IFT 615 – Intelligence artificielle

Réseaux bayésiens dynamiques

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

Sujets couverts

- Qu'est-ce qu'un réseau bayésien dynamique (RBD)?
- Types d'inférence dans un RBD
- Cas particulier des modèles de Markov cachés

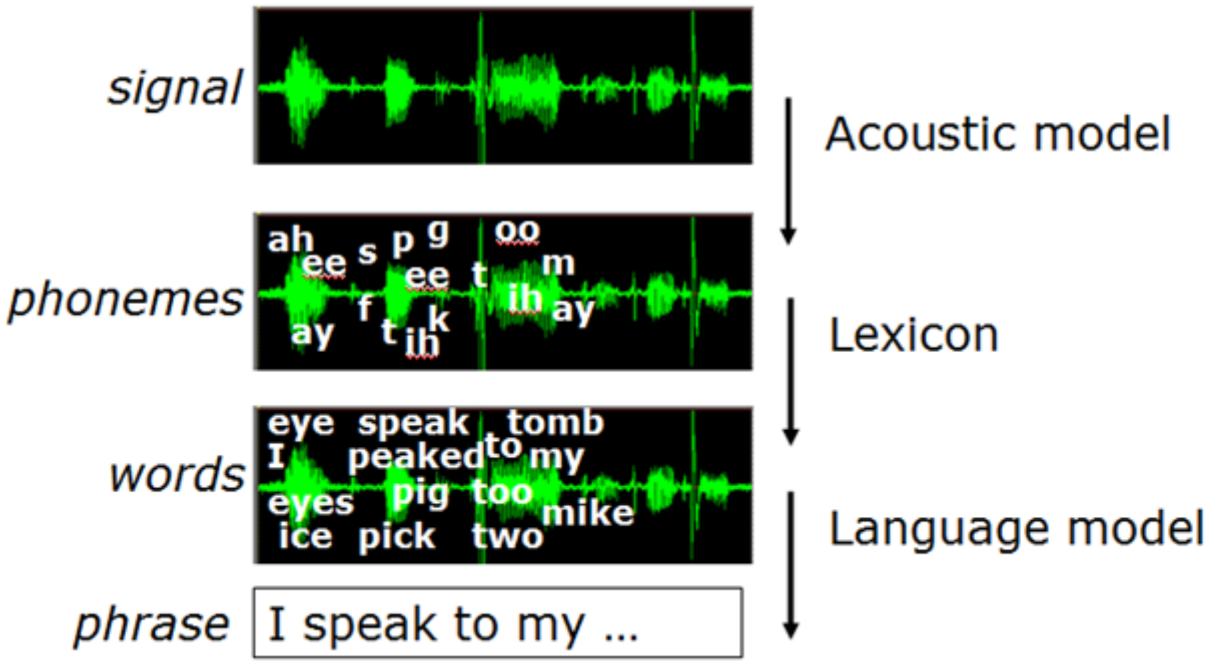
Réseaux bayésiens dynamiques (RBD)

- Comment modéliser des situations dynamiques?
  - les changements dynamiques peuvent être vus **comme une séquence d'états**, chaque état représentant la situation à un instant *t* donné
  - X_t: ensemble des **variables non observables** (**cachées**) décrivant l'état au temps *t*
  - E_t: ensembles de **variables observées** (*evidence*) au temps *t*

- Le terme dynamique réfère au dynamisme du système qu'on veut modéliser et la structure du réseau

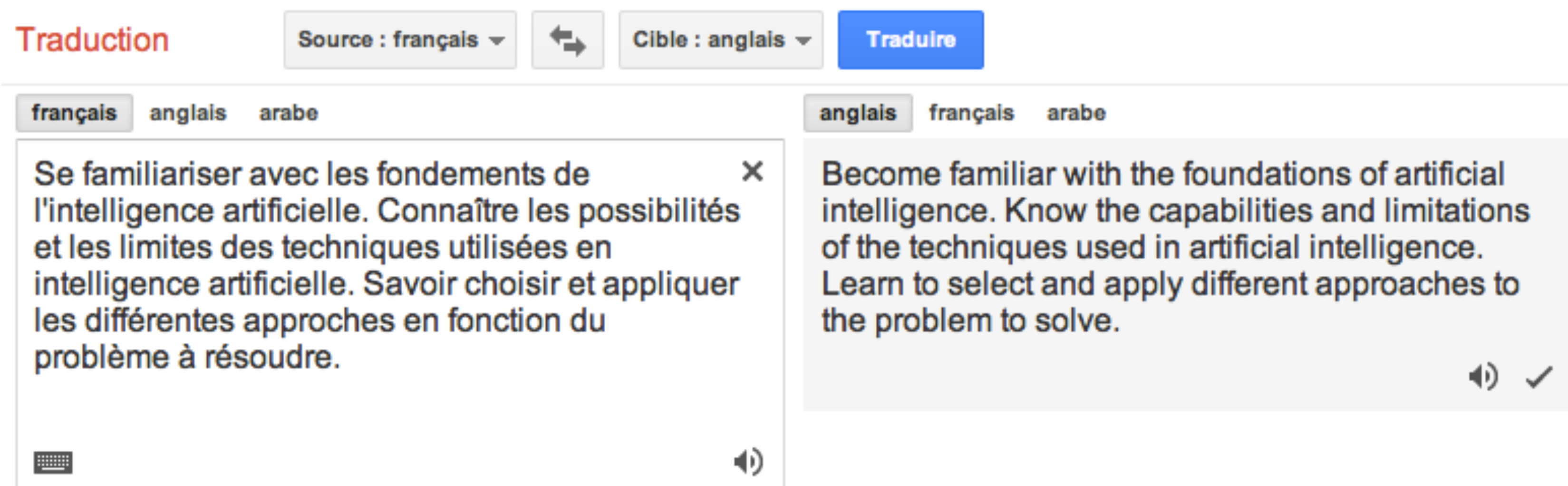
Applications

- **Reconnaissance de la parole**
  - E_t sont les éléments du signal sonore
  - X_t sont les mots prononcés



Applications

- **Traduction automatique**
  - E_t sont les mots en français
  - X_t sont les mots de la traduction en anglais



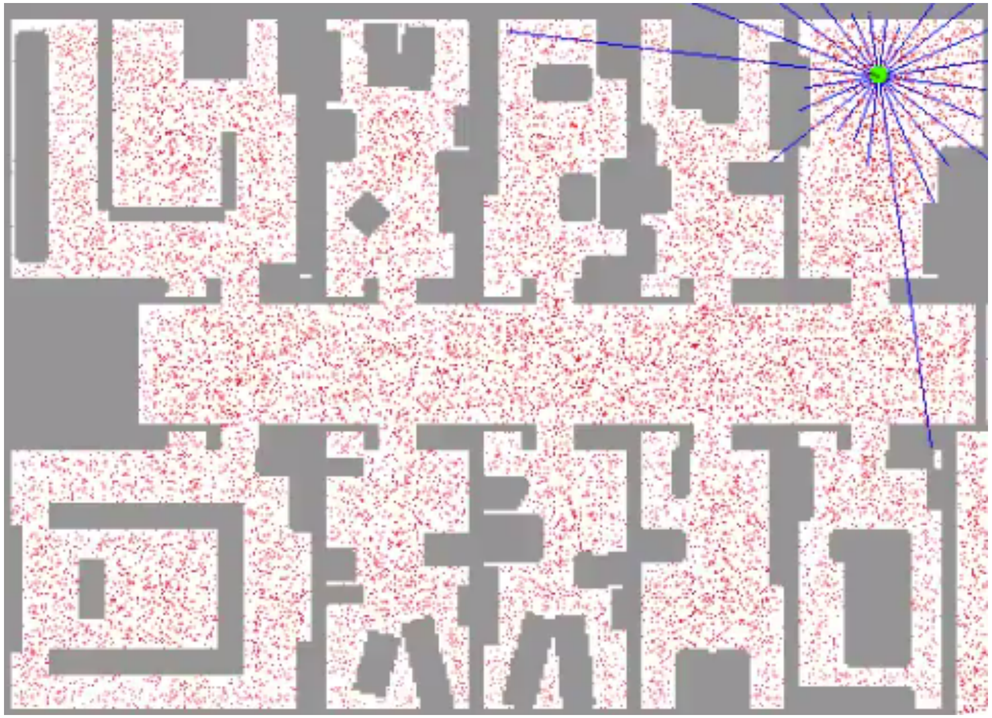
Applications

- **Suivi d'objets (*tracking*)**
  - E_t sont les *frames* de la vidéoe
  - X_t sont l'information sur la position d'un/des objet(s)

http://www.youtube.com/watch?v=fRowYlxKt7s

Applications

- **Localisation de robots**
  - E_t sont l'information fournie par les capteurs du robot
  - X_t sont l'information sur la position du robot

