IFT 615 – Intelligence artificielle

Raisonnement probabiliste

Hugo Larochelle
Département d'informatique
Université de Sherbrooke
http://www.dmi.usherb.ca/~larocheh/cours/ift615.html

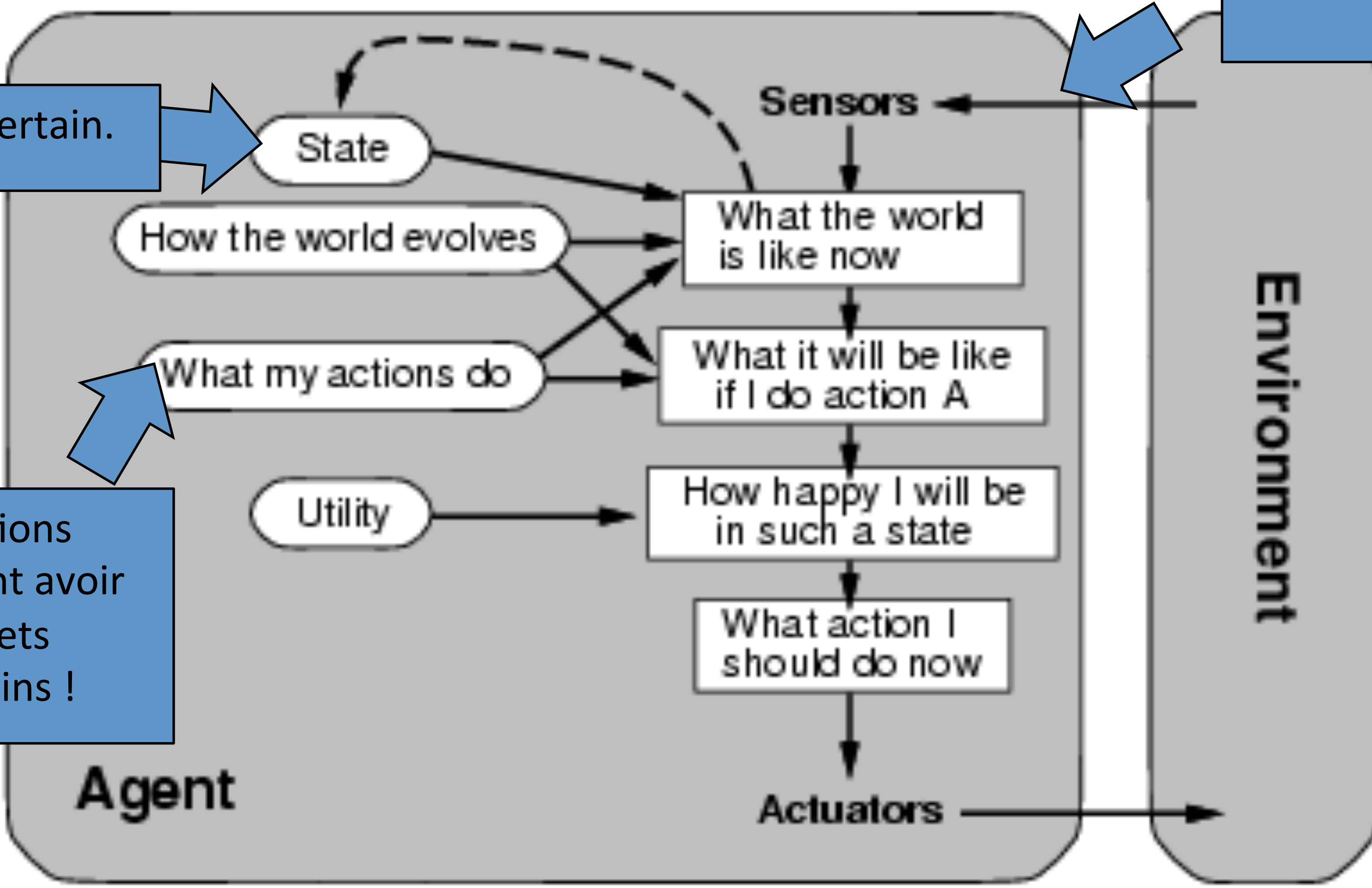
Sujets couverts

- Introduction au raisonnement probabiliste
  - Raisonnement avec incertitude
  - Théorie des probabilités: syntaxe et sémantique
  - Inférences simples
  - Indépendance entre des variables aléatoires
  - Règle de Bayes
  - Illustration avec le monde des wumpus

Rappel: Utility-based agents

Les capteurs peuvent être bruités…

État incertain.



Les actions peuvent avoir des effets incertains !

Incertitude

- Soit A_t l'action d'aller à l'aéroport t minutes avant le départ de l'avion
- A_t me permettra-t-il d'arriver à temps?
- Problèmes:
  - observabilité partielle (conditions routières, etc.)
  - senseurs bruités (annonces du trafic, etc.)
  - incertitude dans l'effet des actions (crevaisons, pannes, etc.)
  - immense complexité pour modéliser les actions et le trafic
- Un raisonnement purement logique:
  - risque de tirer des conclusions erronées
    - « A_{25} me permettra d'arriver à temps » (impossible de faire cette garantie)
  - risque de tirer des conclusions peu exploitables du point de vue de la prise de décision
    - « A_{25} me permettra d'arriver à temps, s'il ne pleut pas, s'il n'y a pas d'accident, si mes pneus ne crèvent pas, etc. »
    - « A_{1440} me permettra presque certainement d'arriver à temps, mais je devrai passer une nuit à l'aéroport. »

Méthodes pour le raisonnement avec incertitude

- **Logique non-monotone (*nonmonotonic/default logic*)**
  - supposer que ma voiture n'aura pas de crevaison
  - supposer que A_{25} suffit **à moins** d'information (*evidence*) contradictoire
  - enjeux:
    - Quelles hypothèses sont raisonnables?
    - Comment gérer les contradictions?
- **Règles de production avec facteurs de certitude**
  - A_{25} → 0.4 *ArriveATemps*
  - *Arroseur* → 0.99 *PelouseMouillée*
  - *PelouseMouillé* → 0.7 Pluie
  - enjeux:
    - Problèmes avec les combinaisons de règles pour faire des déduction. Par exemple: Arroseur causes Pluie !?

