Raisonnement probabiliste

1

Plan

- Réseaux bayésiens
- Inférence dans les réseaux bayésiens
  - Inférence exacte
  - Inférence approximative

2

Réseaux bayésiens

- Au chapitre d'intro sur les pbs, on a vu que
  - Les distributions de probabilités jointes complètes pouvaient répondre à toutes les questions, mais qu'elles pouvaient être fort coûteuses en computation.
  - L'indépendance et l'indépendance conditionnelle permettaient de réduire le nombre de probabilités à spécifier pour définir une distribution de probabilités jointes complète.
- Dans ce chapitre, on va voir les réseaux bayésiens qui permettent de
  - représenter les dépendances entre les variables
  - donner une spécification concise des distributions de probabilités jointes complètes.

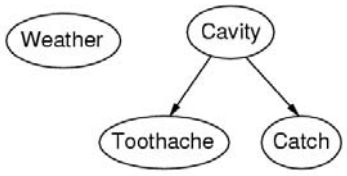
3

Syntaxe des réseaux bayésiens

- Un réseau bayésien est un graphe orienté acyclique où chaque nœud est annoté d'une table de probabilités conditionnelles.
- Plus spécifiquement, il contient:
  - Un ensemble de variables aléatoires.
  - Un ensemble de liens orientés connectant deux nœuds. S'il y a un lien entre le nœud *X* et *Y*, on dit que *X* est le *parent* de *Y*.
  - Chaque nœud a une distribution de probabilité conditionnelle $\boldsymbol{P(X_i|Parents(X_i))}$ qui quantifie l'effet des parents sur le nœud.

4

Exemple du dentiste



- *Weather* est indépendante des autres variables
- *Toothache* et *Catch* sont conditionnellement indépendantes sachant *Cavity*.
  - Il n'y a aucun lien direct entre *Toothache* et *Catch*.

5

Judea Pearl—Pionnier des RB

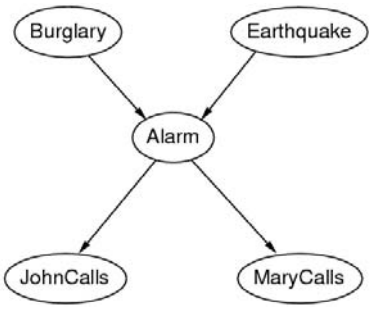
- **Judea Pearl named 2011 winner of Turing Award**
- UCLA professor cited for pioneering work in extending our understanding of artificial intelligence

6

Exemple de l'alarme

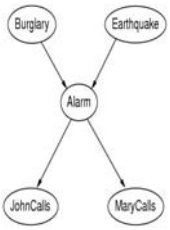
- Vous avez une nouvelle alarme à la maison qui
  - sonne lorsqu'il y a un cambriolage;
  - sonne parfois lorsqu'il y a un tremblement de terre.
- Vous avez deux voisins qui vous appellent au bureau s'ils entendent l'alarme.
  - John appelle tout le temps quand il entend l'alarme, mais parfois il confond le téléphone avec l'alarme.
  - Mary aime écouter de la musique forte et parfois elle n'entend pas l'alarme.
- Sachant qui a appelé, quelle est la probabilité qu'il y ait un cambriolage ?

Exemple de l'alarme

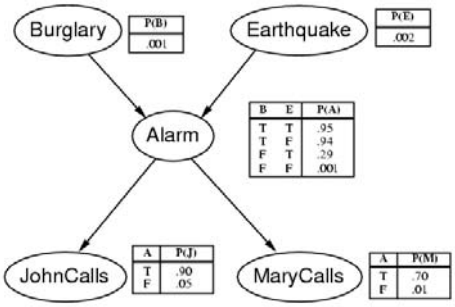


Exemple de l'alarme

- La topologie du réseau reflète un ensemble de relations d'indépendances conditionnelles
  - *Bulgary* et *Earthquake* affectent directement la probabilité de déclenchement d'une alarme
  - Le fait que John ou Mary appelle ne dépend que de l'alarme. John et Mary perçoivent pas directement le cambriolage ou les tremblements de terre mineurs.



Exemple de l'alarme



Spécification concise

- Une table de probabilité conditionnelle (TPC) pour une variable booléenne X_i, avec k parents booléens, a 2^k rangées.
- Chaque rangée a un nombre p *(reflétant une pb)* pour X_i = *True*
  - Le nombre pour X_i = *False* est $1 - p$
- Si le nombre maximal de parents est k, alors le réseau demande $O(nk)$ nombres.
  - Linéaire en n, au lieu de $O(2^n)$ pour la table conjointe complète
- Pour l'exemple (alarme), ça donne 10 nombres (soit 5 variables fois 2) au lieu de $2^5 = 32$ pour la table complète.

