République Algérienne démocratique et populaire
Ministère de l'enseignement supérieur et de la Recherche scientifique
Université M'hamed Bougara Boumerdés

Faculté des Sciences

Département Mathématiques

Mémoire De Fin D'étude Du Seconde Cycle Pour L'obtention Du Diplôme De Master

Chaine de Markov Image Cachée

Spécialité : Mathématiques appliquées

Option : Mathématique Financière

- **Réalisé par :**
- Chelouche Lynda
- Bousadi Nawel
- Bougaba Rachida

Président : O.Benamara.
Examinateur : H.Chemrik
Encadreur : G.Larabi

Soutenu : 2020/2021

Remerciement

En premier lieu, nous remercie Dieu tout puissant de nous avoir aidés à mener bien notre mémoire de fin d'étude.

Nous adresses nos remerciements particulièrement à mon encadreur Madame Larabi.G qui nous a assidument accompagne tout au long de ce projet, ainsi que pour son attention portée sur notre évolution dans le travail et pour sa collaboration.

Nous remercions également tous les professeurs qui ont contribués de près ou de loin à notre formation universitaire, sans oublier toute personne qui nous a aidés à mener à terme notre projet.

Nous présentons en fin nos remerciements aux membres du jury O.Benamara et H.Chemrik qui nous feront l'honneur d'évaluer et juger notre travail, merci à tous.

Dédicace lynda

Je dédie ce mémoire :

A ma très chère mère qui m'a soutenu et encouragé durant ces années d'études.

A mon chère père qui a toujours été à mes côtés.

A mes frères, à mes sœurs et ceux qui ont partagé avec moi tous les moments d'émotion lors de la réalisation de ce travail.

A tous mes amis qui m'ont toujours encouragé.

A toutes les personnes qui me connaissent de près ou de loin.

Dédicace Nawel

Je dédie ce mémoire à mes chers parents ma mère et mon père pour leur patience, leur amoure, leur soutien et leurs encouragements.

A mes frères.

A mes amies et mes camarades.

Sans oublier tous les professeurs que ce soit du primaire, du moyen, du secondaire ou de l'enseignement supérieur.

Dédicace Rachida

Tout au début, je tiens à remercier le bon dieu de m'avoir donné du courage et de la foi afin de réaliser ce modeste travail.

A ma famille, elle qui m'a doté d'une éducation digne, son amour a fait de moi ce que je suis aujourd'hui.

Particulièrement à ma chère maman et mon chère papa, pour leur patience, leur soutien et leur encouragement.

A mes frères et mes belles sœurs, pour leurs conseils précieux tout au long de mes études

A tous mes amies.

A tous ceux que j'aime et ceux qui m'aiment.

Table des matières

Introduction générale

Les chaines de Markov ont une place prépondérante dans le domaine des probabilités. Ils permettent de modéliser l’évolution dynamique d’un système aléatoire. L’étude de ces chaines a connu de nombreux champs d’applications, et s’étend sur plusieurs domaines, notamment en génétique. Néanmoins en écologie, réseaux, mathématiques financières, simulation……, ce qui nécessite un développement continuel de la théorie qui les régit. Ce mémoire à pour objet d’exposé un des problèmes théoriques des chaines de Markov. Il s’agit de vérifier si l'image d'une chaine de Markov cachée est une chaine de Markov cachée.

Le premier chapitre est divisé en trois parties, la première partie aborde des notions fondamentales de base sur les ensembles, les applications, mesures et probabilité. Dans la deuxième partie, on rappelle également quelques notions sur les chaines de Markov à temps discret, des définitions générales (équations de chapman-kolmogorov, classifications des états…). Troisième partie nous allons introduire les modèles de Markov cachées, leurs algorithmes, les trois problèmes fondamentaux, évaluation, décodage et l’apprentissage.

Dans deuxième chapitre on a deux parties, la première partie aborde les chaines de Markov images avec des conditions nécessaire et suffisantes, vérifier si l'image de la chaine de Markov par la fonction f conserve le caractère markovien quand cette fonction a une forme particulière (injective, surjective ou bijective). La deuxième partie est consacré pour les chaines de Markov image cachée et à la description des conditions nécessaires et suffisantes pour que l'image d'une chaine de Markov cachée est une chaine de Markov cachée.

Le dernier chapitre est consacré à la simulation des résultats précédents en utilisant le programme "python", a fin de facilité l'application numérique des résultats théorique étudiés.

Chapitre 1

Chaine de Markov Cachée

I. Notions fondamentales

1. Notion sur les ensembles

1.1 Ensembles

Un ensemble peut se définir par :

•la liste de ses éléments, par exemple :

$$IN = \{0, 1, 2, ..., 10^6, ..., n, ...\}$$

•la propriété que vérifient ses éléments, par exemple :

$$E = \{x \in \mathbb{R},\ x^2 + x - 1 = 0\}$$

Quand x est élément de E, c'est-à-dire qu'il est dans la liste des éléments de E, ou qu'il vérifie la propriété caractéristique de E, on dit que x appartient à E, noté : $x \in E$

1.2 Sous ensemble

On dit que F est inclus dans E, noté $F \subset E$, si et seulement si :

$$\forall x \in F \Rightarrow x \in E.$$

F est ainsi un sous-ensemble de E, ou une partie de E, on note $P(E)$, l'ensemble des parties de E c'est ensemble de tous les sous-ensembles de E.

1.3 Intersection, Réunion et Complémentaire

•Définition : L'intersection de 2 ensembles $A\ et\ B$ est :

$$A \cap B = \{x, x \in A\ et\ x \in B\}$$

•Définition : la réunion de 2 ensembles A et B est :

$$A \cup B = \{x, x \in A\ ou\ x \in B\}$$

•Définition : Si $A \subset B$, le complémentaire de A dans B est :

$$\mathrm{C}_{\mathrm{B}}^{\mathrm{A}} = \{x, x \in B\ et\ x \notin A\}$$

2. Notions sur les applications

2.1 Une fonction (ou une application)

Une application $f: E \to F$ (de E $dans$ F) est définie par un sous-ensemble $G_f \subseteq E \times F$ tel que pour tout $x \in E$, il existe au plus $y \in F$ tel que $(x, y) \in G_f$, on note $y = f(x)$.

2.2 Image directe, tiré en arrière.

- **Image directe**

Soient E, F deux ensembles, $f: E \to F$ une application et A une partie de E. On appelle image directe de A par f l'ensemble des images des éléments de A (voir figure 1), i.e. la partie de F: $f(A) = \{f(x) \mid x \in A\} = \{y \in F \mid \exists x \in A, y = f(x)\}$

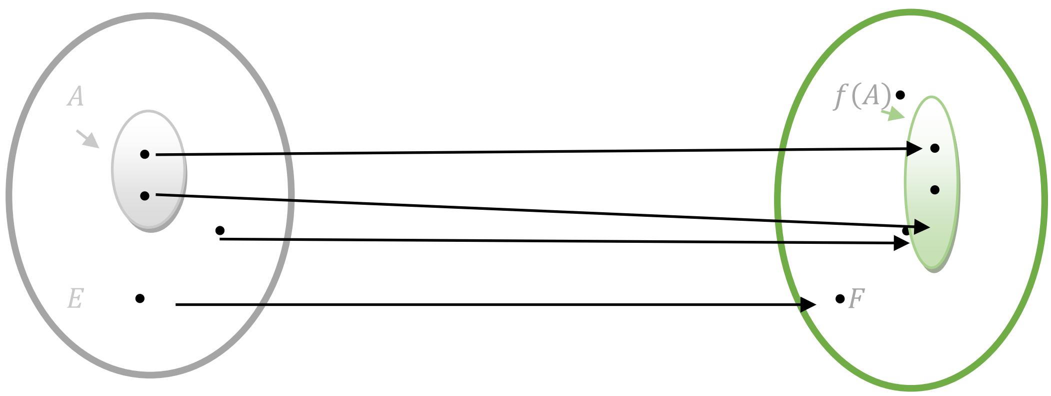


Figure 1.1 - Image direct d'une partie A par une application f

- **Image réciproque (tiré en arrière)**

Soit E et F deux ensembles, $f: E \to F$ une application et B une partie de F. On appelle tiré en arrière de B par f l'ensemble des antécédents des éléments de B (voir figure 2), i .e. la partie de E :

$f^{-1}(B) = \{x \in E \mid f(x) \in B\}$. [3]

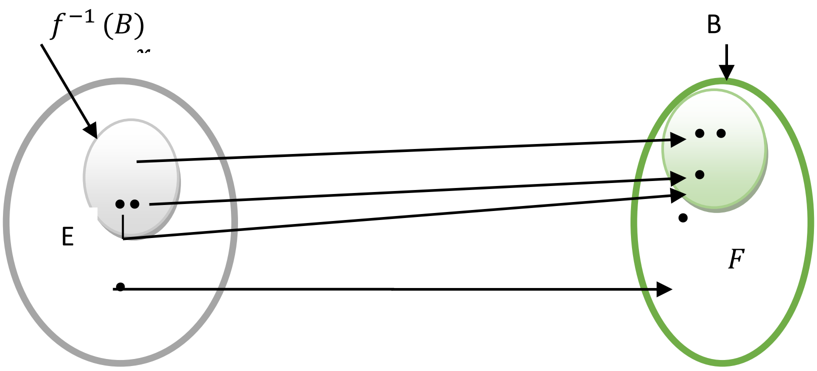


Figure 1.2 -tiré en arrière d'une partie *B* par une application *f*

2.3 Injection, surjection, Bijection

2.3.1 Injection

Soit E, F deux ensembles, $f : E \to F$ une application. On dit que f est injective (ou est une injection) si : $\forall (x, y) \in E^2, f(x) = f(y) \Rightarrow x = y$.

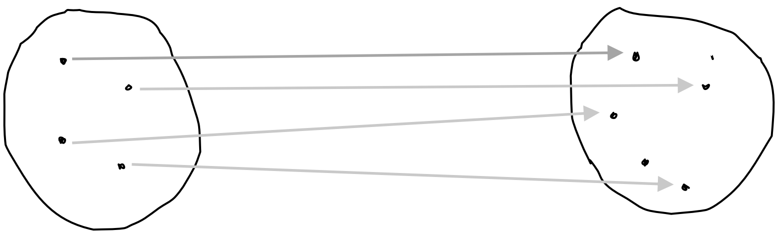


Figure 1.3 -Exemple d'application injective.

2.3.2 Surjection

Soit E, F deux ensembles, $f: E \to F$ une application. On dit que f est surjective (ou est réalise une surjection) si :$\forall y \in F, \exists x \in E, y = f(x)$.

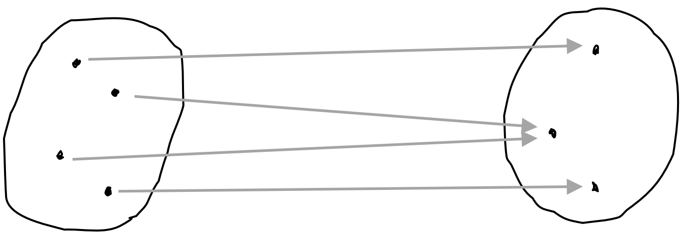


Figure 1.4 -exemple d'application surjective

2.3.3 Bijection

Une application bijective(ou qui réalise une bijection) est une application injective et surjective. Soit $E\ et\ F$ deux ensembles, une application $f: E \to F$ est donc bijective si et seulement si :

$$\forall y \in F, \exists! \, x \in E, y = f(x). \text{ [3]}$$

2.4 Application réciproque

L'application $f: E \to F$ est bijective si et seulement si il existe une application unique

$g: E \to F$ telle que : $fog = Id_F$ et $gof = Id_E$
L'application réciproque de f, notée f^{-1}.

3. Notion sur les espaces mesurables

3.1 Ensembles mesurables

3.1.1 Tribu ou σ-algèbre-Espace mesurable :

Soient E un ensemble, T une famille de parties de E (i.e. $T \subset P(E)$). La famille T est une tribu (on dit aussi une σ-algèbre) sur E si elle vérifie les propriétés suivantes :

• $\emptyset \in T$

• T est stable par union dénombrable, c'est-à-dire que pour toute famille dénombrable $(A_n)_{n \in N}$ d'éléments de T, on a $\bigcup_{n \in N} A_n \in T$.

• T est stable par passage au complémentaire, c'est-à-dire que pour tout $A \in T$, on a

$A^C \in T$ (On rappelle que $A^C = (E \backslash A)$).

- (E, T) est appelé espace mesurable.

Exemples de tribus sur E :

$\{\emptyset, T\}\ et\ P(E)$ Sont des tribus sur E.

3.1.2 Sous tribu

Soient τ_1, τ_2 deux tribus sur Ω, on dit que τ_2 est une sous tribu de $\tau_1$si et seulement si :

$\forall A \in \tau_2,\ \ A \in \tau_1$. [1]

3.1.3 Tribu engendrée

Soient E un ensemble et $C \subset P(E)$. On appelle tribu engendrée par Cla plus petite tribu contenant C, c'est-à-dire la tribu $\tau(C)$ est l'intersection de toutes les tribus sur E contenant C (cette intersection est non vide car $P(E)$ est une tribu contenant C).

3.1.4 Tribu image directe

Si (Ω_2, τ_2) est un espace mesurable, $\Omega_1$est un ensemble et $f\colon \Omega_1 \to \Omega_2$ est une application, l'image directe de τ_1 par f est la tribu $f^*(\tau 1)$ sur Ω_2 définie par :

$$f^*(\tau 1) = \{E_2 \subseteq X_2 / f^{-1}(E_2) \in \tau_1\}.\ \ [1]$$

3.1.5 Tribu image réciproque

Soient Ω_1 un ensemble, (Ω_2, τ_2) un espace mesurable et $f: \Omega_1 \rightarrow \Omega_2$ une application, L'image réciproque de τ_2 par f,$f^{-1}(\tau_2) = \{f^{-1}(E_2) \mid E_2 \in \tau_2\}$ est une tribu sur Ω_1. [1]

3.2 Mesure-Mesure de probabilité

Soit (E, τ) un espace mesurable, on appelle mesure une application $m : \tau \rightarrow \overline{\mathbb{R}}_+$

avec $\overline{\mathbb{R}}+ = [0, +\infty]$ vérifiant :

1. $m(\emptyset) = 0$.

2. m est σ-additive, c'est-à-dire que pour toute famille $(A_n)_{n \in N}$ d'éléments de T disjoints deux à deux (i.e. tels que$A_n \cap A_m \; si \; n \neq m$) on a :

$m(\bigcup_{n \in N} A_n) = \sum_{n \in N} m(A_n)$.

m est dite mesure de probabilité si $m(E) = 1$.

3.3 Espace mesuré, espace probabilisé

Soient (E, τ) un espace mesurable, et m une mesure sur τ. Le triplet (E, τ, m) est appelé espace mesuré. De plus si $m(E) = 1$ cet espace est appelé espace probabilisé.

4. Notion de probabilité

4.1 Variable aléatoire réelle

Soit Ω espace d'épreuves et soit τ une tribu définie sur espace Ω. Une variable aléatoire réelle est une fonction $X: \Omega \rightarrow \mathbb{R}$ tel que pour tous a$\epsilon\mathbb{R}$, l'événement $\{X \leq a\} = \{\omega ; X(w) \leq a\}$ peut se voir attribuer une probabilité, c'est-à-dire, $\{X \leq a\} \in \tau$. [6]

4.2 Vecteur aléatoire

Une variable aléatoire multi variée ou vecteur aléatoire est une liste de variables mathématiques dont la valeur est inconnue, soit parce que la valeur ne s'est pas encore produite, soit parce qu'il y a une connaissance imparfaite de sa valeur, les variable individuelles d'un vecteur aléatoire sont regroupées car elles font toute partie d'un même système mathématique, souvent elles représentent des propriétés différentes d'une unité statistique individuelle.

Les vecteurs aléatoires sont souvent utilisés comme implémentation sous-jacente de divers types de variables aléatoires agrégées par exemple : une matrice aléatoire, un arbre aléatoire, une séquence aléatoire, un processus stochastique, etc. Plus formellement, une variable aléatoire multi variée un vecteur colonne $X = (X_1, \ldots, X_n)^T$ ou Ω est l'espace échantillon, τ est le sigma algèbre (la collection de tous les évènements) et

μ est la mesure de probabilités (une fonction renvoyant la probabilité de chaque évènement). [6]

4.3 Filtration

Sur un espace probabilisé (Ω,τ,μ) on appelle filtration toute suite croissante $(F_n)_{n\in N}$ de sous-tribus de F (on pourra prendre pour F la tribu, notée F_∞ engendrée par les tribus $(F_n)_{n\in N}$ Intuitivement, la tribu $F_n$contient tous les événements qui peuvent survenir avant l'instant n. [6]

4.4 Processus stochastique

On peut définir un processus stochastique comme étant une famille $(X_t)_{t\in T}$ de variables aléatoires indexées par le temps t. Les mots processus et stochastique signifient respectivement fonction et aléatoire. Alors qu'une variable aléatoire X associe à chaque $w\epsilon\Omega$ une réalisation X (w), un processus stochastique $(X_t)_{t\in T}$ associe à chaque w une fonction (ou trajectoire) $(X_t(w)_{t\in T})$: $T \to E, t \to X_t(w)$ Où E est l'espace d'arrivée des variables aléatoires X_t.