Méthodes pour le raisonnement avec incertitude

- **Probabilités**
  - modélise la croyance/certitude des agents
    - » les connaissances de l'agent peuvent au mieux donner un degré de croyance dans les faits
  - étant donnée l'information/observation disponible jusqu'ici, A_{25} me permettra d'arriver avec une probabilité de 0.4

Probabilités

- Les assertions probabilistes facilitent la modélisation:
  - **des faits et de règles complexes**: comparée aux règles de production, l'approche est moins sensible à l'impossibilité d'énumérer toutes les exceptions, antécédents ou conséquences de règles

  - **de l'ignorance**: l'approche est moins sensible à l'omission/oubli des faits, de prémisses ou des conditions initiales à un raisonnement

Probabilités

- Perspective **subjective/bayésienne** des probabilités:
  - les probabilités expriment le degré de croyance d'un agent dans des propositions/faits
    - exemple: $P(A_{25}$ | aucun accident rapporté) = 0.06
  - les probabilités ne sont pas des assertions sur ce qui est vrai de façon absolue
  - n'expriment pas forcément des tendances/fréquences d'une situation, mais pourraient être apprises automatiquement à partir d'expériences
  - les probabilités des propositions changent avec l'acquisition de nouvelles informations
    - exemple: $P(A_{25}$ | aucun accident rapporté, 5h du matin) = 0.15
- À l'opposée, il y a la perspective **objective/fréquentiste** des probabilités
  - les probabilités expriment des faits/propriétés sur des objets
  - on peut estimer ces probabilités en observant ces objets à plusieurs reprises
  - les physiciens diront que les phénomènes quantiques sont objectivement probabilistes

Prise de décisions avec incertitude

- Supposons que je crois ceci:
  - $P(A_{25}$ me permet d'arriver à temps | …) = 0.04
  - $P(A_{90}$ me permet d'arriver à temps | …) = 0.70
  - $P(A_{120}$ me permet d'arriver à temps | …) = 0.95
  - $P(A_{240}$ me permet d'arriver à temps | …) = 0.999
  - $P(A_{1440}$ me permet d'arriver à temps | …) = 0.9999
- Quelle action devrais-je choisir?
  - cela dépend de mes **préférences**: manquer l'avion vs. trop d'attente
- **La théorie de l'utilité** est utilisée pour modéliser et inférer avec des préférences
  - une préférence exprime le degré d'utilité d'une action/situation

- **Théorie de la décision** = **théorie des probabilités** + **théorie de l'utilité**

Probabilités: notions de base

- Exprime le degré de croyance
- Commencer avec un ensemble Ω appelé **espace d'échantillonnage**
  - exemple: 6 possibilités si on roule un dé
  - $\omega \in \Omega$ est un **échantillon** (un **état** ou un **événement atomique**)
- Un **modèle de probabilités** est avec une distribution de probabilité $P(\omega)$ pour chaque élément $\omega \in \Omega$, telle que
  - $0 \leq P(\omega) \leq 1$
  - $\Sigma_{\omega \in \Omega} P(\omega) = 1$
- Exemple du dé: $P(1)=P(2)=P(3)=P(4)=P(5)=P(6)=1/6$
- Un **événement** est un sous-ensemble de Ω
- Probabilité d'un événement
  - $P(\mathrm{A})=\Sigma_{\{\omega \in \mathrm{A}\}} P(\omega)$
- Exemple du dé: $P(\text{Dé est} < 4) = P(\omega=1 \cup \omega=2 \cup \omega=3) = 1/6+1/6+1/6 = 1/2$

Variable aléatoire

- Une **variable aléatoire** est une variable décrivant **une partie** des connaissances incertaines (on la note avec une **première lettre majuscule**)
- Chaque variable a un **domaine** de valeurs qu'elle peut prendre
  - on peut voir une variable comme une fonction définie sur l'espace d'échantillonnage et donnant une valeur à chaque échantillon en entrée
- **Types de variables aléatoires**:
  - **Booléennes**: le domaine est {*true*, *false*}
    - » exemple: *Carie* $\in$ {*true*, *false*} (ai-je la carie?)
  - **Discrètes**: le domaine est énumérable
    - » *Météo* $\in$ {*soleil*, *pluie*, *nuageux*, *neige*}
  - **Continues**: le domaine est continu (par exemple, l'ensemble des réels)
    - » exemple: $X = 4.0$, $PositionX \leq 10.0$, $Speed \leq 20.5$
- P induit une **distribution de probabilité** pour chaque variable aléatoire X
  - $P(X=xi) = \Sigma_{\{\omega:\, X(\omega)=xi\}} P(\omega)$
  - exemple du dé: $P(\, NombreImpaire = true) = P(1)+P(3)+P(5) = 1/6+1/6+1/6=1/2$