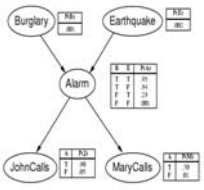
Sémantique des réseaux bayésiens

- Sémantique globale: définit la distribution jointe complète de probabilités comme le produit des distributions conditionnelles locales:

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

Exemple:

$$P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$$
$$= P(j|a)P(m|a)P(a|\neg b \wedge \neg e)P(\neg b)P(\neg e)$$
$$= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.0006$$



Construire des réseaux bayésiens

- Il faut une méthode garantissant qu'une série d'indépendances conditionnelles vérifiées localement induise la sémantique globale requise

Choisir un ordre sur les variables $X_1, \ldots, X_n$
Pour i = 1 à n **Faire**
- Ajouter X_i au réseau
- Sélectionner ses parents dans $X_1, \ldots, X_{i-1}$ tels que $\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})$

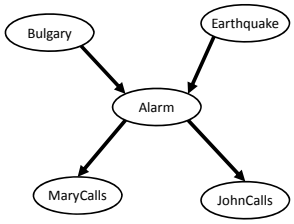
Fin Pour

Il est préférable d'avoir un modèle causal, c'est-à-dire qu'il est mieux d'ajouter la cause « racine » en premier et ensuite les variables qui sont influencées par la cause.

13

Exemple

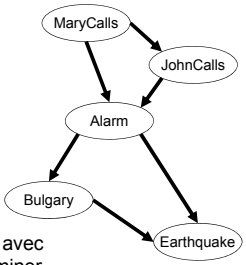
- Supposons que l'on choisit l'ordre B, E, A, M, J



14

Exemple

- Supposons que l'on choisit le mauvais ordre M, J, A, B, E
  - P(J|M) = P(J)? Non
  - P(A|J,M) = P(A|J)? Non
  - P(A|J,M) = P(A|M)? Non
  - P(B|A,J,M) = P(B|A) ? Oui
  - P(B|A,J,M) = P(B) ? Non
  - P(E|B,A,J,M) = P(E|A) ? Non
  - P(E|B,A,J,M) = P(E|A,B) ? Oui



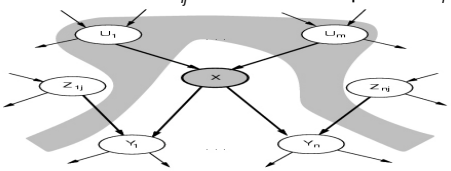
On obtient un réseaux plus complexe avec des probabilités plus difficiles à déterminer.

15

Sémantique des réseaux bayésiens

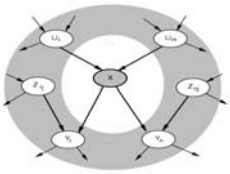
- Sémantique locale: chaque nœud est conditionnellement indépendant de ses non-descendants étant donné ses parents.

X est conditionnellement indépendant de ses non-descendants Z_{ij} étant donné ses parents U_i



Sémantique des réseaux bayésiens (2)

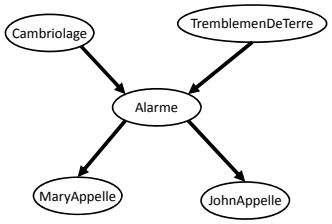
- Chaque nœud est indépendant des autres sachant son « Markov Blanket » (Couverture de Markov—en gris): parent + enfants + parents des enfants.
- Si MB(*X*) = ·*A*, alors *P(A| ·A,B)* = *P(A| ·A)*



- *Autrement dit: la seule connaissance pour prédire le nœud A c'est MB(A)*

Sémantique des réseaux bayésiens (3)

Cambriolage est indépendant de *JeanAppelle* et de *MarieAppelle*, étant *Alarme* et *TemblementDeTerre*



Représentation efficace des distributions

- Les tables de distributions conditionnelles grandissent de manière exponentielle selon le nombre de parents.
  - Ceci est le pire cas lorsque les relations entre les nœuds parents et enfants sont arbitraires.
- Habituellement, la relation peut être décrite par une distribution canonique qui correspond à un certain patron.
  - La table complète peut être définie en nommant le patron et peut-être certains paramètres.

Distribution canonique

- Nœuds déterministes
  - Les valeurs d'un nœud déterministe sont définies exactement par les valeurs de ses parents.
    - $X = f(\text{Parent}(X))$ pour une certaine fonction f.
    - Exemple: Fonction booléenne, le fils = disjonction des parents

$$NordAméricain \Leftrightarrow Canadien \vee Américain \vee Mexicain$$

  - Relation numérique entre des variables continues

$$\frac{\partial Niveau}{\partial t} = fluxEntrant + précipitations - fluxSortant - Évaporation$$

Distribution canonique

- « Noisy-OR » (ou bruité)
  - Utilisé pour décrire les relations incertaines
  - La relation causale entre parent et fils peut être inhibée.
    - Ex: un patient peut avoir la grippe sans avoir de la fièvre.
  - Deux suppositions:
    - Toutes les causes possibles sont listées. (Il peut y avoir un nœud *Autres*).
    - L'inhibition d'un parent est indépendante de l'inhibition des autres parents.