Soit $(X_n)_{n\in N}$ une suite de variables aléatoires réelles définies sur (X, τ, μ). On dit que le processus est adapté à la filtration $(F_n)_{n\in N}$ si pour $n \in N$ la variable aléatoire X_n est F_n- mesurable. [10]

4.5 Marche aléatoire

L'idée de marche aléatoire a été introduite en 1905 par le biostatisticien Karl Pearson[1], Mathématiquement, il s'agit d'un processus stochastique à temps discret, c'est-à-dire une suite de variables aléatoires $(Y_n)_{n\geq 0}$, à valeur dans $\mathbb{Z}^d$ définis par :

$$Y_0 = 0\ \mathbb{Z}^d\ et\ \forall n \geq 1 : Y_n = \sum_{i=1}^{n} X_i$$

Où les pas $(X_i)_{i\geq 1}$ est une suite de variable aléatoire indépendante et identiquement distribuée, définie sur un espace de probabilité (X, τ, μ) à valeurs sur le vecteur unité de $\mathbb{Z}^d$. [4]

4.6 Temps d'arrêt

Sur un espace de probabilité (Ω, τ, μ) muni d'une filtration $(F_n)_{n\in N,}$ une variable aléatoire τ : $\Omega \to \mathbb{N} \cup \{\infty\}$ est appelée un temps d'arrêt pour la filtration $(F_n)_{n\in N}$, si et seulement si,

$$\forall n \in N, \{\tau = n\} \in F_n \Leftrightarrow \forall n \in N, \{\tau \leq n\} \in F_n.$$

Si τ est un temps d'arrêt, on définit la tribu des événements antérieurs à τ en posant

$$F_\tau = \{A \in \tau : A \cap \{\tau \leq n\} \in F_n \text{ pour tout } n\epsilon\mathbb{N}.$$

On obtient bien sur une définition équivalente en remplaçant l'événement $\{\tau \leq n\}$ par l'événement $\{\tau = n\}$. On vérifie immédiatement que F_τ est effectivement une tribu et que τ est F_τ mesurable. [14]

4.7 Probabilité conditionnelle

Soit P une probabilité définie sur (Ω, τ) et soit A et B deux évènement de τ. La probabilité conditionnelle de A étant donné B, notée $P(\mathrm{A}/B)$ est définie par :

$$P(\mathrm{A}/B) = \frac{P(\mathrm{A} \cap B)}{P(B)}$$

Si $P(B) = 0$, alors $P(\mathrm{A} \cap B) = 0$ et le second membre $P(\mathrm{A}/B)$ est une forme indéterminée $\frac{0}{0}$ que l'on prendra égale à 0. On lit $P(\mathrm{A}/B)$ « probabilité de A sachant B ». [5]

II. Chaine de Markov à temps discret

1. Définitions générales

1.1 Introduction

Une chaine de Markov est une suite de variable aléatoire $(X_n, n \in \mathbb{N})$ qui permet de modéliser l'évolution dynamique d'un système aléatoire. X_n représente l'état du système à l'instant n. La propriété fondamentale des chaines de Markov, dite propriété de Markov, est que son évolution future ne dépend du passe qu'au travers de sa valeur actuelle. Autrement dit, conditionnellement a X_n, $(X_0, \dots, X_n)$ et $(X_{n+k}, k \in \mathbb{N}^*)$ sont indépendante. Les applications des chaines de Markov sont très nombreuses (réseau, génétique des populations, mathématique financière, gestion de stock, algorithmes stochastique d'optimisation, simulation,…).

1.2 Chaîne de Markov à temps discret

Une suite $(X_n)_{n\geq 0}$ de variables aléatoires à valeurs dans un ensemble au plus dénombrable E est une chaîne de Markov d'espace d'états E si et seulement si pour tout $n \in \mathbb{N}$, pour tout $x_0, \dots, x_{n+1} \in E$ tels que :

$$P\big(X_n = x_{n,\dots,}X_0 = x_0\big) = P(X_n = x_n, \dots, X_0 = x_0) > 0,$$

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n).$$

Cela s'écrit également :

$$\mathcal{L}\,(X_{n+1} = x_{n+1} | X_n = x_n, \dots\dots, X_0 = x_0) = \mathcal{L}(X_{n+1} = x_{n+1} | X_n = x_n)$$

1.3 Chaînes de Martov homogène

Définition : Une chaîne de Markov $(X_n)_n$ est dite homogène si on a pour tout i et j dans E, $P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i)$.[19]

1.4 Probabilité et matrice de transition

Définition 1 : On appelle probabilité de transition pour aller de l'état i à j, la probabilité

$$p_{i,j} = P(X_{n+1} = j / X_n = i) = (P(X_1 = j / X_0 = i)).$$

Définition 2 : On appelle matrice de transition la matrice $P = (p_{i,j})_{\,i,j\epsilon E}$.

Par exemple pour $E = \{1, 2,3,\dots\}$: $P = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3}\cdots \\ p_{2,1} & p_{2,2} & p_{3,2}\cdots \\ \vdots & \vdots & \vdots\;\cdots \end{pmatrix}$

Remarque: Une chaîne de Markov homogène est représentée par sa matrice de transition.

Lemme : On note v_0 la loi de X_0 $(v_0(i_0)) = P(X_0 = i_0)$. On a alors pour tout $(i_0, \dots, i_n)$ dans E

$$P(X_n = i_n, .., X_0 = i_0) = v_0(i_0) \prod_{k=0}^{n-1} p_{i_k, i_{k+1}}.$$

Démonstration : Par conditionnements successifs :

$P(X_n = i_n, \dots \dots \dots, X_0 = i_0) =$

$= P(X_0 = i_0) P(X_1 = i_1 \mid X_0 = i_0) P(X_2 = i_2 \mid X_1 = i_1, X_0 = i_0) \dots \dots \dots \;\dots$

$P(X_n = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$

$= v_0(i_0) \prod_{k=0}^{n-1} P_{i_n, i_{n+1}}$. Car X_n une chaine de Markov

Proposition 1 : Toute matrice de transition vérifié les propriétés suivantes

- Pour tout couple (i,j) de E, $0 \leq p_{i,j} \leq 1$;
- Pour tout $i \in E$, on a $\Sigma_{j \in E} p_{i,j} = 1$.

Démonstration :

$p_{i,j}$ Sont des probabilités, donc le premier point est évident. Le second point découle du fait qu'on somme les probabilités sur toutes les valeurs possibles d'une variable aléatoire. [8]

Proposition 2 :

Soit p une matrice de transition, alors

- p admet 1 comme valeur propre.
- Le vecteur v ayant toutes ses composantes égales à 1 est un vecteur propre associé à cette valeur propre.

Démonstration : il suffit de multiplier la matrice p par le vecteur v. [8]

1.5 Graphe de transition :

Un graphe de transition est un graphe orienté et pondéré vérifiant :

- Tous les poids appartiennent à l'intervalle [0 ;1].
- Il y a au plus une arête entre deux sommets quelquonques du graphe .
- La somme des poids des arêtes issues d'un sommet est égale à 1. [15]

Exemple: soit la matrice de transitin P tel que :

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

Son graphe de transition est :

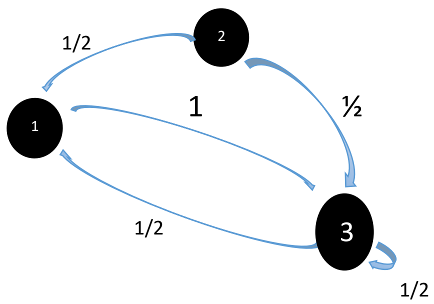


Figure 2.1 -graphe de transition de matrice P

1.6 Loi de probabilité de X_n

Définition : Pour tout entier naturel non nul n, la loi de X_n qu'on note par $\mu(n)$ est donnée par :

$$\mu(n) = \mu(0)P^n$$

En effet,

$$\mu_k(n) = P(X_n = k) = \sum_{i \in E} \mu_i(0)P_{ik}^n$$

$$\mu(n+1) = \mu(n)P$$

Exemple :

Soit $E = \{0.1\}\ et\ \ P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};\ \mu(0) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

$$\mu(1) = \mu(0)P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

$$\mu(2) = \mu(0)P^2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

2. Équation de chapman-Kolmogorov

Soit $P_{i,j}^{(n)}$=$[X_n = j|X_0 = i]$, La probabilitée de passer de i à j en exactement n étapes.

$$\forall\, n \in \mathbb{N}, \forall m \in \mathbb{N} :$$

$$P_{i,j}^{(n+m)} = \sum_{k=0}^{\infty} P(\, X_{n+m} = j | X_n = k) P(\, X_n = j | X_0 = i) = \sum_{k=0}^{\infty} P_{i,k}^{(n)} P_{k,j}^{(m)}$$

En notation matricielle, si $P^{(n)}$est la matrice contenent les $P_{i,j}^{(n)}$, cela donne

$$P^{(n+m)} = P^{(n)} P^{(m)}.$$

En particulier, P^2=PP.

3. Classifications des états

On dit q'un état j est accessible de l'état i s'il existe $n \geq 0$ telle que $p_{ij}^{(n)}$>0. Cela veut dire que si on est dans l'état i, la probabilité d'atteindre j éventuellement n'est pas nulle .Deux états i et j communiquent si chacun est accessible de l'autre. On note cela $i \leftrightarrow j$. La communication est une relation d'équivalence : c'est réflexive, symétrique et transitive. Les classes d'équivalence forment une partition de l'espace d'états E. Ces classes Sont appelées les classes de communication de $(X_n)_n$.

La chaîne est irréductible si tous les états communiquent (une seule classe d'équivalence). Une chaîne de markov n'est pas irréductible est dite réductible. [15]

Exemple :

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

Tous les états communiqent. la chaîne est irréductible.

Exemple :

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Les états ne communiquent pas entre eux. la chaîne et réductible.

3.1 États récurrents et transients

Pour tout état j, désignons par T_j le temps d'atteinte de l'etat j à partir de l'instant 1 ; autrement dit T_j = inf $\{n \geq 1, X_n = j\}$.

Ce temps d'attiente est un temps d'arrêt de la chaîne.

On notera également $N_j = \sum_{n \geq 0} 1_{(X_{n=j})}$ le nombre de passages en j (en comptant le point de départ).

On a en particulier $P_j\left(T_j < \infty\right) = (P_j(N_j > 1)$.

Définition :

On dit que l'état j est récurrent si, partant de l'état j, la probabilité que la chaîne de Markov retourne à l'état j en un temps fini est égale à 1, i.e

$P_j\left(T_j < +\infty\right)$ = $P_j(T_j < +\infty / x_0 = j)$ =1.

Sinon, lorsque $\mathbb{P}\left(T_j < +\infty / X_0 = j\right) < 1$, l'état j est dit transitoire.

Autrement dit si j est récurrent, la chaîne repasse presque sûrement par j en partant de j.

En revanche, un état transitoire est tel que $P\left(T_j < +\infty / X_0 = j\right) > 0$. il y a donc une probabilité strictement positive que la chaîne de Markov ne repasse jamais par l'état j. [12]

3.2 Périodicité d'une chaîne de Markov

Définition :

Soit $j \in E$ On appelle période de j, et on note $d(j)$, le P.G.C.D de tous les entiers $n \geq 0$ pour les quels $P_{jj}^{(n)}$>0 $d(j) = pgcd(n \geq 1, P_{jj}^{(n)}$>0) Si $d(j) = d \geq 2$, on dit que j est périodique de période d.

Si $d(j) = 1$ on dit que j apériodique.

Une chaîne apériodique est une chaine dont tous les états sont apériodiques. [18]

Théorème :

Si i est périodique de periode d et si $i \rightarrow j$ $(j \neq i)$, alors j est aussi périodique de période d : la périodicité est une propriété de classe. [18]

Exemple :

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Son graphe est :

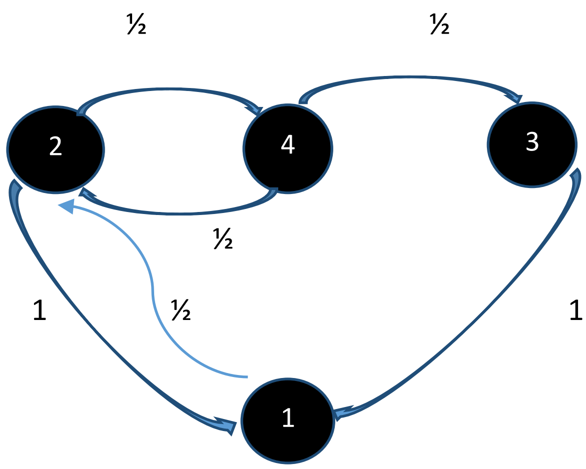


Figure 2.2 - graphe de transition

Nous avons :

$d(1) = pgcd\{n \geq 1, P_{1,1}^{(n)} > 0\}$ =pgsd {2,4,6,8,...},

$d(2)$ = $pgcd$ $\{n \geq 1, P_{2,2}^{(n)} > 0\}$ =pgcd {2,4,6,8,...},

$d(3)$ = $pgcd\{n \geq 1, P_{3,3}^{(n)} > 0\}$ =pgcd {2,4,6,8,...},

$d(4)$ = $pgcd$ $\{n \geq 1, P_{4,4}^{(n)} > 0\}$ =$pgcd$ {2,4,6,8,...},

On constante que la période des états est 2.

III. Les chaines de Markov cachées à temps discret

1. Généralité

Les modèles de Markov cachés (aussi appelés HMM pour Hidden Markov Model), ont été introduits au cours des années 1960 par Baum et ses collègues (théorie de base), ils sont développés par Andrew Markov (étudiant de CHEBYSHEV), pour pouvoir utiliser ces modèles, efficacement, il est nécessaire d'en connaitre les principes qui constituent la théorie des modèles de Markov cachés (MMC).

Les modèles de Markov cachés sont des outils statistiques permettant de modéliser des phénomènes stochastiques, ces modèles sont utilisés dans de nombreux domaines comme la génétique, la finance, l'économie, l'informatique, la météorologie, la biologie, la reconnaissance de la parole, de plus dans le traitement du problème de la segmentation d'image.

Le modèle le plus simple est le modèle de chaine de Markov caché, est un processus aléatoire de variable discrètes évaluées, qui implique un nombre d'état, ces états sont liés par des transitions possibles, chacune avec une probabilité associée et chaque état dispose d'une observation associée, la transition d'état est seulement dépendante de l'état actuel et nom sur les états passés.

La séquence réelle des états n'est pas observable, d'où le nom de caché.

La représentation compacte pour un MMC avec une distribution discrète de probabilité de sortie est donnée par $\lambda = (A, B, \pi)$, qui va être définie dans cette partie.

2. Historique

Les modèles de Markov cachés ont une longue histoire derrière eux, en 1913, les premiers travaux sur les chaines de Markov pour l'analyse du langage permettent à A. Markov de concevoir la théorie des chaines de Markov.

De 1948 à 1951, Shannon conçoit la théorie de l'information en utilisant les chaines de Markov. Dés 1958, les modèles probabilistes d'urnes, le calcul direct du maximum de vraisemblance et l'observation de la suite d'états dans une chaine de Markov, sont réalisés. Mais ce n'est qu'à partir de 1966 avec les travaux de L. E. Baum, que les modèles de Markov cachés, que les algorithmes basiques pour l'estimation des états et des paramètres des modèles, pour les modèles de Markov cachés, sont mis au point. À partir de 1980, ces modèles sont étendus afin d'intégrer la notion de durée variable et densités de probabilités continues multi variables.

Les travaux de A .J. VITERBI et G.D.FORNY ont permis de construire un algorithme efficace et dont la complexité est linéaire, par apport à la longueur de la suite.

3. Chaines de Markov cachées

3.1 Définition :

Une chaine de Markov cachée est une chaine générant un processus stochastique à deux composantes, l'une cachée, l'autre observable tel que :
Un premier processus $X = (X_n)_n$, soit une chaine de Markov homogène, d'espaces d'états E, elle est définie par sa distribution initiale π_0 et sa matrice de transition A. Un deuxième processus$Y = (Y_n)_n$, soit un processus observable définie sur un ensemble Ω et indépendant conditionnellement à X; tel que :

$$P(Y\ /X) = \prod_{n=1}^{N} P(Y_n\ /\ X_n).$$

Proposition :

Soient $(X_n)_{n\in N}$ une chaine de Markov cachée, $(Y_n)_{n\in N}$ une suite de variable aléatoire,

$\forall N \in \mathbb{N}^*$ on a :

$$P(Y_N/X_1, \dots \dots \dots., X_N) = P(Y_N/X_N)$$

Démonstration :

Soient $Y_N \in \Omega, x_1, \dots., x_n \in \mathrm{E}$, on a :

$$P(Y_N = y_N/X_1 = x_1, X_2 = x_2, \dots, X_N = x_N) = \frac{P(Y_N = y_N, X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)}{P(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)}$$

$$= \frac{P(Y_N = y_N, Y_{N-1} \in \Omega, \dots, Y_1 \in \Omega, X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)}{P(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)}$$

$$= \frac{\sum_{y_{N-1}\in\Omega} \dots . \sum_{y_1\in\Omega} P(Y_N = y_N, Y_{N-1} = y_{N-1}, \dots, Y_1 = y_1, X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)}{P(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)}$$

$$= \sum_{y_{N-1}\in\Omega} \dots . \sum_{y_1\in\Omega} P(Y_N = y_N, Y_{N-1} = y_{N-1}, \dots, Y_1 = y_1 / X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)$$

$$= \sum_{y_{N-1}\in\Omega} \dots . \sum_{y_1\in\Omega} P(Y_N = y_N, Y_{N-1} = y_{N-1}, \dots, Y_1 = y_1 / X_N = x_N)$$

$$= \sum_{y_{N-1}\in\Omega} \dots . \sum_{y_1\in\Omega} \prod_{n=1}^{N} P(Y_N = y_N / X_N = x_N)$$

$$= P(Y_N = y_N / X_N = x_N) \times \left[\sum_{y_{N-1}\in\Omega} P(Y_{N-1} = y_{N-1} / X_N = x_N) \dots \sum_{y_1\in\Omega} P(Y_1 = y_1 / X_N = x_N) \right]$$

$$= P(Y_N = y_N / X_N = x_N) \times P(Y_{N-1} \in \Omega / X_N = x_N) \times \cdots \times P(Y_1 \in \Omega / X_N = x_N)$$

$$= P(Y_N = y_N / X_N = x_N).$$

3.2 Graphe d'indépendance d'une chaine de Markov cachée

Les relations de dépendance entre les différentes variables aléatoires d'une chaine de Markov cachée sont schématisées par la figure suivante :

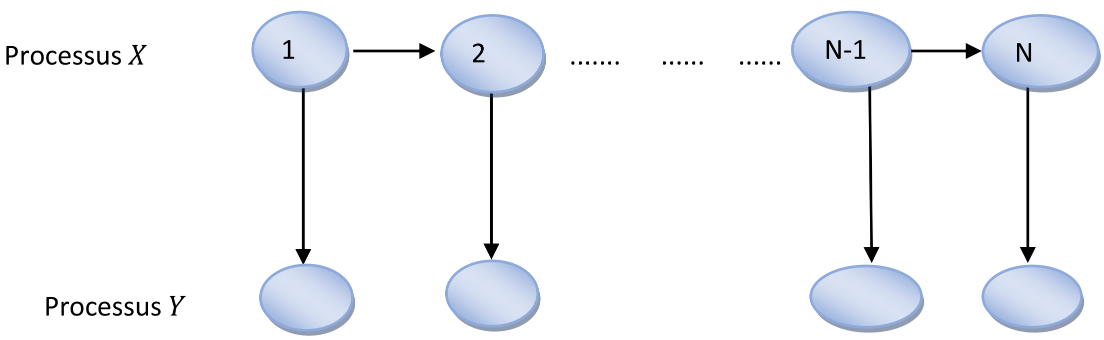


Figure3-1 -graphe d'indépendance d'une chaine de Markov cachée.