Propositions

- Une **proposition** est une assertion de ce qui est vrai, c.-à-d., une assertion sur la valeur d'une variable
  - en d'autre mots, un événement (ensemble d'échantillons ou d'événements atomiques) pour lequel la proposition est vraie
    - exemple: *Carie* = *true*, qu'on va aussi noter carie
- Étant données deux variables booléennes *A* et *B*:
  - l'événement *a* est l'ensemble d'échantillons ω pour lesquels $A(\omega)$ = *true*
  - l'événement $\neg$ *a* est l'ensemble d'échantillons ω pour lesquels $A(\omega)$ = *false*
  - l'événement $a \wedge b$ est l'ensemble des ω pour lesquels $A(\omega)$=*true* et $B(\omega)$=*true*
  - l'événement $a \vee b$ est l'ensemble des ω pour lesquels $A(\omega)$=*true* ou $B(\omega)$=*true*

Propositions

- Souvent nous aurons plusieurs variables aléatoires
  - toutes les variables aléatoires tiennent leur valeur d'un même échantillon ω
  - pour des variables distinctes, l'espace d'échantillonnage est alors le **produit cartésien** des domaines des variables aléatoires
- Un **événement atomique** est donc une spécification complète de l'état du « monde » pour lequel un agent est incertain
  - par exemple, si le « monde » de l'agent est décrit par seulement deux variables aléatoires booléennes (*Carie* et *MalDeDents*), il y a exactement quatre états / événements atomiques possibles:
    - *Carie = false* ∧ *MalDeDents = false*
    - *Carie = false* ∧ *MalDeDents = true*
    - *Carie = true* ∧ *MalDeDents = false*
    - *Carie = true* ∧ *MalDeDents = true*
    - on a donc Ω = { *<true,true>*, *<true,false>*, *<false,ture>*, *<false, false>*}
- Les événements atomiques sont exhaustifs et mutuellement exclusifs

Syntaxe des propositions

- Élément de base: variable aléatoire
- Similaire à la logique propositionnelle

- **Variables aléatoires booléenne**
  - exemple: *DentCariée = true*

- **Variables aléatoires discrètes** (domaines **finis** or **infinis**)
  - exemple: *Météo* = *v*, avec $v \in$ { *soleil*, *pluie*, *nuageux*, *neige* }

- **Variables aléatoires continues** (**bornées** ou **non bornées**)
  - exemple: *Temp*=21.6 (la variable *Temp* a exactement la valeur 21.6)
  - exemple: *Temp* < 22.0 (la variable *Temp* a une valeur inférieure à 22)

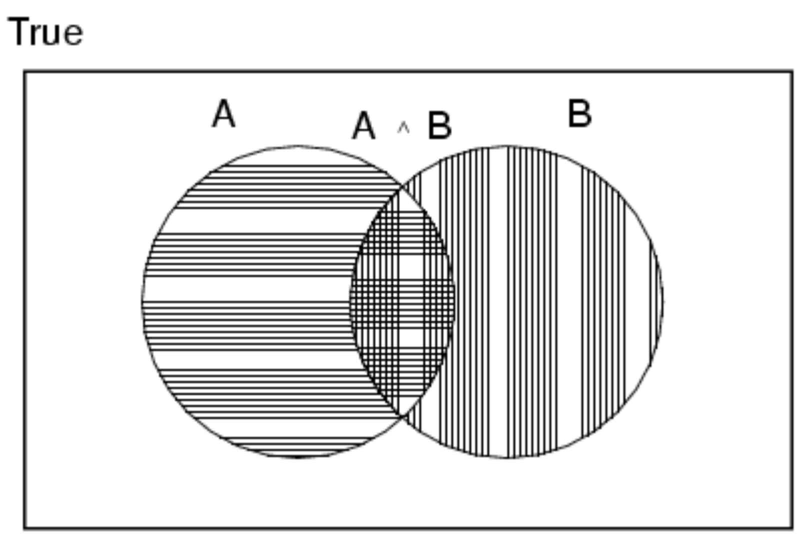
Syntaxe des propositions

- En général, les **propositions élémentaires** sont définies en assignant une valeur ou un intervalle de valeurs aux variables
  - exemple: *Météo = soleil*, *Carie = false* (noté ¬*carie*)

- Les **propositions complexes** sont définies par des combinaisons booléennes
  - exemple: (*Météo = soleil*) ∨ (*Carie = false*)

Axiomes de la théorie des probabilités: Axiomes de Kolmogorov

- Pour toute propositions A, B
  - $0 \leq P(\mathrm{A}) \leq 1$
  - $P(true) = 1$ et $P(false) = 0$
  - $P(\mathrm{A} \vee \mathrm{B}) = P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{A} \wedge \mathrm{B})$



Probabilité a priori/inconditionnelle

- La **probabilité a priori** ou **inconditionnelle** de propositions exprime le degré de croyance dans ces propositions avant l'acquisition de toute (nouvelle) information / observation
  - exemple: $P(Carie = true) = 0.1$ et $P(Météo = soleil) = 0.72$
- La **distribution des probabilités** donne les valeurs de probabilités pour toutes les assignations possibles de valeurs aux variables:
  - exemple: $\mathbf{P}(Météo) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$
  - **fonction de densité de probabilité**: fonction déterminant la probabilité de propositions d'une variable continue
    - » avec la fonction de densité, on peut exprimer la probabilité d'intervalles de valeurs $P(X < xi) = \int_{x<xi} p(X=x)\, dx$
    - » dans le cas continu, on a toujours que $P(X=xi) = 0$

