# Restauration non supervisée des chaînes de Markov couples avec bruitage non gaussien

Taoufik EN-NAJJARY

DEA de Statistiques
Université Pierre et Marie Curie, Paris VI

Institut National des Télécommunications
Stage encadré par le Prof. Wojciech PIECZYNSKI

15 septembre 2001

# Table de matières

# Avant-propos

Le prèsent rapport présente les travaux entrepris par l'auteur dans le cadre d'un stage de recherche faisant partie de sa préparation du *diplôme d'Etudes Approfondies* de statistiques de l'*Université Pierre et Marie Curie*.
Ce stage est effctu au Département *Communications, Images et Traitement de l'Information (CITI)* de l'*Institut national des Télécommunications* sous l'encadrement du Prof. Wojciech PIECZYNSKI.

# Remerciements

Je tiens à remercier mon encadrant Prof. Wojciech PIECZYNSKI pour la considération et l'intérêt qu'il a port à mon travail. Je lui suis trés reconnaissant de m'avoir fait profiter de son expertise en traitement statistiques d'image. Je remercie également M. Stéphane derrode pour son aide informatique. Je tiens également à remecier toute l'équipe du Département Communications, Images et Traitement de l'Information (CITI) pour leur accueil.

# chapitre 1

# introduction

Dans un cadre décisionnel, on dispose d'un échantillon d'apprentissage étiquité, i.e. d'un échantillon Z=(X,Y). Le problème est alors de construire une règle de décision pour classer des futures unités pour lesquelles seules les valeurs $y_i$ seront observées. Il s'agit alors d'un problème courant en reconnaissance statistique des forme ou en diagnostique médicale. De nombreuses méthodes existent. La recherche consiste surtout, à l'heure actuelle, à proposer des techniques répondant à des contextes particuliers et à proposer des méthodes fiables.

**Segmentation et restauration d'image:**

Des mécanismes de dégradation des observations sont souvent inhérents aux problèmes d'images. Dans les problèmes de segmentation, de classification ou de restauration d'image, il s'agit de construire ou de retrouver une image inconnue X lorsque seule une version dégradée Y est observée. Cela relève naturellement des modèles à structure cachée, où les phénomènes aléatoires ont un rôle important. Les images sont constituées d'un ensemble S de pixels qui peuvent prendre une valeur parmi un petit nombre $k$ de couleurs non ordonnées (les classes). Les données mises en jeu sont spatialement localises et induisent l'utilisation de modèles probabilistes spatiaux. Ceux-ci soulève de nombreuses questions de modélisation et d'inférence statistique et n'ont cessé de gagner de l'intérêt. En particulier, le choix de modéles appropriés et l'estimation des paramétres associés aux modèles utilisés sont des questions essentielles pour aller vers une automatisation des algorithmes et tirer tout le profit de la richesse des modèles stochastiques. Ces problèmes, abondamment traités, restent cependant ouverts.

# chapitre 2

# les chaînes de Markov cachées (HMC):

Une chaîne de Markov cachée est un processus à temps discret doublement stochastique composé de deux processus $X = (X_n)_{n\in\mathcal{N}}$ et $Y = (Y_n)_{n\in\mathcal{N}}$. Le processus X est une chaîne de markov et le processus Y est réel. L'appellation cachée signifie que les réalisations de X sont inobservables, le problème est alors de l'estimation de la réalisation de X a partir de la réalisation de Y.

## 2.1 Loi de X

Une suite de variables alèatoires $(X_n)_{n\in\mathcal{N}}$ prenant leurs valeurs dans l'ensemble des classes $\Omega = \{\omega_1, ..., \omega_k\}$ est une chaîne de Markov si elle vérifie pour tout $n \geq 1$ :

$P[X_n = x_n | X_0 = x_0, ..., X_{n-1} = x_{n-1}] = P[X_n = x_n | X_{n-1} = x_{n-1}]$

la loi de $X = (X_n)_{n\in\mathcal{N}}$ est alors déterminée par la loi $\pi$ de $X_0$, dite loi initiale, et la suite des matrices de transition

$a_{ij}^n$=$P[X_{n+1} = \omega_j | X_n = \omega_i]$.

Dans ce travail nous supposerons $c_{ij} = P[X_n = \omega_i, X_{n+1} = \omega_j]$
ne dépendent pas de n.La loi initiale est alors donnèes par

$\pi_i$=$P[X_1 = \omega_i]$=$\sum\limits_{1\leq j\leq k} c_{ij}$

et la matrice de transition,également indépendante de n,donnée par:

$$a_{ij} = P[X_{n+1} = \omega_j | X_n = \omega_i] = \frac{c_{ij}}{\sum\limits_{1\leq j\leq k} c_{ij}}$$

Ainsi, par le théorème de Kolmogorov, les $k^2$ paramètres $(c_{ij})_{1\leq i\leq k, 1\leq j\leq k}$ déterminent la loi du processus X.
La loi de probabilité du vecteur aléatoire $X_n = (X_1, ..., X_n)$ s'ecrit:

$$P[X_1 = \omega_1, X_2 = \omega_2, ..., X_n = \omega_n] = \pi_{i_1} a_{i_1 j_1} ... a_{i_{n-1} i_n}$$

## 2.2 loi du couple (X,Y)

La loi du couple (X,Y) est définie par la loi de X (définie ci-desus ) et les lois de Y conditionnelles à X. Nous supposerons que les variables aléatoires $(Y_n)$ sont indépendantes conditionnellement à X. De plus, la loi de chaque $Y_n$ conditionnelle à X est égale à sa loi conditionnelle à $X_n$. Nous notrons $f_1, ..., f_k$ les densités des lois de $Y_n$ conditionnelles à $X_n = \omega_1, X_n = \omega_2, ..., X_n = \omega_k$ respectivement, ainsi les densités $f_1, ..., f_k$définissent toutes les lois de Y conditionnelles à X.
Dans la suite n sera fixé et nous noterons $X = (X_1, ..., X_n)$, $Y = (Y_1, ..., Y_n)$
les suites finies de variables aléatoire et $x = (x_1, ..., x_n)$,$y = (y_1, ..., y_n)$leurs réalisations.
$f_x$ étant la densité de la loi de Y conditionnelle à X=$x$, la loi de (X,Y) est définie par :

$$h(x, y) = P[X = x] f_x(y)$$

Étant donné l'indépendance des $Y_n$ conditionnellement à X, on a la factorisation suivante:

$$f_x = \prod_{m=1}^{n} f_{i_m}(y_m)$$

Or nous avons vu que la probabilité de $X = x = (\omega_{i_1}, ..., \omega_{i_n})$ s'écrit:

$$P[X = x] = \pi_{i_1} a_{i_1 j_1} ... a_{i_{n-1} i_n}$$

finalement $h$ s'écrit:

$$h(x, y) = \pi_{i_1} f_{\omega_{i_1}}(y_1) a_{i_1 i_2} f_{\omega_{i_2}}(y_2) ... a_{i_{n-1} i_n} f_{\omega_{i_n}}(y_n)$$

## 2.3 loi de X a posteriori

La loi de X a posteriori, i.e. conditionnelle à Y=y, contient toute l'information sur X contenue dans l'observation Y=y, ce qui jout un rôle important pour l'estimation de la réalisation cachée de X à partir de la réalisation observée Y=y. Elle est donnée par :

$$\begin{aligned} P^{Y=y}[X=x] &= P[X=x|Y=y] \\ &= \frac{h(x,y)}{\sum\limits_{x\in\Omega^n} h(x,y)} \end{aligned}$$

on montre que la distribution de X a posteriori est celle d'une chaîne de Markov non stationnaire. Plus précisément, les matrices de transitions de cette distribution sont :

$$\begin{aligned} a_{ij}^{m+1} &= P^{Y=y}[X_{m+1}=\omega_j|X_m=\omega_i] \\ &= \frac{a_{ij}f_j(y_{m+1})\beta_{m+1}(j)}{\sum\limits_{j=1}^{k} a_{ij}f_j(y_{m+1})\beta_{m+1}(j)} \end{aligned}$$

où les $\beta_j$, dite probabilité "backward" est définie dans la paragraphe suivant.

## 2.4 Probabilités Forward-Backward

Nous définissons dans ce paragraphe les probabilités Forward-Backward qui jouent un rôle cruciale aussi bien au niveau de l'estimation des paramètres qu'à celui de la restauration proprement dite.

