Incertitude

8INF878 – Intelligence Artificielle

Kévin Bouchard

Ordre du jour

- Incertitude
- Probabilité
- Syntaxe et sémantique
- Inférence
- Indépendance et règle de Bayes

Incertitude

- Supposons $A_t =$ Se rendre à l'aéroport t minutes avant le vol.
- Est-ce que A_t m'amènera à temps?

- Problèmes:
  1) Partiellement observable (route, autos, etc.)
  2) Bruits dans les capteurs
  3) Incertitude dans la réalisation des actions (crevaison, etc.)
  4) Complexité importante en ce qui attrait à la modélisation et à la prédiction du trafic

Incertitude (suite)

- Une approche purement logique
  1) Risque de mensonge: « A_{25} va me rendre à temps » ou encore…

  2) conduit à des conclusions qui sont trop faibles pour la prise de décision:

  « A_{25} va m'amener à temps s'il n'y a pas d'accident sur le pont et qu'il ne pleut pas et que mes pneus restent intacts, etc. »

  (A_{1224} pourrait raisonnablement être décrit par va me rendre à temps, mais je devrais passer la nuit à l'aéroport …)

Méthodes pour gérer l'incertitude

- Logique non monotone:
  - Assumons que ma voiture n'a pas de crevaison
  - Assumons que A_{25} fonctionne sauf si contredit par une évidence

- Le problème est qu'il est difficile de déterminer quelles hypothèses sont raisonnables et comment gérer les contradictions

- Règles avec facteur arbitraire (fudge factor):

$$A_{25} \longmapsto_{0.3} AtAirportOnTime$$
$$Sprinkler \longmapsto_{0.99} WetGrass$$
$$WetGrass \longmapsto_{0.7} Rain$$

- Problème avec les combinaisons (arroseur cause la pluie?)

Méthodes pour gérer l'incertitude

- Probabilité
  - Selon les évidences disponibles
  - A_{25} m'amènera à temps avec une probabilité de 0.04

- Cardamo (1565) théorie des jeux de hasard

- Attention! La logique floue gère les degrés de vérités, PAS l'incertitude
  - E.g.: Le ciel est bleu est vrai à 0.8

Probabilités

- Les assertions probabilistes résumes les effets de
  - La paresse: incapable d'énumérer les exceptions, les qualifications, etc.
  - L'ignorance: l'absence de faits pertinents, les conditions initiales, etc.

- Probabilités Bayésienne:
  - Les probabilités s'appliquent sur les connaissances d'une personne sur les connaissances
  - E.g.: $P(A_{25} | pas\ d'accidents) = 0.06$

Probabilités

- Ce ne sont pas les revendications d'une «tendance probabiliste» dans la situation actuelle
  - Mais! pourraient être tirés des expériences passées de situations similaires

- Les probabilités de propositions changent avec de nouvelles preuves:
  - E.g.: $P(A_{25}|aucun\ accidents\ signalé, 5h) = 0.15$
  - Analogue à l'état de l'implication logique KB |= α, pas la vérité

Prendre des décisions dans l'incertitude

- Supposons que je crois ceci:

$$P(A_{25}\ m'amène\ à\ temps|\ ...) = 0.04$$
$$P(A_{90}\ m'amène\ à\ temps|\ ...) = 0.7$$
$$P(A_{120}\ m'amène\ à\ temps|\ ...) = 0.95$$
$$P(A_{1440}\ m'amène\ à\ temps|\ ...) = 0.99$$

- Quelle action choisir? Dépend de mes préférences sur manquer un vol vs vivre à l'aéroport
- Decision theory = utility theory + probability theory
  - Rationalité=Maximum Expected Utility!!!!

```
function DT-AGENT(percept) returns an action
  persistent: belief_state, probabilistic beliefs about the current state of the world
              action, the agent's action

  update belief_state based on action and percept
  calculate outcome probabilities for actions,
     given action descriptions and current belief_state
  select action with highest expected utility
     given probabilities of outcomes and utility information
  return action
```

Bases

Bases des probabilités

- Commençons avec l'ensemble Ω, l'espace échantillon
  - E.g.: 6 possibilités au roulement d'un dé
  - $\omega \in \Omega$ est un point d'échantillon ou un événement atomique

- Un espace de probabilités ou un modèle de probabilités est un espace échantillon avec un assignement $P(\omega)\ \forall\ \omega \in \Omega$
  - $0 \leq P(\omega) \leq 1$
  - $\sum_{\omega} P(\omega) = 1$
- E.g.: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$

- Un événement A est n'importe quel sous ensemble de Ω
  - $P(A) = \sum_{\{\omega \in A\}} P(\omega)$
- E.g.: $P(dé < 4) = P(1) + P(2) + P(3) = 1/2$

Variables aléatoires

- Une variable aléatoire est une fonction d'un ensemble de points échantillons
  - E.g. $\mathbb{R}$ ou encore les booléens
  - E.g. impair(1) = Vrai

- P induit une distribution de probabilités pour toute variable aléatoire X:
  - $P(X = x_i) = \sum_{\{\omega : X(\omega) = x_i\}} P(\omega)$

$$P(Impair = vrai) = P(1) + P(3) + P(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

Propositions

- Supposons les variables booléennes A et B:
  - Évènement $a = \{\omega | A(\omega) = Vrai\}$
  - Évènement $\neg a = \{\omega | A(\omega) = Faux\}$
  - Évènement $a \wedge b = \{\omega | A(\omega) = Vrai \wedge B(\omega) = Vrai\}$

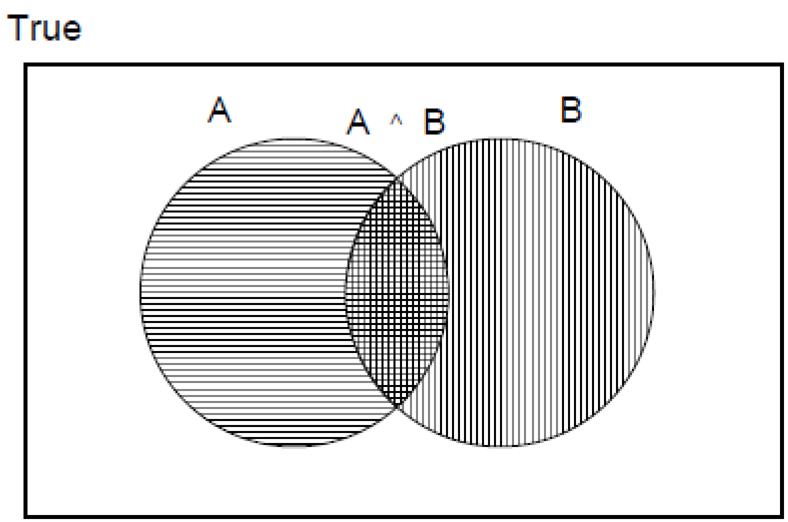
- Souvent dans les applications en IA, l'ensemble des points échantillons (sample points) sont définis par les valeurs d'un ensemble de variables aléatoires
  - I.e.: l'espace échantillon est le produit cartésien des plages de valeurs des variables

Propositions (suite)

- Avec les variables booléennes:
  - Sample points=modèle de logique propositionnelle
  - E.g.: $A = vrai, B = faux$ ou $a \wedge \neg b$
- Proposition = disjonction d'évènements atomiques pour laquelle celle-ci est vraie
  - E.g.: $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b) \Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Pourquoi utiliser les probabilités?