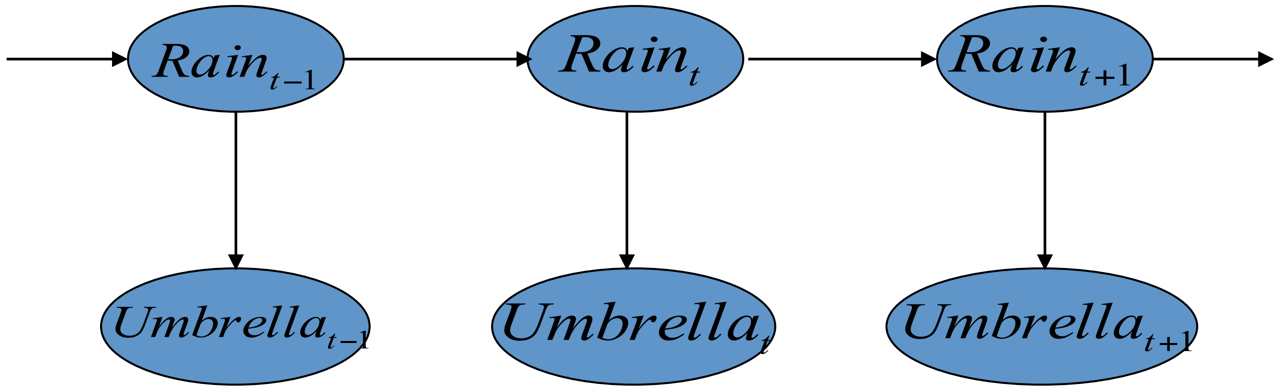


Représentation dans un RBD

- Problème:
  - il faudrait spécifier un grand nombre (même infini) de tables de probabilités conditionnelles, c.-à-d. une pour chaque temps *t*
  - chaque table pourrait impliquer un nombre infini de parents
- Solution:
  1. supposer que les changements dynamiques sont causés par un **processus stationnaire** - les probabilités ne changent pas dans le temps: $\mathbf{P}(X_t \mid Parent(X_t))$ **est la même** pour tous les *t*
  2. supposer que les changements dynamiques sont causés par un **processus markovien** – l'état courant dépend seulement d'un nombre fini d'états précédents
     - » ex.: processus markoviens **du premier ordre**:
       - $\mathbf{P}(X_t \mid X_{0:t-1}) = \mathbf{P}(X_t \mid X_{t-1})$ modèle pour les transitions
  3. supposer que l'observation **dépend uniquement de l'état courant**
     - $\mathbf{P}(E_t \mid X_{0:t}, E_{0:t-1}) = \mathbf{P}(E_t \mid X_t)$ modèle pour les observations/capteurs

Exemple

- « Un gardien de sécurité passe un mois dans un édifice sous-terrain, sans sortir. Chaque jour, son directeur arrive avec ou sans parapluie. Le gardien veut inférer la possibilité qu'il ait plu ou non en fonction des séquences d' observation du parapluie. »
- Modélisation:
  - Variables: $X_t = \{R_t\}$ (pour « *Rain* ») et $E_t = \{U_t\}$ (pour « *Umbrella* »).
  - Dépendances entre les variables (c-.à-d., le RBD):



  - Modèle des transitions: $\mathbf{P}(R_t \mid R_{t-1})$. Modèle d'observation: $\mathbf{P}(U_t \mid R_t)$

Types d'inférence dans un RBD

- **Filtrage** (***filtering***): calcul de l'**état de croyance** (***belief state***), c.-à-d. la distribution a posteriori de la variable cachée la plus récente

$$\mathbf{P}(X_t \mid e_{1:t})$$

  - ex. : quelle est la probabilité qu'il pleuve aujourd'hui ?
  - ex. : quelle est la croyance du robot par rapport à sa position actuelle ?

- **Prédiction**: calculer la distribution a posteriori sur un état futur

$$\mathbf{P}(X_{t+k} \mid e_{1:t}) \text{ où } k > 0$$

  - ex. : quelle est la probabilité qu'il pleuve dans *k* jours ?

Types d'inférence dans un RBD

- **Lissage** (***smoothing***): calculer la distribution a posteriori sur un état passé

$$\mathbf{P}(X_k \mid e_{1:t}) \text{ où } 0 \leq k < t$$

  - ex. : quelle est la probabilité qu'il y ait eu de la pluie hier (k=t-1) ?

- **Explication la plus plausible**: trouver la séquence d'états cachés qui explique le mieux les observations

$$\underset{x_{1:t}}{\text{argmax}}\, P(x_{1:t} \mid e_{1:t}) = \underset{x_{1:t}}{\text{argmax}}\, P(x_{1:t}, e_{1:t}) \,/\, P(e_{1:t}) = \underset{x_{1:t}}{\text{argmax}}\, P(x_{1:t}, e_{1:t})$$

  - ex. : quelle a été la météo la plus probable pour toutes les t dernières journées ?
  - ex. : quelle est la traduction en anglais d'une phrase donnée en français ?
  - ex. : quelle est la phrase qui a été prononcée ?

Chaînes de Markov

- Une **chaîne de Markov** (de premier ordre) est un cas particulier de RBD
  - avec **une seule** variable aléatoire **discrète** S_t dans l'état au temps t

- Le domaine de S_t est souvent un symbole (ex.: un caractère, un mot, etc.)

- Une **distribution a priori** (initiale) de probabilités sur les symboles (états) est spécifiée $\mathbf{P}(S_1)$

- Une **matrice de transition** contenant les probabilités conditionnelles $\mathbf{P}(S_{t+1} \mid S_t)$

