C!

Lacune 2: chainages inadmissibles

- Supposons que nous savons qu'un cambrioleur est présent.

$\Rightarrow$ le détecteur de mouvement sera sûrement déclenché.

$\Rightarrow$ le propriétaire est également présent, par la règle $R2$!

- Problème: nous avons negligé *l'interdépendance*: les cas où le détecteur est declenché par le cambrioleur est justement pas celui où il est declenché par le propriétaire!

Logique probabiliste

Caractériser l'incertitude par la probabilité:

$p(A)$ = *probabilité que la proposition A est vraie*
$p(\neg A) = 1 - p(A)$ = *probabilité que A est fausse*

A devient une variable aléatoire avec valeurs $\{vrai, faux\}$:

- $P(A)$ = distribution $[p(A), p(\neg A)]$.
- Majuscule = distribution, minuscule = probabilité.
- $P(A, B) = [p(A, B), p(A, \neg B), p(\neg A, B), p(\neg A, \neg B)]$.
- $P(X_1, X_2, .., X_n) = [p(X_1, .., X_n), p(\neg X_1, .., X_n)...]$ (2^n valeurs).
- $p(X|Y) = p(X, Y)/p(Y)$ = probabilité que X est vrai étant donnée que Y est egalement vrai
- $p(X|Y) + p(\neg X|Y) = 1$ (mais $p(X|Y) + p(X|\neg Y)$ n'est pas toujours = 1)

Signification de la probabilité

- normalement, probabilité $\simeq$ fréquence.
- mais le raisonnement n'a lieu qu'une seule fois.
- Probabilité $\simeq$ lotterie:

$$p(A) = x \Leftrightarrow \begin{array}{l} \textit{dans un pari qui paie 1 Fr. si A est} \\ \textit{vrai, je suis prêt à mettre} \leq x \textit{ Fr.} \end{array}$$

Raisonnement Bayesien

Observation $\Rightarrow$ la probabilité attribuée à une proposition change:

Proposition	Prob. à priori	Observation	Prob. à postériori
C	$p_0(C) = 0.01$		

Raisonnement Bayesien

Observation $\Rightarrow$ la probabilité attribuée à une proposition change:

Proposition	Prob. à priori	Observation	Prob. à postériori
C	$p_0(C) = 0.01$	M	$p(C\|M) = 0.18$

Raisonnement Bayesien

Observation $\Rightarrow$ la probabilité attribuée à une proposition change:

Proposition	Prob. à priori	Observation	Prob. à postériori
C	$p_0(C) = 0.01$	M	$p(C\|M) = 0.18$
		$\neg M$	$p(C\|\neg M) = 0.0036$

Raisonnement Bayesien

Observation $\Rightarrow$ la probabilité attribuée à une proposition change:

Proposition	Prob. à priori	Observation	Prob. à postériori
C	$p_0(C) = 0.01$	M	$p(C\|M) = 0.18$
		$\neg M$	$p(C\|\neg M) = 0.0036$
P	$p_0(P) = 0.5$		

Raisonnement Bayesien

Observation $\Rightarrow$ la probabilité attribuée à une proposition change:

Proposition	Prob. à priori	Observation	Prob. à postériori
C	$p_0(C) = 0.01$	M	$p(C\|M) = 0.18$
		$\neg M$	$p(C\|\neg M) = 0.0036$
P	$p_0(P) = 0.5$	M	$p(P\|M) = 0.95$

Raisonnement Bayesien

Observation $\Rightarrow$ la probabilité attribuée à une proposition change:

Proposition	Prob. à priori	Observation	Prob. à postériori
C	$p_0(C) = 0.01$	M	$p(C\|M) = 0.18$
		$\neg M$	$p(C\|\neg M) = 0.0036$
P	$p_0(P) = 0.5$	M	$p(P\|M) = 0.95$
		$\neg M$	$p(P\|\neg M) = 0.15$

Raisonnement Bayesien

Observation $\Rightarrow$ la probabilité attribuée à une proposition change:

Proposition	Prob. à priori	Observation	Prob. à postériori
C	$p_0(C) = 0.01$	M	$p(C\|M) = 0.18$
		$\neg M$	$p(C\|\neg M) = 0.0036$
P	$p_0(P) = 0.5$	M	$p(P\|M) = 0.95$
		$\neg M$	$p(P\|\neg M) = 0.15$
	$p_0(P) = 0.01$	M	$p(P\|M) = 0.019$
		$\neg M$	$p(P\|\neg M) = 0.003$

Raisonnement Bayesien

Observation $\Rightarrow$ la probabilité attribuée à une proposition change:

Proposition	Prob. à priori	Observation	Prob. à postériori
C	$p_0(C) = 0.01$	M	$p(C\|M) = 0.18$
		$\neg M$	$p(C\|\neg M) = 0.0036$
P	$p_0(P) = 0.5$	M	$p(P\|M) = 0.95$
		$\neg M$	$p(P\|\neg M) = 0.15$
	$p_0(P) = 0.01$	M	$p(P\|M) = 0.019$
		$\neg M$	$p(P\|\neg M) = 0.003$

Règle de Bayes:

$$p(P|M) = p_0(P)p(M|P)/p(M)$$

$p(M|P)/p(M)$ caractérise l'incertitude de l'inférence.