- Les définitions impliquent que certains événements logiquement liés doivent avoir des probabilités liées
- E.g.: $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



- Finetti (1931): un agent qui mise selon des probabilités qui transgresse ces axiomes peut être forcé de miser de manière à perdre de l'argent indépendamment du résultat.

Probabilités 101

Syntaxes des propositions

- Variables aléatoires propositionnelles ou booléennes
  - E.g.: $Carie$ (Ai-je une carie?)
  - $Carie = vrai$ est une proposition

- Variables aléatoires discrètes (finies ou infinies)
  - E.g.: $Température$ est dans $\{ensoleillée, pluvieuse, neigeuse, nuageuse\}$
  - $Température = pluvieuse$ est une proposition
  - Les valeurs doivent être exhaustive et mutuellement exclusive

- Variables aléatoires continues (bornées ou non)
  - E.g: $Température = 21.3$

- Combinaisons booléennes arbitraires de propositions de base

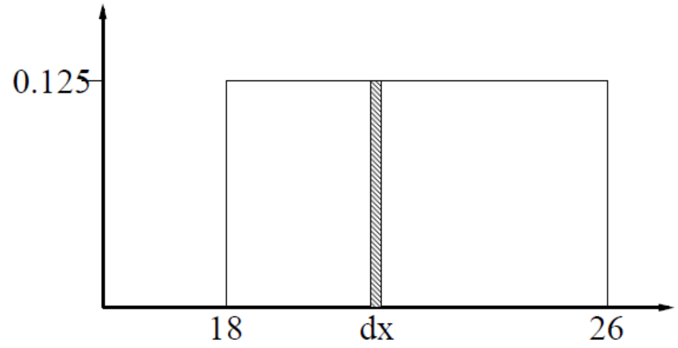
Probabilité antérieure

- Les probabilités antérieures ou inconditionnelles des propositions correspondent à la croyance avant l'arrivée de toute (nouvelle) évidence
  - E.g.: $P(Carie = true) = 0.1$ et
  - $P(Température = neigeuse) = 0.72$

- Une distribution de probabilités donne des valeurs pour toutes les affectations possibles:
  - $P(Température) = \langle 0.08, 0.1, 0.72, 0.1 \rangle$
  - Normalisée -> somme = 1

Probabilités pour variables continues

- La distribution est exprimée comme une fonction paramétrable de valeurs:
  - $P(X = x) = U[18C, 26C](x)$
  - →Densité uniforme entre 18 et 26

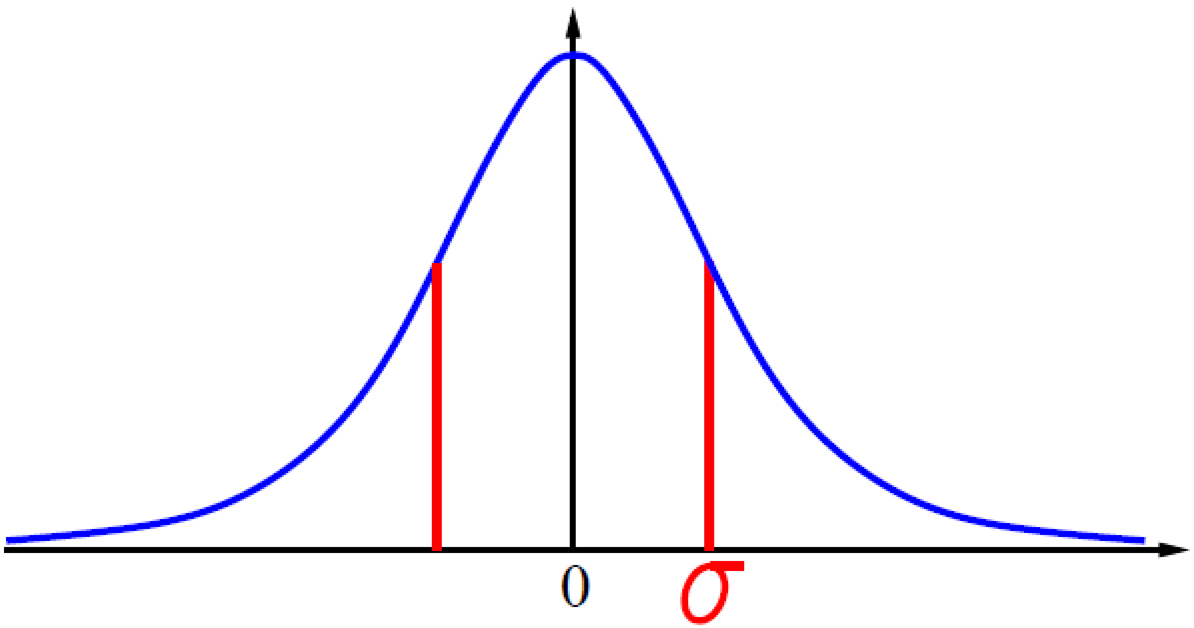


- Ici, P est une densité; son intégrale est 1
- $P(X = 20.5) = 0.125$ signifie

$$\lim_{dx \to 0} \frac{P(20.5 \leq X \leq 20.5 + dx)}{dx} = 0.125$$

Densité Gaussienne

- $P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$



Probabilité antérieure (suite)

- La distribution de probabilités conjointes pour un ensembles de variables aléatoires donne les probabilités de chacun des évènements atomiques
  - I.e.: chaque point échantillon
  - $P(Température, Carie) = 4 \times 2$ matrice de valeurs:

Température	Neigeuse	Pluvieuse	Ensoleillée	nuageuse
Carie=Vrai	0.144	0.02	0.016	0.02
Carie=Faux	0.576	0.08	0.064	0.08

- Chaque question à propos d'un domaine peut être répondu par celle-ci puisque chaque évènement est la somme de points échantillons

Probabilité conditionnelle

- Probabilités conditionnelles ou à posteriori
  - E.g.: $P(carie|maldedent) = 0.8$
  - i.e.: Supposons que je sais MalDeDent=V, que je n'ai pas d'autres informations, Carie=Vrai avec 0.8 de probabilité
  - PAS→Si MalDeDent, 80% de chance de Carie

- Si nous savons plus, par ex. que Carie est vrai
  - $P(carie|maldedent, carie) = 1$

