
Les chaînes de Markov cachées

2.1 : Introduction

Les modèles de Markov cachées *MMCs* sont des modèles statistiques, riches et largement utilisés en traitement du signal. Ils sont développés par *Andrew Markov* (étudiant de *Chebyshev*), et ils sont premièrement, orientés vers des objectifs linguistiques dans des travaux de littérature Russe [29]. Ceux sont des outils efficaces en modélisation des données séquentielles ou ‘*time-series Data*’ [31]. Utilisés par la suite dans des problèmes de reconnaissance de la parole par *Baker*, leur théorie de base est introduite par *Baum* et ses collègues à la fin des années soixante.

Actuellement, ces modèles sont, de plus en plus adoptés en reconnaissance automatique de la parole, pour l’analyse des séquences *ADN*⁴, et dans des problèmes liés à l’écriture et le traitement de texte.

Aussi, leur utilisation en visionique est vaste. Ils sont implémentés en segmentation d’images, reconnaissance des visages, interprétation des gestes, ainsi que la modélisation de l’arrière plan et récemment en traitement vidéo.

⁴ : l’acide nucléique appelé : Acide Désoxyribonucléique

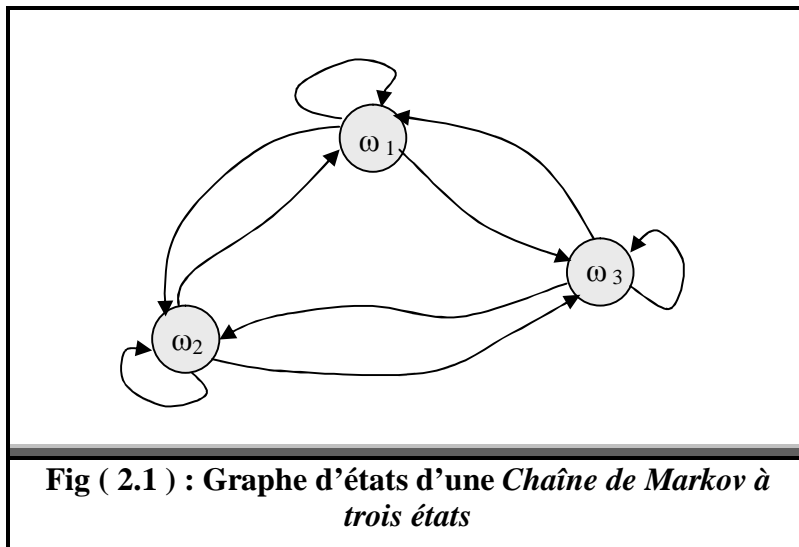
Ce chapitre constitue le support théorique de base de notre étude. Nous allons donc, donner une vue générale de ces modèles, les éléments particuliers intervenant dans leur construction, les problèmes liés à leur utilisation et les solutions usuelles de ces derniers.

2.2 : Chaînes de Markov et extension aux chaînes cachées

2.2.1 : Chaîne de Markov

Supposons $X = (X_1 X_2 \dots X_T)$ une séquence de variables aléatoires prenant leurs valeurs dans un ensemble fini $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ appelé l'espace d'état.

On définit un processus aléatoire, qui à chaque instant, visite l'un des K états possibles. A l'instant suivant, ce système (processus) change d'état, avec possibilité de rester à l'état actuel, suivant un ensemble de probabilités ; **Fig (2.1)** .



Une description probabiliste complète du processus, nécessite la spécification de l'état actuel (à l'instant t), ainsi que tous les états précédents [29, 31].

Pour un processus Markovien du premier ordre, cette spécification est arrondie à l'état actuel et son précédent adjacent. Cela est produit par l'équation suivante :

$$P(X_t = \omega_j / X_{t-1} = \omega_i, \dots, X_1 = \omega_1) = P(X_t = \omega_j / X_{t-1} = \omega_i) \dots \dots \dots (2.1)$$

