CHAINE DE MARKOV
CM

CHAINE DE MARKOV

$\{X_n ; n \in \aleph\}$ une suite de variables aléatoires ayant le même ensemble des états E = {E_1, E_2,…, E_m}

X_n représente l'état au temps t_n.

On suppose la propriété de Markov :

$$\mathbf{P(X_n = E_j / X_0 = E_{i_0}, X_1 = E_{i_1}, ..., X_{n-1} = E_i)}$$
$$\mathbf{= P(X_n = E_j / X_{n-1} = E_i) = a_{ij}}$$

On suppose aussi la propriété d'homogénéité :

$$P(X_n = E_j / X_{n-1} = E_i) = a_{ij} \quad \text{(indépendant de n)}$$

On définit la matrice des probabilités de transitions : $\mathbf{A = [a_{ij}]_{i,j=1,...,m}}$

On définit en plus la table des probabilités de l'état initial :

π=(π_1, π_2, …, π_n) où π_k=Prob(X_0=E_k)

Ainsi, la chaîne de Markov sera caractérisée par :

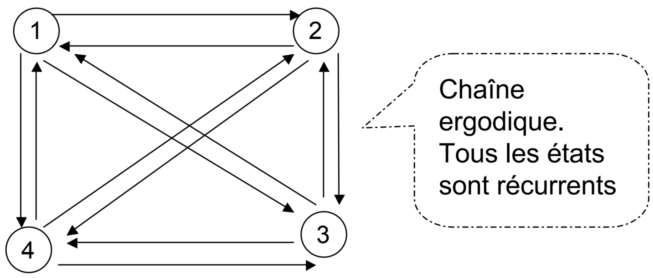
E, A et π

Graphe des probabilités de transitions

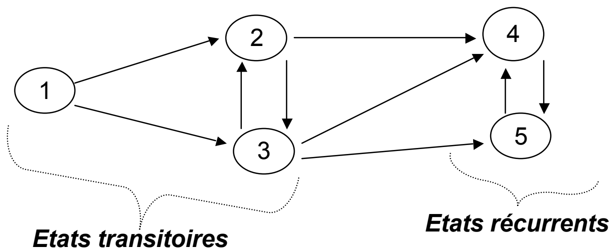
On représente la chaîne par un graphe :

- Ensemble de sommets E = {E_1, E_2,…, E_m}
- Ensemble des arcs défini par : $\{\mathbf{E_i} \rightarrow \mathbf{E_j} / \mathbf{a_{ij}} > 0\}$

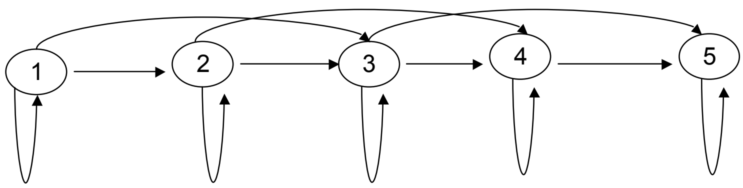
Exemple 1



Exemple 2



Exemple 3



Probabilités des états

On note par $p_n = (p_n(1),..., p_n(j),..., p_n(k))$ la table définissant les probabilités des états au temps t_n : $p_n(j) = \Pr ob\{X_n = E_j\}$

On a alors : $\mathbf{p_{n+1} = p_n A}$

La probabilité d'avoir observer une séquence d'états :

$$E_{i_0}, E_{i_1},..., E_{i_k},..., E_{i_n}$$

Est égale à :

$$\underbrace{p_0(E_{i_0})}_{\text{Probabilité de l'état initial } \pi_{i0}} \underbrace{a_{i_0 i_1} a_{i_1 i_2} ....a_{i_k i_{k+1}} ...a_{i_{n-1} i_n}}_{\text{Probabilités de transitions}}$$

Génération d'une séquence d'états par une chaîne de Markov

La chaîne de Markov peut être considérée comme un générateur de séquences de longueur variable L : $S = E_{i_0} E_{i_1} ...E_{i_k} ...E_{i_L}$

Initialisation :

E_{i_0} est la réalisation de la variable X_0 dont la table de probabilité est p (X_0=E_{i0}).