Probabilité conditionnelle

- Notons que les connaissances moins spécifiques demeurent valides après l'obtention de nouvelles informations, mais ne sont pas toujours utiles

- Elles peuvent donc être simplifiées!
  - $P(carie|maldedent, lesudokuestcomplet) = P(carie|maldedent) = 0.8$

- Ce type d'inférence, sanctionnée par la connaissance du domaine, est crucial

Probabilité conditionnelle

- Définition:

$$P(a|b) = \frac{P(a \land b)}{P(b)} \, if \; P(b) \neq 0$$

- La règle du produit donne une formulation alternative:

$$P(a \land b) = P(a|b)P(b) = P(b|a)P(a)$$

- Une version générale tient pour distributions entières
  - E.g.: $P(Température, Carie) = P(Température|Carie)P(Carie)$
  - (voir comme un ensemble 4X2 d'équations, pas une matrice)

Probabilité conditionnelle

- La règle de chaînage est dérivée de l'application successive de la règle du produit:

$$P(X_1, \dots, X_n) = P(X_1, \dots, X_{n-1})P(X_n | X_1, \dots, X_{n-1})$$
$$= P(X_1, \dots, X_{n-2})P(X_{n-1} | X_1, \dots, X_{n-2})P(X_n | X_1, \dots, X_{n-1})$$
$$= \cdots$$
$$= \prod_i^n P(X_i | X_1, \dots, X_{i-1})$$

Inférence par l'énumération

- Commençons avec la distribution conjointe:

	MalDeDent		¬MalDeDent	
	Prise	¬Prise	Prise	¬Prise
Carie	.108	.012	.072	.008
¬Carie	.016	.064	.144	.576

- Pour toute proposition ϕ, la somme des évènements atomiques est:

$$P(\phi) = \sum_{\omega:\omega \vDash \phi} P(\omega)$$

Inférence par l'énumération

- Commençons avec la distribution conjointe:

	MalDeDent		¬MalDeDent	
	Prise	¬Prise	Prise	¬Prise
Carie	.108	.012	.072	.008
¬Carie	.016	.064	.144	.576

- Pour toute proposition ϕ, la somme des évènements atomiques est:

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

$$P(maldedent) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inférence par l'énumération

- Commençons avec la distribution conjointe:

	MalDeDent		¬MalDeDent	
	Prise	¬Prise	Prise	¬Prise
Carie	.108	.012	.072	.008
¬Carie	.016	.064	.144	.576

- Pour toute proposition ϕ, la somme des évènements atomiques est:

$$P(\phi) = \sum_{\omega:\omega\vDash\phi} P(\omega)$$

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

Inférence par l'énumération

- Commençons avec la distribution conjointe:

	MalDeDent		¬MalDeDent	
	Prise	¬Prise	Prise	¬Prise
Carie	.108	.012	.072	.008
¬Carie	.016	.064	.144	.576

- Pour toute proposition ϕ, la somme des évènements atomiques est:

$$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$$

$$P(\neg carie|maldedent) = \frac{P(\neg carie \wedge maldedent)}{P(maldedent)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Inférence par l'énumération

	MalDeDent		¬MalDeDent	
	Prise	¬Prise	Prise	¬Prise
Carie	.108	.012	.072	.008
¬Carie	.016	.064	.144	.576

- Le dénominateur peut être considéré comme une constante de normalisation α

$$P(Carie|maldedent) = \alpha P(Carie, maldedent)$$
$$= \alpha[P(Carie, maldedent, prise) + P(Carie, maldedent, \neg prise]$$
$$= \alpha[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$
$$= \alpha\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle$$

- Idée générale: Calculer la distribution sur la variable interrogée en fixant les variables d'évidences et en additionnant les variables cachées

Inférence par l'énumération

- Supposons X l'ensemble des variables.

- Nous voulons la distribution conjointe à postériori des variables interrogeables Y
  - Étant donnée les valeurs spécifiques e pour les variables d'évidences E

- Supposons les variables cachées $H = X - Y - E$

- La sommation requise des entrées conjointes se fait en additionnant les variables cachées:

$$P(Y|E = e) = \alpha P(Y, E = e) = \alpha \sum_{h} P(Y, E = e, H = h)$$

Inférence par l'énumération

- Les termes dans la somme sont des entrées conjointes puisque Y, E et H ensemble épuise les variables aléatoires X

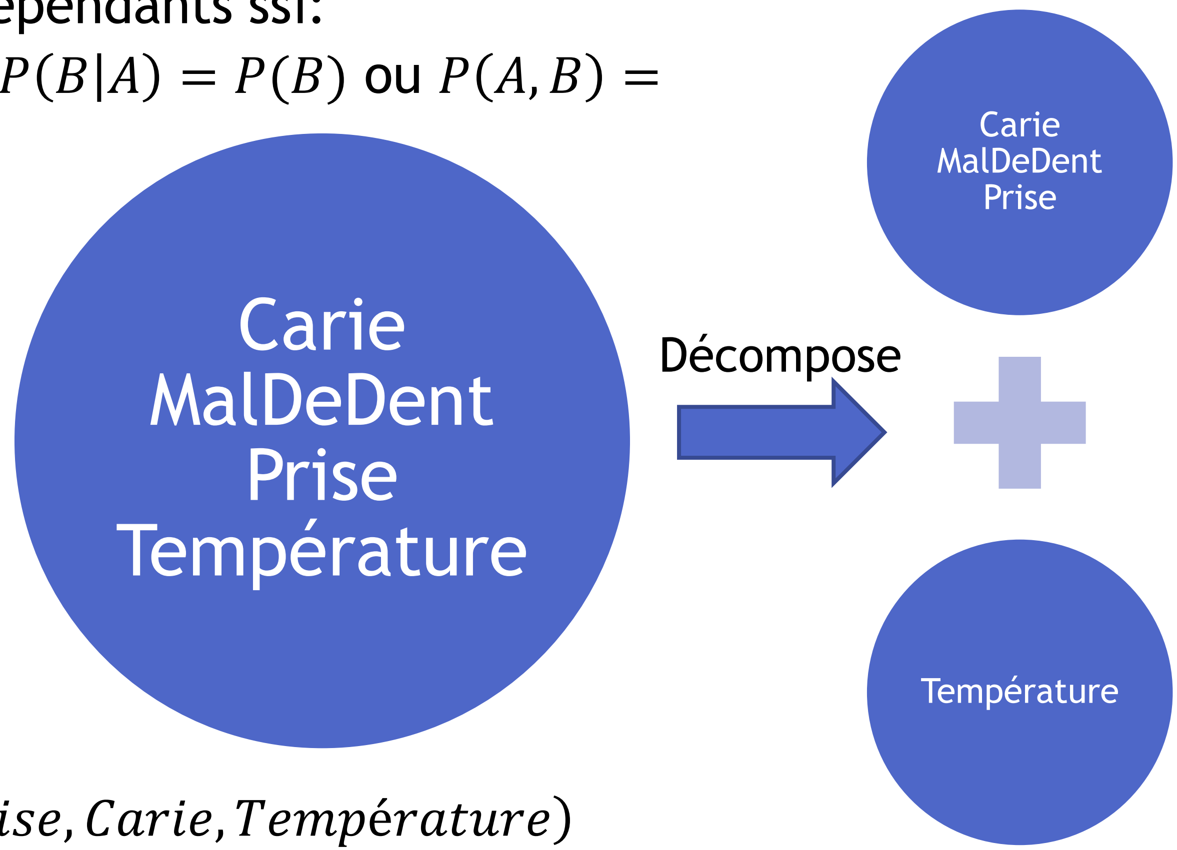
- Problème évident:
  1. Complexité temporelle en pire cas $O(d^n)$ où d est la plus grande arité
  2. Complexité spatiale en pire cas $O(d^n)$ pour l'enregistrement de la distribution
  3. Comment trouver les nombres pour $O(d^n)$ entrées?