Illustration

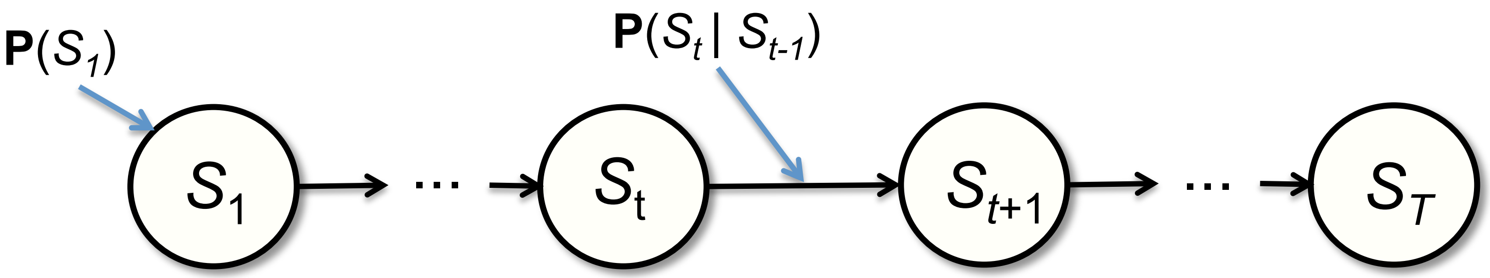


Illustration dans le cas d'une **chaîne finie**

Probabilité de générer une chaîne

produit des probabilités, une pour chaque terme de la séquence

$$P(S_{1:T}) = P(S_1) \prod_{t=2}^{T} P(S_t \mid S_{t-1})$$

distribution initiale de probabilités

une séquence de symboles, allant du temps 1 au temps *T*

modèle de transition

Visualisation d'une chaîne de Markov

Représentation matricielle

Symbole actuel

Prochain symbole	*a*	*b*	*c*
a	.7	.3	0
b	.2	.7	.5
c	.1	0	.5

Représentation graphique

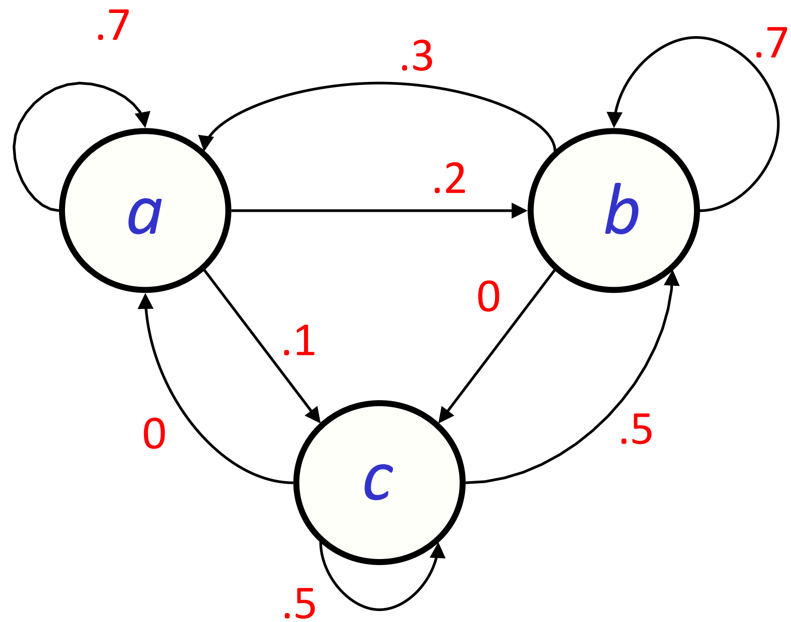


Illustration dans le cas d'une **chaîne infinie** (flux de symboles)

Exemple de chaîne: *ccbbbbaaaaabaabacbabaaa*

Modèle de Markov caché

- Dans une **modèle de Markov caché** (***hidden Markov model*** ou **HMM**):
  - il y a des **variables cachées** H_t et des **variables d'observation** S_t, toutes les deux discrètes
  - la chaîne de Markov est sur les variables cachées H_t
  - le symbole observé (émis) $S_t=s_t$ dépend uniquement de la variable cachée actuelle H_t

Illustration

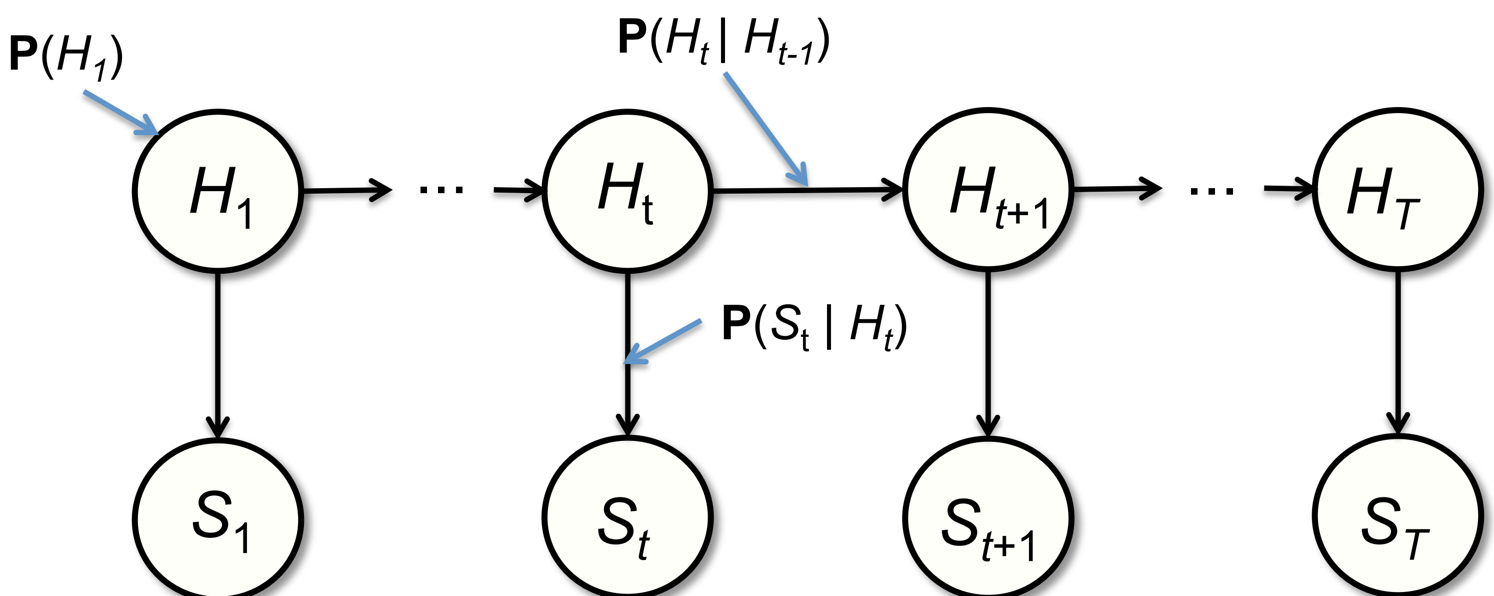


Illustration dans le cas d'une **chaîne finie**

Probabilité de générer une séquence cachée et une séquence visible

distribution initiale
de probabilités

modèle de
transition

$$P(S_{1:T}, H_{1:T}) = P(H_1)\, P(S_1 | H_1) \prod_{t=2}^{T} P(H_t \mid H_{t-1})\, P(S_t \mid H_t)$$

séquence de nœuds cachés et de symboles de sortie

probabilité d'observer le symbole S_t à partir de H_t

Simuler d'un HMM

- Il est facile de générer des observations d'un HMM
  - échantillonner une valeur initiale $H_1 = h_1$ de $\mathbf{P}(H_1)$
  - pour t = 2 jusqu'à T, répéter les deux échantillonnage suivants:
    - » utiliser les probabilités de transition de l'état caché courant pour obtenir un échantillon h_t, sachant l'état caché précédent: $\mathbf{P}(H_t \mid H_{t-1} = h_{t-1})$
    - » utiliser les probabilités de sortie de la variable d'observation étant donné l'état caché courant, pour obtenir le symbole d'observation (émission) s_t: $\mathbf{P}(S_t \mid H_t=h_t)$

- On peut aussi générer la séquence des états cachés d'abord et ensuite générer les observations
  - les variables cachées dépendent uniquement des variables cachées précédentes
  - chaque observation (émission) ne dépendra pas des autres