### 2.4.1 Probabilités Forward

La probabilité backward est définie par : $\alpha_t(i)$=$P[X_t = \omega_i, Y_1 = y_1, ..., Y_t = y_t]$. Cette expression peut se calculer de manière récursive:

- Initialisation: $\pi_i f_i(y_1)$ pour $1 \leq i \leq k$
- Induction: $\alpha_{t+1}(j) = [\sum\limits_{i=1}^{k} \alpha_t(i)a_{ij}]f_j(y_{t+1})$ pour $1 \leq j \leq k$, $1 \leq t \leq n-1$

### 2.4.2 Probabilités Backward

La probabilité backward est définie par : $\beta_t(i) = P[Y_{t+1} = y_{t+1}, ..., Y_n = y_n|X_t = \omega_i]$. Le calcul de $\beta_t$ s'obtient à partir de $\beta_{t+1}$ par une récurence récurssive :

- Initialisation: $\beta_n(i) = 1$ pour $1 \leq i \leq k$
- Induction:

$$\beta_t(i) = [\sum_{j=1}^{k} a_{ij}f_j(y_{t+1})\beta_{t+1}(j)\text{pour}1 \leq i \leq k, 1 \leq t \leq n-1.$$

Le calcul respectif des probabilités forward et backward se heurte à des difficultés d'ordre numérique, Afin de s'affranchir de ce problème, Devijver et al. [4] ont propsé une solution qui consiste à remplacer les probabilitès conjointes par les probabilitès a postsriori. les reformulations de $\alpha_t$ et $\beta_t$ sont les suivantes:

$$\begin{aligned} \alpha_t(i) &= P[X_t = \omega_i | Y_1 = y_1, ..., Y_t = y_t] \\ \beta_t(i) &= \frac{P[Y_{t+1} = y_{t+1}, ..., Y_n = y_n | X_t = \omega_i]}{P[Y_{t+1} = y_{t+1}, ..., Y_n = y_n | Y1 = y_1, ..., Y_t = y_t]} \end{aligned}$$

La procédure de calcul de probabilité forward devient :

- Initialisation:

$$\alpha_1(i) = \frac{\pi_i f_i(y_1)}{\sum_{j=1}^{k} \pi_j f_j(y_1)} \text{pour} 1 \leq i \leq k$$

- Induction:

$$\alpha_{t+1}(i) = \frac{f_i(y_{t+1}) \sum_{j=1}^{k} \alpha_t(j) a_{ij}}{\sum_{l=1}^{k} f_l(y_{t+1}) \sum_{j=1}^{k} \alpha_t(j) a_{ji}} \text{pour} 1 \leq i \leq k, 1 \leq t \leq n-1$$

La procédure de calcul de probabilité backward devient :

- Initialisation: $\beta_n(i) = 1$ pour $1 \leq i \leq k$
- Induction: $\beta_t(i) = \dfrac{\sum_{j=1}^{k} a_{ij} f_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{l=1}^{k} f_l(y_{t+1}) \sum_{j=1}^{k} \alpha_t(j) a_{ji}}$ pour $1 \leq i \leq k$ , $1 \leq t \leq n-1$

## 2.5 Simulation

Dans cette section nous supposons connaître les paramètre caractérisant la chaîne de Markov X (La loi initiale $\pi$ et la matrice de transition $A = (a_{ij})$), ainsi que les moyennes et les variances caractérisant les denstés (qu'on suppose gaussiennes) des lois de $Y_n$conditionnelles à $X_n$ (pour chaque $\omega_i$, $m_i$ (resp $\sigma_i$) est la moyenne (resp la variance) de la densité de $Y_n$ conditionnelle à $X_n = \omega_i$).

### 2.5.1 Simulation de X

La simulation de $X_1$ est obtenue en effectuant un tirage uniforme selon le vecteur de probabilités initiale $\pi$.
En suposons $X_n = \omega_i$ connu, on simulera $X_{n+1}$ en effectuant un tirage uniforme selon le vecteur de probabilités correspondant à la $i^{ieme}$ ligne de la matrice de transition A.

### 2.5.2 Simulation de Y

En supposons $X_n = \omega_i$ connu, on effectue un tirage selon la densité $f_{\omega_i}$ dont les paramètres (la moyenne $m_i$ et la variance $\sigma_i$) sont supposés connus par hypothèse. Et pour cela on utilise le fait que pour $m \in \mathbf{R}$ , $\sigma \in \mathbf{R}_+^*$ et U, V deux variables aléatoires uniformes sur [0,1], la variable aléatoire $m + \sigma\sqrt{(-2\log(U))} * \cos(2\pi V)$ suit la loi $\mathcal{N}$ $(m, \sigma)$ .

## 2.6 Restauration

Pour la restauration de La chaîne de Markov X, nous utiliserons la méthode "mode de la marginale a posteriori" (MPM).
considérons la probabilité que le système soit à l'état $\omega_i$ à l'instant $t$ sachant l'observation Y=y.

$\xi_t(i) = P[X_t = \omega_i | Y = y]$

Cette Probabilité peut s'exprimer simplement en fonction des probabilités forward et backward, en effet nous avons:

$$\begin{aligned} \xi_t(i) &= \frac{P[X_t = \omega_i, Y = y]}{P[Y = y]} \\ &= \alpha_t(i)\beta_t(i) \end{aligned}$$

en effet:

$$\begin{aligned} \xi_t(i) &= \frac{P[X_t = \omega_i, Y = y]}{P[Y = y]} \\ &= \frac{P[Y = y | X_t = \omega_i] P[X_t = \omega_i]}{P[Y = y]} \\ &= \frac{P[Y_1 = y_1, ..., Y_t = y_t | X_t = \omega_i] P[Y_{t+1} = y_{t+1}, ..., Y_n = y_n | X_t = \omega_i] P[X_t = \omega_i]}{P[Y = y]} \end{aligned}$$

$$
\begin{aligned}
&= \frac{P[Y_1 = y_1, .., Y_t = y_t, X_t = \omega_i]P[Y_{t+1} = y_{t+1}, .., Y_n = y_n|X_t = \omega_i]P[Y_1 = y_1, .., Y_t = y_t]}{P[Y_1 = y_1, ..., Y_t = y_t]P[Y = y]} \\
&= P[X_t = \omega_i|Y_1 = y_1, ..., Y_t = y_t]\frac{P[Y_{t+1} = y_{t+1}, ..., Y_n = y_n|X_t = \omega_i]}{P[Y_{t+1} = y_{t+1}, ..., Y_n = y_n|Y_1 = y_1, ..., Y_t = y_t]} \\
&= \alpha_t(i)\beta_t(i)
\end{aligned}
$$

La méthode MPM consiste en “estimation”, pour chaque $t$ de la réalisation “invisible” de $X_t$ par la classe $\omega$ dont la probabilité a posteriori, i.e. conditionnelle à $Y = y$, est maximale, c-à-d:

$$[\hat{X}_t = \omega_i] \Longleftrightarrow [\xi_t(j) = \max_{1 \leq i \leq k} \xi_t(i)]$$

L’algorithme MPM se déroule de la manière suivante:

- Calcul en chaque $t$ et pour chaque $i$
  - la probabilité forward $\alpha_t(i)$;
  - la probabilité backward $\beta_t(i)$;
  - la probabilité a posteriori $\xi_t(i)$.
- Estimation de la réalisation de chaque $X_t$ par l’état qui maximise $\xi_t$.

## 2.7 Estimation des paramètres des modèles cachés

La méthode de restauration exposée dans le paragraphe précédent requirt la connaissance de certains paramètres de la loi du couple (X,Y). Nous avons vu que cette loi est définie par la loi de X, dite a priori, et les lois de conditionnelles à X. Lorsque le nombre de classes est $k$, les paramètres de la loi de X sont au nombre de $k^2$ (ce sont les coefficients $(c_{ij})$ définies précédemment).lorsque les densités $f_1, ..., f_k$ sont gaussiennes (ce que nous supposerons dans la suite), les paramètres définissant les lois de Y conditionnelles à X sont au nombre de $2k$ ($k$ moyennes et $k$ variances). Nous présentant dans cette section deux méthodes d’estimation des paramètres dans les chaînes de Markov cachées :

- L’algorithme Expectation-Maximisation (EM), proposé par Baum et al. [2] connu sous le nom “Forward-Backward Algorithm”.
- L’algorithme Iterative Conditional Estimation (ICE), proposé par Pieczynski [3], [**?**]dont l’application aux chaîne de Markov a été rapidement décrite dans [1].