Indépendance

Indépendance

- A et B sont indépendants ssi:

$P(A|B) = P(A)$ ou $P(B|A) = P(B)$ ou $P(A,B) = P(A)P(B)$



$$P(MalDeDent, Prise, Carie, Température) = P(MalDeDent, Prise, Carie)P(Température)$$

Exercice indépendance

p(intelligent∧ étudie∧préparé)	intelligent		¬intelligent	
	étudie	**¬étudie**	**étudie**	**¬étudie**
préparé	.432	.16	.084	.008
¬préparé	.048	.16	.036	.072

- Q1: Est-ce que *intelligent* est indépendant de *étudie*?
- Q2: Est-ce que *préparé* est indépendant de *étudie*?

Indépendance

- 32 entrées réduites à 12 (8+4)

- Pour n lancer d'une pièce de monnaie: 2^n entrées, représentable en $2n$
- Indépendance absolue!!!
  - Puissante mais rare

- La dentisterie est un domaine complexe avec des centaines de variables
  - Aucune n'est indépendante
  - Quoi faire?

Indépendance conditionnelle

- $P(MalDeDent, Carie, Prise)$ a $2^3 - 1 = 7$ entrées indépendantes
- Si j'ai une carie, la probabilité que le crochet reste pris ne dépend pas de si j'ai un MalDeDent

$$(1) P(prise|maldedent, carie) = P(prise|carie)$$

- La même indépendance tient si je n'ai pas de carie

$$(2) P(prise|maldedent, \neg carie) = P(prise|\neg carie)$$

- Prise est conditionnellement indépendant de MalDeDent si carie

$$P(Prise|MalDeDent, carie) = P(Prise|carie)$$

- Équivalents:

$$P(MalDeDent|Prise, carie) = P(MalDeDent|carie)$$

$$P(MalDeDent, Prise|carie) = P(MalDeDent|carie)P(Prise|carie)$$

Indépendance conditionnelle

- Pour écrire la distribution complète, utilisons la règle de chaînage

$$P(MalDeDent, Prise, Carie)$$
$$= P(MalDeDent|Prise, Carie)P(Prise, Carie)$$
$$= P(MalDeDent|Prise, Carie)P(Prise|Carie)P(Carie)$$
$$= \boldsymbol{P(MalDeDent|Carie)}P(Prise, |Carie)P(Carie)$$

  - I.e.: 2+2+1 = 5 valeurs indépendantes

- La plupart du temps, l'utilisation de l'indépendance conditionnelle réduit la taille de la représentation de la distribution conjointe d'exponentielle en *n* à linéaire en *n*
- **L'indépendance conditionnelle est la plus simple et plus robuste forme de connaissances à propos des environnements incertains**

39

Règle de Bayes

Règle de Bayes

- Règle du produit: $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$
  - Règle de Bayes: $P(a|b) = \frac{P(b|a)P(a)}{P(b)}$

- Ou sous la forme distribution:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} = \alpha P(X|Y)P(Y)$$

Règle de Bayes

- Elle est utile pour évaluer la probabilité diagnostique d'une probabilité causale:

$$P(Cause|Effet) = \frac{P(Effet|Cause)P(Cause)}{P(Effet)}$$

- E.g.: Supposons M, la méningite et S avoir le cou raide:

$$P(s|m) = 0.8 \quad P(s) = 0.1 \qquad P(m) = = .0001$$

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 * 0.0001}{0.1} = 0.0008$$

- Note: La probabilité d'avoir la méningite à posteriori est encore très petite!

Naive Bayes Classifier

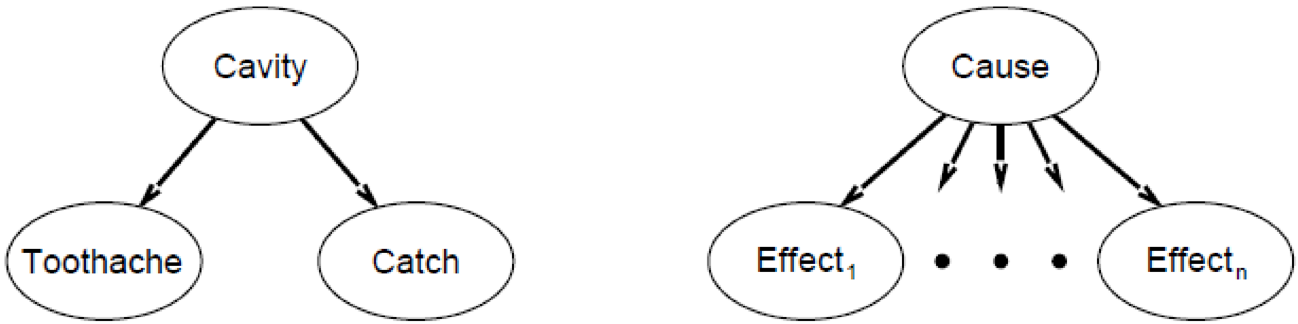
- Base de connaissances:
  - Évidences: $E_1, \dots, E_m$
  - Hypothèses: $H_1, \dots, H_n$
  - E_i, H_j sont binaires
  - Les hypothèses sont mutuellement exclusives (pas d'overlap)
  - Elles sont exhaustives (couvre tous les cas possibles)

- Cas (cases): évidences pour une instance particulière $(E_1, \dots, E_4)$
- But: Trouver l'hypothèse H_i avec la plus haute probabilité à postériori
  - $Max_i P(H_i | E_1, \dots, E_k)$

Naive Bayes Classifier

$$P(Carie|maldedent \wedge prise)$$
$$= \alpha P(maldedent \wedge prise|Carie)P(Carie)$$
$$= \alpha P(maldedent|Carie)P(prise|Carie)P(Carie)$$

$$P(Cause, Effet_1, \dots, Effet_n) = P(Cause)\prod_i P(Effet_i|Cause)$$



- Le nombre total de paramètres est linéaire en *n*

Monde de la forêt enchantée

Monde de la forêt enchantée

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 V **OK**	2,2	3,2	4,2
1,1 **OK**	2,1 V **OK**	3,1	4,1