Probabilité a priori/inconditionnelle

- La **distribution conjointe** de probabilités pour un ensemble de variables donne la probabilité pour chaque événement atomique décrit par ces variables
  - exemple: la distribution conjointe $P(Météo, Carie)$ est une matrice 4×2:

Météo =	*soleil*	*pluie*	*nuageux*	*neige*
Carie = true	0.144	0.02	0.016	0.02
Carie = false	0.576	0.08	0.064	0.08

Probabilité a posteriori/conditionnelle

- La **probabilité conditionnelle** ou **a posteriori** tient compte des nouvelles informations/observations disponibles
  - exemple: $P(carie \mid malDeDents) = 0.8$
    - » C.-à-d., étant donné que la seule chose que je sais est *MalDeDents* = *true*
  - si on constate qu'un patient a mal aux dents et aucune autre information n'est encore disponible, la probabilité qu'il ait une carie est de 0.8
- Si on en apprend plus, (par exemple, on découvre une carie), on a:
  - $P(carie \mid malDeDents, carie) = 1$
- Toutes les nouvelles informations ne sont pas pertinentes, donc on peut simplifier:
  - exemple: $P(carie \mid malDeDents, CanadiensOntGagné = true) =$
    $P(carie \mid malDeDents) = 0.8$

Probabilité a posteriori/conditionnelle

- Définition de la **probabilité conditionnelle**:
  - $P(a \mid b) = P(a \wedge b) / P(b)$ si $P(b) \neq 0$
  - la probabilité de *a*, étant donné (que tout ce qu'on sait est) *b*
- Formulation équivalente (**règle du produit**):
  - $P(a \wedge b) = P(a|b)\ P(b) = P(b|a)\ P(a)$
- Il existe une version plus générale pour les distributions de probabilité
  - exemple: $\mathbf{P}(\mathit{Météo}, \mathit{Carie}) = \mathbf{P}(\mathit{Météo} \mid \mathit{Carie})\ \mathbf{P}(\mathit{Carie})$
- La **règle de chaînage** (***chain rule***) est obtenue par une application successive de la règle du produit:
  - $$\begin{aligned} \mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n \mid X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n-1} \mid X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n \mid X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &= \prod_{\mathrm{i}=1..\mathrm{n}} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1}) \end{aligned}$$

Inférence par énumération

- Commencer avec la distribution conjointe des probabilités:

	malDeDents		***¬ malDeDents***	
	croche	*¬croche*	*croche*	*¬croche*
carie	0.108	0.012	0.072	0.008
¬ carie	0.016	0.064	0.144	0.576

- Pour chaque proposition ϕ, faire une somme sur les événements atomiques pour lesquels elle est vraie: $P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

Inférence par énumération

- Commencer avec la distribution conjointe des probabilités:

	malDeDents		***¬ malDeDents***	
	croche	*¬croche*	*croche*	*¬croche*
carie	0.108	0.012	0.072	0.008
¬ carie	0.016	0.064	0.144	0.576

- Pour chaque proposition ϕ, faire une somme sur les événements atomiques pour lesquels elle est vraie: $P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

- $P(malDeDents) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inférence par énumération

- Commencer avec la distribution conjointe des probabilités:

	malDeDents		¬ *malDeDents*	
	croche	¬*croche*	*croche*	¬*croche*
carie	0.108	0.012	0.072	0.008
¬ *carie*	0.016	0.064	0.144	0.576

- Pour chaque proposition ϕ, faire une somme sur les événements atomiques pour lesquels elle est vraie: $P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

- $P(carie \vee malDeDents) = 0.108 + 0.012 + 0.072 + 0.008$
  $+ 0.016 + 0.064 = 0.28$

Inférence par énumération

- Commencer avec la distribution conjointe des probabilités:

	malDeDents		***¬ malDeDents***	
	croche	*¬croche*	*croche*	*¬croche*
carie	0.108	0.012	0.072	0.008
¬ carie	0.016	0.064	0.144	0.576

- On peut aussi calculer les probabilités conditionnelles:

$$P(\neg\, carie \mid malDeDents) = \frac{P(\neg\, carie \wedge malDeDents)}{P\,(malDeDents)}$$

$$= (0.016+0.064) \,/\, (0.108 + 0.012 + 0.016 + 0.064)$$

$$= 0.4$$

Normalisation

	malDeDents		¬ *malDeDents*	
	croche	¬*croche*	*croche*	¬*croche*
carie	0.108	0.012	0.072	0.008
¬ *carie*	0.016	0.064	0.144	0.576

- Le dénominateur peut être vu comme une constante de normalisation α
- **P**(*Carie* | *malDeDents*) = α **P** (*Carie*, *malDeDents*)

  = α [**P** (*Carie*, *malDeDents*, *croche*) +

  P (*Carie*, *malDeDents*, ¬ *croche*)]

  $= \alpha\, [\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$

  $= \alpha\, \langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$

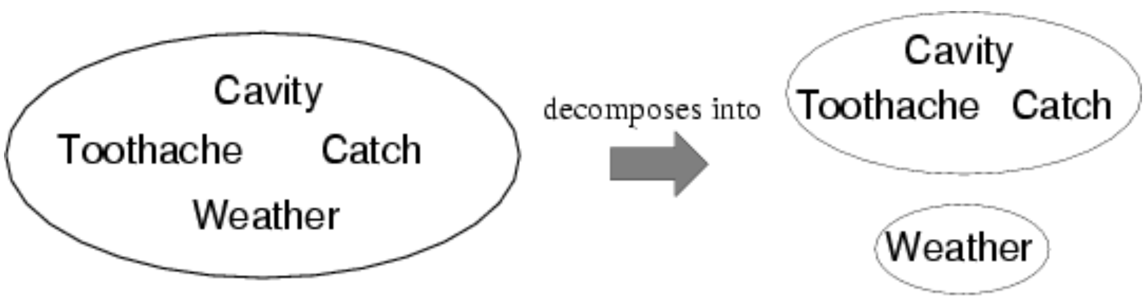
  avec $\alpha = 1 / P(malDeDents) = 1/(.108 + .012 + .016 + .064) = 1/0.2 = 5.$
- Idée générale: calculer la contribution de la **variable de requête** en fixant **les variables d'observation** et en faisant la somme sur les **variables cachées**