3.3 Les éléments d'une Chaine de Markov cachée

Formellement un modèle de Markov caché λ est caractérisé par les éléments Suivants :

K: Le nombre d'états cachés du modèle, on dénote les états individuels par

$$\mathrm{E} = \{e_1, e_2 , \dots , e_K\}$$

M : Le nombre de symboles observé, on dénote les symboles individuels par

$$\Omega = \{ \omega_1 , \omega_2 , \dots , \omega_M\}$$

$A = \{a_{ij}\}$: Matrice de distribution des probabilités de transition, avec :

$$a_{ij} = P\left(X_n = x_j \; / X_{n-1} = x_i \right) \qquad 1 \leq i , j \leq K$$

- $B = \{b_j(y_n)\}$: matrice de distribution des probabilités d'émission avec :

$$b_j(y_n) = P(Y_n = y_n / X_n = x_j)$$

- $\prod = \{\pi_i\}$: un vecteur de distribution des probabilités initiales avec :

$$\pi_i = P(X_1 = x_i), 1 \leq i \leq K$$

N : Longueur de la séquence d'observations Y, avec y_n est le symbole observé à l'instant n .

En résumé, la chaine de Markov cachée est un quintuplé λ qui se définie par

$$\lambda = (K, M, A , B, \Pi).$$

Remarque :

La notation des MMC est très souvent réduite au triplet $\lambda = (A , B, \Pi)$. Car A est une matrice $K \times K$, et B une matrice $K \times M$.

Exemple :

Soit le modèle $\lambda = (A , B, \Pi)$ comprenant quatre états {Printemps, Eté, Hiver, Automne}, chacun permettant d'observer Ω telle que :

$\Omega = \{'Nuage', \quad 'Pluie' , \quad 'Soleil'\}$

$A = \{a_{i\,j}\}$

$B = \{b_j(y)\}$

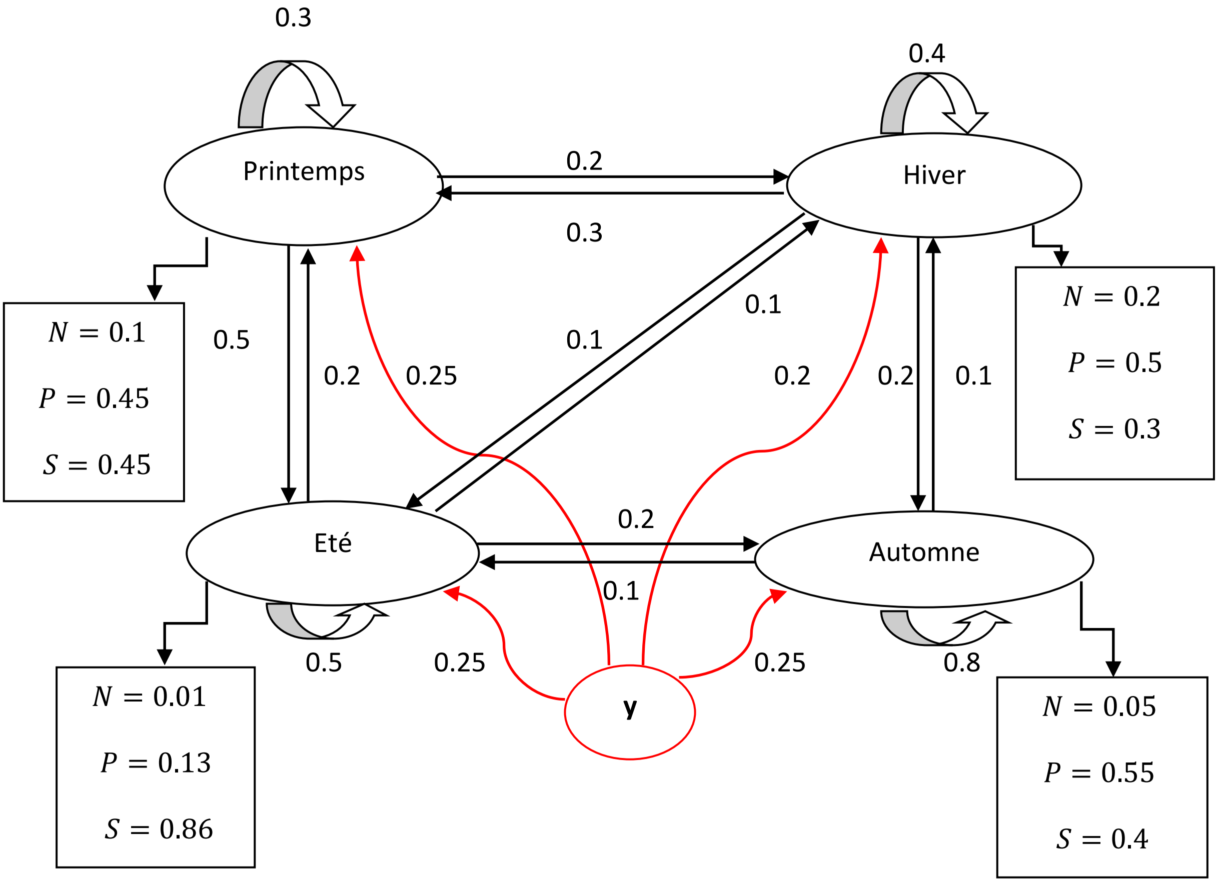


Figure3-2-graphe d'un Modèle de Markov caché [11].

$\lambda = (A, B, \Pi)$ Avec :

$$A = \begin{bmatrix} 0.3 & 0.5 & 0 & 0.2 \\ 0.2 & 0.5 & 0.2 & 0.1 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0.3 & 0.1 & 0.2 & 0.4 \end{bmatrix}, \quad B = \begin{bmatrix} 0.1 & 0.45 & 0.45 \\ 0.01 & 0.13 & 0.86 \\ 0.05 & 0.55 & 0.4 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}, \quad \Pi = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

3.4 Différents types de chaines de Markov cachées

Il existe deux types principaux de topologie des modèles de Markov cachés, le modèle ergodique et le modèle gauche-droite.

3.4.1 Modèle ergodique

Modèle ergodique est sans contrainte : est un cas spécifique, où le modèle est dit complètement connecté, tous les a_{ij} sont positifs, la matrice A est pleine, cela permet d'atteindre n'importe quel état en passant par n'importe quel chemin.

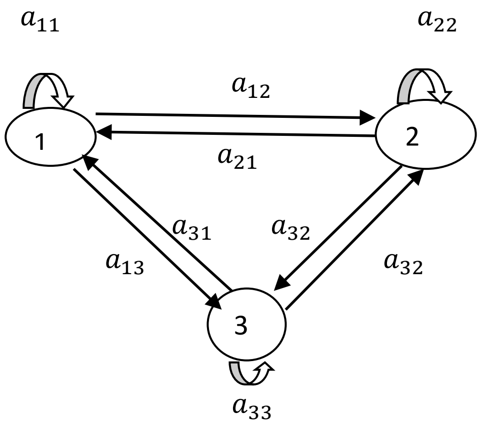


Figure 3-3 -Modèle de chaine de Markov caché ergodique -[17]

3.4.2 Modèle gauche-droite

Modèle gauche- droite est un modèle contenant des contraintes résultant de la mise à zéro de certaines valeurs a_{ij} parce qu'ils n'autorisant aucune transition d'un état vers un autre d'indice inférieur, les états qui se succèdent ont donc des indices égaux ou supérieur aux précédents. Une fois dans le dernier état, le système est condamné à rester c'est pourquoi la probabilité initial du premier état est posée égale à 1, les autres étant égales à 0. Donc la propriété fondamentale de ce genre est :

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

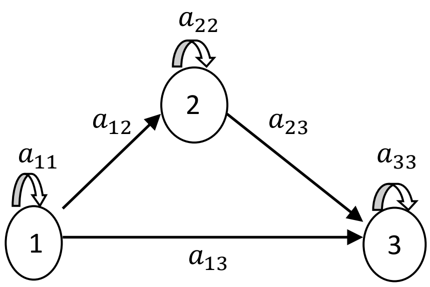


Figure3-4 -Modèle de chaine de Markov caché gauche-droite.

4. Lois de probabilités liées aux chaines de Markov cachées

4.1 Les probabilités jointes d'une chaine de Markov cachée

Soit $X = (X_n)_n$ une chaine de Markov cachée, supposée homogène et stationnaire, et soit $Y = (Y_n)_n$ le vecteur d'observations, les déférentes probabilités jointes liées decettechaine, peuvent s'écrire en fonction des probabilités jointes notées C_{ij}. [20]

4.1.1 Les probabilités jointes C_{ij}

La probabilité jointe C_{ij} donne la probabilité d'être à l'état i à l'instant n et à l'état j à l'instant suivant, elle exprimé par :

$$C_{ij} = P(X_n = x_i \, , X_{n+1} = x_j)$$

4.1.2 Les probabilités initiales π_i

La mesure π_i termine la probabilité que le modèle soit à l'état i au départ, elle peut s'écrire en fonction des probabilités jointes C_{ij} comme suite :

$$\pi_i = P(X_1 = i) = \sum_{1 \le j \le k} c_{ij}$$

4.1.3 Les probabilités de transition a_{ij}

La probabilité a_{ij} représente le saut du processus de la classe x_i à la classe x_j entre deux temps successifs, elle est écrite comme suite :

$$a_{ij} = P(X_{n+1} = x_j \,/\, X_n = x_i) = \frac{c_{ij}}{\sum_{1 \le j \le k} c_{ij}}$$

4.1.4 La loi de X

C'est la probabilité à priori de X, noté soit $P(X)$ est définie par la connaissance des probabilité initiales π_i de X_1, et ses distributions de transition a_{ij}, elle s'écrit par

$$P(X) = P\left(X_1 = x_{X_1}, X_2 = x_{X_2}, \dots \dots \dots, X_N = x_{X_N}\right)$$

$$= \pi_{x_1} \cdot a_{x_1 x_2} \cdot a_{x_2 x_3} a_{x_3 x_4} \cdot \dots \dots . \cdot a_{x_{N-1} x_N} \text{ . [7]}$$

4.1.5 Loi du couple (X, Y)

La loi de X étant définie ci-dessus il reste à définir la loi de Y conditionnellement à X. On suppose que les variables aléatoires (Y_n) sont indépendantes conditionnellement à X.
On note $f_1, f_2, \dots, f_k$ les densités des lois de Y_n conditionnellement à $X_1 = x_1$, $X_2 = x_2, \dots, X_n = x_n$ respectivement.
Dans la suite N sera fixé et on note $X = (X_1, X_2, \dots \dots \dots, X_N)$ et $Y = (Y_1, Y_2 =, \dots \dots \dots, Y_N)$ les suites finies de variables aléatoires et $x = (x_1, x_2, \dots \dots \dots, x_n)$ et $y = (y_1, y_2, \dots \dots \dots, y_n)$ respectivement leurs réalisations.
f_x etant la densité de la loi de Y conditionnelle à $X = x$, la loi de (X, Y) est définie par :

$$h(x,y) = P(X = x) \cdot f_x(y)$$

qui est sa densité par rapport au produit de la mesure de dénombrement sur Ω^N par la mesure de Lebesgue sur R^{n}.

On peut écrit :

$$\mathrm{P(X)} = \pi_{i_1} a_{i_1 i_2} \dots \dots a_{i_{N-1} i_N}$$

Etant donné l'indépendance des Y_m conditionnellement à X, on a la factorisation suivante :

$$f_x(y) = \prod_{n=1}^{N} f x_n(y_n)$$

Finalement on peut écrire:

$$h(x,y) = \pi_{i_1} f_{x_{i\,1}}(y_1) a_{i_1 i_2} f_{x_{i\,2}}(y_2) \dots \dots a_{i_{n-1} i_n} f_{x_{i\,N}}(y_N).\ [7]$$

4.1.6 Les probabilités de FORWARD et BACKWARD

La définition des probabilités FORWARD -BACKWARD qui jouent un rôle crucial aussi bien au niveau de l'estimation des paramètres qu'à celui de la restauration proprement dite.

4.1.6.1 Procédure FORWARD :

$\alpha_t(i)$ Désigne la probabilité suivante :

$$\alpha_t(i) = P(X_t = x_i\,, Y_1 = y_1, Y_2 = y_2, \dots\dots, Y_t = y_t)$$

Cette expression peut se calculer de manière récursive :

On a :

$$\alpha_1(i) = \pi_i f_i(y_1)\ \ , 1 \le j \le k$$

$$\alpha_{t+1}(j) = P\big(X_{t+1} = x_j\,, Y_1 = y_1, Y_2 = y_2, \dots\dots, Y_t = y_t, Y_{t+1} = y_{t+1}\big)$$

$$= \sum_{1 \le i \le k} P(X_{t+1} = x_j\,, X_t = x_i Y_1 = y_1, Y_2 = y_2, \dots\dots, Y_t = y_t\,, Y_{t+1} = y_{t+1})$$

$$= \sum_{1 \le i \le k} P(X_{t+1} = x_j\,, Y_{t+1} = y_{t+1} / \, X_t = x_i) \cdot P(X_t = x_i, Y_1 = y_1, Y_2 = y_2, \dots\dots, Y_t = y_t\,)$$

D'où la relation dite d'induction :

$$\alpha_{t+1}(j) = \left(\sum_{1 \le i \le k} \alpha_t(i)\, a_{ij} \right) \cdot f_j(y_{t+1}), \quad 1 \le j \le k$$

4.1.6.2 Procédure BACKWARD :

La probabilité BACKWARD $\beta_t(i)$ est définie par :

$$\beta_t(i) = P(Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, \ldots\ldots, Y_T = y_T / X_t = x_i)$$

Les probabilités de BACKWARD se calculent par récurrence descendante alors :

$$\beta_t(i) = \sum_{1 \le j \le k} P(Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, \ldots\ldots, Y_T = y_T / X_t = x_i)$$

$$= \sum_{1 \le j \le k} P\big(X_{t+1} = x_j, Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, \ldots\ldots, Y_T = y_T / X_t = x_i\big)$$

$$= \sum_{i \le j \le k} a_{ij} \cdot P\big(Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, \ldots\ldots, Y_T = y_T / X_{t+1} = x_j\big)$$

$$= \sum_{i \le i \le k} a_{ij} \cdot f_j(y_{t+1}) \cdot P\big(Y_{t+2} = y_{t+2}, Y_{t+3} = y_{t+3}, \ldots\ldots, Y_T = y_T / X_{t+1} = x_j\big)$$

Par induction elle devient :

$$\beta_t(i) = \sum_{i \le j \le k} a_{ij} \cdot f_j(y_{t+1}) \cdot \beta_{t+1}(j)$$

Remarque :

Notons que la loi de Y peut se calculer de manières diverses, par exemple :

$$P(Y = y) = \sum_{i \le i \le k} P(X_t = x_i, Y = y)$$

Ou bien :

$$P(Y = y) = \sum_{i \le i \le k} \alpha_T(i)$$

Ou encore :

$$P(Y = y) = \sum_{i \le i \le k} \pi_i \cdot f_i(y_1) \cdot \beta_1(i).$$

4.2 Les probabilités a posteriori d'une chaine de Markov cachée

Le calcule respectif des probabilités FORWARD et BACKWARD se heurte à des difficultés d'ordre numérique,en effet, les quantités figurant dans les expressions de $\alpha_t(i)\ et\ \beta_t(i)$ sont très petites et la présence des sommes dans ces expressions ne permet pas d'utiliser le logarithme. Pour éviter ce problème de dépassement de capacité, Divijver et ses collègues, ont proposé une formulation, qui consiste à remplacer les probabilités conjointes par les probabilités a posteriori.

4.2.2 Les probabilités a posteriori marginale

Considérons la probabilité que le système soit à l'état x_i à l'instant t sachant l'observation

$Y = y$, notée $\xi_t(i)$, tel que :

$$\xi_t(i) = P(X_t = x_i \,/\, Y = y)$$

Cette probabilité peut s'exprimée en fonction des probabilités de FORWARD et BACKWARD comme suit :

$$\xi_t(i) = \frac{P(X_t=\omega_i\,, Y=y)}{P(Y=y)} = \alpha_t(i)\beta_t(i). \text{ [7]}$$

4.2.3 Les probabilités jointes conditionnelles :

Notée $\psi_t(i,j)$, représentant la probabilité d'être à l'instant t dans la classe ω_i et à l'instant suivant dans la classeω_j, tout en connaissant la séquence d'observation Y, elle s'écrit :

$$\psi_t(i,j) = P\left(X_t = \omega_i\,, X_{t+1} = \omega_j / Y = y\right) = \frac{P\left(X_t = \omega_i\,, X_{t+1} = \omega_j, Y = y\right)}{P(Y = y)}$$

Cette probabilité peut s'exprimée en fonction des probabilités de FORWARD et BACKWARD comme suit :

$$\psi_t(i,j) = \frac{\alpha_t(i)\cdot a_{ij}\cdot f_j(y_{t+1})\cdot \beta_{t+1}(j)}{\sum_{1\le l\le k} f_l(y_{t+1}) \sum_{1\le j\le k} \alpha_t(j)\cdot a_{jl}}. \text{ [7]}$$

5. Les trois problèmes de chaine de Markov cachée

Trois problèmes de bases sont soulevés par l'utilisation des chaines de Markov cachées, qui vont être abordés dans les parties suivantes :

5.1 Evaluation

Etant donné la séquence d'observation Y et le modèle de paramètre $\lambda = (A, B\,, \pi)$, calculer la vraisemblance d 'une séquence d'observation $P(Y/\lambda)$?

On veut trouver la probabilité d'une séquence d'observations Y sachant le modèle λ, donc on détermine la mesure de vraisemblance $P(Y/\lambda)$.

5.1.1 Méthode directe

Rappelons que :

$$P(Y_n / X_n) = b_{X_n}(Y_n) \; \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(1)$$

Et

$$P(Y, X / \lambda) = P(Y / X, \lambda) \cdot P(X / \lambda) \ldots\ldots\ldots\ldots\ldots..(2)$$

Une manière immédiate, pour calculer cette probabilité, est d'énumérer toutes les séquences possibles de longueur N.