- $C_{ij} = Vrai\ ssi\ [i,j]$ contient une crevasse
- $V_{ij} = Vrai\ ssi\ [i,j]$ est venteuse
- Inclut uniquement $V_{1,1}, V_{1,2}, V_{2,1}$ dans le modèle de probabilités

Spécification du modèle probabiliste

- La distribution conjointe complète est $P\left(C_{1,1},\ldots,C_{4,4},V_{1,1},V_{1,2},V_{2,1}\right)$

- Appliquons la règle du produit: $P\left(V_{1,1},V_{1,2},V_{2,1}|C_{1,1},\ldots,C_{4,4}\right)P\left(C_{1,1},\ldots,C_{4,4}\right)$
  - (Pour obtenir un $P(Effet|Cause)$)

- Premier terme: 1 si crevasse adjacente à case venteuse, 0 sinon
- Second terme: Les crevasses sont placées aléatoirement avec probabilité de 0.2 par case:

$$P\left(C_{1,1},\ldots,C_{4,4}\right)=\prod_{i,j=1,1}^{4,4}P\left(C_{i,j}\right)=0.2^{n}\times 0.8^{16-n}$$

- Pour *n* crevasses

Observations et requêtes

- Nous savons que:

$$v = \neg v_{1,1} \wedge v_{1,2} \wedge v_{2,1}$$
$$connue = \neg c_{1,1} \wedge \neg c_{1,2} \wedge \neg c_{2,1}$$

- La requête est $P\left(c_{1,3} \middle| connue, v\right)$
- Définissons $Inconnue = C_{i,j} \ \notin Connue \cup C_{1,3}$

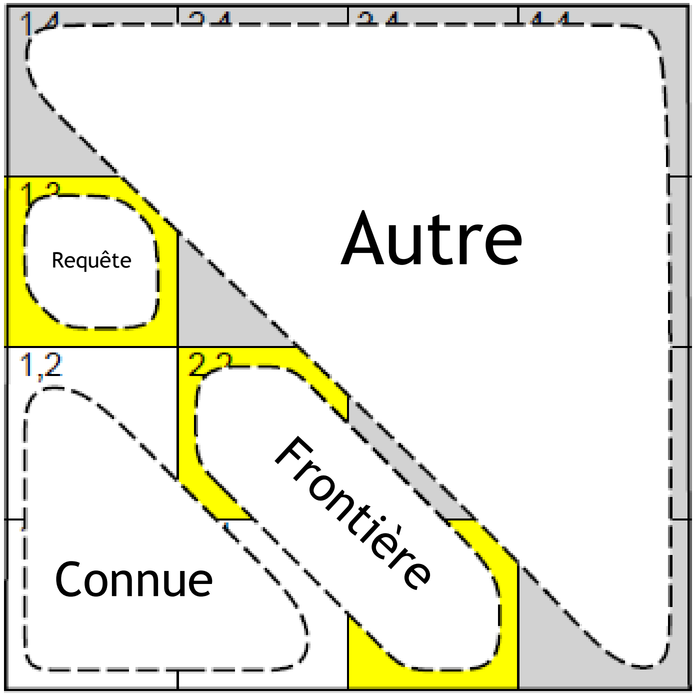
- Pour l'inférence par énumération nous avons

$$P\left(C_{1,3} | connue, v\right)$$
$$= \alpha \sum_{inconnue} P(C_{1,3}, inconnue, connue, v)$$

- Exponentiel avec le nombre de tuiles!

En utilisant l'indépendance conditionnelle

- Intuition: les observations sont conditionnellement indépendantes des autres tuiles étant donné les tuiles voisines cachées



- Définissons $Inconnue = Frontière \cup Autre$
  $$P(v|C_{1,3}, Connue, Inconnue) = P(v|C_{1,3}, Connue, Frontière)$$
- Manipulons la requête dans une forme qui nous permettra d'utiliser ceci!

En utilisant l'indépendance conditionnelle

$$P(C_{1,3}|connue, v) = \alpha \sum_{inconnue} P(C_{1,3}, inconnue, connue, v)$$

$$= \alpha \sum_{inconnue} P(v|connue, C_{1,3}, inconnue)P(C_{1,3}, connue, inconnue)$$

$$= \alpha \sum_{Frontière} \sum_{Autre} P(\,v|connue, frontière, autre)P(C_{1,3}, connue, frontière, autre)$$

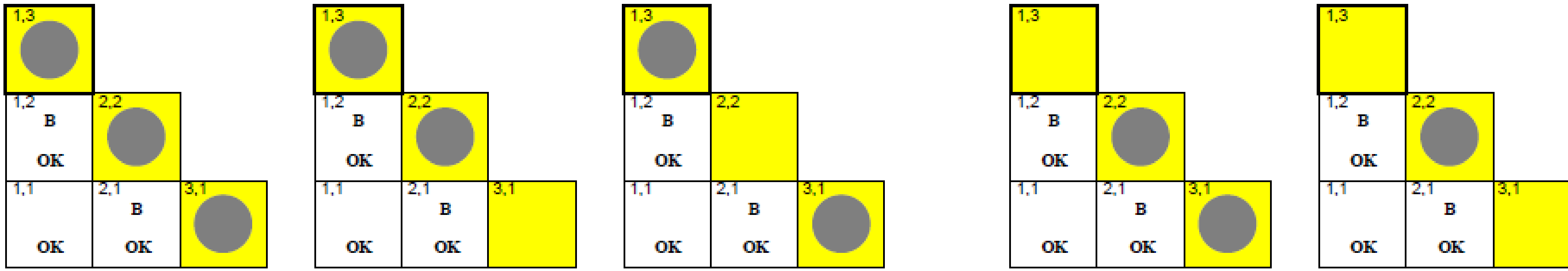
$$= \alpha \sum_{Frontière} \sum_{Autre} P(\,v|connue, frontière)P(C_{1,3}, connue, frontière, autre)$$

$$= \alpha \sum_{Frontière} P(v|connue, C_{1,3}, frontière) \sum_{Autre} P(C_{1,3}, connue, frontière, autre)$$

$$= \alpha \sum_{Frontière} P(v|connue, C_{1,3}, frontière) \sum_{Autre} P(C_{1,3})P(connue)P(frontière)P(autre)$$

$$= \alpha P(connue)P(C_{1,3}) \sum_{Frontière} P(v|connue, C_{1,3}, frontière)P(frontière)$$