Exemple « Noisy-OR »

- La grippe, le rhume et la malaria causent de la fièvre.
- Avec une relation « Noisy-OR », on peut définir toute la table en spécifiant seulement les trois probabilités d'inhibition suivantes :

$$P(\neg fever|cold, \neg flu, \neg malaria) = 0.6$$
$$P(\neg fever|\neg cold, flu, \neg malaria) = 0.2$$
$$P(\neg fever|\neg cold, \neg flu, malaria) = 0.1$$

22

Exemple « Noisy-OR »

- Toutes les causes sont supposées listées
- L'inhibition de chaque parent est indépendante de celle de tous les autres parents. Par ex: ce qui empêche Malaria de provoquer de la fièvre est indépendant de ce qui empêche grippe de provoquer la fièvre.
- Donc: Fièvre est faux si tous ses parents vrais sont inhibés.

23

Exemple « Noisy-OR »

$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	0.6
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Ex:

24

Exemple « Noisy-OR »

- Ex : $P(\neg fever \mid \neg cold, flu, malaria) = P(\neg fever \mid \neg cold, \neg flu, malaria) P(\neg fever \mid \neg cold, flu, \neg malaria)$
  $= 0.2 \times 0.1 = 0.02$
- Le nombre de probabilités à définir est linéaire ($O(k)$ au lieu de $O(2^k)$ si k parents) selon le nombre de parents au lieu d'être exponentiel.

25

Variables continues

- Plusieurs problèmes du monde réel contiennent des quantités continues: hauteur, poids, température, argent, etc.
- Avec des variables continues, on ne peut pas définir des probabilités conditionnelles pour chacune des valeurs possibles.
- Deux solutions
  - Utiliser la discrétisation (perte de précision et très grande table)
  - Définir des densités de probabilités avec un nombre fini de paramètres.

26

Exemple

- Un consommateur achète des fruits dépendamment du coût (*Cost*), qui lui dépend de la taille de la cueillette (*Harvest*) et s'il y a eu une subvention du gouvernement (*Subsidy*).

- Deux cas:
  - Variable continue avec des parents continus et discrets
    - Ex: *Cost*
  - Variable discrète avec des parents continus
    - Ex: *Buys*

27

Variable fils continue

- Pour la variable *Cost*, il faut spécifier **P**(*Cost*|*Harvest*,*Subsidy*).
- Pour le parent discret (*Subsidy*), on a qu'à énumérer les valeurs possibles:
  - **P**(*Cost*|*Harvest*,*subsidy*) et **P**(*Cost*|*Harvest*,¬*subsidy*)
- Pour la variable continue (*Harvest)*, on spécifie une fonction de distribution pour la variable *Cost* en fonction de la variable *Harvest*.

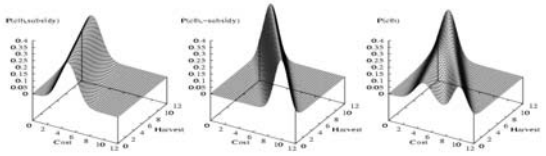
28

Variable fils continue

- La plus utilisée est la fonction linéaire gaussienne.

$$P(c|h, subsidy) = N(a_t h + b_t, \sigma_t^2)(c) = \frac{1}{\sigma_t\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c-(a_t h + b_t)}{\sigma_t}\right)^2\right)$$

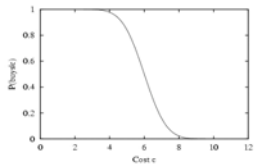
- La moyenne de *Cost* varie linéairement avec *Harvest*, la variance est fixe.



29

Variable fils discrète

- Pour la probabilité de *Buys* sachant *Cost*, on peut utiliser une fonction « probit ».



- Ou à la distribution « logit » qui utilise la fonction de sigmoïd.

30

Sommaire

- Les réseaux bayésiens sont une manière naturelle de représenter les dépendances causales.
- C'est une représentation compact des distributions jointes.
- Généralement facile à construire
- Les distributions canoniques sont une manière compacte de représenter les tables de probabilités conditionnelles.
- Pour les variables continues, on peut utiliser des fonctions de distribution.