Les probabilités conditionnelles $P(X_t = \omega_j / X_{t-1} = \omega_i)$ sont dites des probabilités de **Transition**, et sont notées a_{ij}^t . a_{ij}^t ne dépendra pas du temps si elle est la même à deux temps

déférents t et t' . Dans ce cas, les a_{ij}^t seront notées tout simplement les a_{ij} ; constituant ainsi une matrice carrée d'ordre K ; appelée **Matrice de Transition**, et nommée A . Signalons que K est le nombre d'états possibles, dans lesquels peut évoluer le processus. On écrit :

$$a_{ij} = P(X_t = \omega_j / X_{t-1} = \omega_i) \dots\dots\dots (2.2)$$

et la matrice de transition sera:

$$A = [a_{ij}] = [P(X_t = \omega_j / X_{t-1} = \omega_i)] \quad , 1 \leq i, j \leq K \dots\dots\dots (2.3)$$

Les éléments a_{ij} vérifient les propriétés stochastiques suivantes :

$$\begin{aligned} 1/ \quad & 0 \leq a_{ij} \leq 1 \quad \forall i, j \\ 2/ \quad & \sum_{j=1}^K a_{ij} = 1 \quad i = 1, \dots, K \dots\dots\dots \end{aligned} \quad (2.4)$$

Désignons par $P(t)$, le vecteur stochastique défini par $P(X_t = \omega_i)$ avec : $i = 1, \dots, K$, ainsi $P(1)$ sera le vecteur des distributions initiales (à l'instant $t=1$). On note ce vecteur par π , tel que chacun de ses éléments est donné par:

$$\pi_i = P(X_1 = \omega_i) \dots\dots\dots (2.5)$$

De même, les π_i ; éléments de π , vérifient les propriétés stochastiques précédentes :

$$\begin{aligned} 1/ \quad & 0 \leq \pi_i \leq 1 \quad \forall i \\ 2/ \quad & \sum_{i=1}^K \pi_i = 1 \quad \forall i \dots\dots\dots \end{aligned} \quad (2.6)$$

Notons donc, qu'une chaîne de Markov est entièrement définie par sa matrice de transition A et ses distributions initiales, données par le vecteur π .

Cela en effet, peut être résumée par la définition ci-après :

Définition :

Chaîne de Markov

Une *chaîne de Markov* est un processus stochastique $X = (X_t)_{t=1}^T$ tel que :

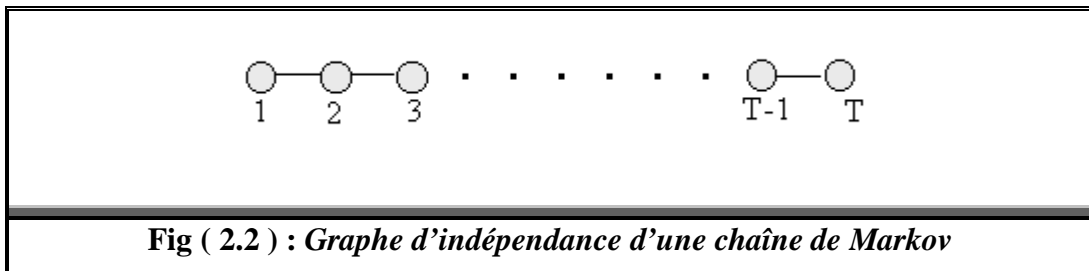
$$\forall t, \quad P(X_t / X_{t-1}, X_{t-2}, \dots, X_1) = P(X_t / X_{t-1}).$$

$$\pi_i = P(X_1 = \omega_i)$$

De plus, si $P(X_t = \omega_j / X_{t-1} = \omega_i) = a_{ij}, \forall t$, si ces probabilités appelées *probabilités de transition* sont indépendantes de t , alors la chaîne est dite homogène.

2.2.2 : Graphe d'indépendance d'une chaîne de Markov

La chaîne de Markov appartient à la famille des signaux pouvant être représentés par des graphes ; où les cercles sont étiquetés par les états possibles, et les arcs désignent les transitions valables entre ces états. Le graphe d'indépendance d'une chaîne de Markov est donné par la **Fig (2.2)** :



La figure ci-dessus représente un processus Markovien observable. Ses sorties donnent directement les états, dont chacun est attaché à un événement physique observable [31].