Soit $X = (X_1, \ldots\ldots\ldots, X_N)$ une séquence d'état fixée selon la propriété d'indépendance conditionnelle des observations, la probabilité $P(Y, X / \lambda)$ est donnée par :

$$P(Y / X, \lambda) = \prod_{n=1}^{N} P(Y_n / X_n, \lambda) \ldots\ldots\ldots\ldots\ldots\ldots..(3)$$

Comme :

$$P(Y_n / X_n) = b_{X_n}(Y_n)$$

Alors la formule (3) devient comme suite :

$$P(Y / X, \lambda) \quad = \prod_{n=1}^{N} b_{X_n}(Y_n) \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Ou encore :

$$P(Y / X, \lambda) = b_{X_1}(Y_1) \cdot b_{X_2}(Y_2) \cdot \ldots \cdot b_{X_{N-1}}(Y_{N-1}) \cdot b_{X_N}(Y_N) \ldots\ldots\ldots(5)$$

D'autre part on a :

$$P(X / \lambda) = \pi_{X_1} \cdot \prod_{n=1}^{N-1} a_{\,X_n\, X_{n+1}} \quad \ldots\ldots\ldots\ldots\ldots(6)$$

La probabilité d'une séquence d'observation Y sachant λ est donnée donc par la somme de toutes les probabilités jointes, associées aux combinaisons d'états possibles, on écrit :

$$P(Y / \lambda) = \sum_{sur\ tous\ les\ X} P(Y / X\ \lambda) \cdot P(X / \lambda) \ldots\ldots\ldots(7)$$

En substituant (5) et (6) dans (7), on obtient la mesure de vraisemblance comme suit :

$$P(Y/\lambda) = \sum_{sur\ tous\ les\ X} \pi_{X_1} \cdot b_{X_1}(Y_1) \cdot a_{X_1 X_2} \cdot b_{X_2}(Y_2) \cdot a_{X_2 X_3} \cdot \ldots \cdot b_{X_{T-1}}(Y_{T-1}) \cdot a_{X_{N-1} X_N} \cdot b_{X_N}(Y_N) \ldots \ldots \ldots (8)$$

Remarque :

Le calcul précédant de $P(Y/\lambda)$, se fait appel à un nombre de multiplication de l'ordre de $2NK^N$. Cette méthode est très couteuse en temps de calcul.

Exemple :

Pour un nombre d'états K=2, et une séquence d'observations de N=100 de longueur, le nombre totale d'opérations est de l'ordre de $(2 \cdot 100 \cdot 2^{100}) = 100 \cdot 2^{101}$.

5.1.2 Algorithme FORWARD :

La procédure FORWARD : [16]

$$\alpha_t(i) = P(Y_1, Y_2, \ldots . Y_t, X_t = x_i / \lambda) \ldots \ldots \ldots .. (9)$$

Pour présenter rapidement cet algorithme, il est nécessaire de définir les variables FORWARD :

Pour $\boldsymbol{i = 1 \ldots \ldots \ldots \ldots K \; et \; j = 2 \ldots \ldots N}$:

$$\begin{cases} \boldsymbol{\alpha_1(i) = p(Y_1 = y_1, X_1 = x_i / \lambda)} \\ \boldsymbol{\alpha_t(i) = p(Y_1 = y_1, \ldots \ldots .., Y_t = y_t, X_t = x_i / \lambda)} \end{cases}$$

On remarque alors que la relation de récurrence suivante est vérifiée pour tout$\boldsymbol{t = 1 \ldots . N - 1}$ et $\boldsymbol{j = 1 \ldots K}$

$$\boldsymbol{\alpha_{t+1}(j) = \sum_{i=1}^{k} a_t(i)\, a_{i,j} b_j(y_{t+1})}$$

De plus, on a pour$\boldsymbol{P(Y = y/\lambda) = \sum_{i=1}^{k} a_T(i)}$.

1- Initialisation :

$$\boldsymbol{Pour\ \ i = 1 \text{ à } K\ fair}$$

$$\boldsymbol{\alpha_1(i) = \pi_i\, b_i(y_1)}$$

$$\boldsymbol{Fin\ pour}$$

2- Itération :

$$\boldsymbol{Pour\ \ t = 1 \text{ à } N - 1\ faire}$$

$$\boldsymbol{Pour\ j = 1 \text{ à } K\ \ faire}$$

$$\boldsymbol{\alpha_{t+1}(j) = \left[\sum_{i=1}^{k} \alpha_t(i)\, a_{i,j}\right] b_j(y_{t+1})}$$

$$\boldsymbol{Fin\ pour}$$

$$\boldsymbol{Fin\ pour}$$

3- Terminaison :

-Algorithme de FORWARD-[16].

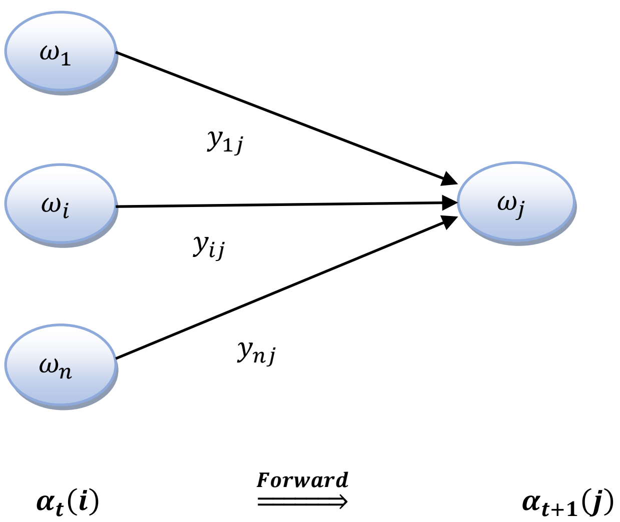


Figure 3-5- séquence d'opérations pour FORWARD. [11]

5.1.3 Algorithme BACKWARD :

Procédure BACKWARD :

$$\beta_t(i) = P(Y_{t+1}\ ,Y_{t+2}\ ,\dots .Y_N/X_t = x_i,\lambda)\ \dots\dots\dots..(10)$$

Cet algorithme permet aussi de calculer la vraisemblance, Les variables de BACKWARD sont définies par :

Pour $\boldsymbol{i = 1 \dots\dots\dots\dots .K\ \ et\ \ t = 1 \dots\dots .N - 1}$

$$\begin{cases} \boldsymbol{\beta_N(i) = 1} \\ \boldsymbol{\beta_t(i) = p(Y_{t+1} = y_{1+1}, \dots\dots\dots ,Y_N = y_N\ ,X_t = x_i/\ \lambda)} \end{cases}$$

Pour tout $\boldsymbol{i = 1 \dots\dots\dots\dots .K\ \ et\ \ t = 1 \dots\dots .N - 1}$ **,** Les relations suivantes sont vérifiées :

$$\boldsymbol{\beta_t(i) = \sum_{i=1}^{k} a_{i,j}\beta_{t+1}(i)\, b_i(y_{t+1})}$$

$$\boldsymbol{P(Y = y\ /\ \lambda) = \sum_{i=1}^{k} \pi_i b_i(y_1)\ \beta_1(i)}$$

1- Initialisation :

$$\boldsymbol{Pour\ \ i = 1 \text{ à } K\ fair}$$

$$\boldsymbol{\beta_N(i) = 1}$$

$$\boldsymbol{Fin\ pour}$$

2- Itération :

$$\boldsymbol{Pour\ \ t =\ \ N - 1 \text{ à } 1\ faire}$$

$$\boldsymbol{Pour\ i = 1 \text{ à } K\ \ faire}$$

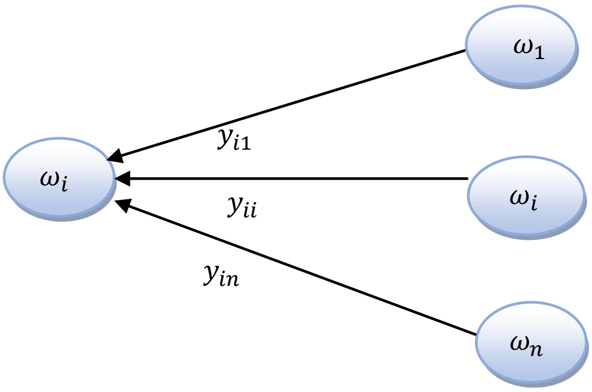
$$\boldsymbol{\beta_t(i) = \sum_{j=1}^{K} a_{i,j}\, \beta_{t+1}(j)\, b_j(y_{t+1})}$$

$$\boldsymbol{Fin\ pour}$$

$$\boldsymbol{Fin\ pour}$$

3- Terminaison :

--



$$\boldsymbol{\beta_t(i)} \quad \xLeftarrow{\textbf{Backward}} \quad \boldsymbol{\beta_{t+1}(j)}$$

Figure 3-6- séquence d'opérations pour BACKWARD [11]

Exemple : [2] [21]

Soit le modèle $\lambda = (A, B, \pi)$ comprenant trois états 1,2,3 chacun permettant d'observer un symbole de l'alphabet {a, b} avec :

$$A = \begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0 & 0.3 & 0.7 \\ 0 & 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \\ 0 & 1 \end{pmatrix}, \pi = \begin{pmatrix} 0.6 \\ 0.4 \\ 0 \end{pmatrix}$$

Le calcul de α pour la suite d'observations (a a b b) est :

$\alpha_1(1) = \pi_1 b_1(a) = 0.6 \times 1 = 0.6$, $\alpha_1(1) = \pi_1 b_1(b) = 0.6 \times 0 = 0$

$\alpha_1(2) = \pi_2 b_2(a) = 0.4 \times 0.5 = 0.2$, $\alpha_1(2) = \pi_2 b_2(b) = 0.4 \times 0.5 = 0.2$

$\alpha_1(3) = \pi_3 b_3(a) = 0 \times 0 = 0$, $\alpha_1(3) = \pi_3 b_3(b) = 0 \times 1 = 0$

$$\alpha_2(1) = (\alpha_1(1)a_{11} + \alpha_1(2)a_{21} + \alpha_1(3)a_{31}) \times b_1(a)$$

$$= (0.6 \times 0.3 + 0.2 \times 0 + 0 \times 0) \times 1 = 0.18$$

$$\alpha_2(2) = (\alpha_1(1)a_{12} + \alpha_1(2)a_{22} + \alpha_1(3)a_{32}) \times b_2(a)$$

$$= (0.6 \times 0.5 + 0.2 \times 0.3 + 0 \times 0) \times 0.5 = 0.18$$

$$\alpha_2(3) = (\alpha_1(1)a_{13} + \alpha_1(2)a_{23} + \alpha_1(3)a_{33}) \times b_3(a)$$

$$= (0.6 \times 0.2 + 0.2 \times 0.7 + 0 \times 0) \times 0 = 0$$

$$\alpha_3(1) = (\alpha_2(1)a_{11} + \alpha_2(2)a_{21} + \alpha_2(3)a_{31}) \times b_1(b)$$

$$= (0.18 \times 0.3 + 0.18 \times 0 + 0) \times 0 = 0$$

$$\alpha_3(2) = (\alpha_2(1)a_{12} + \alpha_2(2)a_{22} + \alpha_2(3)a_{32}) \times b_2(b)$$

$$= (0.18 \times 0.5 + 0.18 \times 0.3 + 0) \times 0.5 = 0.072$$

$$\alpha_3(3) = (\alpha_2(1)a_{13} + \alpha_2(2)a_{23} + \alpha_2(3)a_{33}) \times b_3(b)$$

$$= (0.18 \times 0.2 + 0.18 \times 0.7 + 0) \times 1 = 0.162$$

$$\alpha_4(1) = (\alpha_3(1)a_{11} + \alpha_3(2)a_{21} + \alpha_3(3)a_{31}) \times b_1(b)$$

$$= (0 \times 0.3 + 0.072 \times 0 + 0.162 \times 0) \times 0 = 0$$

$$\alpha_4(2) = (\alpha_3(1)a_{12} + \alpha_3(2)a_{22} + \alpha_3(3)a_{32}) \times b_2(b)$$

$$= (0 \times 0.5 + 0.072 \times 0.3 + 0.162 \times 0) \times 0.5 = 0.0108$$

$$\alpha_4(3) = (\alpha_3(1)a_{13} + \alpha_3(2)a_{23} + \alpha_3(3)a_{33}) \times b_3(b)$$

$$= (0 \times 0.2 + 0.072 \times 0.7 + 0.162 \times 1) \times 1 = 0.2124$$

Donc : $P(a\ a\ b\ b/\lambda) = \sum_{1\leq i\leq 3} \alpha_4(i) = 0.223$

Le calcul de β pour la suite d'observations (a a b b) est :

$\beta_N(i) = 1 \Rightarrow \beta_4(i) = 1$ Alors :

$$\beta_4(1) = 1$$

$$\beta_4(2) = 1$$

$$\beta_4(3) = 1$$

$Pour\ t = 3$

$$\beta_3(1) = (a_{11}b_1(b)\beta_4(1) + a_{12}b_2(b)\beta_4(2) + a_{13}b_3(b)\beta_4(3))$$

$$= (0.3 \times 0 \times 1 + 0.5 \times 0.5 \times 1 + 0.2 \times 1 \times 1) = 0.45$$

$$\beta_3(2) = (a_{21}b_1(b)\beta_4(1) + a_{22}b_2(b)\beta_4(2) + a_{23}b_3(b)\beta_4(3))$$

$$= (0 + 0.3 \times 0.5 \times 1 + 0.7 \times 1 \times 1) = 0.85$$

$$\beta_3(3) = (a_{31}b_1(b)\beta_4(1) + a_{32}b_2(b)\beta_4(2) + a_{33}b_3(b)\beta_4(3))$$

$$= (0 + 0 + 1 \times 1 \times 1) = 1$$

$Pour\ t = 2$:

$\beta_2(1) = (a_{11}b_1(b)\beta_3(1) + a_{12}b_2(b)\beta_3(2) + a_{13}b_3(b)\beta_3(3))$

$= (0.3 \times 0 \times 0.27 + 0.5 \times 0.5 \times 0.85 + 0.2 \times 1 \times 1) = 0.4125$

$\beta_2(2) = (a_{21}b_1(b)\beta_3(1) + a_{22}b_2(b)\beta_3(2) + a_{23}b_3(b)\beta_3(3))$

$= (0 + 0.3 \times 0.5 \times 0.85 + 0.7 \times 1 \times 1) = 0.8275$

$\beta_2(3) = (a_{31}b_1(b)\beta_3(1) + a_{32}b_2(b)\beta_3(2) + a_{33}b_3(b)\beta_3(3))$

$= (0 + 0 + 1 \times 1 \times 1) = 1$

$Pour\ t = 1$:

$\beta_1(1) = (a_{11}b_1(a)\beta_2(1) + a_{12}b_2(a)\beta_2(2) + a_{13}b_3(a)\beta_2(3))$

$= (0.3 \times 1 \times 0.413 + 0.5 \times 0.5 \times 0.8275 + 0.2 \times 0 \times 1) = 0.330625$

$\beta_1(2) = (a_{21}b_1(a)\beta_2(1) + a_{22}b_2(a)\beta_2(2) + a_{23}b_3(a)\beta_2(3))$

$= (0 + 0.3 \times 0.5 \times 0.8275 + 0) = 0.124125$

$\beta_1(3) = (a_{31}b_1(a)\beta_2(1) + a_{32}b_2(a)\beta_2(2) + a_{33}b_3(a)\beta_2(3))$

$= (0 + 0 + 0) = 0$

Donc : $P(a\ a\ b\ b/\lambda = 0) = \sum_{1\le i\le 3} \pi_i b_i(y_1)\beta_1(i)$

$= 0.6 \times 1 \times 0.331 + 0.4 \times 0.5 \times 0.1241 = 0.223$

Donc : $P(a\ a\ b\ b/\lambda) = 0.223.$

Remarque :

On peut conclure qu'il y a plusieurs méthodes pour calculer $P(Y/\lambda)$:

$$P(Y/\lambda) = \sum_{i=1}^{K} \alpha_N(i)$$

$$P(Y/\lambda) = \sum_{i=1}^{K} \pi_i b_i(y_1)\beta_1(i)$$

$$Pour\ tout \quad 1 \le t \le N \quad , P(Y/\lambda) = \sum_{i=1}^{K} \alpha_t(i)\beta_t(i)$$

5.2 Décodage

Etant donné la séquence d'observations Y comment choisir la séquence des états

$X = P(X_1, X_2, \ldots \ldots X_N)$ qui est optimale ?

On cherche donc à décoder les états qui correspondent de mieux à la séquence l'observation Y.

Alors cette partie est la section dans laquelle on découvre les états cachée. Il existe plusieurs méthodes, parmi ces méthodes algorithme de Viterbi. Il s'agit de déterminer meilleur suite d'état$X$, qui maximise la quantité :

$$P(X_n = x_i \,/ Y, \lambda)$$

La solution de ce dernier est basée sur la technique de programmation dynamique :

$$P(X \,/ Y, \lambda) \Longrightarrow maximiserP(X, Y / \lambda)$$

On définit la variable intermédiaire $\delta_t(i)$ comme la probabilité du meilleur chemin amenant à l'état x_i à l'instant t, en étant guidé par les t première observations.

$$\delta_t(i) = max_{(X_{1,\ldots\ldots,}X_{n-1})} P(X_1, X_2, \ldots \ldots, X_{t-1}, X_t = x_i, Y_1, Y_2 \ldots \ldots \ldots, Y_t / \lambda).$$

Par induction, on peut calculer $\delta_{t+1}(j)$ à partir de$\delta_t(i)$ on trouve :

$\delta_{t+1}(j) = \left[max_i \delta_t(i) \cdot a_{ij}\right] \cdot b_j(Y_{t+1})$[13].

5.2.1 L'algorithme de viterbi

1- initialisation :

$$pour\ \ i = 1 \text{ à } K\ fair$$

$$\delta_1(i) = \pi_i \cdot b_i(y_1)$$

$$\psi_1(i) = 0$$

2- récurrence :

$$pour\ \ t = \ 2 \text{ à } N\ faire$$

$$pour\ j = 1 \text{ à } K\ \ faire$$

$$\Psi_t(j) = Arg \max_{1 \le i \le K}\ [\ \delta_{t-1}(i) a_{ij}\ ]$$

$$\delta_t(j) = \max_{1 \le i \le K}[\ \delta_{t-1}(i) a_{ij}\ ]\ b_j(y_t) = \delta_{t-1}\big(\Psi_t(j)\big) a_{\Psi_t(j)} b_j(y_t)$$

$$Fin\ pour$$

$$Fin\ pour$$

3- terminaison :

$$x_N^* = Arg\ \max_{1 \le i \le K}[\delta_N(i)]$$

$$p^* = \max_{1 \le i \le K}[\delta_N(i)] = \delta_N(x_T^*)$$

Retour en arrière :

$$pour\ t = N - 1 \text{ à } 1\ faire$$

$$x_t^* = \Psi_{t+1}(x_{t+1}^*)$$

$$fin\ pour$$

-Algorithme de veterbi – [16]

Remarque :

La fonction $argmax$ permet de mémoriser l'indice $\boldsymbol{i}$, entre 1 et n, avec lequel on atteint le maximum des quantités $\left[\delta_{t-1}(i) \cdot a_{ij}\right]$.