Illustration

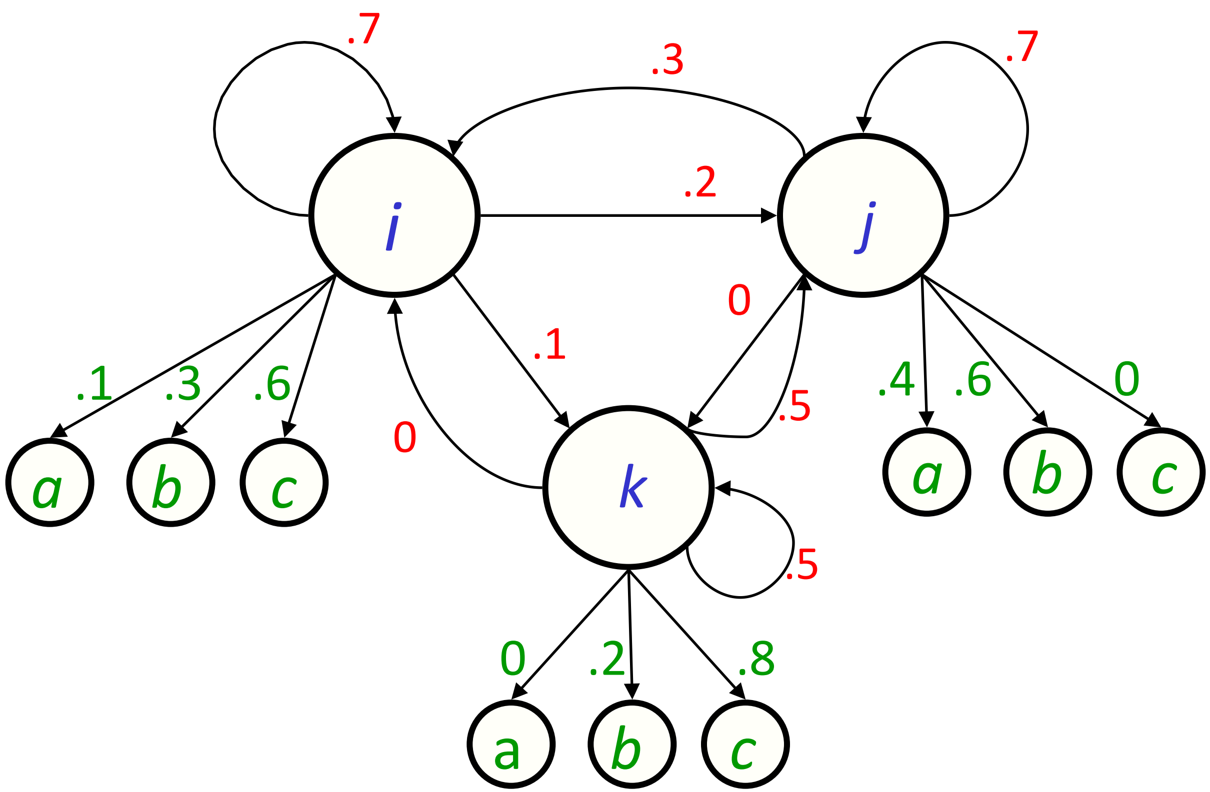


Illustration dans le cas d'une **chaîne infinie**, avec visualisation des valeurs de la variable cachée et la variable d'observation

Chaque **nœud caché** (valeur possible *h* de *H*) a un vecteur de **probabilités de transitions** et un **vecteur de probabilités d'émission (observations)**

Probabilité de générer une séquence visible

- La même séquence de sortie peut être produite par plusieurs séquences cachées différentes
- En fait, il y a un nombre exponentiel de séquences cachées possibles
- Un calcul naïf est donc très inefficace

$$P(S_{1:T}) = \sum_{h_{1:T}} P(H_{1:T} = h_{1:T})\, P(S_{1:T} \mid H_{1:T} = h_{1:T})$$

Programmation dynamique pour HMM

- Une façon plus efficace de calculer la probabilité d'une séquence observée $s_{1:T}$
- Idée: utiliser la **programmation dynamique**
  - on définit $\alpha(i,t) = P(S_{1:t}=s_{1:t}, H_t = i)$
  - on note la récursion

    $$\begin{aligned}\alpha(i,t+1) &= P(S_{1:t+1}=s_{1:t+1}, \mathrm{H}_{t+1} = i) \\ &= \textstyle\sum_j P(S_{1:t+1}=s_{1:t+1}, H_t = j, H_{t+1} = i) \\ &= P(S_{t+1}= s_{t+1} | H_{t+1} = i) \textstyle\sum_j P(\underline{H}_{t+1} = i | H_t = j)\, P(S_{1:t}=s_{1:t}, H_t = j) \\ &= P(S_{t+1}= s_{t+1} | H_{t+1} = i) \textstyle\sum_j P(H_{t+1} = i | H_t = j)\, \alpha(j,t)\end{aligned}$$
  - on a les valeurs initiales

    $$\alpha(i,1) = P(S_1=s_1, \mathrm{H}_1 = i) = P(S_1=s_1 | \mathrm{H}_1 = i)\, P(\mathrm{H}_1 = i) \quad \forall i$$
- Une fois le tableau α calculé, on obtient facilement:

  $$P(S_{1:T}=s_{1:T}) = \textstyle\sum_j P(S_{1:T}=s_{1:T}, \mathrm{H}_T = j) = \sum_j \alpha(j,T)$$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0				
1				

  - initialisation: $\alpha(i,1) = P(S_1{=}s_1, \mathrm{H}_1 = i) = P(S_1{=}s_1 \mid \mathrm{H}_1 = i)\, P(\mathrm{H}_1 = i)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45			
1				

  - initialisation: $\alpha(0,1) = P(S_1=0 \mid H_1 = 0)\, P(H_1 = 0) = 0.9 \times 0.5 = 0.45$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45			
1	0.1			

  - initialisation: $\alpha(1,1) = P(S_1=0 \mid H_1 = 1)\ P(H_1 = 1) = 0.2 \times 0.5 = 0.1$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45			
1	0.1			

  - récursion (t=1): $\alpha(i,t+1) = P(S_{t+1}= s_{t+1} \mid H_{t+1} = i) \sum_j P(H_{t+1} = i \mid H_t = j)\ \alpha(j,t)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45			
1	0.1			

  - récursion: $\alpha(0,2) = P(S_2 = 1 \mid H_2 = 0)\ (\ P(H_2 = 0 \mid H_1 = 0)\ \alpha(0,1) + P(H_2 = 0 \mid H_1 = 1)\ \alpha(1,1))$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45	0.0175		
1	0.1			

  - récursion: $\alpha(0,2) = 0.1\ (0.3 \times 0.45 + 0.4 \times 0.1) = 0.0175$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45	0.0175		
1	0.1			

  - récursion: $\alpha(1,2) = P(S_2 = 1 \mid H_2 = 1)\ (\ P(H_2 = 1 \mid H_1 = 0)\ \alpha(0,1) + P(H_2 = 1 \mid H_1 = 1)\ \alpha(1,1))$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45	0.0175		
1	0.1	0.3		

- récursion: $\alpha(1,2) = 0.8\ (0.7 \times 0.45 + 0.6 \times 0.1) = 0.3$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45	0.0175		
1	0.1	0.3		

- récursion (t=2): $\alpha(i,t+1) = P(S_{t+1} = s_{t+1} \mid H_{t+1} = i) \sum_j P(H_{t+1} = i \mid H_t = j)\ \alpha(j,t)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45	0.0175	0.112725	
1	0.1	0.3		

- récursion: $\alpha(0,3) = 0.9\ (0.3 \times 0.0175 + 0.4 \times 0.3) = 0.112725$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	1	2	3	4
0	0.45	0.0175	0.112725	0.04427
1	0.1	0.3	0.03845	0.02039

- on continue d'appliquer la récursion jusqu'à la fin (t=4)...

Filtrage dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

α(i,t) — i \ t	1	2	3	4
0	0.45	0.0175	0.112725	0.04427
1	0.1	0.3	0.03845	0.02039

  - on peut calculer les probabilités de filtrage

$$P(H_4 = 0 \mid S_1=0, S_2=1, S_3=0, S_4=0) = \frac{P(H_4 = 0, S_1=0, S_2=1, S_3=0, S_4=0)}{\sum_i P(H_4 = \mathrm{i}, S_1=0, S_2=1, S_3=0, S_4=0)}$$

$$= \alpha(0,4) \,/\, (\, \alpha(0,4) + \alpha(1,4)\,)$$

$$= 0.04427 \,/\, (0.04427 + 0.02039)$$

$$\approx 0.6847$$

$$P(H_4 = 1 \mid S_1=0, S_2=1, S_3=0, S_4=0) = 0.02039 \,/\, (0.04427 + 0.02039)$$

$$\approx 0.3153$$

Programmation dynamique pour HMM

- Le calcul des $\alpha(i,t)$ donne un balayage de gauche à droite
- On peut faire la même chose, mais de droite à gauche
  - on définit $\beta(i,t) = P(S_{t+1:T}=s_{t+1:T} \mid H_t = i)$
  - on note la récursion

    $$\begin{aligned}\beta(i,t-1) &= P(S_{t:T}=s_{t:T} \mid \mathrm{H}_{t-1} = i) \\ &= \sum_j P(S_{t:T}=s_{t:T}, \mathrm{H}_t = j \mid H_{t-1} = i) \\ &= \sum_j P(S_t = s_t \mid H_t = j)\, P(\mathrm{H}_t = j \mid H_{t-1} = i)\, P(S_{t+1:T}=s_{t+1:T} \mid H_t = j) \\ &= \sum_j P(S_t = s_t \mid H_t = j)\, P(\mathrm{H}_t = j \mid H_{t-1} = i)\, \beta(j,t)\end{aligned}$$