### 2.7.1 Algorithme EM

#### Principe général

L'algorithme EM est une méthode générale d'estimation des paramètres dans le cas des "données manquantes". De façon générale, considérons un couple de variable aléatoires (X,Y),dont la densité $f_\theta(X,Y)$. Le problème est d'estimer $\theta$ à partir de Y seul. Si X était observable, on pourait envisager l'application du maximum de vraisemblance, quiconsiste en recherche de $\theta$ maximisant $Log[f_\theta(x,y)]$ X n'etant pas observable, on remplace la variable aléatoire $Log[f_\theta(x,y)]$ par $E_\theta[Log[f_\theta(x,y)|Y=y]]$.

On arrive alors à une méthode itérative : la valeur suivante du paramètre $\theta_{k+1}$ est donnée à partir de sa valeur courante $\theta_k$ et Y=y par :

$$E_{\theta_k}[Log[f_{\theta_{k+1}}(X,Y)|Y=y]]=\max_\theta E_{\theta_k}[f_\theta(X,Y)|Y=y]$$

Il existe deux phases dans chaque itération : calcul de $E_{\theta_k}[f_\theta(X,Y)|Y=y]$ et sa maximisation, d'où l'appellation de la méthode.

#### Cas des chaînes de Markov cachées

Nous décrirons ci-dessous les itérations de l'algorithme EM dans le cas du modèle de chaîne de Markov cachées.

On considère la probabilité d'être à l'instant $t$ dans la classe $\omega_i$ et à l'instant suivant dans la classe $\omega_j$ sachant la suite des observations

$$\Psi_t(i,j)=P[X_t=\omega_i, X_{t+1}=\omega_j|Y=y].$$

Cette expression peut s'écrire en fonction des probabilités forward-backward :

$$\Psi_t(i,j)=\frac{\alpha_t(i)a_{ij}f_j(y_{t+1})\beta_{t+1}(j)}{\sum_{l=1}^{k}f_l(y_{t+1})\sum_{o=1}^{k}\alpha_t(o)a_{ol}}.$$

Les formule de la re-estimation des paramètres du modèle à l'itération $q+1$ sont les suivantes:

- les probabilités initiales et la matrice de transition :

$$\pi_i^{q+1}=\frac{1}{n}\sum_{t=1}^{n}\xi_t^q(i)$$

$$a_{ij}^{q+1}=\frac{\sum_{t=1}^{n-1}\Psi_t^q(i,j)}{\sum_{t=1}^{n-1}\xi_t^q(i)}$$

où $\xi_t$ est la loi de X a posteriri définie précédemment.

- Les paramètres définissant la densité de probabilité, pour le cas gaussien considéré, associée à la classe $\omega_i$ est donné par le couple $(\mu_i, \sigma_i)$, $\mu_i$ étant la moyenne de la loi et $\sigma_i$ sa variance. On obtient pour la re-estimation :

$$\mu_i^{q+1} = \frac{\sum_{t=1}^{n} \xi_t^q(i) y_t}{\sum_{t=1}^{n} \xi_t^q(i)}$$

$$(\sigma_i^2)^{q+1} = \frac{\sum_{t=1}^{n} \xi_t^q(i)(y_t - \mu_i^{q+1})^2}{\sum_{t=1}^{n} \xi_t^q(i)}$$

L'algorithme EM se déroule de la façon suivante:

  - Initialisation des paramètres $a_{ij}^0$, $\mu_i^0$, $(\sigma_i^2)^0$
  - A chaque itération $q$:
    * Etape "E":calcul des probabilités suivantes:
      · $\alpha_t(i)$, $\beta_t(i)$
      · Déduction de $\Psi_t^q(i,j)$ à partir de $\alpha_t(i)$ et $\beta_t(i)$ calculée sur la base des paramètres $a_{ij}^q$, $\mu_i^q$,$(\sigma_i^2)^q$
    * Etape "M": calcul des paramètres $\pi_i^{q+1}$, $a_{ij}^{q+1}$, $\mu_i^{q+1}$,$(\sigma_i^2)^{q+1}$ par les formules ci-dessus.

## Remarque:

On peut retenir cet algorithme en se souvenant que la probabilité a priori est re-estimée par la moyenne des probabilités a posteriori, les moyennes sont re-estimées par les moyennes des observations "pondérées par les probabilités a posteriori", et les variances sont re-estimés par les variances empiriques en considérant des observations "pondérées par les probabilités a posteriori".

### 2.7.2 Algorithme ICE

#### Principe général

De façon générale, considérons un couple de variables aléatoires (X,Y),dont la loi $P_{(X,Y)}$ dépend d'un paramètre $\theta$. Nous cherchons une méthode d'estimation de $\theta$ à partir de Y.

Considérons un estimateur $\hat{\theta} = \hat{\theta}(X,Y)$ de $\theta$ à partir de (X,Y). L'idée de l'ICE est de rechercher l'approximation de $\hat{\theta} = \hat{\theta}(X,Y)$ dans l'ensemble de variables aléatoires fonctions de Y, le seul obseravable [?].

La meilleure approximation, au sens de l'erreur quadratique moyenne, est l'espérence conditionnelle.L'espérence conditionnelle $E[\hat{\theta}|Y]$ de $\hat{\theta} = \hat{\theta}(X, Y)$ par rapport à Y dépend du paramtre $\theta$. En effet, c'est la projection orthogonale par rapport au produit scalaire de $L^2$, qui dépend de $\theta$, sur l'ensemble de variables aléatoire fonctions de Y.

Lorsqu'on cherche à calculer, ou à approcher $E[\hat{\theta}|Y]$, on est ainsi amené à considérer la valeur " courante" $\theta_k$ du paramètre, la valeur suivante est donnée par $\theta_{k+1} = E_k[\hat{\theta}|Y]$, où $E_k$ est l'espérence conditionnelle correspondant à $\theta_k$. On arrive à la démarche itérative suivante :

- on considère une valeur initiale $\theta_0$
- on pose :

$$\theta_{k+1} = E_k[\hat{\theta}|Y = y] \tag{2.1}$$

Lorsque $E_k[\hat{\theta}|Y = y]$ n'est pas explicitement calculable mais la simulation des réalisation de X selon la loi conditionnelle à Y possible, (1.1) peut être approcher par :

$$\theta_{k+1} = \frac{1}{N}[\hat{\theta}(x_1, y) + ... + \hat{\theta}(x_N, y)]$$

**Cas des chaînes de marckov cachées**

La procédure ICE est applicable au problème de de l'estimation des paramètres dans le cas des chaînes de Markov cachées. Nous examinons successivement le problème de l'estimation des paramètres définissant la loi de X, et celui de l'estimation des paramètres du bruit.

*Estimation des paramètes de la loi de X*
Rapplons la définition des $c_{ij}$:

$$c_{ij} = P[X_m = \omega_i, X_{m+1} = \omega_j]$$

Supposons, conformément au principe général de ICE, que X est observable. Le paramètre de transtion $c_{ij}$ peut alors être estimé par :

$$\hat{c}_{ij}(X) = \frac{1}{n-1}\sum_{t=1}^{n-1} 1_{[X_t=\omega_i, X_{t+1}=\omega_j]}$$

qui est la fréquence empirique. Afin d'obtenir la valeur suivante du paramètre $c_{ij}$ on considère, conformément au principe général de ICE, l'espèrence conditionnelle de $\hat{c}_{ij}$:

$$c_{ij}^{q+1} = E_{\theta^q}[\hat{c}_{ij}(X)|Y = y]$$

ce qui donne :

$$c_{ij}^{q+1} = \frac{1}{n-1}\sum_{t=1}^{n-1} P^q[X_t = \omega_i, X_{t+1} = \omega_j | Y = y]$$

la matrice de transition correspondante à l'itération $q+1$ est donc :

$$\begin{aligned} a_{ij}^{q+1} &= \frac{c_{ij}^{q+1}}{\sum_{j=1}^{k} c_{ij}^{q+1}} \\ &= \frac{\sum_{t=1}^{n-1} P^q[X_t = \omega_i, X_{t+1} = \omega_j | Y = y]}{\sum_{j=1}^{k}\sum_{t=1}^{n-1} P^q[X_t = \omega_i, X_{t+1} = \omega_j | Y = y]} \end{aligned}$$

ou encore, en reprenant les notations ci-dessus :

$$a_{ij}^{q+1} = \frac{\sum_{t=1}^{n-1} \Psi_t^q(i,j)}{\sum_{t=1}^{n-1} \xi_t^q(i)}$$

**Remarque:**
Les formules ci-dessus de la re-estimation des paramètres de la loi de X sont les mêmes que celles obtenus par la méthode EM. Contrairement à EM, la méthode ICE n'utilise pas la notion de vraisemblance et le calcul reste valable quel que soit la forme de bruit.