Inférence par énumération

- En général on veut calculer la probabilité conjointe a posteriori sur un ensemble de **variables de requête** *X* étant donné les valeurs *e* pour les **variables d'observation** *E*
- Soit *Y* l'ensemble des **variables cachées** (non encore observées), *X* la valeur recherchée, et *E* l'ensemble des variables d'observation
- On obtient la probabilité pour la requête $\mathbf{P}(X \mid E = \mathrm{e})$ en faisant une sommation sur les variables cachées:
  - $\mathbf{P}(X \mid E = e) = \alpha\, \mathbf{P}(X, E = \mathrm{e}) = \alpha\, \Sigma_y\, \mathbf{P}(X, e, y)$
- Les termes dans la somme sont des probabilités conjointes étant donné que *X*, *E* et *Y* pris ensembles couvrent toutes les variables aléatoires
  - complexité en temps: O(dn), avec d la taille du plus grand domaine des variables et n le nombre de variables de requête et cachées
  - complexité en espace: O(dn), pour stocker la distribution

Indépendance

- Les variables *A* et *B* sont indépendantes si et seulement si
  - $\mathbf{P}(A|B) = P(A)$ ou
  - $\mathbf{P}(B|A) = P(B)$ ou
  - $\mathbf{P}(A, B) = P(A)\ P(B)$
- Exemple: $P(MalDeDents, Croche, Carie, Météo)$
  $= P(MalDeDents, Croche, Carie)\ P(Météo)$



- 32(=2^3*4) entrées réduites à 12;
  pour n variables indépendantes, $O(2^n) \rightarrow O(n)$

Indépendance

- L'indépendance totale est puissante mais rare
  - l'indépendance entre les variables permet de réduire la taille de la distribution des probabilités et rendre les inférences plus efficaces
  - mais il est rare d'être dans une situation où toutes les variables sont réellement indépendantes
- La dentisterie est un domaine avec un grand nombre de variables, mais très peu d'entre elles sont indépendantes. Que faire?

Indépendance conditionnelle

- Si j'ai une carie, la probabilité que la sonde accroche dans la dent ne dépend pas du fait que j'aie mal à la dent ou non:
  - **P**(*Croche* | *MalDeDents*, *carie*) = **P**(*Croche* | *carie*)
- Même chose si je n'ai pas la carie:
  - **P**(*Croche* | *MalDeDents*, ¬*carie*) = **P**(*Croche* | ¬*carie*)
- *Croche* est **conditionnellement indépendante** de *MalDeDents* étant donné *Carie*:
  - **P**(*Croche* | *MalDeDents*, *Carie*) = **P**(*Croche* | *Carie*)
- Formulations équivalentes:
  - **P**(*MalDeDents* | *Croche* , *Carie*) = **P**(*MalDeDents* |*Carie*)
  - **P**(*MalDeDents*, *Croche* | *Carie*) = **P**(*MalDeDents* |*Carie*) **P**(*Croche*|*Carie*)

Indépendance conditionnelle

- Réécrivons la distribution conjointe en utilisant la **règle de chaînage** (***chain rule***):

  P(*MalDeDents*, *Croche*, *Carie*)

  = **P**(*MalDeDents* | *Croche*, *Carie*) **P**(*Croche*, *Carie*)

  = **P**(*MalDeDents* | *Croche*, *Carie*) **P**(*Croche* | *Carie*) **P**(*Carie*)

  = **P**(*MalDeDents* | *Carie*) **P**(*Croche* | *Carie*) **P**(*Carie*)
- C-à-d., 2 + 2 + 1 = 5 **paramètres individuels/distincts**

- Dans des cas idéals, l'exploitation de l'indépendance conditionnelle réduit la complexité de représentation de la distribution conjointe de exponentielle ($O(2^n)$) en linéaire ($O(n)$)
- En raisonnement probabiliste, l'indépendance conditionnelle est le concept de représentation des connaissances le plus basique et utile

Règle de Bayes

- Règle du produit: $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$
  - **Règle de Bayes:** $P(a \mid b) = P(b \mid \mathrm{a})\, P(a) \,/\, P(b)$
  - $\mathbf{P}(Y|X) = \mathbf{P}(X|Y)\,\mathbf{P}(Y) \,/\, \mathbf{P}(X) = \alpha\mathbf{P}(X|Y)\,\mathbf{P}(Y)$ [pour les distributions]
- Utile pour calculer/interroger une probabilité **diagnostique** à partir d'une probabilité **causale**:
  - $\mathbf{P}(\mathit{Cause}|\mathit{Effect}) = \mathbf{P}(\mathit{Effect}|\mathit{Cause})\,\mathbf{P}(\mathit{Cause}) \,/\, \mathbf{P}(\mathit{Effect})$
- Exemple: soit *m* (méningite), *s* (*stiff neck* / nuque raide)
  - $P(s|m)=0.5$, $P(m)=1/50000$ et $P(s)=1/20$.
  - $P(m|s) = P(s|m)\,P(m) \,/\, P(s) = 0.5 \times 0.00002 \,/\, 0.05 = 0.0002$
- Règle diagnostique: effets observés $\Rightarrow$ causes cachées
- Règle causale: causes cachées $\Rightarrow$ effets observées