Inférence exacte dans les réseaux bayésiens

- On vise maintenant à calculer la distribution de probabilité a posteriori d'un ensemble de variables de requêtes, étant donnée un événement observé, c'est-à-dire certaines assignations de valeurs à des variables d'évidence.
  - X : variable de question/requête
  - **E**: l'ensemble des variables d'évidence
  - e: un événement particulier
  - **Y**: l'ensemble des variables cachées
- L'ensemble complet des variables est:

$$\mathbf{X} = \{X\} \cup \mathbf{E} \cup \mathbf{Y}$$

Inférence exacte dans les réseaux bayésiens

- Une question/requête typique: $\mathbf{P}(\boldsymbol{X}|\boldsymbol{e})$
- Dans l'exemple du cambriolage, on pourrait observer l'événement: $JohnCalls = true$ et $MaryCalls = true$. Par la suite, on pourrait se demander s'il y a eu un cambriolage.

$$\mathbf{P}(Burglary|JohnCalls = true, MaryCalls = true) = \langle 0.284, 0.716 \rangle$$

Rappel du chapitre 13: Inférence utilisant des distributions conjointes complètes

Marginalisation ou sommation partielle : On peut écrire la règle de marginalisation générale suivante pour tout ensemble de variables Y et Z.

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y}, \mathbf{z})$$

La règle de conditionnment nous permet aussi d'écrire:

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y}|\mathbf{z})P(\mathbf{z})$$

Finalement, une requête du type $\mathbf{P}(X|\mathbf{e})$ peut être évaluée comme suit:

$$\mathbf{P}(X|\mathbf{e}) = \alpha\mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

34

Inférence par énumération

- Comme les réseaux bayésiens donnent la représentation complète de la table de distribution jointe, alors on peut utiliser la formule suivante, vue au chapitre 13.

$$\mathbf{P}(X|\mathbf{e}) = \alpha\mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

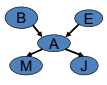
- Si on reprend l'exemple précédent où les variables cachées sont *Earthquake* et *Alarm*.

$$\mathbf{P}(B|j, m) = \alpha\mathbf{P}(B, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(B, e, a, j, m)$$

35

Inférence par énumération

- On peut réécrire la formule en utilisant les entrées des tables de probabilités conditionnelles du réseau bayésien. Pour *Burglary* = *true*, on obtient:



$$P(b|j, m) = \alpha \sum_{e} \sum_{a} P(b)P(e)P(a|b, e)P(j|a)P(m|a)$$

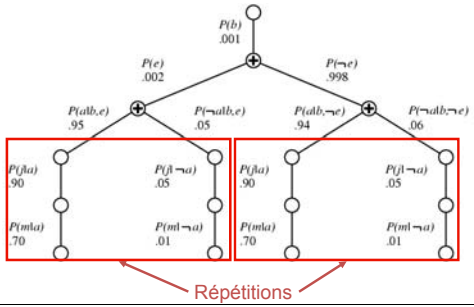
- En simplifiant, on obtient:

$$P(b|j, m) = \alpha\, P(b) \sum_{e} P(e) \sum_{a} P(a|b, e)P(j|a)P(m|a)$$

36

Inférence par énumération

- Arbre de calcul:



Inférence par énumération

- En effectuant les calculs, on obtient:

$$P(b|j,m) = \alpha \times 0.00059224$$

- Si on fait la même chose pour *Burglary* = *false et qu'on fait la somme*, on obtient:

$$\mathbf{P}(B|j,m) = \alpha \langle 0.00059224, 0.0014919 \rangle \approx \langle 0.284, 0.716 \rangle$$

- Même si les deux appellent, il n'y a que 28% des chances qu'il y est eu un cambriolage.
- La complexité en temps de l'inférence par énumération est de $O(2^n)$.

Inférence par élimination de variables

- Améliore l'algorithme par énumération en évitant les calculs répétés.
- La somme est effectuée de la droite vers la gauche.
- Exemple cambriolage:

$$\mathbf{P}(B|j,m) = \alpha \underbrace{\mathbf{P}(B)}_{B} \sum_{e} \underbrace{P(e)}_{E} \sum_{a} \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

Facteurs

Exemple

$$\mathbf{P}(B|j,m) = \alpha \underbrace{\mathbf{P}(B)}_{B} \sum_e \underbrace{P(e)}_{E} \sum_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

- Pour le facteur *M*, on enregistre les probabilités, étant donné chaque valeur de a, dans un vecteur à deux éléments.

$$\mathbf{f}_M(A) = \begin{pmatrix} P(m|a) \\ P(m|\neg a) \end{pmatrix}$$

- On fait la même chose pour *J*.
- Pour le facteur *A*, on obtient une matrice de 2 x 2 x 2, $f_A(A,B,E)$

40

Exemple

$$\mathbf{P}(B|j,m) = \alpha \underbrace{\mathbf{P}(B)}_{B} \sum_e \underbrace{P(e)}_{E} \sum_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