En utilisant l'indépendance conditionnelle



0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16 | 0.8 x 0.2 = 0.16 | 0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16

$$P\left(C_{1,3} \middle| connue, v\right)$$
$$= \alpha' \langle 0.2(0.04 + 0.16 + 0.16), 0.8(0.04 + 0.16) \rangle$$
$$\approx \langle 0.31, 0.69 \rangle$$

$$P\left(C_{2,2} \middle| connue, v\right) \approx \langle 0.86, 0.14 \rangle$$

51

Raisonnement probabiliste

Réseau Bayésien

- Un réseau Bayésien est un modèle de représentation de connaissances qui se traduit par un graphe orienté acyclique

- Les nœuds y correspondent à des variables aléatoires, et les arcs à des influences de nature **causale** ou **conditionnelle**

- Dans cette structure, on émet l'hypothèse que les connaissances disponibles permettent de définir, sur l'ensemble des variables aléatoires, une mesure de probabilité unique

Définition formel d'un réseau bayésien

- Formellement, un réseau bayésien est constitué de quatre principaux éléments

1. Un ensemble V de **variables aléatoires** constituant les noeuds du réseau

2. Un ensemble de relations $r \in R$, où $r \subseteq V \times V$
   - Ces relations sont dirigées
   - $r(v_1, v_2)$ définit un arc partant du noeud v_1 et allant vers le noeud v_2

Définition formel d'un réseau bayésien

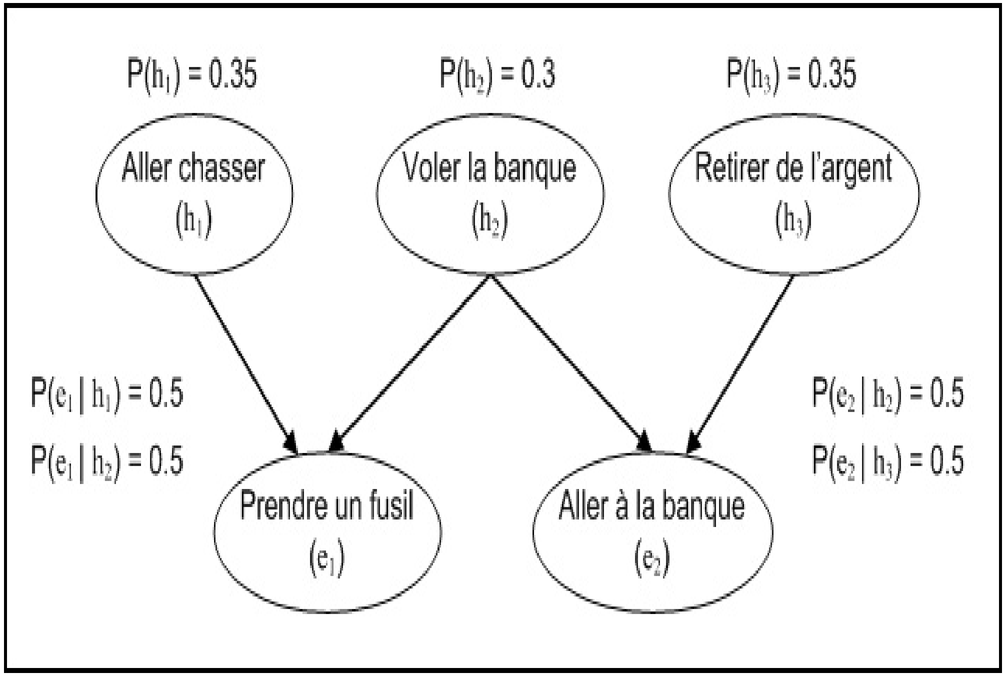
3. Une table *TPI* décrivant la probabilité initiale des nœuds racines
   - C'est l'ensemble des **noeuds parents** n'étant influencés par aucun autre nœud
   - Ces nœuds possèdent seulement des arcs descendants (des arcs allant de ces nœuds vers d'autres)

4. Une table *TPC* décrivant la probabilité conditionnelle des nœuds non-racines
   - C'est-à-dire les **noeuds enfants** étant influencés par au moins un autre noeud et possédant un ou plusieurs arcs ascendants (des arcs partant d'un autre noeud et allant vers celui-ci)

Exemple de reconnaissance d'AVQ

- Dans cet exemple, chacun des plans automotivés de la hiérarchie est représenté par un noeud racine spécifiant une hypothèse de haut niveau

- L'ensemble des hypothèses possibles pour la reconnaissance est formé de H = {h1, h2, h3}



- Une **probabilité initiale** est attribuée à chacune de ces hypothèses, soit
- P(h1) = 0.35, P(h2) = 0.3 et P(h3) = 0.35.

Encoder la librairie avec un réseau bayésien

- Les nœuds non-racines représentent des actions de bas niveau qui influenceront (conditionnent) la probabilité des hypothèses de haut niveau

- On note ici que toutes ces hypothèses de haut niveau sont considérées **mutuellement indépendantes et disjointes**
  - La **somme des probabilités** est donc égale à 1

Définir la reconnaissance en termes de processus de décision bayésien

- En démarrant d'une librairie de plans représentée sous forme d'un réseau bayésien, il est possible de définir l'activité de reconnaissance de plans en termes de **processus de décision bayésien**

- Pour ce faire, on doit d'abord admettre trois postulats de départ:

  1. Les hypothèses h_i sont mutuellement exclusives (**disjointes**)
  2. Elles sont **exhaustives** (la librairie de plans est complète)
  3. Il y a une indépendance conditionnelle des observations par rapport aux hypothèses

- Ces postulats sont nécessaires à l'application du théorème de Bayes dans le processus de décision

Définir la reconnaissance en termes de processus de décision bayésien

- Un processus de décision bayésien prend en entrée une ou plusieurs **observations** *e*

- Ces observations doivent nécessairement correspondre à des noeuds **non-racines** du réseau
  - Des actions basiques

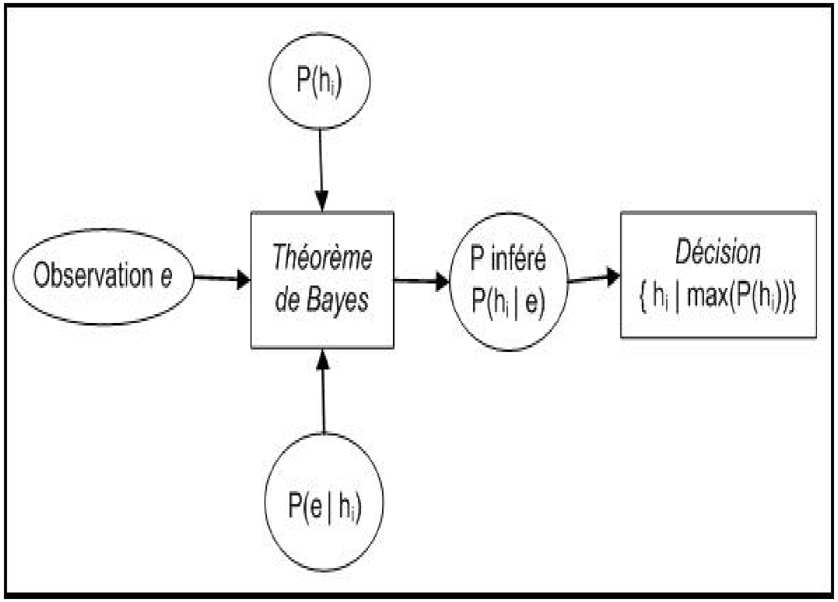
Définir la reconnaissance en termes de processus de décision bayésien