La probabilité jointe d'une séquence d'état $X = (X_1 X_2 \dots X_T)$ est facilement calculée pour une chaîne de Markov. Elle est donnée par un produit de probabilités comme suit :

$$P(X_1, X_2, \dots, X_T) = P(X_1) \cdot P(X_2 / X_1) \cdot P(X_3 / X_1, X_2) \cdot \dots \cdot P(X_T / X_1, X_2, \dots, X_{T-1}) \dots \quad (2.7)$$

En intégrant la propriété de Markov donnée par la formule (2.1), l'équation (2.7) devient :

$$P(X_1, X_2, \dots, X_T) = P(X_1) \cdot P(X_2 / X_1) \cdot P(X_3 / X_2) \cdot \dots \cdot P(X_T / X_{T-1}) \dots \dots \dots (2.8)$$

Cette dernière est condensée donnant l'écriture suivante:

$$P(X_1, X_2, \dots, X_T) = \pi_{X_1} \cdot \prod_{t=1}^{T-1} a_{X_t, X_{t+1}} \dots \dots \dots (2.9)$$

Nous citons un petit exemple qui met en valeur la notion de la Chaîne de Markov, et permet de mesurer la probabilité d'une séquence d'états bien définie. Remarquons que cette mesure de probabilité est évaluée par un simple produit.

Exemple

Utilisation de chaîne de Markov dans une prédiction météo :

Assumons que chaque jour, le temps est dans l'un des trois états suivants :

- Etat ω_1 : pluvieux.
- Etat ω_2 : nuageux.
- Etat ω_3 : ensoleillé.

On donne la matrice de transition A par:

$$A = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

et le vecteur initial π par : $\pi = [0 \ 0 \ 1]$.

La question posée est :

Sachant qu'aujourd'hui est ensoleillé, et avec un tel modèle, quelle est la probabilité que le temps suit, dans les sept (07) jours qui suivent, la séquence d'observation O donnée par :

$$O = (\omega_3, \omega_3, \omega_3, \omega_1, \omega_1, \omega_3, \omega_2, \omega_3) \quad \text{pour } t = 1, 2, \dots, 8 .$$

Une telle probabilité est exprimée par $P(O / \text{mod èle})$ tel que :

$$P(O / \text{mod èle}) = P(\omega_3, \omega_3, \omega_3, \omega_1, \omega_1, \omega_3, \omega_2, \omega_3 / \text{mod èle}) \quad \text{avec } P(O / \text{mod èle}) \text{ dénote}$$

la probabilité de la séquence d'observations O sachant ce modèle.

En utilisant la propriété de la chaîne de Markov marquée par (2.8), la formule précédente

$$P(O / \text{mod èle}) = P(\omega_3) \cdot P(\omega_3 / \omega_3) \cdot P(\omega_3 / \omega_3) \cdot P(\omega_1 / \omega_3) \cdot P(\omega_1 / \omega_1) \cdot P(\omega_3 / \omega_1) \\ \cdot P(\omega_2 / \omega_3) \cdot P(\omega_3 / \omega_2)$$

$$\begin{aligned} \text{devient:} &= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\ &= (1) \cdot (0.8) \cdot (0.8) \cdot (0.1) \cdot (0.4) \cdot (0.3) \cdot (0.1) \cdot (0.2) \\ &= 1.536 \cdot 10^{-4} \end{aligned}$$

2.2.3 : Chaîne de Markov cachée

Maintenant, on ajoute des observations ; on considère que X est un processus caché, et on souhaite déterminer ses états par l'intermédiaire du vecteur d'observations Y .

Donc, un modèle de Markov caché; noté *MMC* est un processus stochastique double [31], avec :

- ✓ Un premier processus X caché, constitué d'un ensemble d'états connectés entre elles par des probabilités de transition.
- ✓ Un deuxième processus Y observable, constitué d'un ensemble de sorties; dites observations. Chaque sortie peut être émise par n'importe quelle état caché, conformément à une fonction densité de probabilité de sortie.

On peut annoncer la définition ci après :

Définition :