Inférence probabiliste

Déduction:

$p(M)$, *régle* $M \Rightarrow C$

Probabilité à-posteriori $p(C)$ se calcule comme suit:

$$p(C) = p(C|M) \cdot p(M) + p(C|\neg M)(1 - p(M))$$

où $p(M)$ caractérise la certitude de la condition et

$$p(C|M) = p(C, M)/p(M) = p_0(C)p(M|C)/p(M)$$

ou $p(M|C)/p(M)$ caractérise l'incertitude de la règle.

Contre-factuels

Différence par rapport aux CF:
le calcul prend en compte $p(C|\neg M)$

Permet de modéliser la probabilité à-priori:

- elevée: $p(P|\neg M)$ est elevée, P reste probable même si M n'est pas vrai.
- faible: $p(C|\neg M)$ est petit, alors il faut que M soit assez certain pour rendre C une conclusion probable.

Chaînage des inférences

Exemple 1:
P(roprietaire) $\Rightarrow$ M(ouvement) $\Rightarrow$ A(larme)

Chaînage des inférences

Exemple 1:
P(roprietaire) $\Rightarrow$ M(ouvement) $\Rightarrow$ A(larme)

- $P(P) = [0.9, 0.1]$,
  $P(M|P) = [0.84, 0.16]$,
  $P(M|\neg P) = [0.06, 0.94]$
  $\Rightarrow p(M) = 0.9 \cdot 0.84 + 0.1 \cdot 0.06 = 0.762$

Chaînage des inférences

Exemple 1:
P(roprietaire) $\Rightarrow$ M(ouvement) $\Rightarrow$ A(larme)

- $P(P) = [0.9, 0.1]$,
  $P(M|P) = [0.84, 0.16]$,
  $P(M|\neg P) = [0.06, 0.94]$
  $\Rightarrow p(M) = 0.9 \cdot 0.84 + 0.1 \cdot 0.06 = 0.762$
- $P(A|M) = [1, 0]$,
  $P(A|\neg M) = [0.01, 0.99]$
  $\Rightarrow p(A) = 0.762 \cdot 1 + 0.238 \cdot 0.01 = 0.764$

Chaînage des inférences

Exemple 1:
P(roprietaire) $\Rightarrow$ M(ouvement) $\Rightarrow$ A(larme)

- $P(P) = [0.9, 0.1]$,
  $P(M|P) = [0.84, 0.16]$,
  $P(M|\neg P) = [0.06, 0.94]$
  $\Rightarrow p(M) = 0.9 \cdot 0.84 + 0.1 \cdot 0.06 = 0.762$
- $P(A|M) = [1, 0]$,
  $P(A|\neg M) = [0.01, 0.99]$
  $\Rightarrow p(A) = 0.762 \cdot 1 + 0.238 \cdot 0.01 = 0.764$
- inférence correcte.

Chaînage des inférences

Exemple 2:

P(roprietaire) $\Rightarrow$ M(ouvement) $\Rightarrow$ C(ambrioleur)

Chaînage des inférences

Exemple 2:
P(roprietaire) ⇒ M(ouvement) ⇒ C(ambrioleur)

- $P(P) = [0.9, 0.1]$,
  $P(M|P) = [0.84, 0.16]$,
  $P(M|\neg P) = [0.06, 0.94]$
  $\Rightarrow p(M) = 0.9 \cdot 0.84 + 0.1 \cdot 0.06 = 0.762$

Chaînage des inférences

Exemple 2:

P(roprietaire) $\Rightarrow$ M(ouvement) $\Rightarrow$ C(ambrioleur)

- $P(P) = [0.9, 0.1]$,
  $P(M|P) = [0.84, 0.16]$,
  $P(M|\neg P) = [0.06, 0.94]$
  $\Rightarrow p(M) = 0.9 \cdot 0.84 + 0.1 \cdot 0.06 = 0.762$
- $P(C|M) = [0.18, 0.82]$,
  $P(C|\neg M) = [0.0036, 0.99]$
  $\Rightarrow p(C) = 0.762 \cdot 0.18 + 0.238 \cdot 0.0036 \simeq 0.138$