Exemple : [22]

Soit le modèle $\lambda = (A, B, \pi)$ comprenant deux états, la séquence d'observation

$Y = 2,1,3$, avec : $A = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$, $B = \begin{pmatrix} 0.4 & 0.3 \\ 0.1 & 0.6 \\ 0.5 & 0.1 \end{pmatrix}$, $\pi = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}$

$K = 2 \; et \; N = 3$

$\delta_1(i) = \boldsymbol{\pi_i} \cdot \boldsymbol{b_i}(\boldsymbol{y_1})$

$\boldsymbol{Alors}$:

é$\boldsymbol{tape}$ $\mathbf{1}$:

$$\delta_1(1) = \boldsymbol{\pi_1} \cdot \boldsymbol{b_1}(\mathbf{2}) = \mathbf{0.6 \times 0.1 = 0.06}$$

$$\delta_1(2) = \boldsymbol{\pi_2} \cdot \boldsymbol{b_2}(\mathbf{2}) = \mathbf{0.4 \times 0.6 = 0.24}$$

$avec \psi_1(i) = 0$ **C.-à-d. :**

$\psi_1(1) = 0$

$\psi_1(2) = 0$

é$\boldsymbol{tape}$ $\mathbf{2}$:

$\boldsymbol{pour\ t} = \mathbf{2}$

$$\delta_2(1) = \max_{1 \le i \le 2} \begin{Bmatrix} \boldsymbol{\delta_1(1) a_{11}} \\ \boldsymbol{\delta_1(2) a_{21}} \end{Bmatrix} \cdot \; \boldsymbol{b_1}(\boldsymbol{y_2})$$

$$= \max_{1 \le i \le 2} \begin{Bmatrix} \mathbf{0.06 \times 0.7} \\ \mathbf{0.24 \times 0.4} \end{Bmatrix} \cdot \mathbf{0.4}$$

$$= \max_{1 \le i \le 2} \begin{Bmatrix} \mathbf{0.042} \\ \mathbf{0.096} \end{Bmatrix} \cdot \mathbf{0.4 = 0.0384}$$

$\boldsymbol{alors}$: $\psi_2(1) = 2$

$$\delta_2(2) = \max_{1 \le i \le 2} \begin{Bmatrix} \boldsymbol{\delta_1(1) a_{12}} \\ \boldsymbol{\delta_1(2) a_{22}} \end{Bmatrix} \cdot \; \boldsymbol{b_2}(\boldsymbol{y_2})$$

$$= \max_{1 \le i \le 2} \begin{Bmatrix} \mathbf{0.06 \times 0.3} \\ \mathbf{0.24 \times 0.6} \end{Bmatrix} \cdot \mathbf{0.3}$$

$$=\max_{1\le i\le 2}\left\{\begin{matrix}\mathbf{0.018}\\ \mathbf{0.144}\end{matrix}\right\}\cdot \mathbf{0.3} = \mathbf{0.0432}$$

$\boldsymbol{alors}$: $\psi_2(2) = 2$

$\boldsymbol{pour\ t = 3}$

$$\delta_3(1) = \max_{1\le i\le 2}\left\{\begin{matrix}\boldsymbol{\delta_2(1)a_{11}}\\ \boldsymbol{\delta_2(2)a_{21}}\end{matrix}\right\}\cdot\ \boldsymbol{b_1(y_3)}$$

$$=\max_{1\le i\le 2}\left\{\begin{matrix}\mathbf{0.0384\times 0.7}\\ \mathbf{0.0432\times 0.4}\end{matrix}\right\}\cdot \mathbf{0.5}$$

$$=\max_{1\le i\le 2}\left\{\begin{matrix}\mathbf{0.02688}\\ \mathbf{0.01728}\end{matrix}\right\}\cdot \mathbf{0.5} = \mathbf{0.01344}$$

$\boldsymbol{alors}$: $\psi_3(1) = 1$

$$\delta_3(2) = \max_{1\le i\le 2}\left\{\begin{matrix}\boldsymbol{\delta_2(1)a_{12}}\\ \boldsymbol{\delta_2(2)a_{22}}\end{matrix}\right\}\cdot\ \boldsymbol{b_2(y_3)}$$

$$=\max_{1\le i\le 2}\left\{\begin{matrix}\mathbf{0.0384\times 0.3}\\ \mathbf{0.0432\times 0.6}\end{matrix}\right\}\cdot \mathbf{0.1}$$

$$=\max_{1\le i\le 2}\left\{\begin{matrix}\mathbf{0.01152}\\ \mathbf{0.02592}\end{matrix}\right\}\cdot \mathbf{0.1} = \mathbf{0.002592}$$

$\boldsymbol{alors}$: $\psi_3(2) = 2$

é$\boldsymbol{tape\ 3}$:

$$\boldsymbol{x_3^*} = \boldsymbol{Arg}\max_{1\le i\le 2}\left\{\begin{matrix}\boldsymbol{\delta_3(1)}\\ \boldsymbol{\delta_3(2)}\end{matrix}\right\} = \max_{1\le i\le 2}\left\{\begin{matrix}\mathbf{0.01344}\\ \mathbf{0.002592}\end{matrix}\right\} = \mathbf{1}$$

$$\boldsymbol{p^*} = \max_{1\le i\le k}[\boldsymbol{\delta_T(i)}] = \boldsymbol{\delta_T(x_T^*)} = \boldsymbol{\delta_3(1)} = \mathbf{0.01344}$$

é$\boldsymbol{tape\ 4}$:

$$\boldsymbol{x_2^*} =: \psi_3(\boldsymbol{x_3^*}) = \psi_3(1) = 1$$

$$\boldsymbol{x_1^*} =: \psi_2(\boldsymbol{x_2^*}) = \psi_2(1) = 2$$

Donc la séquence d'états cachés la plus probable est **2 1 1** avec une vraisemblance

$\boldsymbol{P(O/\lambda)} = \mathbf{0.01344}$.

5.3 Ré-estimation ou d'apprentissage :

Etant donné un ensemble de séquences d'observations $Y = \{y^1, \dots\dots, y^m\}$et un modèle initiale $\lambda = (A, B, \pi)$, comment ré-estimer les différents paramètres du modèle afin de maximiser la vraisemblance de l'ensemble d'apprentissage $p(Y/\lambda)$?

Pour résoudre ce type de problème, en utilisant la procédure itérative telle que celle de BAUM-WELCH.

Comme on suppose les séquences d'apprentissages tirées indépendamment, l'algorithme de BAUM-WELCH permet d'obtenir ainsi :

$$\lambda^* = \underset{\lambda}{\operatorname{argmax}} \left[p(Y/\lambda) = \prod_{k=1}^{m} p(Y^k/\lambda) \right]$$

L'idée de l'application est donc d'utilisé une procédure de ré-estimation qui affine le modèle petit à petit selon les étapes suivantes:

-Choisir un ensemble initial de paramètreλ_0.

-Calculer λ_1 à partir deλ_0.

- Répéter ce processus jusqu'à un critère de fin. [9]

Pour chaque étape p d'apprentissage, on dispose de λ_p et on cherche un λ_{p+1} qui doit Vérifier :

$$p(Y/\lambda_{p+1}) \geq p(Y/\lambda_p)$$

C'est à dire :

$$\prod_{k=1}^{m} p(Y^k/\lambda_{p+1}) \geq \prod_{k=1}^{m} p(Y^k/\lambda_p)$$

On note :

$$\xi_t(i,j) = p(X_t = x_i, X_{t+1} = x_j / Y, \lambda)$$

Avec la règle de bayes :

$$\xi_t(i,j) = \frac{p(X_t = x_i, X_{t+1} = x_j, Y/\lambda)}{p(Y/\lambda)}$$

Par définition de FORWARD-BACKWARD, on en déduit :

$$\xi_t(i,j) = \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(y_{t+1}) \cdot \beta_{t+1}(j)}{p(Y/\lambda)}$$

On définit aussi la quantité $\psi_t(i)$ comme la probabilité d'être dans l'état caché x_i à l'instant t, étant donné la séquence d'observation Y.

$$\psi_t(i) = p(X_t = x_i \,/\, Y, \lambda)$$

$$\psi_t(i) = \sum_{j=1}^{k} p\big(X_t = x_i\,, X_{t+1} = x_j / \,Y, \lambda\big) = \frac{\sum_{j=1}^{k} p(X_t = x_i\,, X_{t+1} = x_j, Y/\,\lambda)}{p(Y/\lambda)}$$

On a la relation :

$$\psi_t(i) = \sum_{j=1}^{k} \xi_t(i,j) = \frac{\alpha_t(i) \cdot \beta_t(j)}{p(Y/\lambda)}$$

$$\sum_{t=1}^{N-1} \xi_t(i,j) : \textit{nombre des transitions de l'état } x_i \text{ à } \textit{l'état } x_j$$

$$\sum_{t=1}^{N-1} \psi_t(i) : \textit{nombre des transitions sortantes de } x_i$$

La ré-estimation des paramètres $(A, B\,, \Pi\,)$ du modèle markovien λ est telle que :

$$\bar{\pi}_i = \textit{probabilité d'étre dans l'état } x_i \text{ à } \textit{l'instant } t = 1.$$

$$\bar{\pi}_i = \psi_1(i) \quad 1 \le i \le \mathrm{N}$$

$$\bar{a}_{ij} = \frac{\textit{nombre de transition de l'état } x_i \textit{ vers l'état } x_j}{\textit{nombre de fois ou l'on quitte } x_i}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \psi_t(i)} \qquad 1 \le i \le \mathrm{N} \; et \; 1 \le j \le N$$

$$\bar{b}_j(N) = \frac{\textit{nombre de fois ou l'on est dans l'état } x_j \textit{ en observant le symbole } y_k}{\textit{nombre de fois ou l'on est dans l'état } x_j}$$

$$\bar{b}_j(N) = \frac{\sum_{t=1, Y_t = y_N}^{N} \psi_t(j)}{\sum_{t=1}^{N} \psi_t(j)}$$

Ces formules ont été établies par la procédure d'EM à l'apprentissage des paramètres HMM, la suite des modèles construites par l'algorithme de BAUM-WELCH vérifie la relation cherchée : $p(Y\,/\lambda_{p+1}) \ge p(Y\,/\lambda_p)$

5.3.1 Algorithme BAUM-WELCH :

$faire$

$Appliquer\ les\ algorithmes\ \ Forward\ et\ Backward$

$pour\ t = 1\ \text{à}\ T\ faire$

$pour\ i = 1\ \text{à}\ N\ faire$

$pour\ j = 1\ \text{à}\ N\ faire$

$calculer\ \xi_t(i,j)$

$fin\ pour$

$calculer\ \psi_t(i)$

$fin\ pour$

$fin\ pour$

$r\acute{e} - estimer\ \ \lambda = (A, B, \Pi\ )$

$tantque\ (il\ ya\ augmentation\ \ de\ p(Y/\lambda))$

Algorithme de BAUM- WELCH-[16]

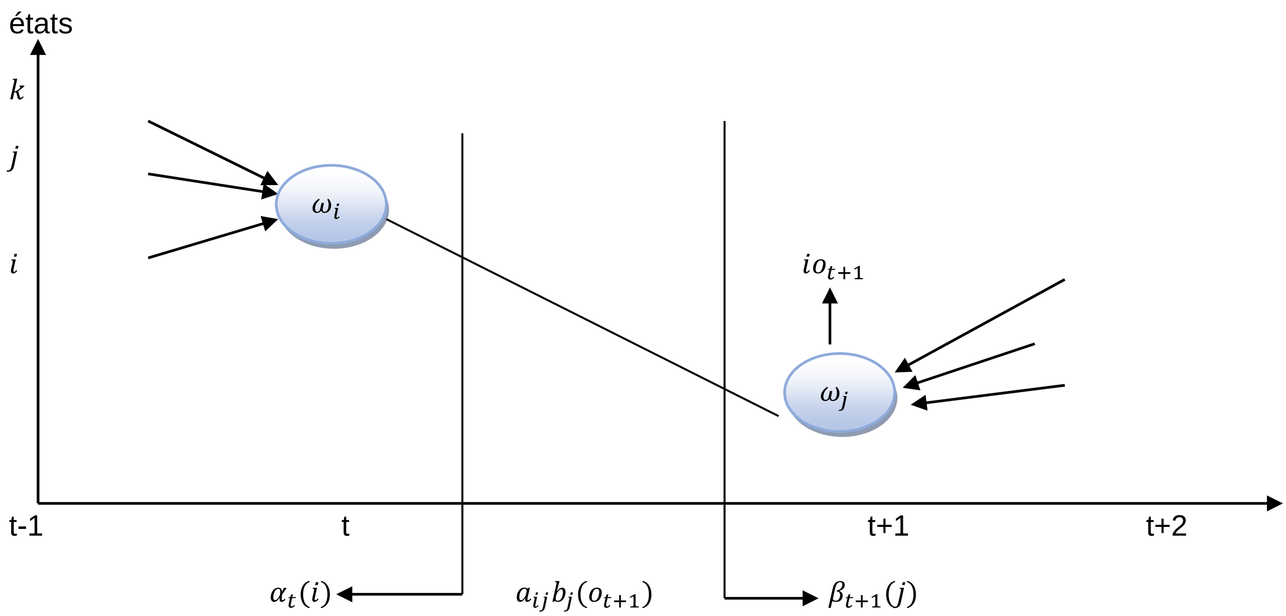


Figure 3-7-séquence d'opération pour BAUM-WELCH-[11]

Exemple : [21]

Soit le modèle $\lambda = (A, B, \pi)$ comprenant trois états 1,2,3 chacun permettant d'observer un symbole de l'alphabet {a, b} avec :

$$Y = (a\ a\ b\ b), \qquad A = \begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0 & 0.3 & 0.7 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \\ 0 & 1 \end{pmatrix}, \qquad \pi = \begin{pmatrix} 0.6 \\ 0.4 \\ 0 \end{pmatrix}$$

Premièrement on applique les algorithmes de FORWARD et BACKWARD Les valeurs de $\alpha_t(i)$:

$$\begin{cases} \alpha_1(1) = 0.6 \\ \alpha_1(2) = 0.2 \\ \alpha_1(3) = 0 \end{cases}$$

$$\begin{cases} \alpha_2(1) = 0.18 \\ \alpha_2(2) = 0.18 \\ \alpha_2(3) = 0 \end{cases}$$

$$\begin{cases} \alpha_3(1) = 0 \\ \alpha_3(2) = 0.072 \\ \alpha_3(3) = 0.162 \end{cases}$$

$$\begin{cases} \alpha_4(1) = 0 \\ \alpha_4(2) = 0.0108 \\ \alpha_4(3) = 0.2124 \end{cases}$$

Les valeurs de $\beta_t(i)$:

$$\begin{cases} \beta_1(1) = 0.330625 \\ \beta_1(2) = 0.124125 \\ \beta_1(3) = 0 \end{cases}$$

$$\begin{cases} \beta_2(1) = 0.4125 \\ \beta_2(2) = 0.8275 \\ \beta_2(3) = 1 \end{cases}$$

$$\begin{cases} \beta_3(1) = 0.45 \\ \beta_3(2) = 0.85 \\ \beta_3(3) = 1 \end{cases}$$

$$\begin{cases} \beta_4(1) = 1 \\ \beta_4(2) = 1 \\ \beta_4(3) = 1 \end{cases}$$

$P(Y/\lambda) = 0.223$

Ré-estimation des matrices$(A, B\ et\ \Pi)$

Ré-estimation de A:

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \psi_t(i)}$$

$$\bar{a}_{11} = \frac{\sum_{t=1}^{T-1} \xi_t(1,1)}{\sum_{t=1}^{T-1} \psi_t(1)} = \frac{\sum_{t=1}^{4-1=3} \xi_t(i,j)}{\sum_{t=1}^{3} \psi_t(i)} = \frac{\sum_{t=1}^{4-1=3} \alpha_t(1) \cdot a_{11} \cdot b_1(y_{t+1}) \cdot \beta_{t+1}(1)}{\sum_{j=1}^{3} \quad \alpha_t(1) \cdot a_{1j} \cdot b_j(y_{t+1}) \cdot \beta_{t+1}(j)}$$

$$\bar{a}_{11} = (0.6 \times 0.3 \times 1 \times 0.4125 + 0.18 \times 0.3 \times 0 \times 0.45 + 0 \times 0.3 \times 0 \times 1)/(0.6 \times 0.3 \times 1 \times 0.4125 + 0.6 \times 0.5 \times 0.5 \times 0.8275 + 0.6 \times 0.2 \times 0 \times 1 + 0.18 \times 0.3 \times 0 \times 0.45 + 0.18 \times 0.5 \times 0.5 \times 0.85 + 0.18 \times 0.2 \times 1 \times 1)$$

$$\bar{a}_{11} = 0.07425/0.272625 = 0.272352$$

$$\bar{a}_{12} = (0.6 \times 0.5 \times 0.5 \times 0.8275 + 0.18 \times 0.5 \times 0.5 \times 0.85 + 0)/0.272625$$

$$\bar{a}_{12} = 0.162375/0.272625 = 0.595598$$

$$\bar{a}_{13} = (0.6 \times 0.2 \times 0 \times 1 + 0.18 \times 0.2 \times 1 \times 1 + 0)/0.272625$$

$$\bar{a}_{13} = 0.036/0.272725 = 0.132050$$

$$\bar{a}_{21} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \psi_t(i)} = \frac{\sum_{t=1}^{4-1=3} \xi_t(2,1)}{\sum_{t=1}^{3} \psi_t(2)} = \frac{\sum_{t=1}^{4-1=3} \alpha_t(2) \cdot a_{21} \cdot b_1(y_{t+1}) \cdot \beta_{t+1}(1)}{\sum_{t=1}^{3} \alpha_t(2) \cdot a_{2j} \cdot b_j(y_{t+1}) \cdot \beta_{t+1}(j)}$$

$$\bar{a}_{21} = (0.2 \times 0 \times 1 \times 0.4125 + 0.18 \times 0 \times 0 \times 0.45 + 0.072 \times 0 \times 0 \times 1)/(0.2 \times 0 \times 1 \times 0.4125 + 0.2 \times 0.3 \times 0.5 \times 0.8275 + 0 + 0.18 \times 0.3 \times 0.5 \times 0.85 + 0.18 \times 0.7 \times 1 \times 1 + 0 + 0.072 \times 0.3 \times 0.5 \times 1 + 0.072 \times 0.7 \times 1 \times 1)$$

$$\bar{a}_{21} = 0/0.234975 = 0$$

$$\bar{a}_{22} = (0.2 \times 0.3 \times 0.5 \times 0.8275 + 0.18 \times 0.3 \times 0.5 \times 0.85 + 0.072 \times 0.3 \times 0.5 \times 1)/0.234975$$

$$\bar{a}_{22} = 0.058575/0.234975 = 0.249282$$

$$\bar{a}_{23} = (0 + 0.18 \times 0.7 \times 1 \times 1 + 0.072 \times 0.7 \times 1 \times 1)/0.234975$$

$$\bar{a}_{23} = 0.1764/0.234975 = 0.750718$$

$$\bar{a}_{31} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \psi_t(i)} = \frac{\sum_{t=1}^{4-1=3} \xi_t(3,1)}{\sum_{t=1}^{3} \psi_t(3)} = \frac{\sum_{t=1}^{4-1=3} \alpha_t(3) \cdot a_{31} \cdot b_1(y_{t+1}) \cdot \beta_{t+1}(1)}{\sum_{t=1}^{4-1=3} \alpha_t(3) \cdot a_{3j} \cdot b_j(y_{t+1}) \cdot \beta_{t+1}(j)}$$

$$\bar{a}_{31} = 0 + 0 + 0/0 + 0 + 0.162 = 0$$

$\bar{a}_{32} = 0 + 0 + 0/0 + 0 + 0.162 = 0$

$\bar{a}_{33} = 0 + 0 + 0.162 \times 1 \times 1 \times 1/0 + 0 + 0.162 = 1$

Donc :