- Par la suite, on se base sur la **probabilité initiale** $P(h_i)$ attribuée à chacune des hypothèses de plans possibles
- Ainsi que sur la **probabilité conditionnelle** $P(e \mid h_i)$ attribuée à chacune des observations moyennant une hypothèse
- Le processus de reconnaissance consiste à **réviser la distribution des probabilités** à travers les hypothèses du réseau en appliquant le théorème de Bayes

Définir la reconnaissance en termes de processus de décision bayésien

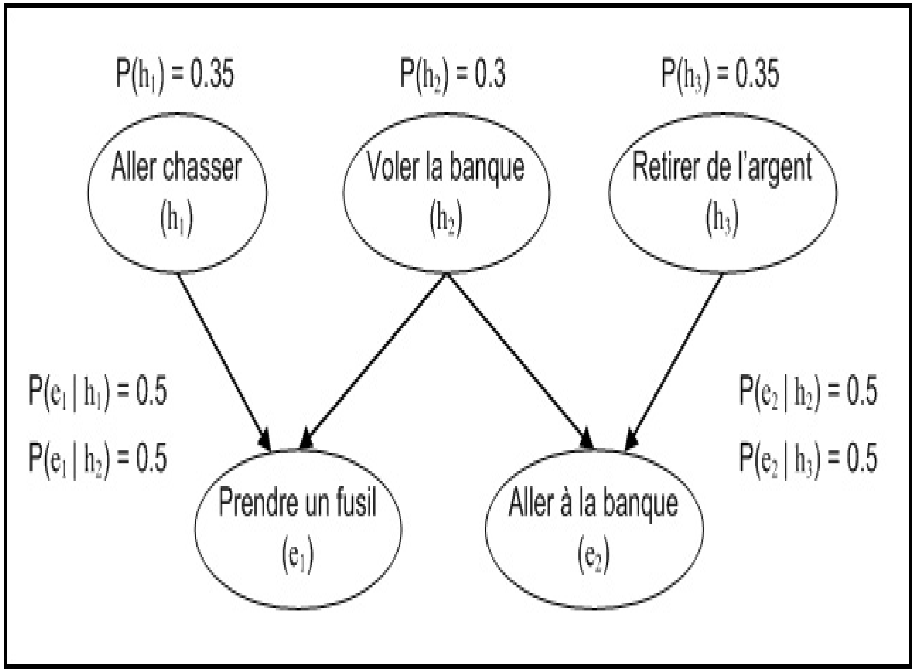
$$P(h_i \mid e_1 \wedge \ldots \wedge e_m) = \frac{P(e_1 \mid h_i) * P(e_m \mid h_i) * P(h_i)}{\sum_{j=[1,n]} P(e_1 \mid h_j) * \ldots * P(e_m \mid h_j) * P(h_j)}$$

- Où *m* correspond au nombre d'observations recueillies jusqu'à maintenant
- Et *n* au nombre d'hypothèses de haut niveau du réseau
- En d'autres termes, on cherche le plan h_i possédant la **probabilité maximale** $max(P(h_i))$, pour $i = 1$ à n.



Exemple applicatif de l'approche bayésienne

- Dans cette librairie, trois hypothèses de plans possibles sont en compétition



- Soit les plans *Aller chasser*, *Voler la banque* et *Retirer de l'argent*, représentés respectivement par h_1, h_2 *et* h_3.

Exemple applicatif de l'approche bayésienne

- La probabilité initiale attribuée aux plans *h1* et *h3* est légèrement **plus élevée** (0.35) que celle attribuée à *h2* (0.3)
- Ces deux plans sont légèrement priorisés au départ par rapport à *h2*

- On peut supposer ici que cette priorité tient compte du fait que la personne observée a plus de chances d'effectuer une activité honnête qu'une activité criminelle

- On note que la **distribution probabiliste est uniforme**, et donc $P(h1) + P(h2) + P(h3) = 1$
- Elle devra demeurer uniforme après chaque cycle d'inférence

Exemple applicatif de l'approche bayésienne

- L'agent observe l'action e_1
- Le système devra effectuer la révision de la probabilité de chacune des hypothèses, suivant le modèle de raisonnement bayésien
- Il doit d'abord estimer la probabilité révisée de **la première hypothèse h_1**

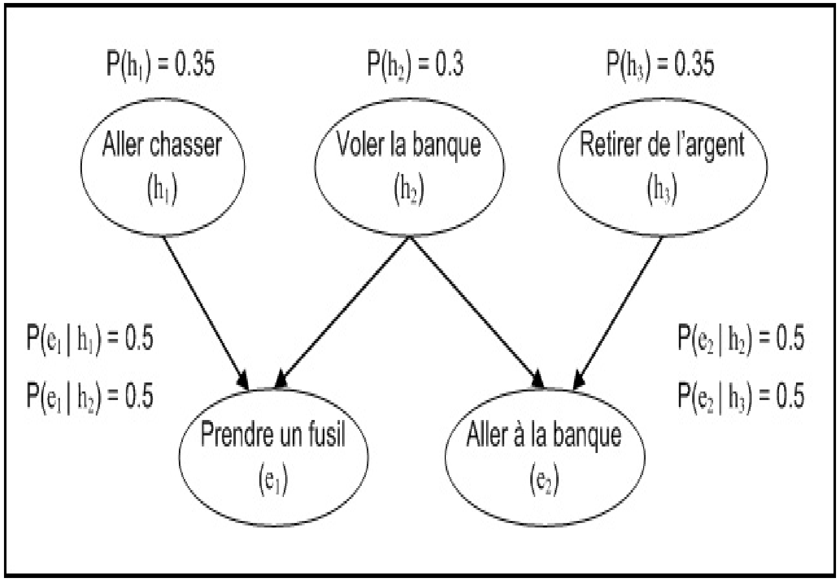
$$P(h_1 \mid e_1) = \frac{P(e_1 \mid h_1) * P(h_1)}{(P(e_1 \mid h_1) * P(h_1)) + (P(e_1 \mid h_2) * P(h_2)) + (P(e_1 \mid h_3) * P(h_3))}$$

$$P(h_1 \mid e_1) = \frac{0.5 * 0.35}{(0.5 * 0.35) + (0.5 * 0.3) + (0 * 0.35)}$$

$$P(h_1 \mid e_1) = \frac{0.175}{0.325}$$

$$P(h_1 \mid e_1) = 0.54$$

Nouvelle probabilité de h_1
Après le cycle d'inférence.