Chaîne de Markov cachée

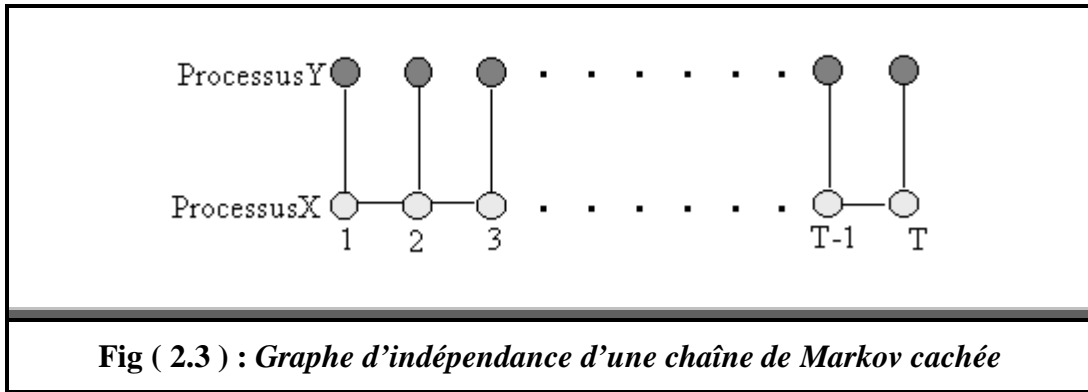
Une *chaîne de Markov cachée* 'CMC' est un processus stochastique double (X, Y) [35], tel que :

- $X = (X_t)_{t=1}^T$, soit une chaîne de Markov homogène. Elle est définie par sa distribution initiale π et sa matrice de transition A .

- $Y = (Y_t)_{t=1}^T$, soit un processus observable et indépendant conditionnellement à X ; tel que : $P(Y / X) = \prod_t P(X_t / Y_t)$.

2.2.4 : Graphe d'indépendance d'une chaîne de Markov cachée

Graphiquement, on représente une chaîne de Markov cachée par la figure **Fig (2.3)** .
On dit qu'on a une dépendance ponctuelle de Y relativement à X .



2.3 : Lois d'observation

Plusieurs approches ont été successivement proposées dans le choix des lois d'observation rattachées aux états cachés du modèle. Si l'espace d'observation est discret et fini, chaque distribution $b_j(.)$ est entièrement définie par la reconnaissance de la probabilité d'émission de chacun des points de l'espace. $\{y_1, y_2, \dots, y_M\}$. La valeur de M , donne la taille de cet espace d'observation. L'ensemble des lois d'observation est $B = \{b_j(.)\}_{1 \leq j \leq K}$, et entièrement caractérisé par le vecteur $\{b_j(m); 1 \leq j \leq K; 1 \leq m \leq M\}$.

Chaque paramètre $b_j(m)$ détermine la probabilité que le point y_m soit émis par la distribution $b_j(.)$, l'ensemble des probabilités $b_j(m)$ vérifiant, pour chaque état ω_j du modèle la contrainte :

$$\forall j: \sum_{1 \leq m \leq M} b_j(m) = 1$$

Lorsque l'espace d'observation est continu, plusieurs choix sont possibles pour les fonctions densités de probabilité, et sont en général basés sur des gaussiennes. L'ensemble des lois d'observation $F = \{f_j(.)\}_{1 \leq j \leq K}$ est alors caractérisé par la connaissance des K densités de probabilité f_1, f_2, \dots, f_K associées aux K classes considérées.

2.4 : Eléments d'une chaîne de Markov cachée

A fin d'utiliser une chaîne de Markov cachée, il est indispensable de spécifier les caractéristiques ci-après. On parle d'éléments de cette chaîne :

- ✓ K : Nombre d'états dans le modèle. Chaque état a une signification physique déterminée, selon l'application prise. On dénote les états individuels par $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$. L'état, à l'instant t est noté X_t .
- ✓ M : Nombre de symboles observés. Ils correspondent aux sorties physiques du système modélisé. On dénote les symboles individuels par $\{y_1, y_2, \dots, y_M\}$. Celui, à l'instant t est noté Y_t . La structure, elle aussi, est imposée par le phénomène étudié.

- ✓ $A = [a_{ij}]$: matrice de transition, avec :

$$a_{ij} = P(X_t = \omega_j / X_{t-1} = \omega_i), \quad 1 \leq i, j \leq K.$$

- ✓ $\pi = [\pi_i]$: probabilités initiales de transition, avec :

$$\pi_i = P(X_1 = \omega_i), \quad 1 \leq i \leq K.$$

- ✓ $B = \{b_j(y_t)\}$: probabilités d'émission, tel que :

$b_j(y_t)$: est la probabilité d'émettre le symbole y_t à l'instant t , sachant que l'état à ce moment est $X_t = \omega_j$. On écrit :

$$b_j(y_t) = P(Y_t = y_t / X_t = \omega_j)$$

- ✓ T : Longueur de la séquence d'observations Y ; avec Y_t est le symbole observé à l'instant t .