$\bar{a}_{11} \; = 0.272352$

$\bar{a}_{12} \; = 0.595598$

$\bar{a}_{13} \; = \; 0.132050$

$\bar{a}_{21} \; = 0$

$\bar{a}_{22} \; = 0.249282$

$\bar{a}_{23} \; = 0.750718$

$\bar{a}_{31} \; = 0$

$\bar{a}_{32} \; = 0$

$\bar{a}_{33} \; = 1$

Ré-estimation de Π:

$$\bar{\pi}_i = \frac{\alpha_1(i) \cdot \beta_1(i)}{P(Y/\lambda)}$$

$$\bar{\pi}_1 = \frac{\alpha_1(1) \cdot \beta_1(1)}{P(Y/\lambda)} = \frac{0.6 \times 0.331}{0.223} = 0.890$$

$$\bar{\pi}_2 = \frac{\alpha_1(2) \cdot \beta_1(2)}{P(Y/\lambda)} = \frac{0.2 \times 0.1241}{0.223} = 0.1113$$

$$\bar{\pi}_3 = \frac{\alpha_1(3) \cdot \beta_1(3)}{P(Y/\lambda)} = \frac{0 \times 0}{0.223} = 0$$

Ré-estimation de B:

$$\bar{b}_j(N) = \frac{\sum_{t=1, Y_t = y_N}^{N} \; \psi_t(j)}{\sum_{t=1}^{N} \; \psi_t(j)}$$

$\bar{b}_1(1) = (0.6 \times 0.331 + 0.18 \times 0.413)/(0.6 \times 0.331 + 0.18 \times 0.413 + 0 \times 0.27 + 0 \times 1)$

$\bar{b}_1(1) = 1$

Le procédé est alors répété pour toutes les valeurs $\bar{b}_i(j)$, nous obtenons les résultats suivants

$\bar{b}_1(1) = 1$

$\bar{b}_2(1) = 0.707049$

$\bar{b}_3(1) = 0$

$\bar{b}_1(2) = 0$

$\bar{b}_2(2) = 0.292951$

$\bar{b}_3(2) = 1$

Résultats de l'algorithme de Baum-Welch sur 3 itérations

Itération numéro 1 :

Matrice A : $\begin{pmatrix} 0.27259 & 0.5954 & 0.13201 \\ 0 & 0.2451 & 0.74 \\ 0 & 0 & 1 \end{pmatrix}$,

Matrice B : $\begin{pmatrix} 1 & 0 \\ 0.7070 & 0.293 \\ 0 & 1 \end{pmatrix}$,

Matrice Π : $\begin{pmatrix} 0.890 \\ 0.1113 \\ 0 \end{pmatrix}$

Matrice Alpha : $\begin{pmatrix} 0.6 & 0.2 & 0 \\ 0.18 & 0.18 & 0 \\ 0 & 0.072 & 0.162 \\ 0 & 0.0108 & 0.2124 \end{pmatrix}$

Matrice Beta : $\begin{pmatrix} 0.330625 & 0.124125 & 0 \\ 0.4125 & 0.8275 & 1 \\ 0.45 & 0.85 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

Matrice Γ : $\begin{pmatrix} 0.888777 & 0.111223 & 0 \\ 0.332661 & 0.667339 & 0 \\ 0 & 0.274194 & 0.725806 \\ 0 & 0.048387 & 0.951613 \end{pmatrix}$

$P(Y/\lambda) = 0.381487$

Itération numéro 2 :

Matrice A : $\begin{pmatrix} 0.152756 & 0.774101 & 0.073144 \\ 0 & 0.10346 & 0.896534 \\ 0 & 0 & 1 \end{pmatrix}$,

Matrice B : $\begin{pmatrix} 1 & 0 \\ 0.837402 & 0.162598 \\ 0 & 1 \end{pmatrix}$,

Matrice Π : $\begin{pmatrix} 0.970538 \\ 0.029462 \\ 0 \end{pmatrix}$

Matrice Alpha : $\begin{pmatrix} 0.888777 & 0.078640 & 0 \\ 0.212060 & 0.388140 & 0 \\ 0 & 0.070580 & 0.323348 \\ 0 & 0.005154 & 0.376333 \end{pmatrix}$

Matrice Beta : $\begin{pmatrix} 0.416582 & 0.142920 & 0 \\ 0.275778 & 0.810874 & 1 \\ 0.306531 & 0.823745 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

Matrice Γ : $\begin{pmatrix} 0.970538 & 0.029462 & 0 \\ 0.174986 & 0.825014 & 0 \\ 0 & 0.152403 & 0.847597 \\ 0 & 0.013511 & 0.986489 \end{pmatrix}$

$P(Y/\lambda) = 0.603545$

Itération numéro 3 :

Matrice A: $\begin{pmatrix} 0.044302 & 0.938471 & 0.017227 \\ 0 & 0.020091 & 0.979909 \\ 0 & 0 & 1 \end{pmatrix}$,

Matrice B : $\begin{pmatrix} 1 & 0 \\ 0.954970 & 0.045030 \\ 0 & 1 \end{pmatrix}$, Matrice Π: $\begin{pmatrix} 0.996770 \\ 0.003230 \\ 0 \end{pmatrix}$

Matrice Alpha : $\begin{pmatrix} 0.970538 & 0.024671 & 0 \\ 0.148255 & 0.631273 & 0 \\ 0 & 0.029281 & 0.576801 \\ 0 & 0.000493 & 0.603052 \end{pmatrix}$

Matrice Beta : $\begin{pmatrix} 0.619858 & 0.079009 & 0 \\ 0.188106 & 0.911900 & 1 \\ 0.199011 & 0.913358 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

Matrice Γ : $\begin{pmatrix} 0.996770 & 0.0003230 & 0 \\ 0.046206 & 0.953794 & 0 \\ 0 & 0.044311 & 0.955689 \\ 0 & 0.000816 & 0.999184 \end{pmatrix}$

$P(Y/\lambda) = 0.878811$

Chapitre 2

Chaine de Markov Image Cachée

IV. Chaînes de Markov images

L'objectif de cette partie est de répondre à la question suivante :
Est-ce-que l'image d'une chaîne de Markov est une chaîne de Markov ?
Dans le cas général, l'image d'une chaîne de Markov par une fonction mesurable f, n'est pas une chaîne de Markov.

1. L'image d'une chaîne de Markov

Dans les propositions suivantes on montre sous certaines conditions que l'image d'une chaine de Markov est une chaine de Markov.

1.1 Cas f bijective (injective)

Proposition :

Soit $(X_n)_n$ est une chaîne de Markov de matrice de transition $P = (P_{ij})$, et de loi initial Π et $f : E \to F$ une fonction bijective telle que : $Z_n = f(X_n)_n$. Alors $(Z_n)_n$est une chaîne de Markov de matrice de transition $Q = P$ et de loi initiale égale à la loi initiale de X_n.

Autrement dit, la bijection de $f$est une condition suffisante pour que l'image d'une chaîne de Markov soit une chaîne de Markov.

Preuve :

Soient$z_0, z_1, \dots, z_n, z_{n+1} \in F. on\ a\ Z_n = f(X_n)$ donc :

$$P(Z_{n+1} = z_{n+1}/Z_n = z_n, \dots, Z_0 = z_0)$$

$$= P(X_{n+1} = f^{-1}(z_{n+1})/X_n = f^{-1}(z_n), \dots, X_0 = f^{-1}(z_0))$$

$$= P(X_{n+1} = f^{-1}(z_{n+1})/X_n = f^{-1}(z_n), \dots, X_0 = f^{-1}(z_0)) \dots \text{(1)}$$

$$= P(X_{n+1} = f^{-1}(z_{n+1})/X_n = f^{-1}(z_n)) \dots \text{(2)}$$

$$= P(Y_{n+1} = z_{n+1}/Y_n = z_n)$$

(1) : car f est bijective

(2) : car X_n est une chaine de Markov

Remarque : le résultat reste vrai si f est injective.

Exemple :

Soit X_n une chaine de Markov de matrice de transition P et de distribution initiale Π.

$$P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}$$

$$\Pi = \left(\frac{2}{5}, \frac{1}{5}, \frac{2}{5}\right)$$

Et soit f est une fonction injective tel que f : $\{1\ , 2, 3\} \to \{1, 2, 3{,}4\}$ tel que :(1)=3, $f(2)$=1, $f(3)$=2.

$(Z_n)_n$ est une chaine de Markov et sa matrice de transition Q tel que :

$Q_{11} = P(Z_{n+1} = 1/Z_n = 1) = P(X_{n+1} = 2/X_n = 2) = P_{22} = 0.$

$Q_{12} = P(Z_{n+1} = 2/Z_n = 1) = P(X_{n+1} = 3/X_n = 2) = P_{23} = \frac{1}{2}.$

$Q_{13} = P(Z_{n+1} = 2/Z_n = 1) = P(X_{n+1} = 1/X_n = 2) = P_{21} = \frac{1}{2}.$

$Q_{21} = P(Z_{n+1} = 1/Z_n = 2) = P(X_{n+1} = 2/X_n = 3) = P_{32} = 1.$

$Q_{22} = P(Z_{n+1} = 2/Z_n = 2) = P(X_{n+1} = 3/X_n = 3) = P_{33} = 0.$

$Q_{23} = P(Z_{n+1} = 3/Z_n = 2) = P(X_{n+1} = 1/X_n = 3) = P_{31} = 0.$

$Q_{31} = P(Z_{n+1} = 1/Z_n = 3) = P(X_{n+1} = 2/X_n = 1) = P_{12} = 1.$

$Q_{32} = P(Z_{n+1} = 2/Z_n = 3) = P(X_{n+1} = 3/X_n = 1) = P_{13} = 0.$

$Q_{33} = P(Z_{n+1} = 3/Z_n = 3) = P(X_{n+1} = 1/X_n = 1) = P_{11} = 0.$

$$Q = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

1.2 Cas f Surjective

On a vu que si la fonction f est injective ou bijective l'image d'une chaine de Markov est une chaine de Markov mais le cas la fonction f est surjective on ne peut rien dire sur le caractère markovien de suite $f(X_n)$, car dans le cas général, la surjection n'assure pas cette propriété

Proposition 1 :

Soit $(X_n)_n$ chaine de Markov et soit $f : E \to F$ une application mesurable surjective et

$Z_n = f(X_n)$.

Si $\forall\, i,j \in E$,$\forall z \in F$: $\sum_{k\in E/f(k)} P_{ik} = \sum_{k'\in E/f(k')} P_{jk'}$ tel que $f(i) = f(j) = z$. Alors

$(Z_n)_n$ est une chaine de Markov, de probabilités de transition

$Q(x,z) = \sum_{j\in E/f(j)=z} P_{ij}$ Tel que $f(i) = x$

Preuve :

Nous allons prouver que la matrice Q est stochastique :

On a:

$$\sum_{z\in F} Q(x,z) = \sum_{z\in \mathrm{F}} P(Z_{n+1} = z/Z_n = x) = \sum_{z\in F} \frac{P(Z_{n+1} = z/Z_n = x)}{P(Z_n = X)}$$

$$= \frac{1}{(Z_n=x)} \sum_{z\in F} P(Z_{n+1} = z/Z_n = x) = \frac{P(Z_n=x)}{P(Z_n=x)} = 1.$$

Alors Q est une matrice stochastique.

- $(Z_n)_n$ est une chaine de Markov si et seulement si :

$$\forall z_0, z_1, \dots, z_{n+1} \in F$$

$$P(Z_{n+1} = z_{n+1}/Z_n = z_n, \dots, Z_0 = z_0) = P(Z_{n+1} = z_{n+1}/Z_n = z_n).$$

On pose

$f^{-1}(\{z_n\}) = \{x_{in} \in E/f(x_i) = z_0\}\} = \{x_{i_n}, l_{n-1} \le i_n \le l_n\}$.

$f^{-1}(\{z_{n+1}\}) = \{x_{i_{n+1}}, l_n < i_{n+1} \le l_{n+1}\}$

$$P(Z_{n+1} = z_{n+1}/Z_n = z_n, \dots, Z_0 = z_0) = \frac{\sum_{i_{n+1}=1}^{l_{n+1}} . \sum_{i_n=1}^{l_n} \dots \sum_{i_0=1}^{l_0} P(X_{n+1} = x_{i_{n+1}}, \dots, X_0 = x_{i_0})}{\sum_{i_n=1}^{l_n} \dots \sum_{i_0}^{l_0} P(X_n = x_{i_n}, \dots, X_0 = x_{i_0})}$$

$$= \frac{\sum_{i_{n+1}=1}^{l_{n+1}} \sum_{i_n=1}^{l_n} \dots \sum_{i_0=1}^{l_0} P\left(X_{n+1} = x_{i_{n+1}}/X_n = x_{i_n}\right) P\left(X_n = x_{i_n}/X_{n-1} = x_{i_{n-1}}\right) \dots P(X_0 = x_{i_0})}{\sum_{i_n=1}^{l_n} \dots \sum_{i_0}^{l_0} P(X_n = x_{i_n}, \dots, X_0 = x_{i_0})}$$

$$= \frac{\sum_{i_0=1}^{l_0} P\left(X_{n+1} = x_{i_{n+1}}/X_n = x_{i_n}\right) \sum_{i_n=1}^{l_n} \dots \sum_{i_0=1}^{l_0} P\left(X_n = x_{i_n}/X_{n-1} = x_{i_{n-1}}\right) \dots P(X_0 = x_{i_0})}{\sum_{i_n=1}^{l_n} \dots \sum_{i_0}^{l_0} P\left(X_n = x_{i_n}/X_{n-1} = x_{i_{n-1}}\right) \dots P(X_0 = x_{i_0})}$$

$= \sum_{i_{n+1}=1}^{l_{n+1}} P\left(X_{n+1} = x_{i_n+1}/X_n = x_{i_n}\right) \dots (1)$

(1) : Car $P\left(X_{n+1} = x_{i_{n+1}}/X_n = x_i\right) = P\left(X_{n+1} = x_{i_{n+1}}/X_n = x_j\right)$

De même on calcule le second terme :

$$P(Z_{n+1} = Z_{n+1}/Z_n = z_n, \dots, Z_0 = z_0) = \frac{P(Z_{n+1} = z_{n+1}, Z_n = z_n)}{P(Z_n = z_n)}$$

$$= \frac{\sum_{i_{n+1}}^{l_{n+1}} \sum_{i_n=1}^{l_n} P\,(X_{n+1} = x_{i_{n+1}}, X_n = x_{i_n})}{\sum_{i_n=1}^{l_n} P\,(X_n = x_{i_n})}$$

$$= \frac{\sum_{i_{n+1}=1}^{l_{n+1}} P\,(X_{n+1} = x_{i_{n+1}}/X_n = x_{i_n}) \sum_{i_n=1}^{l_n} P\,(X_n = x_{i_n})}{\sum_{i_n=1}^{l_n} P\,(X_n = x_{i_n})}$$

$$= \sum_{i_{n+1}=1}^{l_{n+1}} P\left(X_{n+1} = x_{i_{n+1}}/X_n = x_{i_n}\right) \dots (2)$$

Nous avons (1) = (2), donc $(Z_n)_n$ est bien une chaine de Markov de matrice de transition $Q\,(x, z) = \sum_{j \in E/f(j)=z} P_{ij}$, Tel que $f(i) = x$

En effet, Soit $x, z \in F$

On pose $A = f^{-1}(\{z\}) = \{j_1, j_2, \dots, j_m\}$ et $B = f^{-1}(\{x\}) = \{i_1, i_2, \dots, i_k\}$

$$Q(x, z) = P(Z_{n+1} = z/Z_n = x) = \frac{P(Z_{n+1}=z_{n+1}, Z_n=x)}{P(Z_n=x)}$$

$$= \frac{P(X_{n+1} \in A, Z_n \in B)}{P(X_n \in B)} = \frac{\sum_{r=1}^{m} \sum_{s=1}^{k} P(X_{n+1}=j_r, X_n=i_s)}{\sum_{s=1}^{k} P(X_n=i_s)}$$

$$= \frac{\sum_{r=1}^{m} \sum_{s=1}^{k} P(X_{n+1} = j_r/X_n = i_s).\,P(X_n = i_s)}{\sum_{s=1}^{k} P(X_n = i_s)}$$

$$= \frac{\sum_{r=1}^{m} \sum_{s=1}^{k} P(X_{n+1}=j_r/X_n=i_s).\sum_{s=1}^{k} P(X_n=i_s)}{\sum_{s=1}^{k} P(X_n=i_s)}$$

$$= \sum_{r=1}^{m} P(X_{n+1} = j_r/X_n = i_s) = \sum_{r=1}^{m} P_{i_s j_r}.$$

Exemple 1 :

Soit f une fonction surjective tel que :

$f: \{1, 2, 3\} \to \{0,1\}$ Avec : $f(1) = 0, f(2) = f(3) = 1.$

Et soit $(X_n)_n$ une chaine de Markov de loi initiale $\Pi(0) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ et de matrice de transition,

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

$Q_{00} = P(Z_{n+1} = 0/Y_n = 0) = P(X_{n+1} = 1/X_n = 1) = P_{11} = 0.$

$Q_{01} = P(Z_{n+1} = 1/Y_n = 0) = P(X_{n+1} \in \{2, 3\}/P(X_{n+1} \in \{2,3\}/X_n = 1))$

$$= \frac{P(X_{n+1} \in \{2, 3\}/X_n = 1)}{P(X_n = 1)} = \frac{P(X_{n+1} = 2,\ X_n = 1) + P(X_{n+1} = 3, X_n = 1)}{P(X_n = 1)}$$

$$= \frac{P(X_n = 1). P(X_{n+1} = 2/X_n = 1) + P(X_n = 1)P(X_{n+1} = 3/X_n = 1)}{P(X_n = 1)}$$

$$= \frac{P(X_n = 1)[P_{12} + P_{13}]}{P(X_{n=1})}$$

$$= \frac{1}{2} + \frac{1}{2} = 1.$$

$Q_{10} = P(Z_{n+1} = 1/Z_n = 0) = P(X_{n+1} = 1/X_n \in \{2,3\})$

$$= \frac{P(X_{n+1} = 1, X_n \in \{2, 3\})}{P(X_n \in \{2, 3\})} = \frac{P(X_{n+1} = 1, X_n = 2) + P(X_{n+1} = 1, X_n = 3)}{P(X_n = 2) + P(X_n = 3)}$$

$$= \frac{P(X_n = 2). P(X_{n+1} = 1/X_n = 2) + P(X_n = 3)P(X_{n+1} = 1/X_n = 3)}{P(X_n = 1)}$$

$$= \frac{\frac{11}{34} + \frac{11}{34}}{\frac{1}{3} + \frac{1}{3}} = \frac{2}{12} . \frac{3}{2} = \frac{1}{4}$$

$Q_{11} = P(Z_{n+1} = 1/Z_n = 1) = P(X_{n+1} \in \{2, 3\}/X_n \in \{2,3\})$

$$= \frac{P(X_{n+1} \in \{2,3\}, X_n \in \{2,3\})}{P(X_n \in \{2,3\})}$$

$$= \frac{P(X_{n+1} = 2/X_n = 2) + P(X_{n+1} = 2/X_n = 3)P(X_{n+1} = 3/X_n = 2) + P(X_{n+1} = 3/X_n = 3)}{P(X_n = 2) + P(X_n = 3)}$$

$$= \frac{P(X_n = 2). P(X_{n+1} = 2/X_n = 2) + P(X_n = 3)P(X_{n+1} = 2/X_n = 3)P(X_n = 2). P(X_{n+1} = 3/X_n = 2) + P(X_n = 3)P(X_{n+1} = 3/X_n = 3)}{P(X_n = 2) + P(X_n = 3)}$$

$$= \frac{\frac{1}{3}\frac{1}{2} + \frac{1}{3}\frac{1}{4} + \frac{1}{3}\frac{1}{2}}{\frac{1}{3} + \frac{1}{3}} = \frac{6}{12} \times \frac{3}{2} = \frac{3}{4}$$

$$Q = \begin{pmatrix} 0 & 1 \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

Exemple 2 :

Soit $(X_n)_{n \geq 0}$ une chaine de Markov sur $E = \{1, 2, 3\}$ de matrice de transition P

tel que : $P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ et de loi initiale $\Pi(0) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

Soit $f: E = \{1, 2, 3\} \rightarrow F = \{1, 2\}$ donnée par : $f(1) = f(2) = 1$ et $f(3) = 2$.