Génération de l'élément E_{ik}

Ei_k est la réalisation de la variable conditionnelle P(X_k=?/X_{k-1}=Ei_{k-1}) dont la répartition de probabilité est la table :

$$\left(\mathbf{a_{i_{k-1}1}, a_{i_{k-1}2}, ..., a_{i_{k-1}n}}\right)$$

On a alors : $p(S) = \underbrace{P(X_0 = E_{i_0})}_{\pi_{i0}} a_{i_0 i_1} \times a_{i_1 i_2} \times ... \times a_{i_{L-1} i_L}$

CHAINE DE MARKOV CACHEE
CMC

Exemple

Exemple : On dispose de 3 urnes contenant des boules blanches, noires et grises dans des proportions différentes.

On note par $P(B/u_i)$, $P(N/u_i)$, $P(G/u_i)$ les probabilités de tirer une boule blanche, noir ou grise de l'urne u_i (i=1,2,3).

Un opérateur **invisible** choisi, à des instants réguliers, une urne et tire une boule de cette urne et annonce la couleur. S'il choisi **T** urnes il annoncera (à un observateur) la séquence des **T** couleurs tirés notée $\mathbf{O_T}$.

Ainsi, au temps **t**, l'opérateur choisi une urne u_t et tire une boule de couleur o_t de cette urne. On suppose que le tirage d'une boule au temps t est indépendant du tirage de la boule au temps t-1.

Exemple (suite)

On fait l'hypothèse que l'opérateur choisi la séquence des urnes suivant un modèle d'une **chaîne de Markov** ayant pour paramètres **{π, *A*}** et dont les états sont les urnes. Cette chaîne sera dite **cachée** (puisque l'opérateur est invisible) .

Ainsi :
La séquence $\mathbf{U_T}=u_1u_2....u_T$ sera dite la **séquence des états cachée** (car invisible) et elle est générée par la chaîne de Markov cachée.

la séquence $\mathbf{O_T}=o_1o_2....o_T$ sera dite la **séquence des observations** qui sera annoncée à l'observateur.

CMC : définition

- M : le nombre des symboles observables.
- N : le nombre des états cachées.
- $V=\{v_1,v_2,\ldots,v_M\}$: l'ensemble des symboles observables.

chaîne de Markov

- $Q=\{q_1,q_2,\ldots,q_N\}$: l'ensemble des états cachées.
- $\Pi=[\pi_i]$: la table de probabilités de l'état initial de la CM.
- $A=[a_{ij}]$: matrice $N*N$ des probabilités de transition de la CM.
- $B=[b_i(v_k)]$: matrice $N*M$ avec $b_i(v_k)=P(v_k/q_i)$ (probabilité de l'émission de l'état v_k par l'état q_i).
- T : taille d'une séquence d'observations.

<u>Notation</u> : les paramètres du modèle seront notés par :

$$\boldsymbol{\lambda=(\pi , A , B)}$$

Génération d'une séquence d'observations

Une séquence d'observations $O_T=o_1o_2\ldots o_T$ est générée de la manière suivante :

1. Faire t=1, Choisir un état initial selon la distribution Π.
2. i_t étant l'indice de l'état au temps t, choisir o_t selon la table de probabilité de l'émission de l'état i_t (la table $b_i()$).
3. Choisir l'état $\mathbf{q}_{\mathbf{i}_{t+1}}$ selon les probabilités de transitions $[a_{i_t i_{t+1}}]$ Faire t=t+1.
4. Si t<T alors **retourner à l'étape 3** sinon **Fin**.

$I_T=i_1i_2\ldots i_T$: la séquence des indices des états cachées.

Trois problèmes

Problème 1 :

Etant donné la séquence d'observations $O_T=o_1o_2\ldots o_T$ et le modèle de paramètre $\lambda=(\pi,A,B)$, comment calculer $P(O_T/\lambda)$?

Problème 2 :

Etant donné la séquence d'observations $O_T=o_1o_2\ldots o_T$ comment choisir la séquence des états sous-jacentes $i_1i_2\ldots i_T$ et qui est optimal suivant un certain critère ?