Chaînage des inférences

Exemple 2:
P(roprietaire) $\Rightarrow$ M(ouvement) $\Rightarrow$ C(ambrioleur)

- $P(P) = [0.9, 0.1]$,
  $P(M|P) = [0.84, 0.16]$,
  $P(M|\neg P) = [0.06, 0.94]$
  $\Rightarrow p(M) = 0.9 \cdot 0.84 + 0.1 \cdot 0.06 = 0.762$
- $P(C|M) = [0.18, 0.82]$,
  $P(C|\neg M) = [0.0036, 0.99]$
  $\Rightarrow p(C) = 0.762 \cdot 0.18 + 0.238 \cdot 0.0036 \simeq 0.138$
- pas juste: la probabilité de C ne devrait pas augmenter (prob. à priori $= 0.01$).

Exemple 2 (suite)

$P(C, M|P) \neq p(C|M)P(M|P)$!

Calcul correct tient compte des dépendances entre P, M et C:

$$\begin{aligned} p(C) &= \underbrace{p(C|M,P)}_{=0}\underbrace{p(M,P)}_{=0.76} \\ &\quad + \underbrace{p(C|\neg M,P)}_{=0}\underbrace{p(\neg M,P)}_{=0.14} \\ &\quad + \underbrace{p(C|M,\neg P)}_{=0.99}\underbrace{p(M,\neg P)}_{=0.006} \\ &\quad + \underbrace{p(C|\neg M,\neg P)}_{=0.01}\underbrace{p(\neg M,\neg P)}_{=0.094} \\ &= 0.9 \cdot 0.006 + 0.01 \cdot 0.094 = 0.00634 \end{aligned}$$

Complexité

Pour être sur de faire le bon calcul, on devrait tenir compte des dépendances partout.

Exemple:

- 20 propositions (variables aléatoires) $V_1, .., V_{20}$
- calculer $p(X)$ demande la distribution $P(X|V_1, .., V_{20})$
  $P(X|V_1, .., V_{20}) = \{p(X|V_1, .., V_{20}), p(X|\neg V_1, .., V_{20}), ..\}$
- $\Rightarrow 2^{20} \approx 1'000'000$ valeurs!

Q: Comment peut-on réduire les dépendances?

R: En se limitant aux *causes* directes.

Causalité

- P(roprietaire) est une cause de M(ouvement), et
- M(ouvement) est une cause de A(larme);
- alors: $P \to M \to A$ permet la propagation locale des probabilités!

Causalité

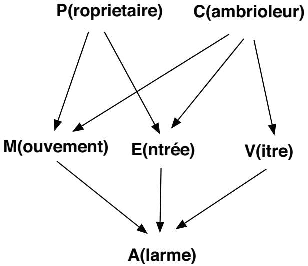
- P(roprietaire) est une cause de M(ouvement), et
- M(ouvement) est une cause de A(larme);
- alors: $P \rightarrow M \rightarrow A$ permet la propagation locale des probabilités!

mais:

- C(ambrioleur) est une cause de M(ouvement), et
- P(roprietaire) est une cause de M(ouvement);
- mais si C explique déjà M, P n'est pas plus probable
- donc: $C \rightarrow M \leftarrow P$ n'admet pas la propagation locale des probabilités!

Graphe des influences causales

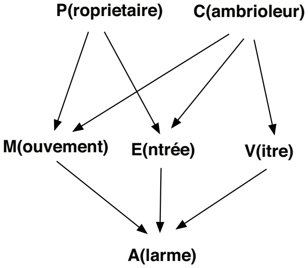
Formalisation comme graphe:



$x \to y$: x est une cause de y

Graphe des influences causales

Formalisation comme graphe:

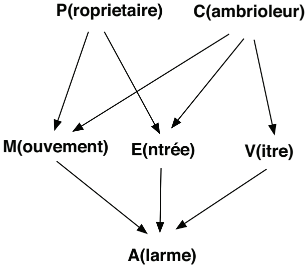


$x \rightarrow y$: x est une cause de y

Possible: $C \vdash E \vdash A$

Graphe des influences causales

Formalisation comme graphe:



$x \rightarrow y$: x est une cause de y

Possible: $C \vdash E \vdash A$

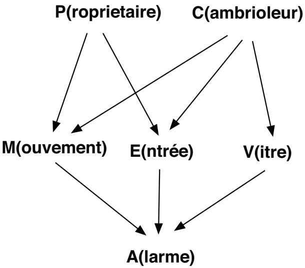
Impossible: $C \vdash M \vdash P$, $E \vdash A \vdash V$, ...