On a: $E = \{1, 2, 3\} et F = \{1, 2\},\ f: E \rightarrow F\ avec : f(1) = f(2) = 1.$

Pour $z = 1 \in F$ et $x = 1,\ x' = 2 \in E: f(1) = f(2).$

Donc l'égalité : $P_1\big(f(X_1)\big) = P_2\big(f(X_1)\big) = 1$ est une condition nécessaire pour que $(Z_n)_{n\in N}$ soit une chaine de Markov

On calcule le premier terme :

$P_1(f(X_1) = 1) = P_1(Z_1 = 2)$

$$= P(Z_1 = 2/X_0 = 1) = \frac{P(X_1 = 3, Y_0 = 1)}{P(X_0 = 1)} = 1$$

On calcule le second terme

$P_2(f(X_1) = 2) = P_2(Z_1 = 2)$

$$= P(Z_1 = 2/X_1 = 2) = P(X_1 = 3/X_1 = 2) = 0$$

Donc on conclure que $(Z_n)_{n\in N}$n'est pas une chaine de Markov

V. Chaines de Markov Images Cachées

Nous arrivons à l'objet même de ce mémoire. La question qui se pose à présent est : Dans le cas général, est-ce que l'image d'une chaine de Markov cachée est une chaine de Markov cachée ?

Dans la suite nous donnons des conditions liées à la fonction f et à la matrice de transition P. Pour que ce résultat reste vrai.

Dans toute la suite (X, Y) une chaine de Markov cachée.

1. Les conditions nécessaires et suffisante

1.1 Cas d'une fonction bijective (condition suffisante)

La bijection de f est une condition suffisante pour que l'image d'une chaine de Markov cachée soit une chaine de Markov cachée, comme le montre la proposition suivante.

Proposition :

Si f bijective, (Y, Z) est une chaine de Markov cachée de probabilité de transition

$$IP_{a,b} = P_{f^{-1}(a).f^{-1}(b)} \text{ et } IB_z(y) = P_{f^{-1}(z).}(y)$$

Preuve :

Comme la fonction f est bijective on a : $f(X_n) = Z_n \Leftrightarrow X_n = f^{-1}(Z_n)$

$P(Y_1 = y_1, Y_2 = y_2, \dots \dots .., Y_N = y_N / f(X_1) = z_1, f(X_2) = z_2, \dots .., f(X_N) = z_N)$

$= P(Y_1 = y_1, Y_2 = y_2, .., Y_N = y_N / X_1 = f^{-1}(z_1), X_2 = f^{-1}(z_2), \dots, X_N = f^{-1}(z_N))$...(1)

$$= \prod_{n=1}^{N} P(Y_n = y_n / X_n = f^{-1}(z_n)) \dots .. (2)$$

$$= \prod_{n=1}^{N} P(Y_n = y_n / Z_n = z_n)$$

(1) : car f est bijective, donc inversible.

(2) : car (Y/X) est une chaine de Markov cachée.

Donc, le couple (Y, Z) est une chaine de Markov cachée.

Exemple :

$(X_n)_{n\in N}$ est une chaine de Markov cachée, soit f une fonction injective telle que :

$f: E \to F = \{s, t, u\}$, Avec : $f(1) = s$, $f(2) = t$, $f(3) = u$.

La matrice de transition $A = \begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0 & 0.3 & 0.7 \\ 0 & 0 & 1 \end{pmatrix}$

La matrice d'émission $B = \begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \\ 0 & 1 \end{pmatrix}$

f Bijective, donc (Y, Z) est une chaine de Markov cachée.

Le calcul de $IP_{xy} = P(Z_n = z/Z_{n-1} = x)$

$IP_{s\,s} = P_{11} = 0.3$

$IP_{s\,t} = P_{12} = 0.5$

$IP_{s\,u} = P_{13} = 0.2$

$IP_{t\,t} = P_{22} = 0.3$

$IP_{t\,s} = P_{21} = 0$

$IP_{t\,u} = P_{23} = 0.7$

$IP_{uu} = P_{33} = 1$

$IP_{us} = P_{31} = 0$

$IP_{u\,t} = P_{32} = 0$

Donc d'après les calculs $IP = \begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0 & 0.3 & 0.7 \\ 0 & 0 & 1 \end{pmatrix}$

Et on a $B_j(y) = B = \begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \\ 0 & 1 \end{pmatrix}$

Remarque :

Le résultat reste vraie si f est injective, pas surjective.

1.2 Cas d'une fonction surjective

Proposition :

Soient $f: \mathrm{E} \rightarrow F$ et (Y, X) une chaine de Markov cachée d'espace d'états E, et d'ensemble d'observations Ω si :

- f surjective et $(f(X_n))_n = (Z_n)_{n \in N}$ est une chaine de Markov.
- $\forall\, n \in \mathbb{N}^*, \forall y_n \in \Omega, \forall x_{n-1} \in \mathrm{E}, \forall z_n \in F$

$$P(Y_n = y_n/Z_n = z_n, X_{n-1} = x_{n-1}) = P(Y_n = y_n/Z_n = z_n)$$

Avec $P(Z_n = z_n, X_{n-1} = x_{n-1}) \neq 0,\ P(Z_n = z_n) \neq 0$.

Alors (Y, Z) est une chaine de Markov cachée, appelée Chaine de Markov Image Cachée.

Preuve :

$$P(Y_N = y_N/Z_N = z_N) = \frac{P(Y_N = y_N, Z_N = z_N)}{P(Z_N = z_N)}$$

$$= \frac{P(Y_1 = y_1, Y_2 = y_2, \dots, Y_N = y_N, Z_1 = z_1, Z_2 = z_2, \dots, Z_N = z_N)}{P(Z_1 = z_1, Z_2 = z_2, \dots, Z_N = z_N)}$$

$$= \frac{P(Y_1 = y_1, Y_2 = y_2, \dots, Y_N = y_N, f(X_1) = z_1, f(X_2) = z_2, \dots, f(X_N) = z_N)}{P(f(X_1) = z_1, f(X_2) = z_2, \dots, f(X_N) = z_N)}$$

$$= \frac{P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n, X_1 \in f^{-1}(\{z_1\}), X_2 \in f^{-1}(\{z_2\}), \dots, X_n \in f^{-1}(\{z_n\})}{P(X_1 \in f^{-1}(\{z_1\}), X_2 \in f^{-1}(\{z_2\}), \dots, X_N \in f^{-1}(\{z_N\}))}$$

$$= \frac{\sum_{x_1 \in f^{-1}(\{z_1\})} \sum_{x_2 \in f^{-1}(\{z_2\})} \dots \sum_{x_N \in f^{-1}(\{z_N\})} P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n, X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)}{\sum_{x_1 \in f^{-1}(\{z_1\})} \sum_{x_2 \in f^{-1}(\{z_2\})} \dots \sum_{x_N \in f^{-1}(\{z_N\})} P(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)}$$

$$= \frac{\sum_{x_1 \in f^{-1}(\{z_1\})} \sum_{x_2 \in f^{-1}(\{z_2\})} \dots \sum_{x_N \in f^{-1}(\{z_N\})} P\big(Y_N = y_N/X_1 = x_1, \dots, X_N = x_N) P(X_1 = x_1, \dots, X_N = x_N)\big)}{\sum_{x_1 \in f^{-1}(\{z_1\})} \sum_{x_2 \in f^{-1}(\{z_2\})} \dots \sum_{x_N \in f^{-1}(\{z_N\})} P(X_1 = x_1, \dots, X_N = x_N)}$$

$$= \frac{\sum_{x_1 \in f^{-1}(\{z_1\})} \sum_{x_2 \in f^{-1}(\{z_2\})} \dots \sum_{x_N \in f^{-1}(\{z_N\})} \prod_{n=1}^{N} P\big(Y_N = y_N/X_N = x_N) P(X_N = x_N)\big)}{\sum_{x_1 \in f^{-1}(\{z_1\})} \sum_{x_2 \in f^{-1}(\{z_2\})} \dots \sum_{x_N \in f^{-1}(\{z_N\})} P(X_1 = x_1) P(X_2 = x_2/X_1 = x_1) \dots P(X_N = x_N/X_{N-1} = x_{N-1})}$$

$$= \frac{\sum_{x_1 \in f^{-1}(\{z_1\})} P(Y_1 = y_1/X_1 = x_1) P(X_1 = x_1) \dots \sum_{x_N \in f^{-1}(\{z_N\})} P(Y_N = y_N/X_N = x_N) P(X_N = x_N/X_{N-1} = x_{N-1})}{\sum_{x_1 \in f^{-1}(\{z_1\})} P(X_1 = x_1) \sum_{x_2 \in f^{-1}(\{z_2\})} P(X_2 = x_2/X_1 = x_1) \dots \sum_{x_N \in f^{-1}(\{z_N\})} P(X_N = x_N/X_{N-1} = x_{N-1})}$$

$$= \frac{P(Y_1 = y_1, Z_1 = z_1) P(Y_2 = y_2, Z_2 = z_2) \dots P(Y_N = y_N, Z_N = z_N)}{P(Z_1 = z_1) P(Z_2 = z_2) \dots P(Z_N = z_N)}$$

$$= \prod_{n=1}^{N} P\left(Y_n = \frac{y_n}{Z_n} = y_n\right)$$

Exemple :

$(X_n)_{n\in N}$ est une chaine de markov cachée, soit f une fonction surjective tq :

$f : E \rightarrow F = \{0,1\}$, avec $f(1) = 0, f(2) = f(3) = 1$.

On a la matrice de transition $A = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$

La matrice d'émission $B = \begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \\ 0 & 1 \end{pmatrix}, avec\ l'ensembled'états\ \Omega = \{\mathrm{a}, \mathrm{b}\}$.

Pour la première condition on a :

$P(X_n = 2/X_{n-1} = 1) = P(x_n = 3/X_{n-1} = 1)$ Ou bien $P_{12} = P_{13} = \frac{1}{2}$.

Pour la deuxième condition :

On a $\varphi = P(Z_n = z/Z_{n-1} = x) = \begin{pmatrix} 0 & 1 \\ 1/4 & 3/4 \end{pmatrix}$

Le calcul de $B_j(y)$:

$$B_1(a) = \frac{b_2(a)P_{x2} + b_3(a)P_{x3}}{P_{x2} + P_{x3}}$$

Pour $x = 2$

$$B_1(a) = \frac{b_2(a)P_{22} + b_3(a)P_{23}}{P_{22} + P_{23}} = \frac{0.5 \times 0.5 + 0 \times 0.25}{0.5 + 0.25} = \frac{1}{3}$$

Pour $x = 3$

$$B_1(a) = \frac{b_2(a)P_{32} + b_3(a)P_{33}}{P_{32} + P_{33}} = \frac{0.5 \times 0.25 + 0}{0.25 + 0.5} = \frac{0.5}{3}$$

On remarque que $B_1(a)1_{x=2} \neq B_1(a)1_{x=3}$, donc $f(X_n)$ n'est pas d'une chaine de Markov image cachée.

2. Les éléments d'une chaine de Markov image cachée

Soit(Y, X) une chaine de Markov cachée définie par $\lambda = (K, M, A, B, \pi)$ et soit $f : E \to F$ une application. On note $Z_n = f(X_n)$.

- **Si f bijective (Injective) :**

(Y, Z) est une chaine de Markov cachée définie par $\lambda = (K, M, IA, IB, I\pi)$ tel que

$$\mathrm{IA} = \left(\mathrm{Ia}_{\mathrm{z_i z_j}}\right) 1 \leq \mathrm{i}, \mathrm{j} \leq \mathrm{K}\,, \mathrm{IB} = \left(\mathrm{Ib}_{\mathrm{z_j}}(\mathrm{y_i})\right), 1 \leq \mathrm{j} \leq \mathrm{K}\,, 1 \leq \mathrm{i} \leq \mathrm{M}$$

$$\mathrm{I\pi_i} = \mathrm{P}(\mathrm{Z_1} = \mathrm{z_i}) = \mathrm{P}(\mathrm{X_1} = \mathrm{x_i}), \mathrm{f}(\mathrm{x_i}) = \mathrm{z_i}$$

$$\mathrm{Ia}_{\mathrm{z_i z_j}} = \mathrm{P}\left(\mathrm{Z_n} = \mathrm{z_j}/\mathrm{Z_{n-1}} = \mathrm{x_i}\right) = \mathrm{a_{ij}}$$

$$\mathrm{Ib}_{\mathrm{z_j}}(\mathrm{y_i}) = \mathrm{P}\left(\mathrm{Y} = \mathrm{y_i}/\mathrm{Z_n} = \mathrm{z_j}\right) = \mathrm{P}\left(\mathrm{Y} = \mathrm{y_i}/\mathrm{X_n} = \mathrm{x_j}\right) = \mathrm{b_{x_0}}(\mathrm{y_i})$$

- **Si f surjective non Injective :**

(Y, Z) Est une chaine de Markov cachée définie par $\vartheta = (L, M, IA, IB, I\pi)$ tel que

$$\mathrm{L} < K, IA = \left(\mathrm{Ia}_{\mathrm{z_i z_j}}\right) 1 \leq \mathrm{i}, \mathrm{j} \leq \mathrm{L}$$

$$\mathrm{IB} = \left(\mathrm{Ib}_{\mathrm{z_j}}(\mathrm{y_i})\right), 1 \leq \mathrm{j} \leq \mathrm{L}\,, 1 \leq \mathrm{i} \leq \mathrm{M}$$

$$\mathrm{I\pi_i} = \mathrm{P}(\mathrm{Z_1} = \mathrm{z_i}) = \sum_{\mathrm{f(x)=z_i}} \mathrm{P}(\mathrm{X_1} = \mathrm{x_i}) = \sum_{\mathrm{f(x_i)=z_i}} \mathrm{\pi_{x_i}}$$

3. Lois de probabilité liées aux chaine de Markov image cachée

3.1 Les probabilité jointes

Elle est exprimé par :

$$IC_{ij} = P\left(Z_n = z_i\,, Z_{n+1} = z_j\right) = P\left(f(X_n) = z_i\,, f(X_{n+1}) = z_j\right)$$

3.1.1 f bijective (injective)

$$IC_{ij} = P\left(X_n = f^{-1}(z_i)\,, X_{n+1} = f^{-1}(z_j)\right)$$

3.1.2 f surjective

$$IC_{ij} = P\left(X_n \in f^{-1}(\{z_i\})\,, X_{n+1} \in f^{-1}(\{z_j\})\right)$$

$$= \sum_{x_i \in f^{-1}(\{z_i\})} \sum_{x_j \in f^{-1}(\{z_j\})} P(X_n = x_i\,, X_{n+1} = x_j)$$

$$= \sum_{x_i \in f^{-1}(\{z_i\})} \sum_{x_j \in f^{-1}(\{z_j\})} a_{x_i x_j}$$

3.2 Les probabilités initiales $I\pi_i$

$$I\pi_i = P(Z_1 = z_i) = P(f(X_1) = z_1)$$

3.2.1 f bijective (injective)

$$I\pi_i = P(X_1 = f^{-1}(z_1)), \text{ donc } I\pi = \pi.$$

3.2.2 f surjective

$$I\pi_i = P(X_1 = f^{-1}(\{z_1\})) = \sum_{x_i \in f^{-1}(\{z_i\})} P(X_1 = x_i) = \sum_{x_i \in f^{-1}(\{z_i\})} \pi_{x_i}.$$

3.3 Les probabilités de transition Ia_{ij}

$$Ia_{ij} = P(Z_{n+1} = z_j \,/\, Z_n = z_i)$$

3.3.1 f bijective (injective)

$$Ia_{ij} = P(X_{n+1} = f^{-1}(z_j) \,/\, X_n = f^{-1}(z_i)) = P_{f^{-1}(z_i) f^{-1}(z_j)}$$

3.3.2 f surjective

$$Ia_{ij} = P\big(Z_{n+1} = z_j / X_n = x_i\big),\ avec\ f(x_i) = z_i$$

$$= P(f(X_{n+1}) = z_j \,/\, X_n = x_i) = P(X_{n+1} \in f^{-1}(\{z_j\}) \,/\, X_n = x_i)$$

$$= \sum_{x_j \in f^{-1}(\{z_j\})} P(X_{n+1} = x_j \,/\, X_n = x_i) = \sum_{x_j \in f^{-1}(\{z_j\})} P_{x_i x_j}$$

3.4 La loi de Z

Elle s'écrit par :

$$P(Z_1 = z_1, Z_2 = z_2, \dots, Z_N = z_N) = P(f(X_1) = z_1, f(X_2) = z_2, \dots, f(X_N) = z_N)$$

3.4.1 f bijective (injective)

On pose $x_1 = f^{-1}(z_1), \dots., x_N = f^{-1}(z_N)$

$$P(Z_1 = z_1, Z_2 = z_2, \dots, Z_N = z_N) = P(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)$$

$$= \pi_{x_1} a_{x_1 x_2} a_{x_2 x_3} \dots \dots a_{x_{N-1} x_N}$$

3.4.2 f surjective

$$P(Z_1 = z_1, Z_2 = z_2, \dots, Z_N = z_N) = P(Z_1 = z_1) P(Z_2 = z_2 / Z_1 = z_1) \dots P(Z_N = z_N / Z_{N-1} = z_{N-1})$$

$$= \sum_{f(x_1)=z_1} \pi_{x_1} P(Z_2 = z_2 / X_1 = x_1) \dots . P(Z_N = z_N / X_{N-1} = x_{N-1})$$

$$= \sum_{f(x_1)=z_1} \pi_{x_1} \sum_{f(x_2)=z_2} P(X_2 = x_2 / X_1 = x_1) \dots \sum_{f(x_N)=z_N} P(X_N = x_N / X_{N-1} = x_{N-1})$$

$$= \sum_{f(x_1)=z_1} \pi_{x_1} \sum_{f(x_2)=z_2} \dots \dots \sum_{f(x_N)=z_N} \pi_{x_1} a_{x_1 x_2} a_{x_2 x_3} \dots \dots a_{x_{N-1} x_N}.$$

Chapitre 3

Application

VI Application

3.1 Introduction sur python

Le langage de programmation Python a été créé en 1989 par Guido van Russum, aux Pays-Bas. Le nom *Python* vient d'un hommage à la série télévisée *Monty Python's Flying Circus* dont G. Van Rossum est fan. La première version publique de ce langage a été publiée en 1991.