Donc, en utilisant une notation compacte, un *MMC* est noté par $\lambda = (\pi, A, B)$ désignant les paramètres complets d'un modèle de Markov caché.

2.5 : Types principaux des modèles *MMC*s

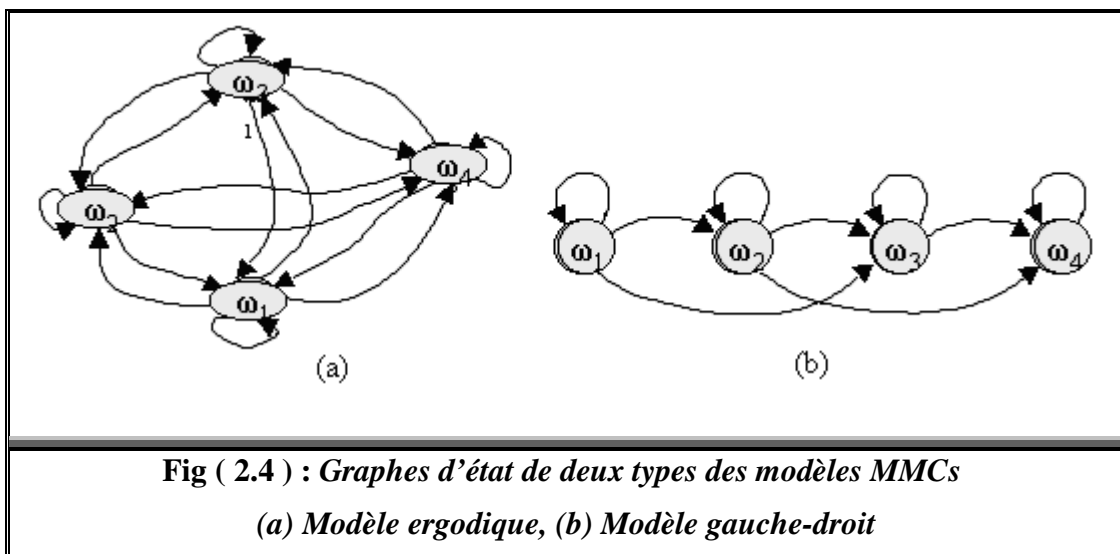
A toute chaîne de Markov est associé un graphe d'états, représentatif des possibilités d'évolution séquentielle du processus modélisé. Sa topologie dépend directement des positions relatives des zéros dans la matrice A de transition. La structure d'un *MMC* diffère donc, selon le phénomène à étudier. Parmi celles les plus utilisées, on cite deux types importants : « *MMC ergodique* » et « *MMC gauche-droit* » [31,33], tout deux présentés par la figure : **Fig (2.4)** .

Celui ergodique de **Fig (2.4.a)**, est un cas spécifique ; où le modèle est dit complètement connecté, et toutes les possibilités d'évolution sont permises a priori. Autrement dit, tous les a_{ij} sont positifs, et la matrice A est pleine. Cela permet d'atteindre n'importe quel état en passant par n'importe quel chemin.

Tandis que un *MMC* à matrice A triangulaire supérieure appelé « *gauche-droit* » ou '*Left-Right*' est illustré dans **Fig (2.4.b)**. Celui-ci est particulièrement adapté à la modélisation d'évolution temporelle de processus (parole, gestes). Ce type se dote de la propriété d'accroissance de l'indice d'état relativement au temps (pas de retour arrière). Il est notamment choisi pour modéliser les signaux dont les caractéristiques changent avec le temps. La propriété fondamentale de ce genre est :

$$a_{ij} = 0, \text{ si } i < j, \quad \text{et } \pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases}$$

Il revient de souligner que le graphe d'états d'un *MMC* n'est pas à confondre avec le graphe d'indépendance, lequel représente le modèle « déroulé » dans le temps, en soulignant graphiquement la dépendance ponctuelle des observations relativement aux états.



2.6 : Les trois problèmes liés aux *MMCS*

Etant donné un *MMC*, la plus part des applications sont réduites à résoudre les trois problèmes essentiels liés à une chaîne de Markov cachée [29, 30, 31, 32]. Autrement dit, pour une modélisation à base de chaînes de Markov cachées, il faut répondre aux trois questions suivantes :

Soit λ un MMC donné, et soit $Y = (Y_1, Y_2, \dots, Y_T)$ une séquence d'observation donnée :

2.6.1 : Evaluation

Le problème d'évaluation est celui permettant de répondre à la question suivante:

Calculer $P(Y/\lambda)$?