M, E, A sont *bloquants*: destination de plusieurs flèches.

Ordre Causale

La causalité implique un *ordre* des variables:

$$cause > effet \Leftrightarrow parent > descendant$$



noeud	parents	descendants
P	$\{\}$	$\{M, E\}$
C	$\{\}$	$\{M, E, V\}$
M	$\{P, C\}$	$\{A\}$
E	$\{P, C\}$	$\{A\}$
V	$\{C\}$	$\{A\}$
A	$\{M, E, V\}$	$\{\}$

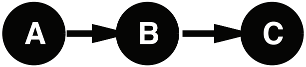
Prix de Turing

En 2012, le prix de Turing a été decerné à Judea Pearl (UCLA):

> *For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.*

que nous allons voir maintenant.

Indépendance conditionelle

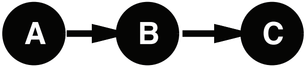


Indépendance de A et C:

$$p(C|A) = p(C|\neg A) = p(C)$$

n'est pas donnée à cause de l'influence à travers B.

Indépendance conditionelle



Indépendance de A et C:

$$p(C|A) = p(C|\neg A) = p(C)$$

n'est pas donnée à cause de l'influence à travers B.

Par contre, nous pouvons profiter de l'indépendance *conditionelle* de A et C étant donné B:

$$p(C|A, B) = p(C|\neg A, B) = p(C|B)$$

et de même pour $\neg B$.

Utilité pour l'inférence

Chaînage:

$$p(C|A) = p(C|A,B) \cdot p(B|A) + p(C|A,\neg B) \cdot p(\neg B|A)$$

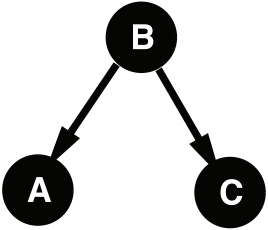
A, C conditionellement indépendants étant donné B:

$$\begin{aligned} p(C|A) &= p(C|B) \cdot p(B|A) + p(C|\neg B) \cdot (1 - p(B|A)) \\ &= \sum_{b=B,\neg B} p(C|b) p(b|A) \end{aligned}$$

$\Rightarrow$ il suffit de connaître $P(C|B)$ et $P(B|A)$ au lieu de $P(C|A,B)$.
On appele ce chaînage la *marginalisation* de B.

Descendants multiples

Les descendants multiples ne mettent pas en cause l'indépendance conditionelle:



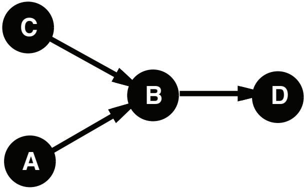
$P(A|B, C) = P(A|B)$, pas besoin de connaitre distribution jointe!

Causes multiples $\Rightarrow$ dépendance conditionelle

Inverse de l'indépendence: A et C sont indépendants, mais deviennent dépendants si B est connu:

$$\begin{aligned} p(A|C) &= p(A|\neg C)(= p(A)) \\ p(A|B,C) &\neq p(A|B,\neg C) \end{aligned}$$

Verifiée si A et C sont les causes d'un effet commun B:



Intuition: Si A explique déjà B, C devient moins probable.
$\Rightarrow$ il faut la distribution jointe: $p(B|A,C)$

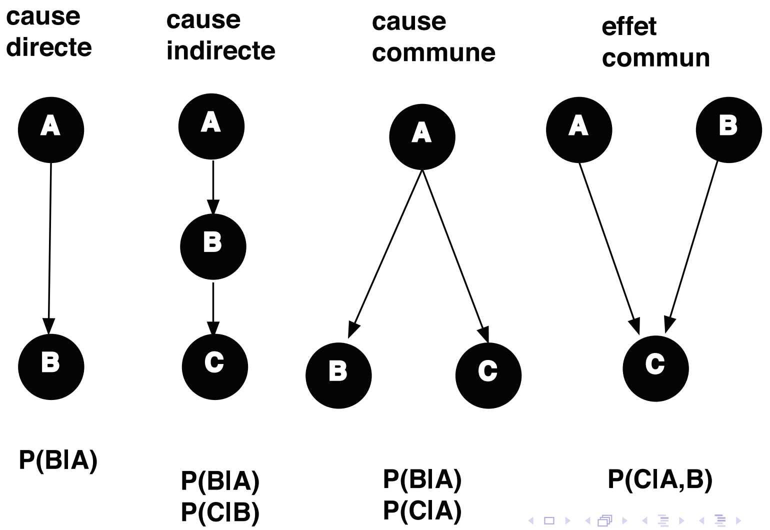
Exemple

- A = Entrée
  B = Alarme
  C = Vitre
- Normalement, aucune dépendence entre A et C.
- Si B, alors A **et** C deviennent plus probable.
- Si A est vrai, cela explique B et la probabilité de C redescend à son niveau antérieur.