La dernière version de Python est la version 3. Plus précisément, la version 3.7 a été publiée en juin 2018. La version 2 de Python est désormais obsolète et cessera d'être maintenue après le 1er janvier 2020. Dans la mesure du possible éviter de l'utiliser.

La Python Software Foundation est l'association qui organise le développement de Python et anime la communauté de développeurs et d'utilisateurs.

Ce langage de programmation présente de nombreuses caractéristiques intéressantes :

- Il est multiplateforme. C'est-à-dire qu'il fonctionne sur de nombreux systèmes d'exploitation : Windows, Mac OS X, Android, iOS, depuis les mini ordinateurs Raspberry Pi jusqu'aux supercalculateurs.
- Il est gratuit. Vous pouvez l'installer sur autant d'ordinateurs que vous voulez
- C'est un langage de haut niveau. Il demande relativement peu de connaissance sur le fonctionnement d'un ordinateur pour être utilisé.
- C'est un langage interprété. Un script Python n'a pas besoin d'être compilé pour être exécuté, contrairement à des langages comme C ou le C++.
- I est orienté objet. C'est-à-dire qu'il est possible de concevoir en Python des entités qui miment celles du monde réel (une cellule, une protéine, un atome,…, etc.) avec un certain nombre de règles de fonctionnement et d'interactions.
- Il est relativement simple à prendre en main.
- Enfin, il est très utilise en bionformatique et plus généralement en analyse de données.

Toutes ces caractéristiques font que Python est désormais enseigne dans de nombreuse formations, depuis l'enseignement secondaire jusqu'à l'enseignement supérieur.

3.2 Définition des donnée

Importation des bibioteque Numpy

pour les calcule scientifique

```
Entrée [1]: import numpy as np
```

Définition des données

```
Entrée [4]: """
    A: matrice de Transitions
    B: matrice de probabilite
    PI: distribution initiale
"""
A=np.array([[0.5, 0.3, 0.2],[0.4, 0.2, 0.4],[0.0, 0.3, 0.7]])
B=np.array([[0.9, 0.1],[0.6, 0.4],[0.2, 0.8]])
PI=np.array([0.218, 0.273, 0.509])
np.set_printoptions(formatter={'float': "{: 7.2f}".format})

print(f"A\n{A}")
print(f"B\n{B}")
print(f"PI\n{PI}")
```

```
A
[[   0.50    0.30    0.20]
 [   0.40    0.20    0.40]
 [   0.00    0.30    0.70]]
B
[[   0.90    0.10]
 [   0.60    0.40]
 [   0.20    0.80]]
PI
[   0.22    0.27    0.51]
```

```
Entrée [350]: """
    X: Etats
    Y: Observations
"""
X=np.array([2,1,2])
Y=np.array([1,1,0])
print(X)
print(Y)
```

3.3 Définition d'une chaine de Markov

Chaine de Markov

Entrée [351]:
```
"""
    I: distributions initiale
    T: matrice de transitions
    n: 3
"""
I=np.array([2000, 8000])
T=np.array([[0.8, 0.2],[0.1, 0.9]])
n=3
```

Entrée [352]:
```
def chaine_markov(I, T, n):
    T_n = T
    for i in range(1,n):
        T_n = np.dot(T, T_n)
    return np.dot(I, T_n)
```

Entrée [353]:
```
p=chaine_markov(I, T, n)
print(p)
```

```
[ 2876.0000  7124.0000]
```

3.4 Définition une chaine de Markov cachée

Chaine de Markov Cachée

Entrée [354]:
```
"""
    Fonctions qui retourne la probabilité
"""
def chaine_markov_cachée(PI,A,B,X,Y):
    p=PI[X[0]]*B[X[0],Y[0]]
    for i in np.arange(0, len(X)-1):
        p=p*A[X[i],X[i+1]]*B[X[i+1],Y[i+1]]
    return p
```

Entrée [355]:
```
p=chaine_markov_cachée(PI,A,B,X,Y)
print(f"p(X={X}, Y={Y}={p}")
```

```
p(X=[2 1 2], Y=[1 1 0]=0.00390912
```

3.4.1 Forward algorithme

Farword algorithm :

Entrée [356]:
```
def forward(Y, A, B, PI):
    alpha = np.zeros((Y.shape[0], A.shape[0]))
    alpha[0, :] = PI * B[:, Y[0]]

    for t in range(1, Y.shape[0]):
        for j in range(A.shape[0]):
            alpha[t, j] = alpha[t - 1].dot(A[:, j]) * B[j, Y[t]]

    return alpha

alpha = forward(Y, A, B, PI)
np.set_printoptions(formatter={'float': "{: 7.4f}".format})
print(f"Affichage du resultat final de la matrice alpha:\n{alpha}")
```

```
Affichage du resultat final de la matrice alpha:
[[ 0.0218  0.1092  0.4072]
 [ 0.0055  0.0602  0.2665]
 [ 0.0241  0.0562  0.0423]]
```

3.4.2 Backward algorithme

Farword algorithm :

Entrée [356]:
```
def forward(Y, A, B, PI):
    alpha = np.zeros((Y.shape[0], A.shape[0]))
    alpha[0, :] = PI * B[:, Y[0]]

    for t in range(1, Y.shape[0]):
        for j in range(A.shape[0]):
            alpha[t, j] = alpha[t - 1].dot(A[:, j]) * B[j, Y[t]]

    return alpha

alpha = forward(Y, A, B, PI)
np.set_printoptions(formatter={'float': "{: 7.4f}".format})
print(f"Affichage du resultat final de la matrice alpha:\n{alpha}")
```

```
Affichage du resultat final de la matrice alpha:
[[ 0.0218  0.1092  0.4072]
 [ 0.0055  0.0602  0.2665]
 [ 0.0241  0.0562  0.0423]]
```

3.4.3 viterbi algorithme

Viterbi Algorithm

Entrée [358]:

```python
import numpy as np
from numba import jit

@jit(nopython=True)
def viterbi(A, C, B, O):
    """
    paramètres:
        A (np.ndarray): Matrice de transition
        C (np.ndarray): distribution initial
        B (np.ndarray): Matrice de probabilité
        O : sequence d'observation

    Returns:
        S_opt : sequence optimale
        D : matrice de probabilité accumulée
        E : matrice de tracking
    """
    I = A.shape[0]    # Number of states
    N = len(O)  # Length of observation sequence

    # Initialize D and E matrices
    D = np.zeros((I, N))
    E = np.zeros((I, N-1)).astype(np.int32)
    D[:, 0] = np.multiply(C, B[:, O[0]])

    # Compute D and E in a nested loop
    for n in range(1, N):
        for i in range(I):
            temp_product = np.multiply(A[:, i], D[:, n-1])
            D[i, n] = np.max(temp_product) * B[i, O[n]]
            E[i, n-1] = np.argmax(temp_product)
```

```
    # Backtracking
    S_opt = np.zeros(N).astype(np.int32)
    S_opt[-1] = np.argmax(D[:, -1])
    for n in range(N-2, -1, -1):
        S_opt[n] = E[int(S_opt[n+1]), n]

    return S_opt, D, E


# Apply Viterbi algorithm
S_opt, D, E = viterbi(A, PI, B, Y)
#
print("sequence d'observation:   Y = ", Y)
print('La sequence des etats optimal: S = ', S_opt)
np.set_printoptions(formatter={'float': "{: 7.4f}".format})
print('D =', D, sep='\n')
np.set_printoptions(formatter={'float': "{: 7.0f}".format})
print('E =', E, sep='\n')
```

```
sequence d'observation:   Y =  [1 1 0]
La sequence des etats optimal: S =  [2 2 1]
D =
[[ 0.0218  0.0044  0.0176]
 [ 0.1092  0.0489  0.0410]
 [ 0.4072  0.2280  0.0319]]
E =
[[1 1]
 [2 2]
 [2 2]]
```

3.4.4 Baum-Welch algorithme

baum_welch Algorithm

```
Entrée [359]: def baum_welch(V, a, b, initial_distribution, n_iter=100):
    M = a.shape[0]
    T = len(V)

    for n in range(n_iter):
        alpha = forward(V, a, b, initial_distribution)
        beta = backward(V, a, b)

        xi = np.zeros((M, M, T - 1))
        for t in range(T - 1):
            denominator = np.dot(np.dot(alpha[t, :].T, a) * b[:, V[t +
            for i in range(M):
                numerator = alpha[t, i] * a[i, :] * b[:, V[t + 1]].T *
                xi[i, :, t] = numerator / denominator

        gamma = np.sum(xi, axis=1)
        a = np.sum(xi, 2) / np.sum(gamma, axis=1).reshape((-1, 1))

        # Add additional T'th element in gamma
        gamma = np.hstack((gamma, np.sum(xi[:, :, T - 2], axis=0).resha

        K = b.shape[1]
        denominator = np.sum(gamma, axis=1)
        for l in range(K):
            b[:, l] = np.sum(gamma[:, V == l], axis=1)

        b = np.divide(b, denominator.reshape((-1, 1)))

    return (a,b)
```

```
Entrée [360]: a,b = baum_welch(Y, A, B, PI)
np.set_printoptions(formatter={'float': "{: 7.4f}".format})
print(f"a={a}\n b={b}")
```

```
a=[[ 1.0000  0.0000  0.0000]
 [ 1.0000  0.0000  0.0000]
 [ 0.0000  1.0000  0.0000]]
 b=[[ 1.0000  0.0000]
 [ 0.0000  1.0000]
 [ 0.0000  1.0000]]
```

3.5 chaine de Markov image cachée

Chaines de Markov cachées Image

Entrée [361]:

```
"""
    E: Ensemble d'entré
    F: Ensemble de sortie
"""
E=[1,2,3]
F=[1,2,3]
```

Entrée [362]:

```
len_e=len(E)
len_f=len(F)
c=0
if len_e == len_f:
    c=1
    print("f est une fonction bijective, Yn est une chaine de Markov ima
elif len_e < len_f:
    c=2
    print("f est une fonction injective, Yn est une chiane de Markov ima
elif len_e > len_f:
    c=3
    print("f est une fonction surjective, il faut verifier le CNS")
```

```
f est une fonction bijective, Yn est une chaine de Markov image
```

3.5.1 Fonction pour calcule $Q = IP$

Fonction pour calcul de IP

si f est ''injective'' ou ''bijective'', on peut conclure directement que $Yn = f(Xn)n$ est une chaine de Markov.

Par contre si f est surjective, ce n'est pas suffisant pour conclure, il faut pour cela vérifier une autre CNS.

Entrée [363]:

```
"""
    X: ensemble des entecedante
    Y: ensemble des image Yi associe au antecedent
    on a f(1)=1, f(2)=2, f(3)=2
"""
X=[1,2,3]
Y=[1,2,2]
f_x={"1":1, "2":2, "3":2}
```

Entrée [364]:
```
def PtoI_P(E,F,f_x,P):
    I_P=np.zeros(P.shape)
    for i in np.arange(0, P.shape[0]):
        for j in np.arange(0, P.shape[1]):
            z_j=j+1
            z_i=i+1
            x_i=0
            x_j=0
            for key in f_x:
                if int(f_x[key])==z_i:
                    x_i=int(key)
                if int(f_x[key])==z_j:
                    x_j=int(key)

            I_P[i,j]=P[x_i-1,x_j-1]

    return I_P
```

Si f est surjectif

Entrée [365]:
```
p1=0;
p2=0;

for i in np.arange(0,len_e-1):
    if Y[i]==Y[i+1]:
        l=i
        c=i+1
i_p=PtoI_P(X,Y,f_x,A)
print(f"voici la matrice de transition:\n{A}")
print(f"voici la matrice de transition de (Zn )n\n{i_p}")

for i in np.arange(0,len_e-1):
    p1=p1+i_p[l,i]
    p2=p2+i_p[c,i]
print("\n Conclusion:")
if p1==p2:
    print(f"Yn est une chaine de Markov image pour z={Y[l+1]} et x={l+1}
    h=1
else:
    print("Yn n'est pas une chaine de markov image")
```

```
voici la matrice de transition:
[[ 0.5000  0.3000  0.2000]
 [ 0.4000  0.2000  0.4000]
 [ 0.0000  0.3000  0.7000]]
voici la matrice de transition de (Zn )n
[[ 0.5000  0.2000  0.2000]
 [ 0.0000  0.7000  0.7000]
 [ 0.0000  0.7000  0.7000]]

 Conclusion:
Yn est une chaine de Markov image pour z=2 et x=2 et x'=3.
```

Conclusion

Conclusion :

Conclusion générale

Après les études décrites, on peut répondre à quelques questions et fournir des explications sur le sujet.

L'image d'une chaine de Markov par une fonction f dans le cas général, n'est pas forcément une chaine de Markov. Il existe différentes conditions nécessaires et suffisantes qui permettent à la fonction $f(X_n)$ de conserver le caractère markovien de la chaine de Markov à temps discret $(X_n)_n$

- f est injective.
- f est bijective.
- f est surjective.

Le cas où f est surjective, en général, n'implique pas le caractère markovien de $f(X_n)$, mais avec des conditions particulière posées sur la matrice de transition on peut obtenir ce caractère.

La matrice de transition image Q associée à la chaine de Markov $(Z_n)_n$ est la même que la matrice de transition P de X_n si f est bijective.

Si f est seulement surjective, les éléments de Q sont donnes en fonction des éléments de P et de la loi initiale X_0.

De même dans le cas général l'image d'une chaine de Markov cachée n'est pas une chaine de Markov cachée, et cela à cause de la dépendance de f des probabilités conditionnelles de Y par rapport à Z.

La simulation de ces résultats avec le programme python permet de réaliser numériquement les opérations traitées manuellement dans la partie théorique.

Bibliographie:

[1] A.Giroux. (2004). Mesure et intégration. Dans A. Giroux, *Mesure et intégration* (P. canada).

montréal : université de Montréal

[2] : Antoine, C.Laurent, M.(2014).Apprentissage artificiel Concepts et algorithme(2e

édition,803pages).EDITIONS EYROLLES.

[3] Arnoud Bodin et al. (2016). Cours de mathématique. Dans A. B. al, *Cours de mathématique.*lille : université de Lille

[4] Bergère., B. (s,d.) Marche et scènes aléatoires. Dans B. Bergère, *Marche et scènes aléatoires*.Université paris

[5] Brémaud., P. (2009). Initiation aux probabilités et aux chaines de markov. Dans P.

Brémaud*, initiation aux probabilités tt aux chaines de Markov* (éd. 2e). spinger.

[6] Brémaud., P. (2009). Markov chains Gibbs Fields, Monte Carlo Simulation and Queues. Dans P. Brémoud, *Markov Chains Gibbs fields, monte Carlo Simulation and Queues*. Springer.

[7] Btissam BENMILOUD, Wojciech PIECZYNSKI« Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images . » Traitement du Signal, vol. 12, no. 5, pp. 433–454, 1995

[8] chabonal., J.-J. R.-M.-L. (2012). Chaine de Markov. Dans J.-J. R.-M.-L. chabonal, chaines

Jean-jacquesRuch –Marie-Line chabanol

[9]: Christian GAGNE, 'Modèles de Markov caches (Apprentissage et reconnaissance)',

université de laval, 2010

[10]Coupier., D. (s.d.). *Processus stochastique*. Dans D.coupier, Processus stochastique. Lille :Poly techlille

[11] :Dequier JEROME, 'Chaînes de Markov et applications', Centre D'enseignement De

Grenoble ,2005

[12] Eisc -106/208-chaine de Markov F.simatos 20 mai 2019

[13]: Fatima zahraa AMARA, Meriem SERIARI,"la reconnaissance des battements cardiaques par les HMM, université de Tlemcen, 2011

[14] Ledoux., P. B. (2007). *Probabilité*. Dans P. B. Ledoux, probabilité france : EDP sciences.

Bibliographie :

[15] les mathématiques du chapitre (dans l'ordre de leur apparition) : les ruses Andrei Markov (1856, Fénelon sainte- marie PCIPSI) Conception polycopie : Claudie Hassen forder

[16]:L. R. Rabinier, A tutorial on hidden Markov models and selected applications in speech recognition. In Procedings of the IEEE,Vol.77, pages 257 286, 1989.

[17] :Miloud SEDIRA, 'Application des Modèles de Markov Cachés Dans la Classification des Défauts de Machines Tournantes', laboratoire LMPA, université Ferhat Abbas Sétif Algérie, 2011.

[18] Nassim, T. (2017). Chaine de Markov à temps discret. Dans T. Nassim, chaine de Markov àtemps discret .Bejaia : université de Bejaia

[19] préparation à l'agrégation bordeaux 1 année 2012-2013

[20]: s.aouragh.2006 URL HTT:// thesis.univ-biskra.dz/2288/4/chapitre%202.pdf

[21] :SAHRAEIAN,Sayed Mohammad Ebrahim et YOON,Byung-Jun."Anovel low-complexity HMM similarity measure" . IEEE Signal Processing Letters, 2011,vol.18,no 2,p.87-90.

Grenoble ,2005

[22] :YannisKorilis, Christian St-Jean, Dave DeBarr, Bob Carpenter, Jennifer Chu-Carroll , Modèles de Markov Cachés, (2012)