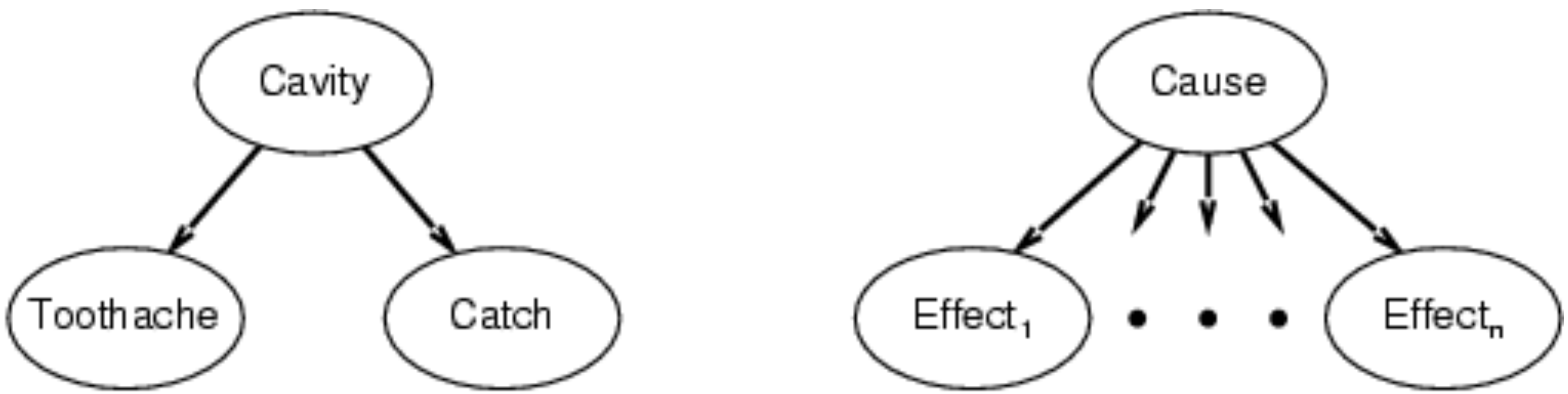
Règle de Bayes et indépendance conditionnelle

- $\mathbf{P}(Carie \mid MalDeDents \wedge Croche)$

  $= \alpha\, \mathbf{P}(MalDeDents \wedge Croche \mid Carie)\, \mathbf{P}(Carie)$

  $= \alpha\, \mathbf{P}(MalDeDents \mid Carie)\, \mathbf{P}(Croche \mid Carie)\, \mathbf{P}(Carie)$

- Exemple d'un **modèle de Bayes simple** (***naive Bayes classifier***):
  - $\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause)$

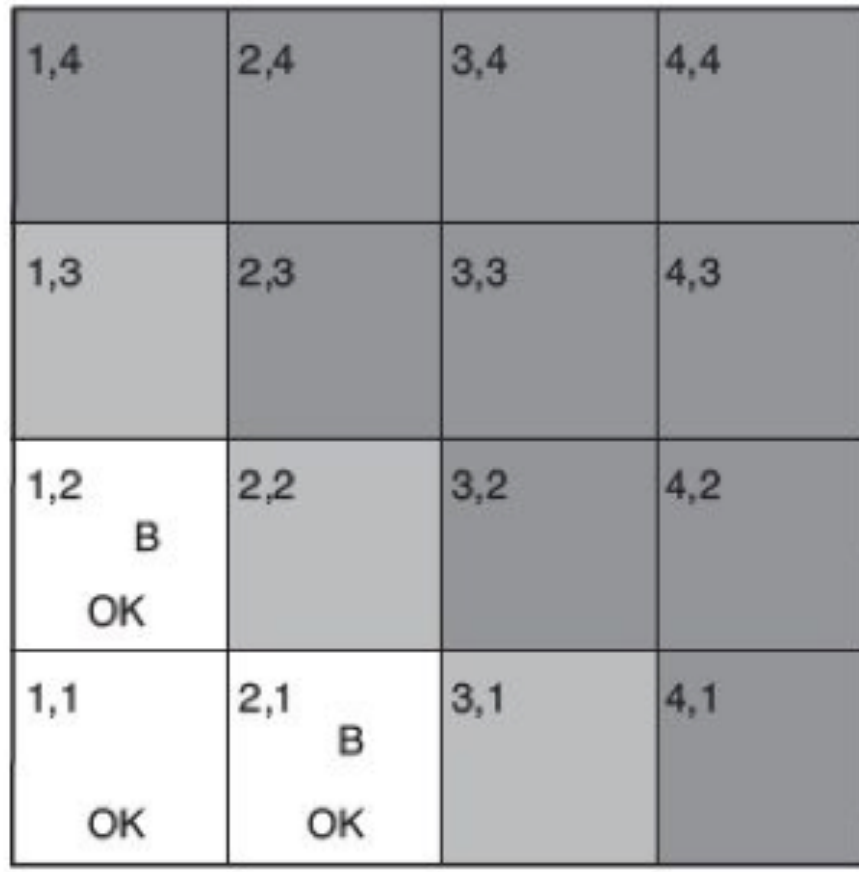


Le monde des Wumpus

Problème: calculer la probabilité que [1,3] contiennent une fosse?