Pour le bruit gaussien les paramètres des densités conditionnelles sont donnés par les moyennes $\mu_i$ et les variances $\sigma_i$. Suivant le principe de ICE, supposons X observable. Les paramtres $\mu_i$ et $\sigma_i$ peuvent alors être estimés par les moyennes et les variances empiriques :

$$\hat{\mu}_i(X,Y) = \frac{\sum_{j=1}^{n} Y_j 1_{[X_j=\omega_i]}}{\sum_{j=1}^{n} 1_{[X_j=\omega_i]}}$$

$$\hat{\sigma}_i^2(X,Y) = \frac{\sum_{j=1}^{n} (Y_j - \hat{\mu}_i)^2 1_{[X_j=\omega_i]}}{\sum_{j=1}^{n} 1_{[X_j=\omega_i]}}$$

Les espérences conditionnelles des estimateurs $\hat{\mu}_i(X,Y)$, $\hat{\sigma}_i^2(X,Y)$ ne sont pas calculables explicitement, on peut cependant utiliser l'approximation stochastique. En effet, la simulation des réalisation de X selon sa loi a posteriori est possible.
La démarche à chaque itération $q$ est la suivante :

- Simulation de N réalisation de X selon sa loi a posteriori : $(X^1, ..., X^N)$ (chaque $X^i$ est un n-uple).
- Calcul, pour chaque $1 \leq i \leq k$ et $1 \leq j \leq k$, de $\hat{\mu}_i^j$ et $(\hat{\sigma}_i^j)^2$ ( à partir de $X^j$,$y$) par les formules ci-dessus.
- Calcul, pour chaque $1 \leq i \leq k$ de $\mu_i^{q+1}$ et $(\sigma_i^{q+1})^2$ par :

$$\mu_i^{q+1} = \frac{1}{N}(\mu_i^1, ..., \mu_i^N)$$

$$(\sigma_i^{q+1})^2 = \frac{1}{N}((\sigma_i^1)^2, ..., (\sigma_i^N)^2$$

**Remarque**

ICE est une méthode d'estimation à la fois déterministe (les paramètres $a_{ij}$) et stochastique (les paramètres $\mu_i$ et $\sigma_i$ ). Grâce à sa phase stochastique ICE s'avère moins sensible au choix des paramètres de l'initialisation que EM. Cette situation est précisée dans le tableaux ci-dessous.

| Parmètres à estimer | EM | ICE |
|---|---|---|
| $(c_{ij})$ | Déterministe | Déterministe |
| $(\mu_i\ ,\sigma_i^2)$ | Déterministe | Stochastique |

## 2.8 Résultats des expérimentations

Nous présentons maintenant quelques résultat concernant la simulation de X et Y dams le cas de deux classes ($k$=2) en utilisant les résultats précédent pour certaines valeurs de paramtres, puis la restauration de X à partir des paramètres de Y, selon l'algorithme MPM.
Les chaines sont constituées de 25000 éléments (pour la restauration, nous avons utilisé les probabilités " Backward " et " Forward " normalisées).
Dans cette étude X est simuler avec les paramètres suivantes:

$$C_{k,l} = \begin{pmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{pmatrix}$$

### 2.8.1 Résultats des expérimentations dans le cas supervisé

Le mot "supervisé" signifie qu'on utilise les vrai paramètres pour la restauration de X. Le bruit est supposé gaussien de moyenne $\mu$ et variance $\sigma$. On peut distinguer trois types de bruits selon leurs moyennes et variances :

1. $\mu_1 \neq \mu_2$, $\sigma_1 = \sigma_2$ :discrimination par les moyennes.
2. $\mu_1 = \mu_2$, $\sigma_1 \neq \sigma_2$ :discrimination par les variances
3. $\mu_1 \neq \mu_2$, $\sigma_1 \neq \sigma_2$:discrimination par les moyennes et les variances.

Le tableau suivant résume les résultats des expérimentations.

| $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Taux d'erreur |
|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 3.65% |
| 1 | 1 | 1 | 2 | 6.8% |
| 2 | 3 | 1 | 2 | 6.17% |

### 2.8.2 Résultats des expérimentations dans le cas non supervisé

Le mot "non supervisé" signifie qu'on utilise des paramètres estimé à partir de la version dégradé Y pour la restauration de X.
Nous présentons les taux d'erreur de la restauration avec les trois types de bruitages decrit ci-dessus, en utilisant les valeurs des paramètres estimé dans un cas par la méthode ICE, et dans un autre Cas par la méthode EM.

#### Initialisation des algorithmes

Notons M la moyennes et $\Sigma$ l'écart-type de l'échantillon. Pour $k$ nombre de classe, les paramétres sont initialisés de la façon suivante :

- Initialisation des paramétres de modèle (loi a periori):
  - $\forall 1 \leq i \leq \pi_i^0 = \frac{1}{k}$
  - $\forall 1 \leq i, j \leq k$ avec $i \neq j$ $a_{ij}^0 = \frac{1}{2(k-1)}$
  - $\forall 1 \leq i \leq k$ $a_{ii}^0 = 0.5$
- Initialisation des moyennes du bruit:
  - pour $k$ pair et $\forall i$ tel que $0 \leq i \leq \frac{k}{2}$
    - $\mu_{i+1}^0 = M - (\frac{k}{2} - i)\frac{\Sigma}{2}$
    - $\mu_{k-i}^0 = M + (\frac{k}{2} - i)\frac{\Sigma}{2}$
  - Pour $k$ impair et $\forall 0 \leq i \leq k$:
    - $\mu_{i+1}^0 = M - (\frac{k}{2} - i)\frac{\Sigma}{2}$
- Initialisation des variances du bruit:
  - $\forall 1 \leq i \leq k\sigma_i = \Sigma$

**Résultats dans le cas des paramètres estimer par la méthode ICE**

Le tableau suivant nous présente les taux d'erreur de la restauration, les paramètres de bruit sont estimer avec la méthode ICE, sur le tableau n'apparait que les vrais paramètres du bruit :

| $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Taux d'erreur |
|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 5.94% |
| 1 | 1 | 1 | 2 | 9.98% |
| 2 | 3 | 1 | 2 | 11.3% |

**Résultats dans le cas des paramètres estimer par la méthode EM**

Le tableau suivant présente les taux d'erreur de la restauration, les paramètres du bruit sont estims avec la méthode EM, sur le tableau n'apparait que les vrais paramètres du bruit :

| $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Taux d'erreur |
|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 6.11% |
| 1 | 1 | 1 | 2 | 10.18% |
| 2 | 3 | 1 | 2 | 11.2% |

# chapitre 3

# Chaînes de Markov couples (PMC)

Une chaîne de Markov couple est un processus $Z = (Z_1, ..., Z_N)$ ( N fixé ) avec $Z_n = (X_n, Y_n)$ la variable aléatoire couple à l'instant n, où $X_n \in \Omega = [1, ..., k]$ avec $k$ le nombre d'etats, et $Y_n$ dans $\mathbf{R}$ , si sa distribution est définie par

$$p(z) = \frac{p(z_1, z_2)...p(z_{N-1}, z_N)}{p(z_2)...p(z_{N-1})} \tag{3.1}$$

Rappellons que que dans le cas d'une chaîne de Markov caché (HMC),nous avons:

$$p(z) = p(x, y) = p(x_1)f_{x_1}(y_1)p(x_2|x_1)f_{x_2}(y_2)...p(x_N|x_{N-1})f_{x_N}(y_N)$$