- Il faut maintenant faire la somme du produit des trois facteurs. La barre sur le *A*, indique que l'on a fait la somme pour *A*.

$$\begin{aligned} \mathbf{f}_{\bar{A}JM}(B,E) &= \sum_a \mathbf{f}_A(a,B,E) \times \mathbf{f}_J(a) \times \mathbf{f}_M(a) \\ &= \mathbf{f}_A(a,B,E) \times \mathbf{f}_J(a) \times \mathbf{f}_M(a) \\ &+ \mathbf{f}_A(\neg a,B,E) \times \mathbf{f}_J(\neg a) \times \mathbf{f}_M(\neg a) \end{aligned}$$

- La multiplication utilisée est: « pointwize product » (produit point par point).

41

Exemple

$$\mathbf{P}(B|j,m) = \alpha \underbrace{\mathbf{P}(B)}_{B} \sum_e \underbrace{P(e)}_{E} \sum_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(j|a)}_{J} \underbrace{P(m|a)}_{M}$$

- Le facteur et la sommation sur *E* sont calculés de la même manière.

$$\begin{aligned} \mathbf{f}_{\bar{E}\bar{A}JM}(B) &= \mathbf{f}_E(e) \times \mathbf{f}_{\bar{A}JM}(B,e) \\ &+ \mathbf{f}_E(\neg e) \times \mathbf{f}_{\bar{A}JM}(B,\neg e) \end{aligned}$$

- Finalement, on obtient:

$$\mathbf{P}(B|j,m) = \alpha \mathbf{f}_B(B) \times \mathbf{f}_{\bar{E}\bar{A}JM}(B)$$

42

Produit point par point

- Le « pointwize product » de deux facteurs $f_1$et f_2 donne un nouveau facteur f dont les variables sont l'union des variables de f_1 et f_2.
- Exemple:

A	B	$f_1(A,B)$
T	T	.3
T	F	.7
F	T	.9
F	F	.1

A	C	$f_2(A,C)$
T	T	.2
T	F	.8
F	T	.6
F	F	.4

A	B	C	$f_3(A,B,C)$
T	T	T	$.3 \times .2$
T	T	F	$.3 \times .8$
T	F	T	$.7 \times .2$
T	F	F	$.7 \times .8$
F	T	T	$.9 \times .6$
F	T	F	$.9 \times .4$
F	F	T	$.1 \times .6$
F	F	F	$.1 \times .4$

43

Variables inutiles

- Considérons: $\boldsymbol{P}(J|b)$

$$P(J|b) = \alpha P(b) \sum_e P(e) \sum_a P(a|b,e) P(J|a) \sum_m P(m|a)$$

La somme sur M donne 1, donc M est inutile.

- Théorème: Y est inutile sauf si $Y \in Ancetres(\{X\} \cup \mathbf{E})$
- Ici: $\boldsymbol{X = J}$, $\boldsymbol{E = B}$

$\boldsymbol{Ancestes(X \cup E) = \{A, E, B\}}$ donc, M est inutile.

44

Inférence approximative dans les réseaux bayésiens

- Les méthodes d'inférences exactes que l'on vient de voir ne sont pas utilisables pour de grands réseaux.
- C'est pourquoi on considère des approches approximatives.
- On va voir des algorithmes basés sur l'échantillonnage aléatoire (Monte Carlo) dont la précision va dépendre du nombre d'échantillons.

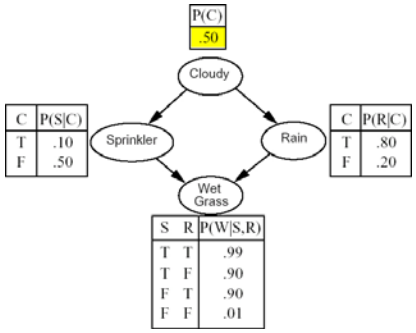
45

Méthodes d'échantillonnage directes

- La forme la plus simple d'échantillonnage aléatoire est de générer des événements sans variable d'évidence.
- La distribution de probabilité à partir de laquelle un échantillon pour une variable est choisi est basée sur les valeurs attribuées aux parents.