1. **Identifier l'ensemble de variables aléatoires nécessaires**:
   - $P_{i,j}$=true ssi il y a une fosse dans [i,j]
   - $B_{i,j}$=true ssi il y a une brise dans [i,j]

Inclure seulement les variables observées $B_{1,1}$, $B_{1,2}$, $B_{2,1}$ dans la distribution des probabilités (modèle)

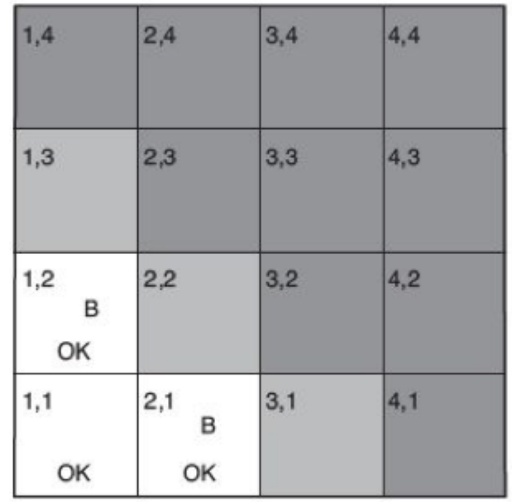


Spécifier la distribution des probabilités

2. **Spécifier la distribution conjointe** ($\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$)

- appliquer la règle du produit: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} | P_{1,1}, \ldots, P_{4,4})\ \mathbf{P}(P_{1,1}, \ldots, P_{4,4})$
(on spécifie une forme $\mathbf{P}(\mathit{Effect} | \mathit{Cause})$)
- premier terme: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} | P_{1,1}, \ldots, P_{4,4})$
  - » probabilité conditionnelle d'une configuration/état de brises, étant donnée une configuration de fosses
  - » 1 si les fosses sont adjacentes aux brises, 0 sinon
- second terme: $\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$
  - » probabilité a priori des configurations des fosses
  - » les fosses sont placées aléatoirement, avec une probabilité de 0.2 par chambre
  - » si $P_{1,1}, \ldots, P_{4,4}$ sont telles qu'il y a exactement n fausses, on aura
$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \Pi_{(i,j)=(1,1) \ldots (4,4)}\ \mathbf{P}(P_{i,j}) = 0.2^{n} * 0.8^{16-n}$

Observations et requête



3. Identifier les observations

- on sait ce qui suit:
  - » $b = \neg\, b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$
  - » $known = \neg\, p_{1,1} \wedge \neg\, p_{1,2} \wedge \neg\, p_{2,1}$

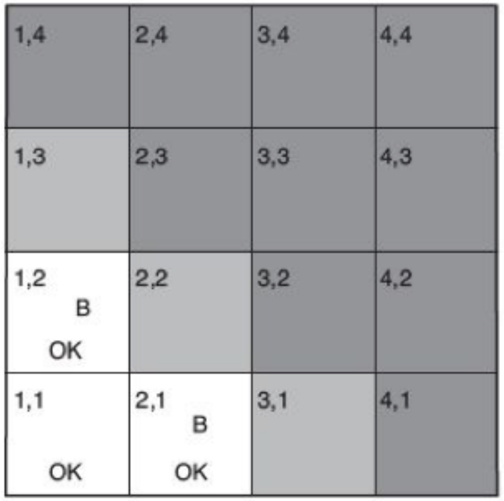
4. Identifier les variables de requête

- y a-t-il une fosse à la position 1,3?
- $\mathbf{P}(P_{1,3} \mid known, b)$?

5. Identifier les variables cachées

- on définit *Unknown* comme étant l'ensemble des variables $P_{i,j}$ autres que celles qui sont connues (*known*) et la variable de requête $P_{1,3}$

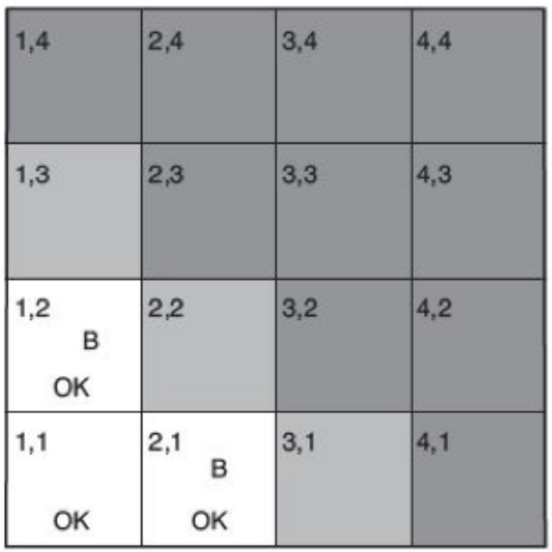
Observations et requête

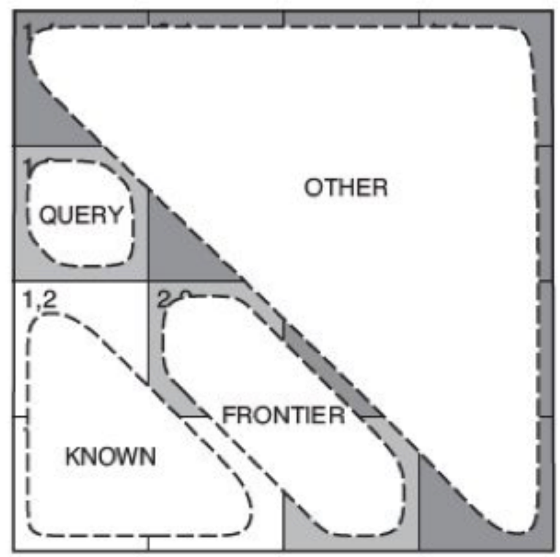


6. **Faire l'inférence**
   - avec l'inférence par énumération, on obtient:
     $\mathbf{P}(P_{1,3} | \, known, b) =$
     $\alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$
   - croît exponentiellement avec le nombre de chambres!
     » avec 12 chambres *unknown*: 2^{12}=4096 termes