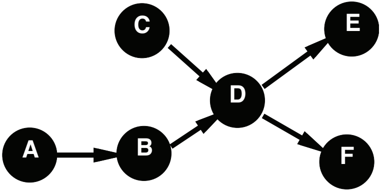
$\Rightarrow$ dépendence entre A et C.

$\Rightarrow$ besoin de connaître $p(B|A, C)$.

Resumé des Structures



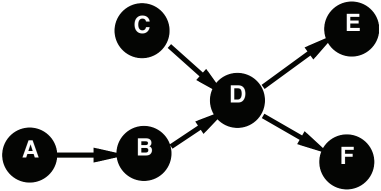
Question



Pour chaque arc $X \rightarrow Y$, on connait $p(Y|X)$, mais rien d'autre.
Quels calculs sont possibles (par chaînage):

1. $p(D|A)$
2. $p(A|C)$
3. $p(E|C)$

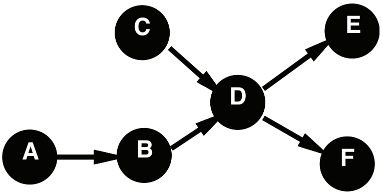
Question



Pour chaque arc $X \rightarrow Y$, on connait $p(Y|X)$, mais rien d'autre.
Quels calculs sont possibles (par chaînage):

1. $p(D|A)$ oui, car $A \rightarrow B \rightarrow D$ est une chaîne causale.
2. $p(A|C)$
3. $p(E|C)$

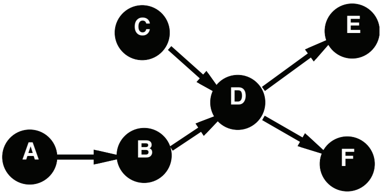
Question



Pour chaque arc $X \rightarrow Y$, on connait $p(Y|X)$, mais rien d'autre.
Quels calculs sont possibles (par chaînage):

1. $p(D|A)$ oui, car $A \rightarrow B \rightarrow D$ est une chaîne causale.
2. $p(A|C)$ non, car D est un noeud bloquant.
3. $p(E|C)$

Question



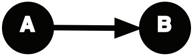
Pour chaque arc $X \rightarrow Y$, on connait $p(Y|X)$, mais rien d'autre.
Quels calculs sont possibles (par chaînage):

1. $p(D|A)$ oui, car $A \rightarrow B \rightarrow D$ est une chaîne causale.
2. $p(A|C)$ non, car D est un noeud bloquant.
3. $p(E|C)$ oui, car $C \rightarrow D \rightarrow E$ est une chaîne causale.

Inférence abductive

- Inférence causes $\vdash$ effets = *déduction*.
- Inférence des causes sur la base des effets = *abduction*.
- L'abduction est très courant: diagnostic, vision, apprentissage, etc.
- Nécessite l'inférence dans le sens invers de la causalité.

Abduction Bayesienne



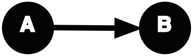
Par la règle de Bayes:

$$p(A|B) = \frac{p(A,B)}{p(B)} = \frac{p(B|A)p(A)}{p(B)} = \alpha p(B|A)p(A)$$

l'inférence dans le sens invers des arcs utilise:

- $p(B|A)$: vraissemblance ("likelihood") de B
- $p(A)$: probabilité à priori.
- $p(B)$: probabilité de l'observation B; difficile à connaitre et souvent exprimé comme facteur α

Déterminer p(B)



Par la règle de Bayes:

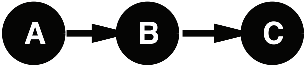
$$p(\neg A|B) = \frac{p(\neg A, B)}{p(B)} = \frac{p(B|\neg A)p(\neg A)}{p(B)}$$

Comme $p(A|B) + p(\neg A|B) = 1$:

$$p(B) = p(B|A)p(A) + p(B|\neg A)p(\neg A)$$

est obtenu par normalisation, et ainsi $\alpha = 1/p(B)$.