Problème 3 :

Comment ajuster les paramètres $\lambda=(\pi,A,B)$ afin de maximiser $P(O_T/\lambda)$?

Problème 1

Calcule de $P(O_T/\lambda)$

Une manière immédiate pour calculer cette probabilité est d'énumérer toutes le séquences $I_T=i_1i_2\ldots i_T$ d'états possibles et de calculer :

$$p(O_T/\lambda) = \sum_{Séquences\ I_T} p(O_T/\lambda, I_T)\, p(I_T/\lambda)$$

Avec : $$P(I_T/\lambda) = \pi_{i_1} a_{i_1 i_2} a_{i_2 i_3} \ldots a_{i_{T-1} i_T}$$

Chaîne de Markov

$$p(O_T / I_T, \lambda) = b_{i_1}(o_1) b_{i_2}(o_2) \ldots b_{i_T}(o_T)$$

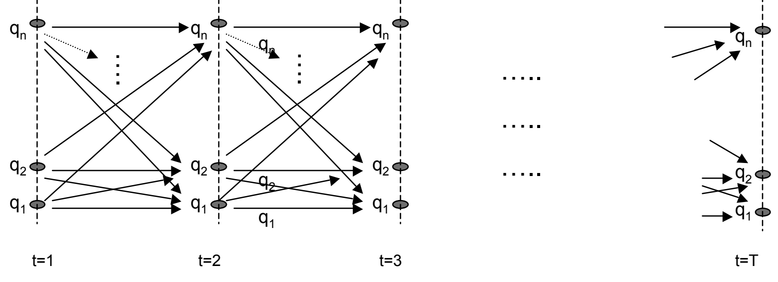
Indépendance des observations

D'où $\mathbf{P(O_T, I_T/\lambda) = \Pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) a_{i_2 i_3} b_{i_3}(o_3) \ldots a_{i_{T-1} i_T} b_{i_T}(o_T)}$

Mais cette formule est très coûteuse en temps de calcule.
Ordre de calcule $2TN^T$

Méthodes Forward et Backward

Ces deux méthodes utilisent implicitement le graphe suivant :



L 'arc $q_i \rightarrow q_j$ existe si et seulement si a_{ij}<>0, il sera valué par a_{ij}

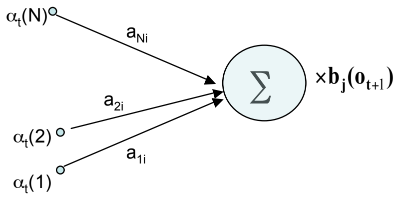
Calcule de $P(O_T/\lambda)$
Méthode Forward

On définit $\alpha_t(i)=P(\overbrace{o_1o_2\ldots o_t}^{\text{Observations partielles}},\ i_t=i\ /\lambda)$

Calcule de α_t

1. $\alpha_1(i)=\pi_i b_i(o_1)$ pour i=1,…,N
2. Pour t=1,2,…,T-1 et pour j=1,…,N

$$\alpha_{t+1}(j)=\left[\sum_{i=1}^{N}\alpha_t(i)a_{ij}\right]b_j(o_{t+1})$$



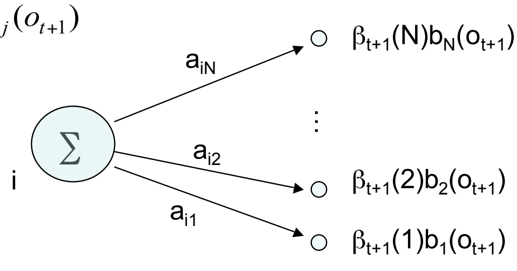
Calcule de $P(O_T/\lambda)$
Méthode Backward

On calcule $\beta_t(i)=P(\overbrace{o_{t+1}o_{t+2}\ldots o_T}^{\text{Observations partielles}}\ /\ i_t=i,\lambda)$

Calcule de α_t

1. $\beta_T(i)=1$ pour i=1,…,N
2. Pour t=T-1,…,1 et pour i=1,2,…,N

$$\beta_t(i)=\sum_{j=1}^{N}\beta_{t+1}(i)a_{ij}b_j(o_{t+1})$$



Calcule de $p(O_T/\lambda)$

On a :

$$p(O_T / \lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

$$p(O_T / \lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1)\beta_1(i)$$

Pour tout t

$$1 \leq t \leq T \quad P(O_T \mid \lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i)$$

Problème 2 : Décodage

Détermination de la séquence cachée « optimale »

Trouver la séquence des états cachées qui explique “au mieux » la séquence des observations.