  - on a les valeurs initiales $\beta(i,T) = 1\ \forall i$

- Une fois le tableau β calculé, on obtient facilement:

$$\begin{aligned}P(S_{1:T}=s_{1:T}) &= \sum_j P(S_{1:T}=s_{1:T}, \mathrm{H}_1 = j) \\ &= \sum_j P(S_{2:T}=s_{2:T} \mid \mathrm{H}_1 = j)\, P(S_1=s_1 \mid H_1 = j)\, P(H_1 = j) \\ &= \sum_j \beta(j,1)\, P(S_1=s_1 \mid H_1 = j)\, P(H_1 = j)\end{aligned}$$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0				
1				

  - initialisation: β(i,4) = 1

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0				1
1				1

  - initialisation: β(i,4) = 1

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0				1
1				1

  - récursion (t=4): $\beta(i,t\text{-}1) = \sum_j P(S_t = s_t \mid H_t = j)\ P(H_t = j \mid H_{t-1} = i)\ \beta(j,t)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0				1
1				1

  - récursion β(0,3) = $P(S_4=0|H_4=0)\ P(H_4=0|H_3=0)$ β(0,4)
    + $P(S_4=0|H_4=1)\ P(H_4=1|H_3=0)$ β(1,4)

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.41	1
1				1

- récursion β(0,3) = 0.9 x 0.3 x 1 + 0.2 x 0.7 x 1 = 0.41

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.41	1
1				1

  - récursion β(1,3) = $P(S_4=0|H_4=0)\ P(H_4=0|H_3=1)\ \beta(0,4)$
    $+\ P(S_4=0|H_4=1)\ P(H_4=1|H_3=1)\ \beta(1,4)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.41	1
1			0.48	1

- récursion β(1,3) = 0.9 x 0.4 x 1 + 0.2 x 0.6 x 1 = 0.48

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t$=0 | $H_t)$	0.9	0.2
$P(S_t$=1 | $H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t$=0 | $H_{t-1})$	0.3	0.4
$P(H_t$=1 | $H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.41	1
1			0.48	1

- récursion (t=3): β(i,t-1) = $\sum_j P(S_t = s_t | H_t = j)\ P(H_t = j | H_{t-1} = i)$ β(j,t)

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0			0.41	1
1			0.48	1

- récursion $\beta(0,2) = P(S_3=0|H_3=0)\, P(H_3=0|H_2=0)\, \beta(0,3) + P(S_3=0|H_3=1)\, P(H_3=1|H_2=0)\, \beta(1,3)$

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0		0.1779	0.41	1
1			0.48	1

  - récursion β(0,2) = 0.9 x 0.3 x 0.41 + 0.2 x 0.7 x 0.48 = 0.1779

Programmation dynamique pour HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

β(i,t)

i \ t	1	2	3	4
0	0.120249	0.1779	0.41	1
1	0.105612	0.2052	0.48	1

  - on continue d'appliquer la récursion jusqu'au début (t=1)...

Lissage avec un HMM

- Les tables α(i,t) et β(i,t) peuvent également être utilisées pour faire du lissage
  - $P(H_k = i \mid S_{1:T}=s_{1:T}) = P(H_k = i, S_{1:k}=s_{1:k}, S_{k+1:T}=s_{k+1:T}) / \Upsilon$ (ϒ est la normalisation)
    $= P(H_k = i, S_{1:k}=s_{1:k})\, P(S_{k+1:T}=s_{k+1:T} \mid H_k = i) / \Upsilon$
    $= \alpha(i,k)\, \beta(i,k) / \Upsilon$

- On peut également faire du lissage sur deux variables cachées adjacentes
  - $P(H_k = i, H_{k+1}= j \mid S_{1:T}=s_{1:T}) = P(H_k = i, H_{k+1}= j, S_{1:k}=s_{1:k}, S_{k+1:T}=s_{k+1:T}) / \Upsilon'$
    $= P(H_k = i, S_{1:k}=s_{1:k})\, P(H_{k+1}= j \mid H_k = i)\, P(S_{k+1}=s_{k+1} \mid H_{k+1} = j)$
    $P(S_{k+2:T}=s_{k+2:T} \mid H_{k+1} = j) / \Upsilon'$
    $= \alpha(i,k)\, \beta(j,k+1)\, P(H_{k+1}= j \mid H_k = i)\, P(S_{k+1}=s_{k+1} \mid H_{k+1} = j) / \Upsilon'$
- À noter que ϒ correspond à une somme sur *i* seulement,
  tandis que ϒ’ est une somme sur *i* **et** *j*

Lissage avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

α(i,t)

i \ t	...	2	...
0	...	0.0175	...
1	...	0.3	...

β(i,t)

i \ t	...	2	...
0	...	0.1779	...
1	...	0.2052	...

  - on peut calculer les probabilités de lissage au temps t=2

$$P(H_2 = 0 \mid S_1{=}0, S_2{=}1, S_3{=}0, S_4{=}0) = \frac{\alpha(0,2)\ \beta(0,2)}{\sum_i \alpha(i,2)\ \beta(i,2)}$$

$$= \alpha(0,2)\ \beta(0,2) / \left(\alpha(0,2)\ \beta(0,2) + \alpha(1,2)\ \beta(1,2)\ \right)$$

$$= 0.0175 \times 0.1779 \ / \ (0.0175 \times 0.1779 + 0.3 \times 0.2052)$$

$$\approx 0.04813$$

$$P(H_2 = 1 \mid S_1{=}0, S_2{=}1, S_3{=}0, S_4{=}0)$$

$$= 0.3 \times 0.2052 \ / \ (0.0175 \times 0.1779 + 0.3 \times 0.2052)$$

$$\approx 0.95186$$

Prédiction dans un HMM

- $\alpha(i,t)$ peut être utilisé pour inférer la distribution de prédiction $\mathbf{P}(H_{t+k} | s_{1:t})$
- On utilise également un programme dynamique
  - on définit $\pi(i,k) = P(H_{t+k} = i | S_{1:t}=s_{1:t})$
  - on note la récursion

    $$\begin{aligned}\pi(i,k+1) &= P(H_{t+k+1} = i | S_{1:t}=s_{1:t}) \\ &= \textstyle\sum_s \sum_j P(H_{t+k+1} = i, H_{t+k} = j, S_{t+k}=s | S_{1:t}=s_{1:t}) \\ &= \textstyle\sum_s \sum_j P(S_{t+k}= s | H_{t+k} = j)\, P(H_{t+k+1} = i | H_{t+k} = j)\, P(H_{t+k} = j | S_{1:t}=s_{1:t}) \\ &= \textstyle\sum_j P(H_{t+k+1} = i | H_{t+k} = j)\, P(H_{t+k} = j | S_{1:t}=s_{1:t}) \sum_s P(S_{t+k}= s | H_{t+k} = j) \\ &= \textstyle\sum_j P(H_{t+k+1} = i | H_{t+k} = j)\, \pi(j,k)\end{aligned}$$
  - on a les valeurs initiales

    $$\pi(i,0) = P(H_t = i | s_{1:t}) = \alpha(i,t) / \textstyle\sum_j \alpha(j,t) \quad \forall i$$
- On pourrait également faire une prédiction de S_{t+k}
  - $$\begin{aligned}P(S_{t+k} = s | S_{1:t}=s_{1:t}) &= \textstyle\sum_j P(S_{t+k}= s | H_{t+k} = j)\, P(H_{t+k} = j | S_{1:t}=s_{1:t}) \\ &= \textstyle\sum_j P(S_{t+k}= s | H_{t+k} = j)\, \pi(j,k)\end{aligned}$$