Où $P(Y/\lambda)$ est la probabilité d'une séquence d'observations Y sachant le modèle λ . C'est le problème nommé aussi, par 'Scoring'. Par exemple, si on dispose de plusieurs modèles compétitifs, cette démarche permet de choisir le meilleur modèle générant cette séquence d'observations Y .

2.6.2 : Optimisation

Dans ce cas la, on se tente de déterminer la démarche convenable, amenant à la réponse de la question posée ainsi:

Comment choisir une séquence d'état $X = (X_1, X_2, \dots, X_T)$, qui est optimale ?

On cherche, donc à décoder les états qui correspondent de mieux à la séquence d'observations Y ; c'est pour quoi on utilise souvent le nom : 'Decoding'.

Alors, cette partie est la section dans laquelle on découvre les états cachés d'un MMC ; ou bien, trouver la séquence d'état *correcte*. Notons que cette dernière n'existe pas ; c'est plutôt une *estimation meilleure*. C'est pourquoi en pratique, on se sert des critères d'optimisation ; d'où le nom *optimisation* donné à ce problème.

2.6.3 : Apprentissage

Ce problème s'occupe d'ajuster les paramètres du modèle λ , dans l'objectif de maximiser la probabilité d'observation annoncée au premier problème; donc :

Trouver $\hat{\lambda}$? tel que : $\hat{\lambda} = \arg \max_{\lambda} P(Y/\lambda)$.

Cela signifie d'adapter les paramètres (π, A, B) du modèle, pour mieux décrire l'application. La séquence d'observation Y utilisée, dans ce cas est appelée *séquence d'apprentissage* ; notée parfois $Y_{Training}$.

Ce dernier problème est crucial, puisqu'il permet de mieux adapter le modèle choisi au phénomène considéré.

2.7 : Solutions des trois problèmes

Dans ce paragraphe, nous allons donner, en bref, les solutions usuelles des trois problèmes ci-dessus selon [31].

2.7.1 : Premier problème : évaluation

On veut trouver la probabilité d'une séquence d'observations Y sachant le modèle λ . C'est déterminer la mesure de vraisemblance $P(Y/\lambda)$.

Rappelons que :

$$P(Y_t / X_t) = b_{X_t}(Y_t) \dots \dots \dots (2.10)$$

et :

$$P(Y, X / \lambda) = P(Y / X, \lambda) \cdot P(X / \lambda) \dots \dots \dots (2.11)$$

Tel que $P(Y, X / \lambda)$: est la probabilité jointe de Y et X sachant ce modèle λ .

La démarche suivie pour mesurer la probabilité $P(Y/\lambda)$ est d'énumérer toutes les séquences possibles de longueur T .

Mais pour le moment, considérons $X = (X_1, X_2, \dots, X_T)$, une séquence d'état fixe. Selon la propriété d'indépendance conditionnelle des observations, la probabilité $P(Y/X, \lambda)$ est donnée par :

$$P(Y / X, \lambda) = \prod_{t=1}^T P(Y_t / X_t, \lambda) \dots \dots \dots (2.12)$$

Accordée à l'équation (2.10), la formule (2.12) devient comme suit :

$$P(Y / X, \lambda) = \prod_{t=1}^T b_{X_t}(Y_t) \dots\dots\dots (2.13)$$

Ou encore :

$$P(Y / X, \lambda) = b_{X_1}(Y_1) \cdot b_{X_2}(Y_2) \cdot \dots \cdot b_{X_{T-1}}(Y_{T-1}) \cdot b_{X_T}(Y_T) \dots\dots\dots (2.14)$$

d'une part.