Utiliser l'indépendance conditionnelle

- Idée de base: les observations sont conditionnellement indépendantes des chambres cachées étant données les chambres adjacentes
  - C.-à-d., les autres chambres ne sont pas pertinentes





- Définir *Unknown* = *Frontier* ∪ *Other*
- $\mathbf{P}(b|P_{1,3}, known, Unknown) = \mathbf{P}(b|P_{1,3}, known, Frontier, Other)$
- Réécrire la probabilité $\mathbf{P}(P_{1,3}|\ known, b)$ pour exploiter cette indépendance

Utiliser l'indépendance conditionnelle

$\mathbf{P}(P_{1,3} | \, known, b) = \alpha \, \Sigma_{unknown} \, \mathbf{P}(P_{1,3}, \, unknown, known, b)$

$= \alpha \, \Sigma_{unknown} \, \mathbf{P}(b | P_{1,3} \, , known, unknown) \, \mathbf{P}(P_{1,3} \, , known, unknown)$

$= \alpha \, \Sigma_{frontier} \, \Sigma_{other} \, \mathbf{P}(b | \, known, P_{1,3} \, , frontier, other) \, \mathbf{P}(P_{1,3} \, , known, frontier, other)$

$= \alpha \, \Sigma_{frontier} \, \Sigma_{other} \, \mathbf{P}(b | \, known, P_{1,3} \, , frontier) \, \mathbf{P}(P_{1,3} \, , known, frontier, other)$

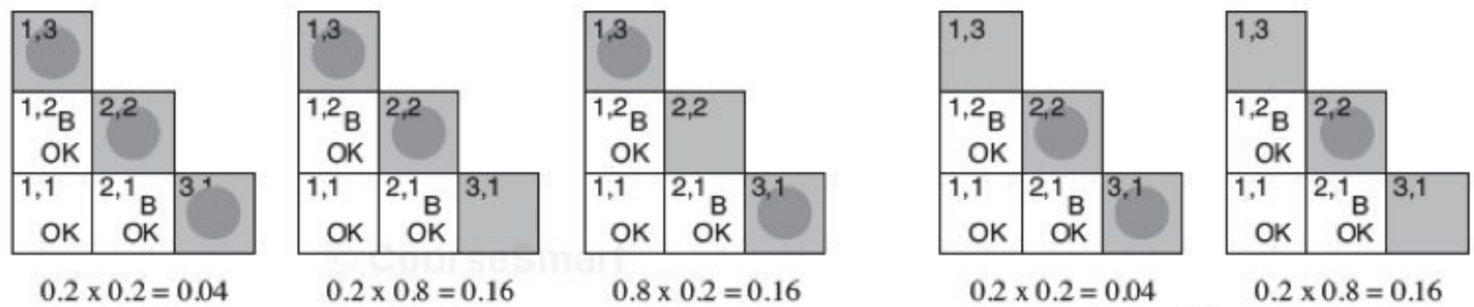
$= \alpha \, \Sigma_{frontier} \, \mathbf{P}(b | \, known, P_{1,3} \, , frontier) \, \Sigma_{other} \, \mathbf{P}(P_{1,3} \, , known, frontier, other)$

$= \alpha \, \Sigma_{frontier} \, \mathbf{P}(b | \, known, P_{1,3} \, , frontier) \, \Sigma_{other} \, \mathbf{P}(P_{1,3} \,) \, P(known) \, P(frontier) \, P(other)$

$= \alpha \, P(known) \, \mathbf{P}(P_{1,3}) \, \Sigma_{frontier} \, \mathbf{P}(b | \, known, P_{1,3} \, , frontier) \, P(frontier) \, \Sigma_{other} \, P(other)$

$= \alpha' \, \mathbf{P}(P_{1,3}) \, \Sigma_{frontier} \, \mathbf{P}(b | \, known, P_{1,3} \, , frontier) \, P(frontier)$

Utiliser l'indépendance conditionnelle



(a) (b)

- Événements cohérents pour les variables $P_{2,2}$ et $P_{3,1}$, montrant $\mathbf{P}(\mathit{frontier})$
- Pour chaque événement:
  a) 3 événements avec $P_{1,3}$= *true*, montrant 2 ou 3 fosses.
  b) 2 événements avec $P_{1,3}$= *false*, montrant 1 ou 2 fosses.

$$\mathbf{P}(P_{1,3} | \mathit{known}, b) = \alpha' <0.2(0.04+0.16+0.16), 0.8(0.04+0.16)>$$

$$\approx <0.31, 0.69>$$

Résumé

- La théorie des probabilités est un formalisme cohérent pour raisonner avec l'incertitude
- Une distribution conjointe spécifie la probabilité pour toutes les variables aléatoires
- On peut répondre à des requêtes en faisant une somme sur les événements atomiques
- Pour les domaines d'application réalistes, on doit trouver une façon de réduire la taille de la distribution conjointe
- L'indépendance et l'indépendance conditionnelles nous fournissent les outils de base pour simplifier les distributions conjointes