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0, S_2=1, S_3=0, S_4=0)$

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0			
1			

- initialisation: $\pi(i,0) = \alpha(i,t) / \sum_j \alpha(j,t)$

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0,S_2=1,S_3=0,S_4=0)$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0			
1			

  - initialisation: π(0,0) = α(0,4) / (α(0,4) + α(1,4))

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0,S_2=1,S_3=0,S_4=0)$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466		
1			

- initialisation: π(0,0) = 0.04427 / (0.04427 + 0.02039) = 0.68466

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0, S_2=1, S_3=0, S_4=0)$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466		
1	0.31534		

- initialisation: $\pi(1,0) = 0.02039 / (0.04427 + 0.02039) = 0.31534$

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0, S_2=1, S_3=0, S_4=0)$

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466		
1	0.31534		

- récursion (k=0): $\pi(i,k+1) = \sum_j P(H_{t+k+1} = i \mid H_{t+k} = j)\, \pi(j,k)$

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0, S_2=1, S_3=0, S_4=0)$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466		
1	0.31534		

  - récursion (k=0): $\pi(0, 1) = P(H_5 = 0 \mid H_4 = 0)\ \pi(0,0) + P(H_5 = 0 \mid H_4 = 1)\ \pi(1,0)$

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0,S_2=1,S_3=0,S_4=0)$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466	0.33154	
1	0.31534		

  - récursion (k=0): π(0, 1) = 0.3 x 0.68466 + 0.4 x 0.31534 = 0.33154

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0,S_2=1,S_3=0,S_4=0)$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466	0.33154	
1	0.31534		

- récursion (k=0): $\pi(1, 1) = P(H_5 = 1 | H_4 = 0)\ \pi(0,0) + P(H_5 = 1 | H_4 = 1)\ \pi(1,0)$

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6=0 \mid S_1=0, S_2=1, S_3=0, S_4=0)$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466	0.33154	
1	0.31534	0.66846	

  - récursion (k=0): π(0, 1) = 0.7 x 0.68466 + 0.6 x 0.31534 = 0.66846

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6{=}0 \mid S_1{=}0, S_2{=}1, S_3{=}0, S_4{=}0)$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466	0.33154	0.36685
1	0.31534	0.66846	0.63315

- on continue d'appliquer la récursion jusqu'à la fin (k=2)...

Prédiction dans un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - on souhaite calculer la distribution de prédiction $P(H_6\text{=}0 \mid S_1\text{=}0, S_2\text{=}1, S_3\text{=}0, S_4\text{=}0)$

Modèle d'observation

	H_t=0	H_t=1
$P(S_t\text{=}0 \mid H_t)$	0.9	0.2
$P(S_t\text{=}1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t\text{=}0 \mid H_{t-1})$	0.3	0.4
$P(H_t\text{=}1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α(i,t)

i \ t	...	4
0	...	0.04427
1	...	0.02039

π(i,k)

i \ k	0	1	2
0	0.68466	0.33154	0.36685
1	0.31534	0.66846	0.63315

  - $P(H_6\text{=}0 \mid S_1\text{=}0, S_2\text{=}1, S_3\text{=}0, S_4\text{=}0) = \pi(0,2) = 0.36685$

Explication la plus plausible avec un HMM

- On peut également éviter une énumération exponentielle
  - exemple avec T=3
    $$\max_{h^*_{1:3}} P(h^*_1)\, P(s_1 | h^*_1)\, P(h^*_2 | h^*_1)\, P(s_2 | h^*_2)\, P(h^*_3 | h^*_2)\, P(s_3 | h^*_3)$$
    $$= \max_{h^*_3} P(s_3 | h^*_3) \max_{h^*_2} P(s_2 | h^*_2)\, P(h^*_3 | h^*_2) \max_{h^*_1} P(h^*_2 | h^*_1)\, P(h^*_1)\, P(s_1 | h^*_1)$$
- Solution: **programmation dynamique**, avec un **max au lieu de la somme**
  - on définit $\alpha^*(\mathrm{i,t}) = P(S_{1:t}=s_{1:t}, H_{1:t-1} = h^*_{1:t-1}, H_t = i)$
  - on note la récursion
    $$\alpha^*(\mathrm{i,t+1}) = \max_j P(S_{1:t+1}=s_{1:t+1}, H_{1:t-1} = h^*_{1:t-1}, H_t = j, H_{t+1} = i)$$
    $$= \max_j P(S_{t+1}=s_{t+1} \mid H_{t+1} = i)\, P(H_{t+1} = i \mid H_t = j)\, P(S_{1:t}=s_{1:t}, H_{1:t-1} = h^*_{1:t-1}, H_t = j)$$
    $$= P(S_{t+1}=s_{t+1} \mid H_{t+1} = i) \max_j P(H_{t+1} = i \mid H_t = j)\, \alpha^*(\mathrm{j,t})$$
  - on a les valeurs initiales: $\alpha^*(\mathrm{i,1}) = P(S_1=s_1 | H_1 = i)\, P(H_1 = i) \quad \forall i$
- On a alors que $P(S_{1:T}=s_{1:T}, H_{1:T} = h^*_{1:T}) = \max_j \alpha^*(\mathrm{j,T})$
- On retrouve $h^*_{1:T}$ à partir de tous les argmax_j

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0				
1				

  - initialisation: $\alpha^*(i,1) = P(S_1=s_1, H_1 = i) = P(S_1=s_1 \mid H_1 = i)\, P(H_1 = i)$

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45			
1				

  - initialisation: α*(0,1) = $P(S_1=0 \mid H_1 = 0)\ P(H_1 = 0)$ = 0.9 x 0.5 = 0.45

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45			
1	0.1			

  - initialisation: $\alpha^*(1,1) = P(S_1=0 \mid H_1 = 1)\, P(H_1 = 1) = 0.2 \times 0.5 = 0.1$

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t\text{=0} \mid H_t)$	0.9	0.2
$P(S_t\text{=1} \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t\text{=0} \mid H_{t-1})$	0.3	0.4
$P(H_t\text{=1} \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45			
1	0.1			

- récursion (t=1): α*(i,t+1) = $P(S_{t+1}\text{=}s_{t+1} \mid H_{t+1} = i)$ $\max_j P(H_{t+1} = i \mid H_t = j)$ α*(j,t)

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45			
1	0.1			

  - récursion: $\alpha^*(0,2) = P(S_2{=}1 \mid H_2{=}0) \max\{P(H_2 = 0 \mid H_1 = 0)\ \alpha^*(0,1),\ P(H_2 = 0 \mid H_1 = 1)\ \alpha^*(1,1)\}$

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45	0.0135		
1	0.1			

  - récursion: $\alpha^*(0,2) = 0.1 \max\{ \underline{0.3 \times 0.45}, 0.4 \times 0.1\} = 0.0135$

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t$=0 | $H_t)$	0.9	0.2
$P(S_t$=1 | $H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t$=0 | $H_{t-1})$	0.3	0.4
$P(H_t$=1 | $H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45	0.0135		
1	0.1			

  - récursion: α*(1,2) = $P(S_2$=1|H_2=1) max{$P(H_2$ = 1|H_1 = 0) α*(0,1), $P(H_2$ = 1|H_1 = 1) α*(1,1)}

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

$\alpha^*(i,t)$

i \ t	1	2	3	4
0	0.45	0.0135		
1	0.1	0.252		

  - récursion: $\alpha^*(1,2) = 0.8 \max\{ \underline{0.7 \times 0.45}, 0.6 \times 0.1\} = 0.252$

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45	0.0135		
1	0.1	0.252		

  - récursion (t=2): $\alpha^*(i,t+1) = P(S_{t+1}=s_{t+1} \mid H_{t+1} = i) \max_j P(H_{t+1} = i \mid H_t = j)\, \alpha^*(j,t)$

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: $S_1=0$, $S_2=1$, $S_3=0$, $S_4=0$

Modèle d'observation

	$H_t=0$	$H_t=1$
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	$H_{t-1}=0$	$H_{t-1}=1$
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	$H_1=0$	$H_1=1$
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45	0.0135	0.09072	
1	0.1	0.252		

  - récursion: $\alpha^*(0,3) = 0.9 \max\{ 0.3 \times 0.0135, \underline{0.4 \times 0.252}\} = 0.09072$

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45	0.0135	0.09072	0.02449
1	0.1	0.252	0.03024	0.01270

- on continue d'appliquer la récursion jusqu'à la fin (t=4)...