Chaînage de l'Abduction



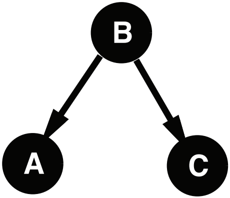
$$\begin{aligned} p(A|C) &= \sum_{b=B,\neg B} p(A|b) \cdot p(b|C) \\ &= \sum_{b=B,\neg B} \frac{p(b|A)p(A)}{p(b)} \cdot \frac{p(C|b)p(b)}{p(C)} \\ &= \frac{p(A)}{p(C)} \sum_{b=B,\neg B} p(b|A) \cdot p(C|b) \end{aligned}$$

$\Rightarrow$ pas besoin de connaître $p(B)$!

$\Rightarrow$ déterminer $p(C)$ par normalisation.

Abduction avec plusieurs conséquences

Problème: il peut y avoir plusieurs effets d'une cause B, donc plusieurs chaînes qui partent de B.



$$\begin{aligned} p(B|A,C) &= \frac{1}{p(A,C)} p(A,B,C) \\ &= \alpha p(A|B,C) \cdot p(B,C) \\ &= \alpha p(A|B) \cdot p(C|B) \cdot p(B) \end{aligned}$$

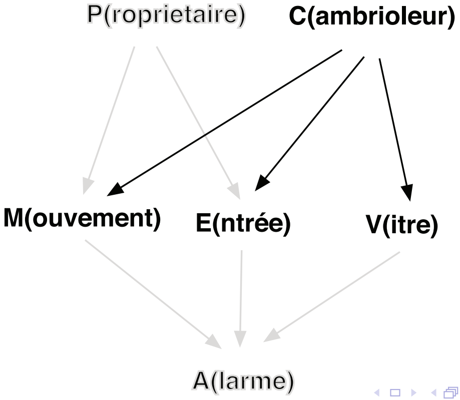
(car A et C indépendants étant donné B)

En général, pour k conséquences $Y_1, .., Y_k$:

$$p(B|Y_1,..,Y_k) = \alpha p(B) \prod_{i=1}^{k} p(Y_i|B)$$

Exemple

Déclencher l'alarme si $p(C|obs) > p(P|obs)$.
Exemple: comparer $p(C|M, \neg E, V)$ et $p(P|M, \neg E, V)$.
Formalisation comme graphe (pour $p(C|M, \neg E, V)$):



Exemple

$$p(C|M,\neg E,V) = \alpha \underbrace{p(C)}_{=0.01} \underbrace{p(M|C)}_{=0.8}(1-\underbrace{p(E|C)}_{=0.5}) \underbrace{p(V|C)}_{=0.5} = 2\cdot 10^{-3}\alpha$$

et

$$p(P|M,\neg E,V) = \alpha \underbrace{p(P)}_{=0.5} \underbrace{p(M|P)}_{=0.9}(1-\underbrace{p(E|P)}_{=0.99}) \underbrace{p(V|P)}_{=0.001} = 4.5\cdot 10^{-6}\alpha$$

avec $\alpha = 1/p(M,\neg E,V)$
$p(C|M,\neg E,V) >> p(P|M,\neg E,V)$: sonner l'alarme!
Pas besoin de calculer α
$(\alpha = 473,\ p(C|M,\neg E,V) = 0.946,\ p(P|M,\neg E,V) = 0.000213)$

Système d'alarme Bayesien

Déclencher l'alarme uniquement quand:

1. il y a une indication (M, E ou V), et
2. C est plus probable que P comme la cause.

Pour vérifier la deuxième condition: comparer $p(C|\cdot)$ à $p(P|\cdot)$: pas besoin de calculer α.

"Naive Bayes"

Souvent, on veut détecter une condition précise sur la base des plusieurs conséquences

Exemple: détecter si un mail est spam à partir de mots contenus dans un message

$\Rightarrow$ pour chaque mot $m \in M$, mésurer $p(m|spam)$ sur un échantillon de messages spam

Trier les messages par

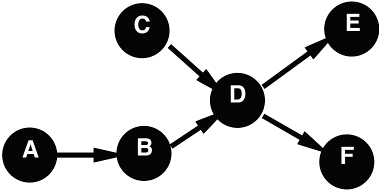
$$\alpha p(spam) \prod_{m \in M} p(m|spam)$$

$\alpha, p(spam)$ sont toujours identiques $\Rightarrow$

$$\prod_{m \in M} p(m|spam)$$

est suffisant comme critère.