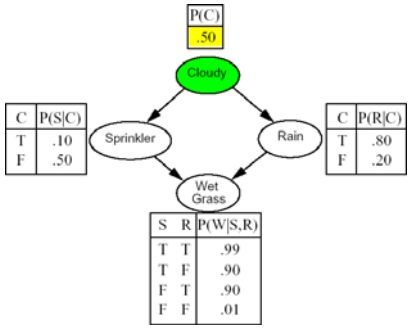
46

Exemple



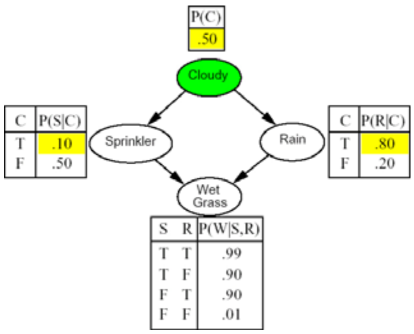
47

Exemple



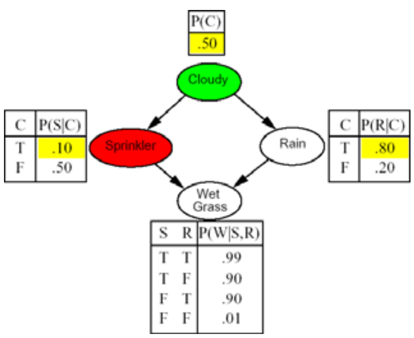
48

Exemple



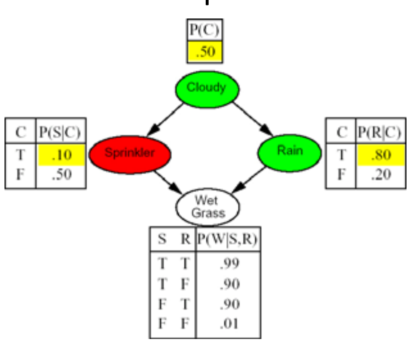
49

Exemple



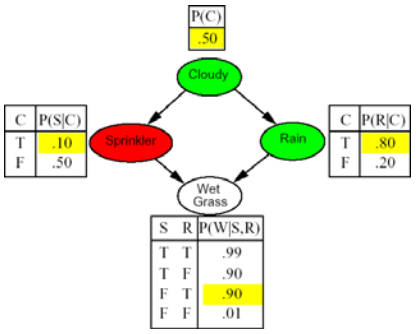
50

Exemple



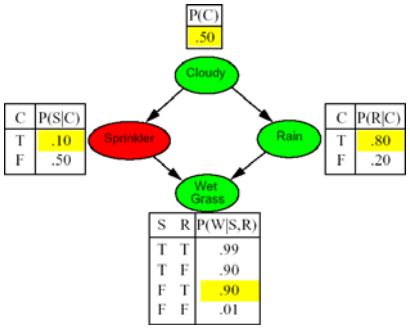
51

Exemple



52

Exemple



53

Estimer la probabilité d'un événement

- On peut estimer la probabilité d'un événement avec la fraction des événements générés aléatoirement qui remplit la condition.
- Par exemple, si on génère 1000 échantillons et que dans 511 d'entre eux, Rain = true, donc on peut faire l'estimation suivante:

$$\hat{P}(Rain = true) = 0.511$$

54

Estimer la probabilité d'un événement

$$S_{PS}(x_1, \ldots, x_n) = \Pi_{i=1}^{n} P(x_i | parents(X_i)$$

$$S_{PS}(x_1, \ldots, x_n) = P(x_i, \ldots, x_n)$$

$$\lim_{N \to \infty} \frac{N_{PS}(x_1, \ldots, x_n)}{N} = S_{PS}(x_1, \ldots, x_n) = P(x_1, \ldots, x_n)$$

$S_{PS}(vrai, faux, vrai, vrai) = 0.5 \times 0.9 \times 0,8 \times 0.9$
$= 0.324$

55

Échantillonnage par rejet

- Utilisé en vue de déterminer les probabilités conditionnelles.
- **Méthode:**
  - Génère des échantillons comme la méthode précédente.
  - Enlève tous les échantillons où les variables d'évidence n'ont pas les bonnes valeurs.
  - Estime la probabilité en comptant parmi les échantillons restants.

56

Échantillonnage par rejet

- Supposons que l'on veut estimer **P**(*Rain*|*Sprinkler* = *true*) en utilisant 100 échantillons.
  - Dans 73 échantillons, *Sprinkler* = *false*, ils sont donc rejetés.
  - Pour les 27 échantillons où *Sprinkler* = *true*:
    - 8 ont *Rain* = *true*
    - 19 ont *Rain* = *false*
  - Donc,

$\mathbf{P}(Rain|Sprinkler = true) \approx \text{NORMALIZE}(\langle 8, 19 \rangle) = \langle 0.296, 0.704 \rangle$

57

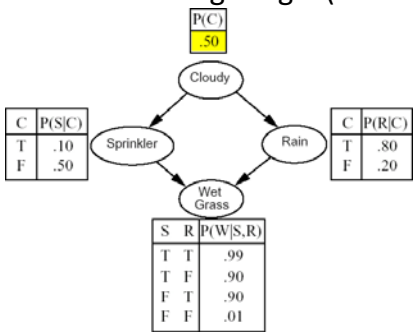
Échantillonnage par rejet

- Le plus gros problème de cette méthode, c'est qu'elle rejette beaucoup d'échantillons.
- Elle génère donc beaucoup d'échantillons inutiles.