Exemple applicatif de l'approche bayésienne

- Maintenant, calculons la nouvelle probabilité de **l'hypothèse h_2**
- L'estimation de $P(h_2 \mid e_1) = (0.5 * 0.3) / ((0.5 * 0.35) + (0.5 * 0.3) + (0 * 0.35))$ correspond à 0.46, soit 46%

- Cette probabilité est **inférieure** à celle attribuée à la première hypothèse h_1

- **Pourtant, la probabilité conditionnelle de l'observation par rapport aux hypothèses h_1 et h_2 est équivalente**, soit 0.5 dans les deux cas

- La différence entre les deux nouvelles probabilités estimées est alors attribuable à la **probabilité initiale** des deux hypothèses
  - Cependant, on note que l'observation e_1 renforce néanmoins la crédibilité de l'hypothèse h_2

Exemple applicatif de l'approche bayésienne

- Finalement, la probabilité de **l'hypothèse h_3**, soit $P(h_3 \mid e_1) = (0 * 0.3) / ((0.5 * 0.35) + (0.5 * 0.3) + (0 * 0.35))$ est égale à 0

- La raison est qu'il n'existe aucune relation causale entre l'observation e_1 et l'hypothèse h_3

- Ça signifie que cette observation ne peut pas avoir été causée par la réalisation du plan h_3

- Le résultat du processus de reconnaissance correspond donc à l'hypothèse la plus probable h_1
  - soit le plan ***Aller chasser***.

Exemple applicatif de l'approche bayésienne

2ième observation : révision de l'hypothèse *h_1*

- Supposons maintenant l'arrivée d'une deuxième observation e_2
  - Soit l'action de bas niveau **Aller à la banque**

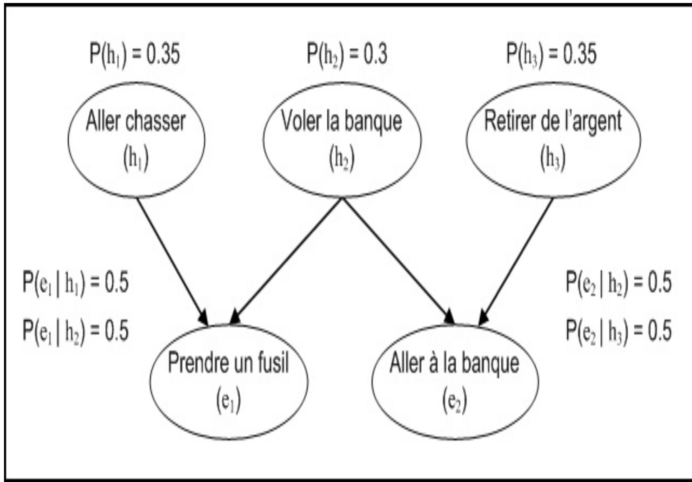
- Le processus de reconnaissance reprend depuis le début
- On estime la probabilité de chacune des hypothèses en fonction des deux observations conjointes e_1 et e_2
- Pour la première hypothèse :

$$P(h_1 \mid e_1 \wedge e_2) = \frac{P(e_1 \mid h_1) * P(e_2 \mid h_1) * P(h_1)}{\sum_{j=[1,3]}(P(e_1 \mid h_j) * P(e_2 \mid h_j) * P(h_j))}$$

$$P(h_1 \mid e_1 \wedge e_2) = \frac{0.5 * 0 * 0.35}{(0.5 * 0 * 0.35) + (0.5 * 0.5 * 0.3) + (0 * 0.5 * 0.35)}$$

$$P(h_1 \mid e_1 \wedge e_2) = \frac{0}{0.075}$$

$$P(h_1 \mid e_1 \wedge e_2) = 0$$



Exemple applicatif de l'approche bayésienne

- La probabilité de **l'hypothèse h_1** est réévaluée à 0

- La raison est qu'on suppose que les observations sont reliées à un seul et même plan

- Il n'existe aucune relation de causalité entre l'hypothèse h_1 et l'observation e_2

- Ce plan ne peut donc pas être à la fois la cause de l'observation e_1 et e_2

Exemple applicatif de l'approche bayésienne

- Le calcul de la probabilité de l'hypothèse h_3 donnera alors le même résultat
- Celle-ci ne permet pas **d'expliquer la première observation e_1**
- Calculons alors la nouvelle probabilité estimée de l'**hypothèse h_2**
- Celle-ci devrait logiquement récolter 100% de la distribution probabiliste

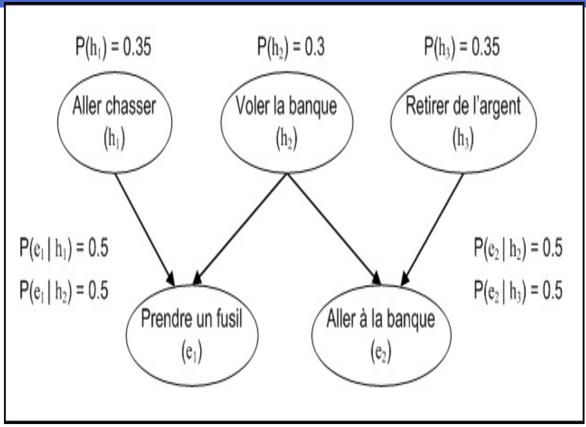
Exemple applicatif de l'approche bayésienne

$$P(h_2 \mid e_1 \wedge e_2) = \frac{P(e_1 \mid h_2) * P(e_2 \mid h_2) * P(h_2)}{\sum_{j=[1,3]}(P(e_1 \mid h_j) * P(e_2 \mid h_j) * P(h_j))}$$

$$P(h_2 \mid e_1 \wedge e_2) = \frac{0.5 * 0.5 * 0.35}{(0.5 * 0 * 0.35) + (0.5 * 0.5 * 0.3) + (0 * 0.5 * 0.35)}$$

$$P(h_2 \mid e_1 \wedge e_2) = \frac{0.075}{0.075}$$

$$P(h_2 \mid e_1 \wedge e_2) = 1$$



- La seule hypothèse pouvant expliquer à la fois l'observation e_1 et e_2 est h_2
- On note que cette distribution est toujours uniforme *$(P(h_1) = 0) + (P(h_2) = 1) + (P(h_3) = 0) = 1$*

- Le résultat du processus de reconnaissance serait le plan h_2 : ***Voler la banque***

Sommaire

- Les probabilités constituent un formalisme rigoureux pour le raisonnement incertain
- Les distributions de probabilités conjointes spécifient la probabilité de chaque évènement atomique
- Les requêtes peuvent être répondues par la somme des évènements atomiques
- Pour les domaines non triviaux, il faut réduire la taille de la distribution conjointe
- L'indépendance et l'indépendance conditionnelle servent d'outils à cette fin