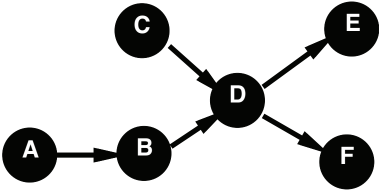
Question



Pour chaque arc $X \rightarrow Y$, on connait $p(Y|X)$, et aussi $p(X)$. Quels calculs sont possibles:

1. $p(A|D)$
2. $p(D|E, F)$
3. $p(B|C, E)$

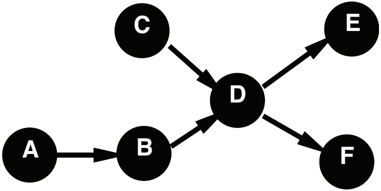
Question



Pour chaque arc $X \rightarrow Y$, on connait $p(Y|X)$, et aussi $p(X)$. Quels calculs sont possibles:

1. $p(A|D)$ oui, car $A \rightarrow B \rightarrow D$ est une chaîne causale.
2. $p(D|E, F)$
3. $p(B|C, E)$

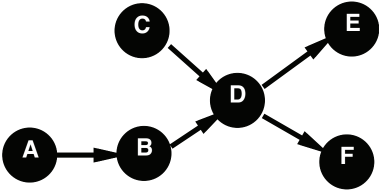
Question



Pour chaque arc $X \rightarrow Y$, on connait $p(Y|X)$, et aussi $p(X)$. Quels calculs sont possibles:

1. $p(A|D)$ oui, car $A \rightarrow B \rightarrow D$ est une chaîne causale.
2. $p(D|E, F)$ oui, car E et F sont effets indépendants de D.
3. $p(B|C, E)$

Question

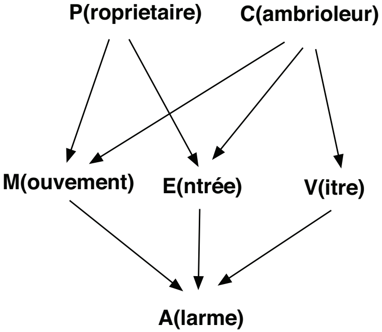


Pour chaque arc $X \rightarrow Y$, on connait $p(Y|X)$, et aussi $p(X)$. Quels calculs sont possibles:

1. $p(A|D)$ oui, car $A \rightarrow B \rightarrow D$ est une chaîne causale.
2. $p(D|E, F)$ oui, car E et F sont effets indépendants de D.
3. $p(B|C, E)$ non, car D est un noeud bloquant.

Vérifier la performance

Système d'alarme Bayesien: chaque capteur a une influence causale sur l'alarme.



Comment calculer la probabilité de détection $p(A|C)$ d'une alarme correct (et $p(A|P)$ d'une fausse alarme)?

Inférence avec chemins multiples

- 3 chemins: $C \to M \to A$, $C \to E \to A$, $C \to V \to A$.
- d-séparateur de x et y: ensemble de noeuds S qui coupent toutes les chemins d'inférence entre x et y.
- d-séparateur rend x et y conditionellement indépendants étant donnée S:

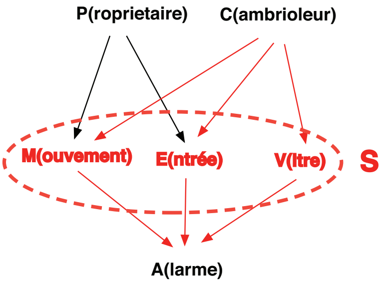
$$Pr(y|S) = Pr(y|S,x)$$

et donc

$$p(y,S|x) = p(y|S,x)p(S|x) = p(y|S)p(S|x)$$

- calculer $p(y|x)$ de $p(y,S|x)$ en marginalisant sur toutes les combinaisons de valeurs de l'ensemble S.

Obtenir $P(A|C)$



- $S = \{M, E, V\}$ est d-séparateur entre $x = C$ et $y = A$.
- calculer $p(A|M, E, V)$ et $p(M, E, V|C)$
  - $P(A|M, E, V)$ doit être connue: complexe!
  - $p(M, E, V|C) = p(M|C)p(E|C)p(V|C)$ peut être derivée.

Complexité de l'inférence

En cas de chemins multiples:

- l'inférence doit considérer toutes les combinaisons de causes dans un d-separateur.

$\Rightarrow$ complexité exponentielle dans sa taille.

- donné par la *largeur* de l'ordre causal (proprieté du graphe).
- complexité exponentielle non seulement en temps, mais aussi en mémoire!

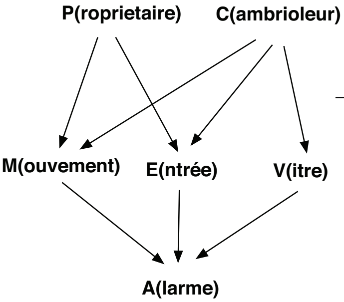
$\Rightarrow$ inférence exacte souvent pas faisable sauf si le graphe est proche d'un arbre.