« Likelihood weighting » (Pondération par vraisemblence)

- Évite l'inefficacité de l'échantillonnage par rejet en générant uniquement des échantillons consistant avec les variables d'évidence.
- Idée pour un algo WEIGHTED-SAMPLE:
  - Fixer les variables d'évidence
  - Échantillonner uniquement sur les autres variables
  - Attribuer un poids aux échantillons (*w*) selon la probabilité que l'événement survienne en accord avec l'évidence.

« Likelihood weighting » (Pondération



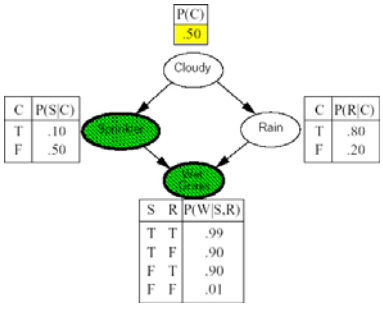
Pondération par vraisemblence(2)

- Requête $\boldsymbol{P}(Rain|Sprinkler{=}true, WetGrass{=}true)$.
- Le processus est comme suit: on fixe *w* à 1
  - Échantillonner à partir de $\boldsymbol{P}(Cloudy) = (0.5, 0.5)$, supposons que ça retourne $true$;
  - $Sprinkler$ est variable d'évidence avec $true$. Dans ce cas, $w \leftarrow w.\boldsymbol{P}(Sprinker{=}true|Cloudy{=}true) = 0.1$
  - Échantillonner à partir de $\boldsymbol{P}(Rain|Cloudy{=}true) = <0.8, 0.2>$, on suppose que ça retourne $true$.
  - $WetGrass$ est une variable d'évidence avec $true$. Dans ce cas, $w \leftarrow w.\boldsymbol{P}(WetGrass{=}true|Sprinkler{=}true, Rain{=}true) = 0.1$ x 0.99

Pondération par vraisemblance(3)

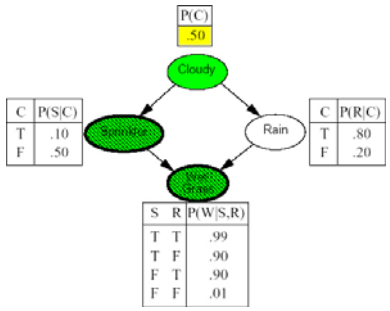
- WEIGHTED-SAMPLE retourne ici l'événement $[true, true, true, true]$ avec un poids de 0.099 qui est compté sous $Rain = true$.
- Le poids est ici faible parce que l'événement décrit un jour nuageux, qui rend improbable le fait que l'arrosage soit en marche.
- Voir le formalisme plus complet au niveau du livre

Pondération par vraisemblance(4)



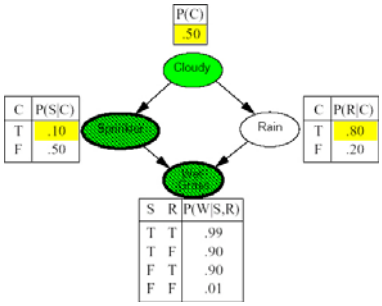
w = 1

Pondération par vraisemblance(5)



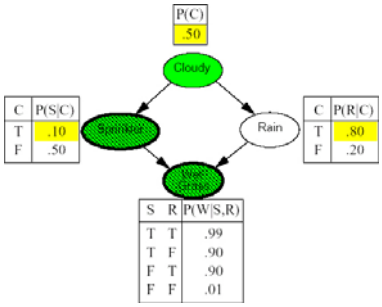
w = 1

Pondération par vraisemblance(6)



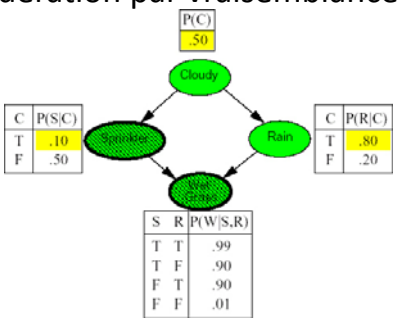
w = 1

Pondération par vraisemblance(7)



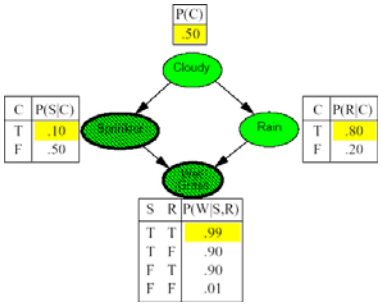
w = 1 * 0.1

Pondération par vraisemblance(8)