D'autre part, rappelons la formule (2.9):

$$P(X / \lambda) = \pi_{X_1} \cdot \prod_{t=1}^{T-1} a_{X_t X_{t+1}}$$

La probabilité d'une séquence d'observations Y sachant λ est donnée donc, par la somme de toutes les probabilités jointes, associées aux combinaisons d'état possibles, donnant chacune une séquence X d'états de même longueur T . On écrit :

$$P(Y/\lambda) = \sum_{\text{sur tous les } X} P(Y/X, \lambda) \cdot P(X/\lambda) \dots\dots\dots (2.15)$$

En substituant (2.9) et (2.14) dans (2.15), on obtient la mesure de vraisemblance comme suit :

$$P(Y/\lambda) = \sum_{\text{sur tous les } X} \pi_{X_1} \cdot b_{X_1}(Y_1) \cdot a_{X_1 X_2} \cdot b_{X_2}(Y_2) \cdot a_{X_2 X_3} \cdot \dots \cdot b_{X_{T-1}}(Y_{T-1}) \cdot a_{X_{T-1} X_T} \cdot b_{X_T}(Y_T) \dots (2.16)$$

remarque ↑ :

Le calcul précédant de $P(Y/\lambda)$, fait appel à un nombre de multiplications de l'ordre de $(2 \cdot T - 1) \cdot K^T$, associé à $(K^T - 1)$ additions [31]. Il semble être difficile à effectuer même pour des petites valeurs de K et T .

A titre d'exemple, pour un nombre d'états de $K = 2$, et une séquence d'observations de $T = 100$ de longueur, le nombre total d'opérations est de l'ordre de $(2 \cdot 100 \cdot 2^{100}) = 100 \cdot 2^{101}$

Cet accès de calcul peut être résolu en utilisant la procédure nommée ‘*Forward-Backward*’. Elle est basée sur la détermination des deux probabilités ‘*Forward*’ ; $\alpha_t(i)$, et ‘*Backward*’ ; $\beta_t(i)$, dont les spécifications et méthodes de calcul seront données au chapitre suivant.

Notons maintenant que :

$$\alpha_t(i) = P(Y_1, Y_2, \dots, Y_t, X_t = \omega_i / \lambda) \dots \dots \dots (2.17)$$

C’est la probabilité de la séquence d’observations partielle jusqu’à l’instant t avec le système est dans la classe ω_i à ce dernier moment t , sachant le modèle λ .

Et :

$$\beta_t(i) = P(Y_{t+1}, Y_{t+2}, \dots, Y_T / X_t = \omega_i, \lambda) \dots \dots \dots (2.18)$$

Cette quantité représente la probabilité de la séquence d’observations partielle débutant à l’instant $t+1$ et allant jusqu’à l’instant final, sachant qu’au moment t , le système est dans la classe ω_i , et aussi sachant le modèle λ .

2.7.2 : Deuxième problème : ‘*decoding*’

Contrairement au premier problème, ou on peut donner une réponse exacte et unique, la solution de celui-ci est non spécifique. La diversité des critères conduit à un choix large de méthodes. Rappelons qu’on cherche la séquence d’état optimale, associée à un vecteur d’observations donnée [29, 31].

Parmi les politiques d’optimisation, l’un s’intéresse à chaque état seul. Il désigne, donc l’état attendu par optimisation individuelle. Cette démarche est basée sur le calcul de la probabilité $\xi_t(i)$; la probabilité d’être à l’état ω_i à l’instant t sachant l’observation complète Y et le modèle λ . La probabilité en question est donnée par :

$$\xi_t(i) = P(X_t = \omega_i / Y, \lambda) \dots \dots \dots (2.19)$$

De même, cette probabilité sera exprimée ultérieurement à l’aide des probabilités ‘*Forward*’ et ‘*Backward*’ ; $\alpha_t(i)$ et $\beta_t(i)$ respectivement.

La séquence d'état optimale est alors déterminée, par l'accumulation des états optimaux. Chaque état vérifie cette caractéristique tout seul et indépendamment des autres. On écrit :

$$X_t \Big|_{optimal} = \arg \max_{1 \leq i \leq K} \xi_t(i), \quad 1 \leq t \leq T \dots\dots\dots (2.20)$$

Un autre néglige les états isolés et exploite la notion « *chemin* ». Il s'intéresse donc, aux chemins partiels optimaux. Un des premiers dits algorithmes, utilisé avec les *MMCs*, est l'algorithme de *Viterbi*. Il détermine ce chemin optimal progressivement, tout en conservant l'optimalité partielle [29, 31, 35]. Pour utiliser ce dernier, une nouvelle probabilité à introduire, est $\delta_t(i)$ avec :

$$\delta_t(i) = \max_{X_1, X_2, \dots, X_{t-1}} P(X_1, X_2, \dots, X_t = \omega_i, Y_1, Y_2, \dots, Y_t / \lambda) \dots\dots\dots (2.21)$$

Cette mesure représente la probabilité jointe maximale correspondante aux sous séquences d'observations et d'état ; (Y_1, Y_2, \dots, Y_t) et $(X_1, X_2, \dots, X_{t-1})$ respectivement, et avec le processus se trouvant actuellement à l'état ω_i .