## 3.1 loi de Z

Soit Z une chaîne de Markov couple (PMC) dont la distribution est donnée par(3.1). On se donne $p(z_n, z_{n+1})$, indépendant de n, qui définit la probilité a priori $p(z_1)$ et la matrice de transition stationnaire $p(Z_2|Z_1)$.
l'objectif est de savoir calculer:

$$p(z_n, z_{n+1}) = p(x_n, x_{n+1})f_{x_n, x_{n+1}}(y_n, y_{n+1}) \tag{3.2}$$

Notons que le cas des chaîne Markov couple apparait comme un cas particulier de PMC,avec:

$$p(z_n, z_{n+1}) = p(x_n, x_{n+1})f_{x_n}(y_n)f_{x_{n+1}}(y_{n+1})$$

### 3.1.1 Calcul de $P(z_n)$

d'après (3.2),on'a:

$$p(z_n) = \int_{\mathbf{R}} \sum_{k=1}^{K} p(x_n, y_n, x_{n+1} = k, y_{n+1}) \mathrm{d}y_{n+1}$$

$$= \sum_{k=1}^{k} p(x_n, x_{n+1} = k) \int_{\mathbf{R}} f_{x_n, x_{n+1}=k}(y_n, y_{n+1}) \mathrm{d}y_{n+1}$$
$$= \sum_{k=1}^{k} p(x_n, x_{n+1} = k) f_{x_n, x_{n+1}=k}(y_n)$$

ce qui se présente comme le mélange de densités mono-dimensionnelles.

### 3.1.2 Calcul de $P(z_n|z_{n+1})$

La matrice de transition de la chaîne de Markov Z s'écrit :
pour $1 \leq n \leq N-1$

$$\begin{aligned} p(z_{n+1}|z_n) &= \frac{p(z_n, z_{n+1})}{p(z_n)} \\ &= \frac{p(x_n, x_{n+1}) f_{x_n, x_n+1}(y_n, y_{n+1})}{\sum_{k=1}^{k} p(x_n, x_{n+1} = k) f_{x_n, x_{n+1}=k}(y_n)} \end{aligned}$$

## 3.2 Simulation de Z

Pour la simulation de Z, i.e. la simulation de X et Y, nous supposons connaître l'ensemble des paramètres caractérisant le processus Z, i.e. les paramètres du modèle $p(x_n, x_{n+1})$ et les paramètres de bruits que nous supposons gaussien.
Lorsque le nombres de classes est $k$ on a:

- Paramètres de modèle : $k^2$ paramètres
- Paramètres de bruit : $f_{x_n, x_{n+1}}(y_n, y_{n+1})$ : $k^2$ densités dans $\mathbf{R}$, pour chaque densité, deux moyennes, deux variances et une covariances, soit $5k^2$ paramètres.

Au total, dans le cadre d'un bruit gaussien, on a $6k^2$ paramètres. $x_n$, $y_n$ etant fixés, on simulera $z_{i+1} = (x_{i+1}, y_{i+1})$ en simulant d'abord $x_{i+1}$, puis $y_{i+1}$.

### 3.2.1 Simulation de $x_1$ et $y_1$

nous avons l'expression de $p(z_1)$:

$$\begin{aligned} p(z_1) &= p(x_1, y_1) \\ &= \sum_{k=1}^{k} p(x_1, x_2 = k) f_{x_1, x_2=k}(y_1) \end{aligned}$$

d'où :

$$\begin{aligned} p(x_1) &= \sum_{i=1}^{k} p(x_1, x_2 = i) \\ p(y_1|x_1) &= \frac{\sum_{l=1}^{k} p(x_1, x_2 = l) f_{x_1,x_2=l}(y_1)}{p(x_1)} \\ &= \sum_{l=1}^{k} \frac{p(x_1, x_2 = l)}{p(x_1)} f_{x_1,x_2=l}(y_1) \end{aligned}$$

La simulation de $x_1$ est obtenue en effectuant un tirage uniforme selon le vecteur de probabilités $p(x_1)$.
Dans le cas de $y_1$, $x_1$ est connu , nous devons effectuer un tirage selon un mélange pondéré de densités mono-dimentionnelles. Pour cela, nous effectuons d'abord un tirage selon les poids $\frac{p(x_1,x_2=l)}{p(x_1)}$ pour $1 \leq l \leq k$. On note le résultat du tirage par $\omega_1$, ensuite nous effectuant un second tirage selon la densité $f_{x_1,\omega_1}$ dont les paramètres sont connus par hypothèses.

### 3.2.2 Simulation de $x_{n+1}$ et $y_{n+1}$

En supposons $x_n$ et $y_n$ connus, on simulera $z_{n+1} = (x_{n+1}, y_{n+1})$ en simulant d'abord $x_{n+1}$ puis $y_{n+1}$ de manière alternative.

#### Simulation de $x_{n+1}$ à partir de $z_n$

La densité de la loi de $x_{n+1}$ se calcule en intégrant $p(z_{n+1}|z_n)$ par rapport a $y_{n+1}$

$$\begin{aligned} p(x_{n+1}|z_n) &= \int_{\mathbf{R}} p(x_{n+1}, y_{n+1}|z_n) \mathrm{d}y_{n+1} \\ &= \frac{p(x_n, x_{n+1}) f_{x_n,x_{n+1}}(y_n)}{\sum_{j=1}^{k} p(x_n, x_{n+1} = j) f_{x_n,x_{n+1}=j}(y_n)}, 1 \leq n \leq N-1 \end{aligned}$$

**Remarque**
Dans le cas classique d'une chaîne de Markov caché, étant donné l'hypothse̊ d'indépendance conditionnelle, nous avons :

$$p(x_{n+1}|z_n) = \frac{P(x_n, x_{n+1})}{p(x_n)}, 1 \leq n \leq N-1.$$

#### Simulation de $y_{n+1}$ à partir de $x_{n+1}$ et $z_n$

La densité de la loi de $y_{n+1}$ conditionnellement à $x_{n+1}$ et à $z_n$ est donné par :

$$p(y_{n+1}|x_{n+1}, z_n) = \frac{p(z_{n+1}|z_n)}{p(x_{n+1}|z_n)}$$

$$= \frac{f_{x_n,x_{n+1}}(y_n, y_{n+1})}{f_{x_n,x_{n+1}}(y_n)}, 1 \leq n \leq N-1$$

Si on note les moyennes et la matrice de variances-covariances de $f_{x_n,x_{n+1}}$ par :

$$m_{x_n,x_{n+1}} = (m^1_{x_n,x_{n+1}}, m^2_{x_n,x_{n+1}})$$

$$\Gamma = \begin{pmatrix} (\sigma^2_{x_n,x_{n+1}})^1 & \rho_{x_n,x_{n+1}} \\ \rho_{x_n,x_{n+1}} & (\sigma^2_{x_n,x_{n+1}})^2 \end{pmatrix}$$

alors $y_{n+1}$ conditionnellement aux $x_{n+1}$ et $z_n$ est gaussien de densité $g_{y_{n+1}|x_{n+1},z_n}$ caractérisé par la moyenne :

$$m_{y_{n+1}|x_{n+1},z_n} = m^2_{x_n,x_{n+1}} - \frac{(y_n - m^1_{x_n,x_{n+1}})\rho_{x_n,x_{n+1}}}{(\sigma^2_{x_n,x_{n+1}})^1}$$

et la variance

$$\sigma_{y_{n+1}|x_{n+1},z_n} = (\sigma^2_{x_n,x_{n+1}})^2 - \frac{\rho_{x_n,x_{n+1}}}{(\sigma^2_{x_n,x_{n+1}})^1}$$

La simulation de $y_{n+1}$ est réalisé par un tirage selon la densité $g_{y_{n+1}|x_{n+1},z_n}$

## 3.3 Estimation des paramètres

Lorsque le nombre de classes est $k$, le modèle couple est définie par $6k^2$ paramètres :

- Paramètres de modèle : $k^2$ paramètres
- Paramètres de bruit : $f_{x_n,x_{n+1}}(y_n, y_{n+1})$ : $k^2$ densités dans $\mathbf{R}$, pour chaque densité, deux moyennes, deux variances et une covariances, soit $5k^2$ paramètres.