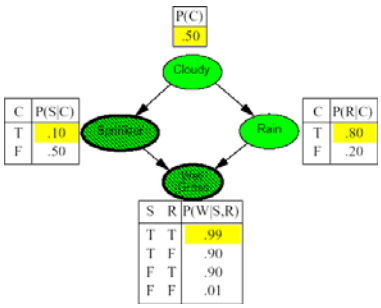
w = 1 * 0.1

Pondération par vraisemblance(9)



w = 1 * 0.1

Pondération par vraisemblance(10)



w = 1 * 0.1 * 0.99 = 0.099

Utilisation de W

$\boldsymbol{W}$

Sample	*Key*					*Weight*
1	$\sim b$	$\sim e$	$\sim a$	$\sim j$	$\sim m$	*0.997*
2	$\sim b$	$\sim e$	$\sim a$	j	$\sim m$	*0.10*
3	$\sim b$	$\sim e$	a	j	m	*0.63*
4	b	$\sim e$	$\sim a$	$\sim j$	$\sim m$	*0.001*

In order to compute the probability of an event that is independent, such as P(*Burglary=true*), we sum the weight for every sample where *Burglary=true* and divide by the sum of all of the weights. For example, in the above data, the only sample where *Burglary=true* is sample 4, with weight 0.001.

Therefore, $P(Burglary=true) = (0.001) / (0.997 + 0.10 + 0.63 + 0.001) = 0.001 / 1.728 = 0.00058$

Similarly $P(a \mid j) = 0.63 / (0.10 + 0.63) = 0.63 / 0.73 = 0.863.$

Pondération par vraisemblance(11)

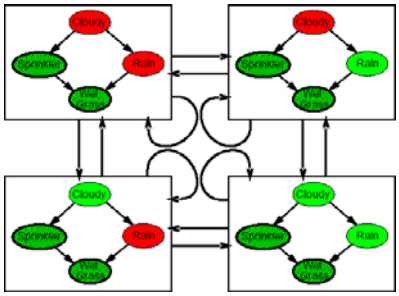
- L'estimation de la probabilité va donc être la somme pondérée des échantillons où ce qui est recherchée est vrai.
- Plus efficace que l'échantillonnage par rejet, mais l'efficacité de la méthode se dégrade si le nombre de variables d'évidence augmente, parce que:
  - la majorité des échantillons vont avoir des petits poids et, donc
  - seulement une minorité d'échantillons vont avoir pratiquement tout le poids total.

Inférence par MCMC

- L'algorithme Markov chain Monte Carlo (MCMC) génère les événements en faisant un changement aléatoire à l'événement précédent.
- L'algorithme maintient donc un état courant où toutes les variables ont une valeur.
- Pour générer le prochain état:
  - Choisir une variable qui n'est pas une variable d'évidence.
  - La distribution de cette variable dépend des valeurs des variables dans son Markov Blanket

$$P(Rain \mid MB(Rain)) = P(Rain \mid Cloudy, Sprinkler, WetGrass)$$

Inférence par MCMC



Avec $Sprinkler = true$ et $WetGrass = true$, il y a quatre états possibles.

Échantillonage via MCMC

- La requête est : $\mathbf{P}(Rain|Sprinkler = true, WetGrass = true)$. Les variables d'observation Sprinkler et WetGrass sont initialisées à leurs variables observées et les variables cachées $Cloud$ et $Rain$ sont initialisées au hasard—par exemple respectivement à $true$ et $false$: L'état initial est donc $[true, true, false, true]$ pour [C,S,R,W].
- Les variables cachées [$Cloudy$ et $Rain$] sont alors échantillonnées itérativement dans un ordre arbitraire

Échantillonnage via MCMC (2)

- à partir de $\mathbf{P}(Cloudy|Sprinkler = true, Rain = false)$ on échantillonne $Cloudy$ on obtient alors $Cloudy = false$; le nouvel état est alors $[false, true, false, true]$
- On échantillonne maintenant $Rain$ à partir de $\mathbf{P}(Rain|Cloudy = False, Sprinkler = True, WetGrass = true)$
  Supposons que cela donne $Rain = true$, le nouvel état est alors $[false, true, true, true]$ et on continue, en échantillonnant $Cloudy$

Inférence par MCMC

- Chaque état visité durant ce processus est un échantillon qui contribue à la variable de requête Rain
- Exemple: si on génère 100 échantillons et que l'on trouve:
  - 20 où $Rain = true$
  - Et 60 où $Rain = false$
- Donc, l'estimation de la distribution est $\mathrm{Normalize}(20{,}60) = (0.25{,}0.75)$.