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45	0.0135	0.09072	0.02449
1	0.1	0.252	0.03024	0.01270

- on trouve le maximum **à la dernière colonne**...

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45	0.0135	0.09072	0.02449
1	0.1	0.252	0.03024	0.01270

  - ... puis on **retrouve le chemin** qui a mené là

Explication la plus plausible avec un HMM

- Exemple: décoder un message binaire avec canal bruité (T=4)
  - message observé: S_1=0, S_2=1, S_3=0, S_4=0

Modèle d'observation

	H_t=0	H_t=1
$P(S_t=0 \mid H_t)$	0.9	0.2
$P(S_t=1 \mid H_t)$	0.1	0.8

Modèle de transition

	H_{t-1}=0	H_{t-1}=1
$P(H_t=0 \mid H_{t-1})$	0.3	0.4
$P(H_t=1 \mid H_{t-1})$	0.7	0.6

Distribution initiale

	H_1=0	H_1=1
$P(H_1)$	0.5	0.5

α*(i,t)

i \ t	1	2	3	4
0	0.45	0.0135	0.09072	0.02449
1	0.1	0.252	0.03024	0.01270

H_1=0 H_2=1 H_3=0 H_4=0

  - ce chemin nous donne la séquence des H_t la plus probable

Filtrage en ligne avec RBD

- Étant donné mon état de croyance actuel $P(X_t \mid e_{1:t})$
  - comment le mettre à jour après une nouvelle observation e_{t+1} en temps réel

- En appliquant la règle de Bayes et l'hypothèse markovienne, nous arrivons à:

$$P(X_{t+1} \mid e_{1:t+1}) = P(X_{t+1} \mid e_{1:t}, e_{t+1})$$

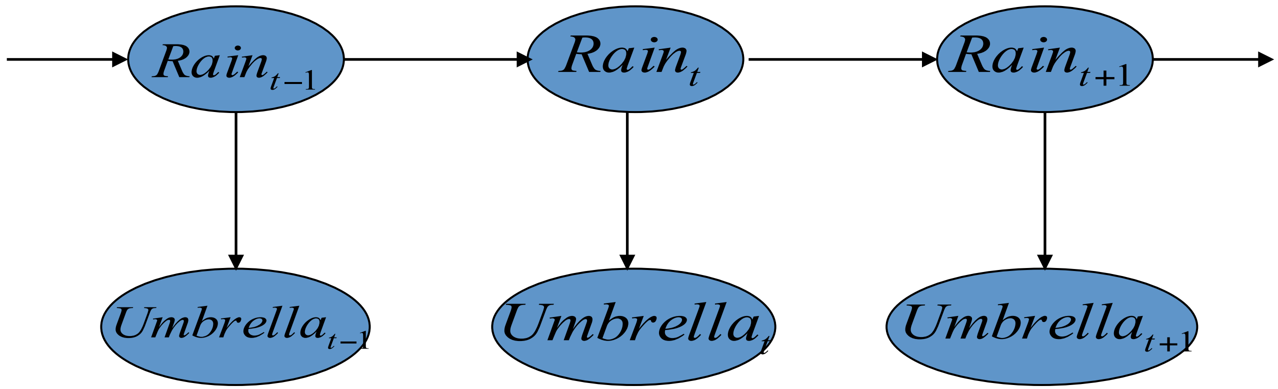
(détails page 572 du manuel de référence)

α: constante de normalisation

$$= \underbrace{P(e_{t+1} \mid X_{t+1})}_{\text{probabilité de la nouvelle observation (disponible dans la table des probabilités)}} \underbrace{\sum_{x_t} P(X_{t+1} \mid x_t)\, P(x_t \mid e_{1:t})}_{\text{prédiction du prochain état en se basant sur notre état de croyance au temps } t} / \alpha$$

Exemple de l'agent de sécurité

- **RBD**:
  - une distribution de **probabilité a priori P**(R_0), par exemple [0.5, 0.5]
  - un **modèle des transition P**($R_t | R_{t-1}$)
  - un **modèle d'observation P**($U_t | R_t$)

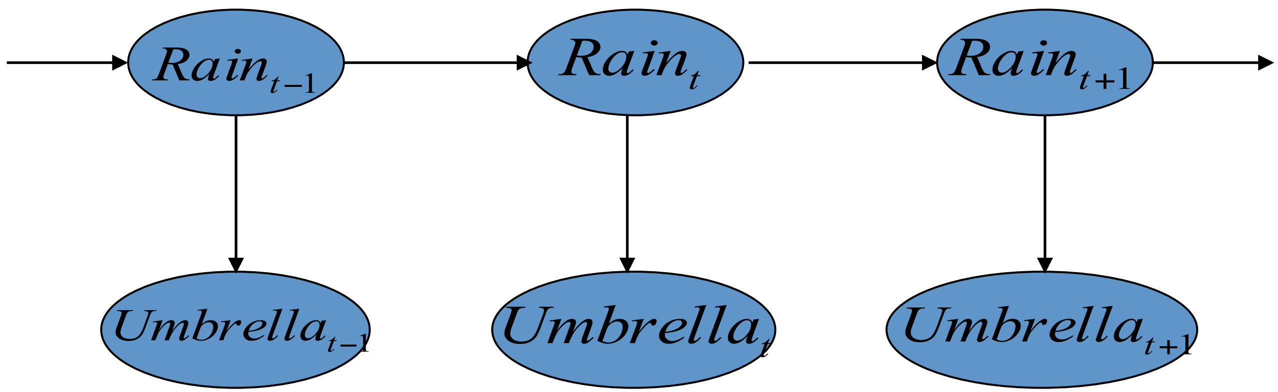


R_{t-1}	P($r_t | R_{t-1}$)
V	0.7
F	0.3

R_t	P($u_t | R_t$)
V	0.9
F	0.2

- **Jour 1**: le parapluie apparait, (U_1=*true* ou u_1)
  - le filtrage de t=0 à t=1 est: $P(R_1 \mid u_1) = P(u_1 \mid R_1)\, P(R_1)\, \alpha$

Exemple de l'agent de sécurité



- **Jour 2**: le parapluie apparait de nouveau, c.-à-d., U_2=*true*
  - le filtrage de t=1 à t=2 est:

$$P(R_2 \mid u_1, u_2) = P(u_2 \mid R_2)\, P(R_2 | u_1) / \alpha$$
$$= (\, P(u_2 \mid R_2) \sum_{r_1} P(R_2 | r_1)\, P(r_1 | u_1)\,) / \alpha$$

Améliorations sur le HMM

- Comment rendre un HMM (ou un RBD en général) plus précis?
  - **augmenter l'ordre du modèle markovien**
    - ex.: $Rain_t$ aurait comme parents, non seulement $Rain_{t-1}$ mais aussi $Rain_{t-2}$ pour un processus markovien du second ordre
    - ceci donnerait des prédictions plus précises
  - **permettre des interactions directes entre la variables d'observation**
    - on pourrait avoir plutôt $\mathbf{P}(E_t \mid X_{0:t}, E_{0:t-1}) = \mathbf{P}(E_t \mid X_t, E_{t-1})$
    - ça peut rendre l'inférence encore plus complexe

Au delà du HMM

- **Filtre de Kalman:** cas où les variables d'observation et cachées ne sont pas discrètes mais sont plutôt réelles
  - voir livre de référence, section 15.4

- **État caché avec structure complexe:** cas où il n'est pas possible de faire une sommation exacte sur toutes les configurations de l'état caché
  - on doit alors approximer l'inférence
  - **filtre particulaire** (*particle filter*): inférence approximative basée sur l'échantillonnage, où on maintien une population stochastique de configurations (particules) de l'état caché
  - à chaque temps *t*, on met à jour notre population de particules en tenant compte des nouvelles observations
  - voir livre de référence, section 15.5.3