Par induction, on trouve :

$$\delta_{t+1}(j) = \left[\max_{X_1, X_2, \dots, X_t} \delta_t(i) \cdot a_{ij} \right] \cdot b_{X_{t+1}=\omega_j}(Y_{t+1}) \dots\dots\dots (2.22)$$

Une fois de plus, on fait une accumulation d'états vérifiant la maximisation, en se servant d'un accumulateur γ (tableau), avec :

$$\gamma_t(j) = \arg \max_{1 \leq i \leq K} [\delta_{t-1}(i) \cdot a_{ij}], \quad \begin{cases} 2 \leq t \leq T \\ 1 \leq j \leq K \end{cases} \dots\dots\dots (2.23)$$

L'algorithme complet de *Viterbi* est donné ultérieurement au chapitre suivant.

2.7.3 : Troisième problème : ‘training’

Ce problème est classé, comme étant, le plus difficile des trois. Sa solution vise à déterminer une méthode capable d’ajuster les paramètres de λ ; c’est maximiser $P(Y_{Training} / \lambda)$ au sens du maximum de vraisemblance [31], dont les paramètres estimés $\hat{\lambda}$ sont donnés comme suit:

$$\hat{\lambda} = \arg \max_{\lambda} P(Y_{Training} / \lambda) \dots \dots \dots (2.24)$$

Dans la littérature, il n’y a pas de méthodes analytiques spécifiques pour maximiser la probabilité précédente. On se sert donc, des procédures itératives. Un algorithme répondu, dans ce domaine est celui de *Baum-Welch*, connu aussi par algorithme ‘*Forward_Backward*’. D’autres utilisent son équivalent, dit *algorithme EM* ; pour ‘*Expectation-Modification*’.

Pour appliquer ces procédures itératives, on a besoin de définir une nouvelle probabilité, nomée $\psi_t(i, j)$, et mesurant le fait d’être à l’état ω_i à l’instant t , et à l’état ω_j à l’instant suivant, sachant la séquence d’observations Y et le modèle λ . On exprime celle ci par :

$$\psi_t(i, j) = P(X_t = \omega_i, X_{t+1} = \omega_j / Y, \lambda) \dots \dots \dots (2.25)$$

Cette nouvelle quantité est écrite en fonction des deux fameuses probabilités ‘*Forward*’ et ‘*Backward*’. Par la suite, on déduit tous les paramètres nécessaires à la mise à jours du modèle (voir chapitre suivant).

remarque \uparrow :

Notons seulement que cette algorithme démarre d’un *MMC* initial connu, dont les paramètres affectent sa performance attendue. Une bonne initialisation, alors est toujours demandée.

2.8 : Conclusion

Les Modèles de Markov cachés *MMCs*, précisément les Chaînes de Markov Cachées, sont donc des outils de modélisation très utiles. Leur choix en vue de résoudre un problème bien défini nécessite une spécification complète de leurs éléments ; à savoir: la matrice de transition, les distributions initiales et les densités d'émission des observations.

Ces modèles peuvent être classés comme des méthodes probabilistes, d'une part. D'autre part, ils appartiennent à la famille des modèles pouvant être représentés graphiquement à l'aide d'arcs et de cercles.

Pour la modélisation d'un phénomène quelconque par ces outils là, il est donc, indispensable de répondre aux trois problèmes essentiels associés à ces modèles, dont nous avons brièvement, nommé et donné les solutions usuelles utilisées. Nous avons aussi, défini les diverses probabilités intermédiaires nécessaires à l'intégration des algorithmes destinés à établir ces solutions.

Ainsi, le chapitre suivant couvre plus largement, les procédures et les algorithmes dédiés à la résolution de ces problèmes, et l'exploitation de ces *MMCs* dans des domaines variés. Par la suite, nous verrons leur intégration en vue de la détection d'objet mobile.