Résolution stochastique

Echantillonage: trouver les probabilités par simulation.
Pour estimer $p(A|C)$:

1. ordonner les noeuds par ordre causale.
2. générer des instances d'états du réseau en commençant par les noeuds sans parents, et suivant les chaînes causales:
    - générer une valeur pour chaque descendant X suivant la distribution de probabilité dès que toutes les parents ont obtenus une valeur.
    - selon la distribution $P(X|parents)$ du réseau.
    - itérer jusqu'à ce que toutes les variables ont une valeur.
3. enregistrer la fréquence de paires de valeurs pour C et A jusquà ce qu'un nombre suffisant de valeurs ont été obtenus.
4. estimer $p(A, C)$ et donc $p(A|C) = p(A, C)/p(C)$ par la fréquence observée.

Exemple



noeud	parents	descendants
P	$\{\}$	$\{M, E\}$
C	$\{\}$	$\{M, E, V\}$
M	$\{P, C\}$	$\{A\}$
E	$\{P, C\}$	$\{A\}$
V	$\{C\}$	$\{A\}$
A	$\{M, E, V\}$	$\{\}$

- Générer d'abord P et C selon leur distribution à priori.
- Ensuite M, E et V suivant les probabilités conditionelles.
- Ensuite A par simulation du raisonnement Bayesien.

Simulation

	P	C	(M,E,V)	A
1	1	0	(1,0,0)	1
2	0	0	(0,0,0)	0
...				
50	0	1	(1,0,1)	1
...				

Estimations des probabilités par la fréquence, par exemple:

$$\Rightarrow p(A|C) = \frac{count(A,C)}{count(C)} = \frac{4}{5} = 0.8$$

Faiblesses de l'échantillonage

- Les états avec faible probabilité (ex: présence du cambrioleur) ont une fréquence très faible.
- Si on s'interesse à $P(A|C)$, alors simuler que des situations ou C est vrai.
- Sous-pondérer les échantillons à haute probabilité, et corriger les fréquences à la fin.
- Utiliser Gibbs sampling: le prochain échantillon est obtenue par modification du précedent (Markov Chain Monte Carlo).

Exemple

	P	C	(M,E,V)	A
1	1	1	(1,0,0)	0
2	0	1	(1,0,0)	1
3	0	1	(1,0,1)	1
4	0	1	(1,0,0)	0
5	1	1	(1,1,1)	1
...				
50	0	1	(1,0,1)	1
...				

Estimations des probabilités par la fréquence, par exemple:

$$\Rightarrow p(A|C) = \frac{count(A, C)}{count(C)} = \frac{42}{50} = 0.84$$

Echantillonage selon Gibbs

- Initialiser les variables de facon arbitraire.
- Fixer les variables connus, pour tous les autres variables x_i itérer:
  1. selectionner une variable x_i.
  2. attribuer une nouvelle valeur selon la distribution $P(x_i|parents, descendants)$
  3. ajouter l'échantillon aux statistiques.

$$P(x_i|parents, descendants) = P(x_i|parents) \prod_{Y_i \in descendants(X_i)} P(y_i|parents(Y_i))$$

Parfois en 2 phases: stabilisation des distributions, et échantillonage par la suite.

Utilisation pour l'inférence

- Application pour l'inférence directe peut être trop inefficace.
- Peut aussi s'utiliser pour "compiler" certains rélations à partir du réseau causal.
- Example pour un diagnostic médical: extraire la distribution conditionelle $p(\text{maladie}|\text{symptômes visibles})$ sans passer par toute la chaîne causale.
- En général: élimination de variables cachées et de cycles pour permettre l'inférence efficace.

Les réseaux Bayesiens en pratique

Utilisés dans des systèmes experts pour:

- diagnostic médical
- diagnostic de réacteurs d'avion
- preuve que fumer $\Rightarrow$ cancer

Fondamental pour des voitures autonomes, filtres spam, etc.

Application: Diagnostic medical par Réseaux Bayesiens

- `Promedas`: système général pour le diagnostic médical.
- Utiilisé comme aide aux medecins dans des hopitaux aux Pays-Bas.
- Application pour un domaine très général.
- Améliore souvent le diagnostic du medecin.

Résumé

- Facteurs de certitude/loqique floue: simple à implémenter, mais ne tiennent pas compte des dépendances.
- Raisonnement probabiliste: résultats corrects, mais problèmes de dépendences.
- Graphes causales pour l'identification des dépendances.
- Limitation: graphes sans chemins multiples!
- Méthode générale: simulation et échantillonage.