### 3.3.1 Probabilité Forward

la probabilité Forward est définie par :

$\alpha_t(i)$=$P[X_t = \omega_i, Y_1 = y_1, ..., Y_t = y_t]$

cette expression peut se calculer de manière récursive :

- Initialisation: $\alpha_1(l) = p(x_1 = l, y_1)$ pour $1 \leq l \leq k$
- Induction: $\alpha_{t+1}(j) = [\sum_{i=1}^{k} \alpha_t(i) p(x_{n+1} = j, y_{n+1}|x_n = i, y_n)$ pour $1 \leq j \leq k$, $1 \leq t \leq n-1$

Normalisation (par analogie avec le travail de Devijver) :

- Initialisatin :

$$\alpha_1(l) = \frac{p(x_1 = k, y_1)}{\sum_{m=1}^{k} p(x_1 = m, y_1)}$$

- Induction :

$$\alpha_{n+1}(l) = \frac{\sum_{i=1}^{k} \alpha_n(i) p(x_{n+1} = l, y_{n+1} | x_n = i, y_n)}{\sum_{m=1}^{k} \sum_{i=0}^{k} \alpha_n(i) p(x_{n+1} = m, y_{n+1} | x_n = i, y_n)}$$

### 3.3.2 Probabilité Backward

la probabilité backward est définie par :

$\beta_t(i)$=$P[Y_1 = y_1, ..., Y_t = y_t | X_t = \omega_i]$

cette expression peut se calculer de manière récursive :

- Initialisation: $\beta_1(l) = 1$ pour $1 \leq l \leq k$
- Induction: $\beta_{t+1}(j) = [\sum_{i=1}^{k} \beta_{t+1}(i) p(x_{n+1} = i, y_{n+1} | x_n = j, y_n)$ pour $1 \leq j \leq k$, $1 \leq t \leq n-1$

Normalisation (par analogie avec le travail de Devijver) :

- initialisatin :$\beta_1(l) = 1$
- Induction :

$$\beta_n(l) = \frac{\sum_{i=1}^{k} \beta_{n+1}(i) p(x_{n+1} = i, y_{n+1} | x_n = l, y_n)}{\sum_{m=1}^{k} \sum_{i=0}^{k} \beta_{n+1}(i) p(x_{n+1} = m, y_{n+1} | x_n = i, y_n)}$$

**Remarque:**
$\alpha_n(i)$ et $\beta_n(i)$ se calculent en utilisant le resultat :

$$p(z_{n+1} | z_n) = \frac{p(x_n, x_{n+1}) f_{x_n, x_n+1}(y_n, y_{n+1})}{\sum_{k=1}^{k} p(x_n, x_{n+1} = k) f_{x_n, x_{n+1}=k}(y_n)}$$

### 3.3.3 Probabilités conjointes de X a posteriori

C'est la probabilité que le système soit à l'instant $t$ dans la classe $k$ et à l'instant suivant dans la classe l sachant la suite des observations. Elle est donne par :

$$\Psi_n(k,l) = p(x_n = k, x_{n+1} = l|y),\ 1 \leq n \leq N-1$$

et se calcule de la manière suivante:

$$\Psi_n(k,l) = \frac{\alpha_n(k)p(x_{n+1} = l, y_{n+1}|x_n = k, y_n)\beta_{n+1}(l)}{\sum_{j=1}^{k}\sum_{m=1}^{k}\alpha_n(j)p(x_{n+1} = m, y_{n+1}|x_n = j, y_n)\beta_{n+1}(j)}$$

### 3.3.4 Matrice de transition de la chaîne $x|y$: $T_n(k,l)$

Notons que la chaîne x/y est de Markov. Sa matrice de transition est donne par :

$$T_{n+1}(i,j) = p(x_{n+1} = j|x_n = i, y),\ \text{pour } i \leq n \leq N-1,$$

qui se calcule à l'aide de la probabilité marginale a posterior et la probabilité conjointe a posterior de la façon suivante:

$$\begin{aligned} T_{n+1}(i,j) &= \frac{\Psi_n(i,j)}{\xi_n(i)} \\ &= \frac{p(x_{n+1} = j, y_{n+1}|x_n = i, y_n)\beta_{n+1}(j)}{\sum_{m=0}^{k} p(x_{n+1} = m, y_{n+1}|x_n = i, y_n)\beta_{n+1}(m)},\ 1 \leq n \leq N-1. \end{aligned}$$

### 3.3.5 Probabilités conjointes de X a priori :$C(i,j)$

Notons que X n'est pas markovienn, mais stationnaire,
la loi de X apriori est définie par:

$$C(i.j) = p(x_n = i, x_{n+1} = j)$$

on peut l'estimer analytiquement par la moyenne des probabilités conjointes a posteriori sans faire appelle à la simulation:

$$C(i,j) = \frac{1}{N-1}\sum_{n=1}^{N-1}\Psi_n(i,j)$$

## 3.4 Restauration

Comme dans le cas des chaînes de Markov cachés nous utiliserons la méthode MPM (mode de la marginale a posteriori) qui consiste en estimation de la réalistion de X à chaque instant $t$ par l'etat $\omega$ qui maximise la probabilité de X conditionnellement à Y.

### 3.4.1 Loi marginale de X a posteriori

La loi de X a posteriori, i.e. conditionnelle à Y=y, contient toute l'information sur X contenue dans l'observation Y=y, ce qui jout un rôle cruciale pour l'estimation de la réalisation cachée de X à partir de la réalisation observée Y=y. Elle est donnée par :

$$\xi_n(k) = p(x_n = k|y) \; 1 \leq n \leq N-1.$$

et se calcule de la manière suivante:

$$\begin{aligned} \xi_n(i) &= \frac{\alpha_n(i)\beta_n(i)}{\sum_{l=1}^{k} \alpha_n(l)\beta_n(l)} \\ &= \sum_{l=1}^{k} \Psi_n(i,l), \; 1 \leq n \leq N-1. \end{aligned}$$

La méthode MPM consiste en "estimation", pour chaque $t$, de la réalisation "invisible" de $X_t$ par la classe $\omega$ dont la probabilité a posteriori, i.e. conditionnelle à $Y = y$, est maximale, c-à-d,

$[\hat{X}_t = \omega_i] \Longleftrightarrow [\xi_t(j) = \max_{1\leq i\leq k} \xi_t(i)].$

L'algorithme MPM se déroule de la manière suivante:

- Calcul en chaque $t$ et pour chaque $i$
  - la probabilité forward $\alpha_t(i)$,
  - la probabilité backward $\beta_t(i)$,
  - la probabilité a posteriori $\xi_t(i)$,
- Estimation de la réalisation de chaque $X_t$ par l'état qui maximise $\xi_t$.

## 3.5 Comparaison de résultat de reastauration de PMC et HMC

Nous présentons quelques résultats concernant la restauration de X, dans le cas de deux classes ($k$=2). Dans un premier temps, nous simulons $X_0$=X et $Y_0$=Y en utilisant les résultats precécédent pour certaines valeurs de paramètres. Puis, dans un deuxième temps, nous retaurons X à partir des paramètres et de $Y_0$, selon l'algorithme MPM. Nous noterons $X_1$ la restauration obtenue avec le MPM couple, et $X_2$ la restauration obtenue avec le modèle clasique. En calculant le taux d'erreur entre$X_0$ et $X_1$ d'une part, et entre $X_0$ et $X_2$ d'autre part, nous obtenons une information sur l'apport du modèle couple par rapport au modèle classique. Les taux d'erreur présentés sont des taux moyens calculés à partir de 100 simulations. les chaînes sont constituées de 10000 éléments. Pour la restauration, nous avons utilisé les prababiltés "Forward" et "Backward" normalisées.