Déterminer : $\arg\max_{I_T} p(I_T \mid O_T)$

Or $\quad p(I_T / O_T, \lambda) = \dfrac{p(I_T, O_T / \lambda)}{p(O_T / \lambda)}$

Déterminer $\quad \Rightarrow \quad \arg\max_{I_T} p(I_T, O_T / \lambda)$

Algorithme de Viterbi

Problème 2 : Algorithme de Viterbi

On définit : $\delta_j(t) = \max_{i_1 \ldots i_{t-1}} p(i_1 \ldots i_{t-1}, o_1 \ldots o_{t-1}, i_t = j, o_t / \lambda)$

La séquence des états cachés $i_1 i_2 \ldots i_{t-1}$ qui maximise la probabilité d'avoir observé :

- les observations jusqu'au temps t-1
- l'état au temps t
- l'observation au temps t

Par définition : $\delta_1(i) = \pi_i b_i(o_1)$

Problème 2 : Algorithme de Viterbi

Etape 1 : pour i=1,…,N $\delta_1(i)=\pi_i b_i(o_1)$ et $\Psi_1(i)=0$

Etape 2 : pour t=2 … T faire

$$\begin{cases} \delta_t(j) = \underset{1\le i\le N}{\mathrm{Max}}[\delta_{t-1} a_{ij}] b_j(o_t) \\ \Psi_t(j) = \underset{1\le i\le N}{\mathrm{Arg\,max}}[\delta_{t-1}(i) a_{ij}] \end{cases}$$

Etape 3 : FIN on tire $\begin{cases} p^* = \underset{1\le i\le N}{\mathrm{Max}}[\delta_T(i)] \\ i_T^* = \underset{1\le i\le N}{\mathrm{Arg\,max}}[\delta_T(i)] \end{cases}$

Etape 4 : (reconstitution de la séquence)

Pour t=T-1,…, 1 $\quad i_t^* = \Psi_{t+1}(i_{t+1}^*)$

Problème 3

Détermination des paramètres

Etant donné un modèle et une séquence d'observation, modifier les paramètres du modèle afin de reproduire au mieux les observations.

On définit :

$$p_t(i,j) = \frac{\alpha_i(t) a_{ij} b_j(o_{t+1}) \beta_j(t+1)}{\sum_{m=1\ldots N} \alpha_m(t)\beta_m(t)}$$

La probabilité de passer de l'état q_i à l'état q_j au temps t

$$\gamma_i(t) = \sum_{j=1\ldots N} p_t(i,j)$$

Probabilité d'être à l'état q_i au temps t

Problème 3

Détermination des paramètres

$$\sum_{t=1}^{T-1} \gamma_t(i)$$

Représente la fréquence relative des transitions à travers l'état q_i

$$\sum_{t=1}^{T-1} p_t(i,j)$$

Représente la fréquence relative des transitions de l'état q_i à l'état q_j

Problème 3 (suit)

Détermination des paramètres

Algorithme

- Initialiser les paramètres $\lambda^0=(\Pi^0,A^0,B^0)$
- Répéter l'étape itérative suivante

Les paramètres courants $\lambda=(\Pi,A,B)$ permettent de calculer les nouveaux paramètres $\hat{\lambda} = (\hat{\Pi}, \hat{A}, \hat{B})$ par :

1. $$\hat{\pi}_i = \gamma_i(1)$$
2. $$\hat{a}_{ij} = \frac{\sum_{t=1}^{T} p_t(i,j)}{\sum_{t=1}^{T} \gamma_i(t)}$$
3. $$\hat{b}_j(v_k) = \frac{\sum_{\{t;o_t=v_k\}} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$