Résumé

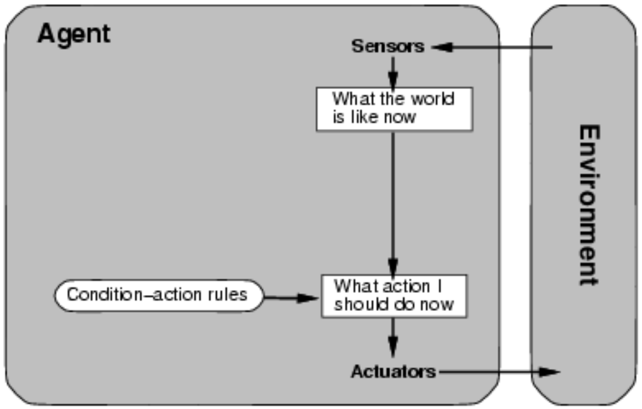
- Un **réseau bayésien dynamique** (RBD) permet de tenir compte de la nature séquentielle d'un environnement

- Un **modèle de Markov caché** (HMM) est un cas particulier de RBD avec
  - une seule variable cachée $X_{\mathrm{t}} = \{H_t\}$ et une seule variable observée $E_{\mathrm{t}} = \{S_t\}$
  - les variables H_t et S_t sont discrètes

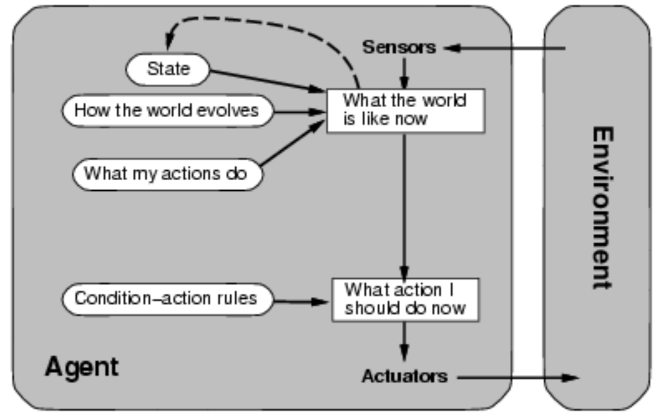
- Il existe des procédures de programmation dynamique efficaces dans un HMM pour faire de l'inférence (**filtrage**, **prédiction**, **lissage**, **explication la plus plausible**)

Raisonnement probabiliste : pour quel type d'agent?

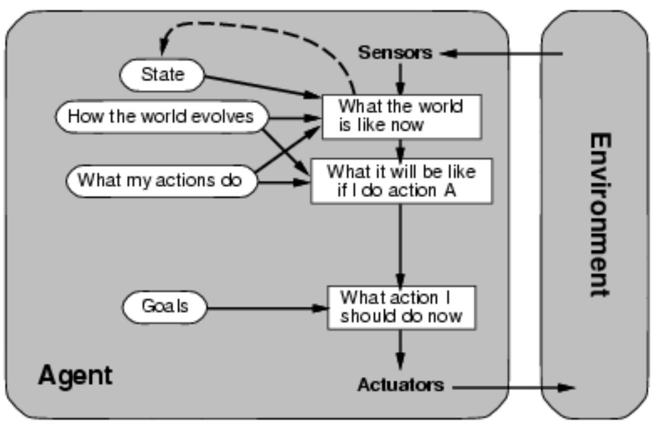
Simple reflex



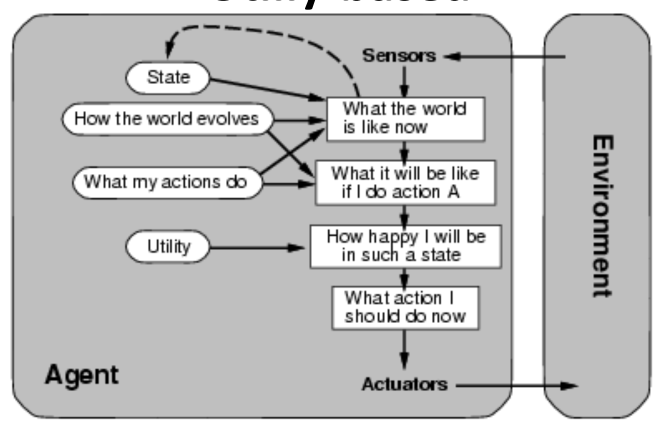
Model-based reflex



Goal-based

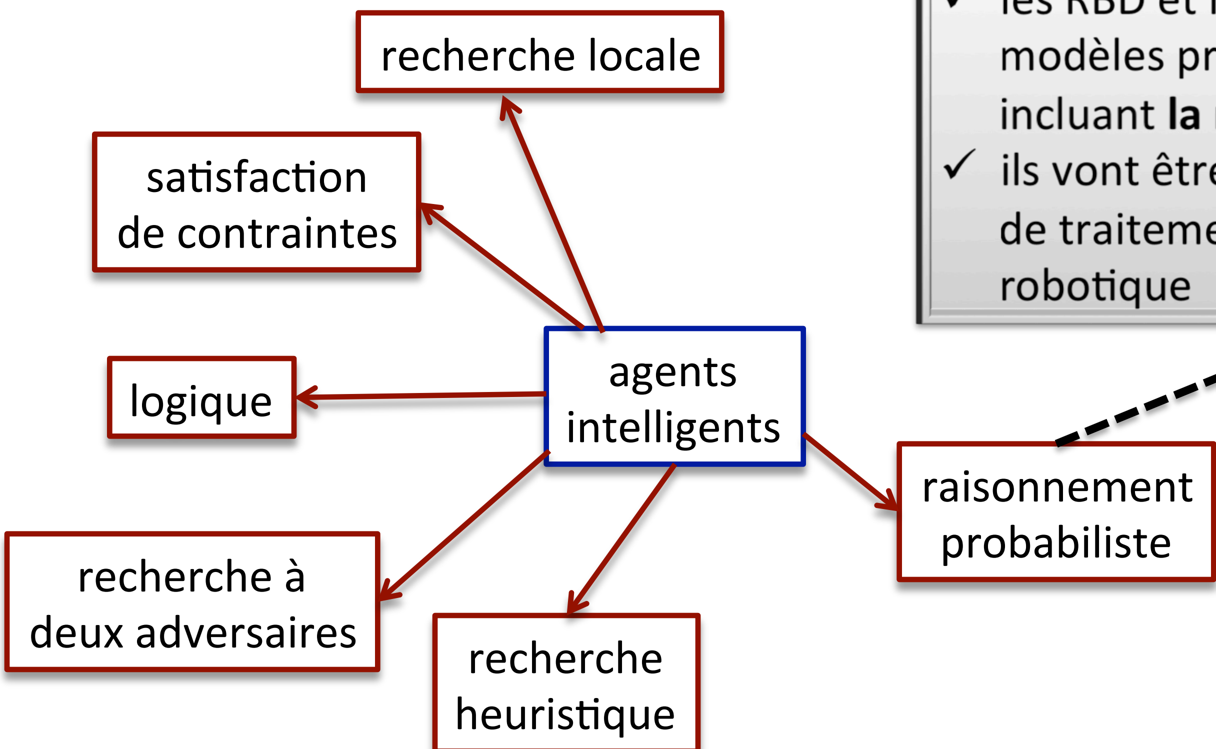


Utiliy-based



Objectifs du cours

Algorithmes et concepts



- ✓ les RBD et le HMM sont les premiers modèles probabilistes que l'on a vu incluant **la notion de temps**
- ✓ ils vont être utiles lorsqu'on va parler de traitement de la langue et robotique

Vous devriez être capable de...

- Distinguer les différents types d'inférence
  - distribution de filtrage
  - distribution de prédiction
  - distribution de lissage
  - explication la plus plausible

- Décrire ce qu'est un modèle de Markov caché
  - connaître les définitions des tableaux α, β, π et α^* (que calcule ces tableaux?)
  - savoir utiliser des tableaux α, β, π et α^* pré-calculés

- Implémenter de l'inférence dans un modèle de Markov caché