### 3.5.1 présentation des paramètres:

Les parmètres du modèle sont fixés à:

$$\mathbf{C_{k,l}} = \begin{pmatrix} 0.48 & 0.02 \\ 0.02 & 0.48 \end{pmatrix}$$

Pour simuler la chaîne avec le modèle couple, nous avons utilisé des bruits gaussien avec les moyennes et les variances suivantes:

| Classe | | Paramètres des lois | | | | |
|---|---|---|---|---|---|---|
| $k$ | $l$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | $\rho$ |
| 0 | 0 | 0 | 0 | 1 | 1 | $\rho_{0,0}$ |
| 0 | 1 | 0.2 | 0.8 | 1 | 1 | $\rho_{0,1}$ |
| 1 | 0 | 0.8 | 0.2 | 1 | 1 | $\rho_{1,0}$ |
| 1 | 1 | 1 | 1 | 1 | 1 | $\rho_{1,1}$ |

et les coefficients de correlation:

| Coefficient de corelation | $\rho_{0,0}$ | $\rho_{0.1}$ | $\rho_{1.0}$ | $\rho_{1.1}$ |
|---|---|---|---|---|
| Expérience 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| Expérience 2 | 0.1 | 0.9 | 0.9 | 0.1 |
| Expérience 3 | 0.9 | -0.9 | -0.9 | 0.9 |

Le modèle classique apparaît comme un cas particulier du modèle couple. Ainsi, pour restaurer la chaine avec le modèle classique, nous avons utilisé le MPM couple avec les paramètres suivants:

| Classe | | Paramètres des lois | | | | |
|---|---|---|---|---|---|---|
| $k$ | $l$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | $\rho$ |
| 0 | 0 | 0 | 0 | 1 | 1 | $\rho_{0,0}$ |
| 0 | 1 | 0 | 1 | 1 | 1 | $\rho_{0,1}$ |
| 1 | 0 | 1 | 0 | 1 | 1 | $\rho_{1,0}$ |
| 1 | 1 | 1 | 1 | 1 | 1 | $\rho_{1,1}$ |

### 3.5.2 Comparaisons des résultats

Le tableau suivant présente les résultats de restauration avec le MPM classique et le MPM couple:

| Taux d'erreur moyen | MPM classique | MPM couple | Gain |
|---|---|---|---|
| Expérience 1 | 14.1% | 11.08% | 21.4% |
| Expérience 2 | 14.06% | 9.08% | 35.41% |
| Expérience 3 | 32.57% | 15.13% | 53.54% |

# chapitre 4

# PMC dans le cas de bruit non gaussien

Le modèle traité au paragraphe précédent:

$$p(Z_i, Z_{i+1}) = p(x_i, x_{i+1}) f_{x_i,x_{i+1}}(y_i, y_{i+1})$$

n'est pas restreint au cadre gaussien.Nous pourrons bien considérer que les densités $f_{x_i,x_{i+1}}$ appartient à d'autres familles de lois, en particulier des lois exponentielles. On a vu dans le chapitre précédent que la connaissance des paramètres du modèle $p(x_i, x_{i+1})$ et les densité des lois de $(y_I, y_{i+1})$ suffit pour faire la simulation et la restauration de la chaîne de Markov couple (X,Y).

**Calcul explicite de** $f_{x_i,x_{i+1}}(y_i, y_{i+1})$
soit $\Gamma$ la matrice des variances-covariances de $(y_i, y_{i+1})$ conditionnellement à $(x_i, x_{i+1})$, on peut supposé que :

$$\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Alors $\Gamma$ peut se décomposer de la manière suivante:

$$\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} a & b \\ b & a \end{pmatrix}$$

avec $a = \cos(\alpha)$ et $b = \sin(\alpha)$ où $\alpha = \frac{\arcsin(\rho)}{2}$
Soit

$$A = \begin{pmatrix} a & b \\ a & b \end{pmatrix}$$

alors on peut écrir $(Y_i, Y_{i+1})$ de la forme suivante:

$$\begin{pmatrix} Y_i \\ Y_{i+1} \end{pmatrix} = A \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

où $T_1$ $T_2$ deux variables aléatoires indépendantes conditionnellement à $(x_i, x_{i+1})$. Nous noterons $g_1$ et $g_2$ les densité des lois de $T_1$ $T_2$ (respectivement) conditionnelles à $(x_i, x_{i+1})$.
notons

$$A_1 = (1,0) \text{ et } A_2 = (0,1)$$

finalement la densité de la loi de $(Y_i, Y_{i+1})$ conditionnellement à $(x_i, x_{i+1})$ s'ecrit:

$$f_{x_i,x_{i+1}}(y_i, y_{i+1}) = \frac{1}{detA} g_1 \left( A_1 A^{-1} \begin{pmatrix} y_i \\ y_{i+1} \end{pmatrix} \right) g_2 \left( A_2 A^{-1} \begin{pmatrix} y_i \\ y_{i+1} \end{pmatrix} \right)$$

## 4.1 Simulation et Restauration avec bruitage exponentielle

Pour la simulation de Z,i.e. la simulation de X et Y, nous supposons connaître l'ensemble des paramètres caractérisant le processus Z, i.e. les paramètres du modèle $p(x_n, x_{n+1})$ et les paramètres de bruits que nous supposons exponentielle .
Lorsque le nombres de classes est $k$ on a :

- paramètres de modèle : $k^2$ paramètres
- paramètres de bruit : $f_{x_n,x_{n+1}}(y_n, y_{n+1})$ : $k^2$ densités dans $\mathbf{R}$, pour chaque densité, deux paramètre $(\lambda^1_{x_i,x_{i+1}}, \lambda^2_{x_i,x_{i+1}})$, soit $2k^2$ paramètres.

Au total, dans dans le cadre d'un bruit gaussien, on a $3k^2$ paramètres. $x_n$, $y_n$ etant fixés, on simulera $z_{i+1} = (x_{i+1}, y_{i+1})$ en simulant d'abord $x_{i+1}$, ensuite, $y_{i+1}$. La simulation est faite d'une façon similaire que dans le cas avec bruitage gaussien. Comme dans le cas de PMC avec bruitage gaussien nous utiliserons la méthode MPM[1]. Les expréssion des prababilités Forward-Backward et des probabilité marginale a posteriori sont les mêmes que dans le cas de PMC avec bruitage gaussien.

## 4.2 Présentation des paramètres

Les parmètres du modèle sont fixés à :

$$\mathbf{C_{k,l}} = \begin{pmatrix} 0.48 & 0.02 \\ 0.02 & 0.48 \end{pmatrix}$$

[1]mode de la marginale a posteriori : qui consiste en estimation de la réalistion de X à chaque instant $t$ par l'etat $\omega$ qui maximise la probabilité de X conditionnellement à Y

Pour simuler la chaîne avec le nouveau modèle, nous avons utilisé des bruits exponentielle avec les paramétres suivantes :

| Classe | | Paramètres des lois | | |
|---|---|---|---|---|
| $k$ | $l$ | $\lambda_1$ | $\lambda_2$ | $\rho$ |
| 0 | 0 | 1 | 1 | $\rho_{0,0}$ |
| 0 | 1 | 1.2 | 1.8 | $\rho_{0,1}$ |
| 1 | 0 | 1.8 | 1.2 | $\rho_{1,0}$ |
| 1 | 1 | 2 | 2 | $\rho_{1,1}$ |

et les coefficients de correlation :

| Coefficient de corelation | $\rho_{0,0}$ | $\rho_{0.1}$ | $\rho_{1.0}$ | $\rho_{1.1}$ |
|---|---|---|---|---|
| Expérience 1 | 0.1 | 0.1 | 0.1 | 0.1 |
| Expérience 2 | 0.1 | 0.9 | 0.9 | 0.1 |
| Expérience 3 | 0.9 | -0.9 | -0.9 | 0.9 |

pour restaurer la chaine avec le PMC avec bruitage gaussien, nous avons utilisé le MPM couple avec les paramètres suivants:

| Classe | | Paramètres des lois | | | | |
|---|---|---|---|---|---|---|
| $k$ | $l$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | $\rho$ |
| 0 | 0 | 1 | 1 | 1 | 1 | $\rho_{0,0}$ |
| 0 | 1 | 0.8 | 0.55 | 0.64 | 0.3 | $\rho_{0,1}$ |
| 1 | 0 | 0.55 | 0.8 | 0.3 | 0.64 | $\rho_{1,0}$ |
| 1 | 1 | 0.5 | 0.5 | 0.25 | 0.25 | $\rho_{1,1}$ |

## 4.3 Comparaisons des résultats

Le tableau suivant présente les résultats de restauration avec MPM avec bruitage gaussien et le MPM avec bruitage exponentielle:

| Taux d'erreur moyen | MPM avec bruit gaaussien | MPM avec bruit exponentielle | Gain |
|---|---|---|---|
| Expérience 1 | 38.61% | 35.17% | 8.9% |
| Expérience 2 | 38.26% | 35.01% | 8.49% |
| Expérience 3 | 33.47% | 29.14% | 12.93% |

# chapitre 5

# Annexes

## 5.1 Quelques expressions sur la distribution de Z

On note que $p(z_m, ..., z_l)$, une distributuion marginale de $p(z)$ conserve la structure d'une PMC :

$$p(z_m, ..., z_l) = \frac{p(z_m, z_{m+1})...p(z_{l-1}, z_l)}{p(z_{m+1})...p(z_{l-1)}}$$

On peut alors décompose la structure d'une PMC de façon suivante:

$$\begin{aligned} p(z_m, ..., z_l) &= p(z_m, ..., z_a, ...z_l) \\ &= \frac{p(z_m, z_{m+1})...p(z_{a-1}, z_a)p(z_a, z_{a+1})...p(z_{l-1}, z_l)}{p(z_{m+1})...p(z_{a-1})p(z_a)p(z_{a+1})...p(z_{l-1)}} \\ &= p(z_m, ..., z_a)\frac{1}{p(z_a)}p(z_a, ..., z_l) \end{aligned}$$

Ensuite, on peut écrire la distribution de Z des deux manières :

$$\begin{aligned} p(z) &= p(z_1, ..., z_N) = p(z_1|z_2)p(z_2, ..., z_N) \\ &\text{et} \\ p(z) &= p(z_1, ..., z_N) = p(z_1, ..., z_{N-1})p(z_N|z_{N-1}) \end{aligned}$$

Par récurence, ceci nous amène à exprimer la distribution de Z des deux façons suivantes :

$$\begin{aligned} p(z) &= p(z_1|z_2)...p(z_{N-2}|z_{N-1})p(z_{n-1}, z_N) \\ &= p(z_{N-1}, z_N)\prod_{n=1}^{N-2} p(z_n|z_{n+1}) = p(z_N)\prod_{n=1}^{N-1} p(z_n|z_{n+1}) \\ \text{et} & \\ p(z) &= p(z_1, z_2)p(z_3|z_2)...p(z_N|z_{N-1}) \\ &= p(z_1, z_2)\prod_{n=2}^{N-1} p(z_{n+1}|z_n) = p(z_1)\prod_{n=1}^{N-1} p(z_{n+1}|z_n) \end{aligned}$$

## 5.2 Markovienté des chaînes $y|x$ et $x|y$

Les distributions $p(y|x)$ et $p(x|y)$ sont de markov.On étudie les deux cas séparément.

### 5.2.1 Markovienté de la chaîne $y|x$

On cherche à montrer que $p(y|x) = p^x(y)$ est une chaîne de markov, i.e. $p^x(y_{n+1}|y_1, ..., y_n) = p^x(y_{n+1}|y_n)$.

$$\begin{aligned} p^x(y_{n+1}|y_1, ..., y_n) &= \frac{p^x(y_1, ..., y_n, y_{n+1})}{p^x(y_1, ..., y_n)} \\ &= \frac{p(x, y_1, ..., y_n, y_{n+1}}{p(x)} \frac{p(x)}{p(x, y_1, ..., y_n, y_n)} \\ &= \frac{p(z_1, ..., z_{n+1}, x_{n+2}, ..., x_N)}{p(z_1, ..., z_n, x_{n+1}, ..., x_N)} \end{aligned}$$

En notant $dy^2 = dy_{n+2}...dy_N$ (mesure sur $\mathbf{R}^{N-n-1}$) et $dy^1 = dy_{n+1}...dy_N = dy_{n+1}dy^2$ (mesure sur $\mathbf{R}^{N-n}$) nous avons:

$$p^x(y_{n+1}|y_1, ..., y_n) = \frac{\int ... \int p(z_1, ..., z_{n+1}, z_{n+2}, ..., z_N)dy^2}{\int ... \int p(z_1, ..., z_n, z_{n+1}, ..., z_N)dy^1}$$

On peut décomposer les distributions à l'intérieur des intégrales($a = n + 1$ pour le numérateur, et $a = n$ pour le dénominateur), et extraire les termes qui sont indépendants des variables d'intégration:

$$\begin{aligned} p^x(y_{n+1}|y_1, ..., y_n) &= \frac{p(z_1, ..., z_{n+1})p(z_n) \int ... \int p(z_{n+1}, ..., z_N)dy^2}{p(z_1, ..., z_n)p(z_{n+1}) \int ... \int p(z_n, ..., z_N)dy^1} \\ &= \frac{p(z_n + 1|z_n, ...z_1)p(z_n, ...z_1)p(z_n)}{p(z_n, ...z_1)p(z_{n+1})} \frac{\int ... \int p(z_{n+1}, z_{n+2}, ..., z_N)dy^2}{\int ... \int p(z_n, z_{n+1}, ..., z_N)dy^1} \\ &= \frac{p(z_n + 1|z_n)p(z_n)}{p(z_{n+1})} \frac{\int ... \int p(z_{n+1}, z_{n+2}, ..., z_N)dy^2}{\int ... \int p(z_n, z_{n+1}, ..., z_N)dy^1} \\ &= p(z_n|z_{n+1}) \frac{\int ... \int p(z_{n+1}, z_{n+2}, ..., z_N)dy^2}{\int ... \int p(z_n, z_{n+1}, ..., z_N)dy^1} \end{aligned}$$

L'expression précédente ne dépend pas de $y_1, ..., y_{n-1}$ .Ainsi,

$$p^x(y_{n+1}|y_1, ..., y_n) = a_y(y_n, y_{n+1}, x_n, ..., x_N)$$

correspond à un terme d'une matrice de tansition $a_y$ d'une chaîne de markov.

### 5.2.2 Markovienté de la chaîne $x|y$

Par calcul similaire au précédent, nous obtenons:

$$p^y(y_{n+1}|y_1,...,y_n) = p(z_n,z_{n+1})\frac{\sum\limits_{(x_{n+2},...x_N)\in\Omega^{N-n-1}} p(z_{n+1},...,z_N)}{\sum\limits_{(x_{n+2},...x_N)\in\Omega^{N-n}} p(z_n,...,z_N)}$$

L'expression précédente ne dépend que de $x_n, x_{n+1}$. Ainsi,

$$p^y(y_{n+1}|y_1,...,y_n) = a_x(x_n,x_{n+1},y_n,...,y_N)$$

correspond à un terme d'une matrice de tansition $a_x$ d'une chaîne de markov.

# chapitre 6

# Perspectives

Dans ce travail, nous avons traiter une modèlisation originale par chaîne de markov couple, pouvant être utilisée dans des problèmes d'estimation des champs aléatoires inobservable. Dans le cadre de traitement d'image, cette modèlisation permet, en particulier, d'effectuer des segmentations statistiques d'images bruitées par des bruits corrélés. Nous avons présenté quelques résultats de restauration par la méthode bayésienne MPM, en comparant avec le modèle classique ( chaînes de markov cachés), les résultats trouvs sont encourageants. L'application du modèle proposé en segmentation d'images, bruitées par du bruit corrélé réelle appartenant à une famille des forme admissibles (par exemple le système de pearson), nécessitrait une méthode d'estimation de mélange généralisé. La recherche de telle méthode constitue une perspective naturelle pour la poursuite de ce travail.

# Bibliographie


[1] B.Benmiloud, W.Pieczynski, " Estimation des paramètres dans les chaîne de markov cachées et segmentation d'images",*traitement du signal*, Vol. 12,No. 5, 1995 pp. 433-454

[2] L.E.Baum,T. Petrie,G. Soules,N. Weiss,"A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains",Ann, math. 41, 1970, pp. 803-806.

[3] B.Braathen, W,Pieczynski, O.Masson,"Global and local methods of unsupervised bayesian segmentation of image",*Machine graphics and vision*,Vol. 2,No 1, 1993,pp. 39-52.

[4] P.A.Devijver,M.Deksel,"Champs aléatoires de Pickard et modilisation d'images digitales",*Traitement du Signale*, vol. 5, N 5, 1983, pp. 131-150

[5] W.Pieczynski,A.-N.Tebbache, "Pairwise Markov random fields and segmentation of textured images",*Machine graphics and vision*,Vol. 9,No 4, 2000,pp. 705-718.

[6] W.Pieczynski,"Pairwise Markov chain and bayesian unsupervised fusion",Proc.3rd Int.Conf.on Information Fusion,FUSION 2000,1 july, 10th-13th,Paris, France, MoD4-24 - MoD4-31.