Chaînes de Markov triplets avec bruit à dépendance longue

Dans ce chapitre, nous introduisons la notion de dépendance longue dans les modèles de Markov triplets. La dépendance longue, appelée également mémoire longue, se traduit par une corrélation persistante entre les marginales d'un processus. Cette persistance peut être due à des phénomènes fractals, comme dans les bruits gaussiens fractionnaires [78]. Elle peut être également due à des phénomènes saisonniers, comme dans les processus de Gegenbauer [30, 53]. Les phénomènes fractals et saisonniers se retrouvent notamment dans les données financières [56, 114], ou dans les images naturelles [66, 69, 92]. Parmi les autres applications de la dépendance longue, on peut citer le traitement des protocoles TCP/IP [87, 61, 120], ou encore la modélisation des précipitations et intempéries [82].

Nous commençons par définir la notion de processus à dépendance longue et donnons les principaux exemples. Ensuite, nous verrons comment modéliser des observations à dépendance longue à l'aide du modèle des "chaînes couples (resp. triplets) partiellement de Markov" introduit dans [98]. En particulier, nous proposons un nouveau modèle de chaîne semi-markovienne cachée par du bruit à mémoire longue. Nous proposons un algorithme ICE original, demandant des aménagements importants de son principe général, permettant d'estimer les paramètres de notre modèle à observations à dépendance longue. Pour finir, nous proposerons la modélisation de la dépendance longue dans des processus non gaussiens à l'aide des copules. Diverses simulations informatiques valident l'intérêt des nouveaux modèles et traitements non supervisés associés.

5.1 Processus à dépendance longue

5.1.1 Définition

Soit $Y = (Y_n)_{n \in \mathbb{Z}}$ un processus réel. On dira qu'il est du second ordre si pour tout $n \in \mathbb{Z}$, Y_n est de carré intégrable, et qu'il est stationnaire du second ordre si sa covariance $\mathbb{E}(Y_n Y_{n-k}) - \mathbb{E}(Y_n)\mathbb{E}(Y_{n-k})$ ne dépend pas de n. On la note alors $\gamma(k)$ et on a $\gamma(k) = \gamma(-k)$. La famille $(\gamma(k))_{k \in \mathbb{Z}}$ sera appelée famille de covariances.

Définition 5.1.1 (Dépendance longue). Soit $Y = (Y_n)_{n \in \mathbb{Z}}$ un processus réel stationnaire du second ordre et de famille de covariances $(\gamma(k))_{k \in \mathbb{Z}}$.

Le processus Y est à dépendance longue s'il existe $\alpha \in [0,1]$ et $C \in \mathbb{R}$ tels que :

$$\lim_{k \to +\infty} k^{\alpha} \gamma(k) = C.$$

On notera alors $\gamma(k) \sim_{+\infty} Ck^{-\alpha}$.

Remarque : La covariance d'un processus à mémoire longue satisfait $\sum_{k \in \mathbb{N}} \gamma(k) = +\infty$. Lorsque $\gamma(k) \sim_{+\infty} Ck^{-\alpha}$ pour $\alpha > 1$, on dit que Y est à dépendance intermédiaire.

Dans les deux sous-sections suivantes, nous donnons les principaux exemples de processus à dépendance longue et leurs propriétés.

5.1.2 Processus auto-similaires et bruits gaussiens fractionnaires

Un bruit gaussien fractionnaire est défini comme le processus d'accroissements d'un mouvement brownien fractionnaire, ce dernier étant défini comme un processus à temps continu possédant des propriétés d'auto-similarité. Nous préciserons ces notions ci-après.

Mouvements browniens fractionnaires et auto-similarité

On donne tout d'abord les définitions de processus auto-similaires et à accroissements stationnaires dont les mouvements browniens fractionnaires sont des cas particuliers.

Définition 5.1.2 (Processus auto-similaire). Un processus réel à temps continu $Z = (Z_t)_{t \in \mathbb{R}}$ est auto-similaire d'indice H > 0 si pour tout a > 0, le processus $(Z_{at})_{t \in \mathbb{R}}$ suit la même loi que $(a^H Z_t)_{t \in \mathbb{R}}$.

L'indice H est appelé indice ou paramètre de Hurst.

Définition 5.1.3 (Processus à accroissements stationnaires). Un processus réel à temps continu $Z = (Z_t)_{t \in \mathbb{R}}$ est à accroissements stationnaires si pour tout $h \in \mathbb{R}$ les variables aléatoires $Z_{t+h} - Z_t$ et $Z_h - Z_0$ ont mêmes lois.

Définition 5.1.4 (Mouvement brownien fractionnaire). Un processus réel à temps continu $B = (B_t)_{t \in \mathbb{R}}$ est un mouvement brownien fractionnaire de paramètre de Hurst H > 0 s'il est gaussien, auto-similaire de paramètre H et à accroissements stationnaires.

Un processus stationnaire du second ordre $Z = (Z_t)_{t \in \mathbb{R}}$ auto-similaire d'indice H et à accroissements stationnaires satisfait les propriétés suivantes :

1.
$$Z_0 = 0;$$

- 2. Si $H \neq 1$, alors $\mathbb{E}(Z_t) = 0$ pour tout $t \in \mathbb{R}$;
- 3. Z_{-t} et $-Z_t$ suivent la même loi;
- 4. Soit $\Gamma_H(s,t) = \mathbb{E}(Z_t Z_s)$ et $\sigma^2 = \mathbb{E}(Z_0^2)$. Si $H \neq 1$, alors :

$$\Gamma_H(s,t) = \frac{\sigma^2}{2} [|t|^{2H} + |s|^{2H} - |t-s|^{2H}];$$

- 5. $H \leq 1$;
- 6. Si H = 1, $Z_t = tZ_1$ pour tout $t \in \mathbb{R}$;
- 7. Si Z est un mouvement brownien fractionnaire de paramètre de Hurst H, alors la dimension de Hausdorff de la trajectoire $(t, Z_t)_{t \in \mathbb{R}}$ est égale à 2 - H (pour la définition de la dimension de Hausdorff, se reporter à l'annexe B).

Bruits gaussiens fractionnaires

Un bruit fractionnaire est le processus d'accroissements d'un processus auto-similaire à accroissements stationnaires, soit :

Définition 5.1.5 (Bruit fractionnaire). Un processus stationnaire du second ordre $Y = (Y_n)_{n \in \mathbb{Z}}$ est un bruit fractionnaire de moyenne m s'il existe un processus auto-similaire à accroissements stationnaires $Z = (Z_t)_{t \in \mathbb{R}}$ tel que :

$$\forall n \in \mathbb{Z}, \ Y_n = Z_{n+1} - Z_n + m.$$

Lorsque le processus Z est un mouvement brownien fractionnaire, Y est appelé "bruit gaussien fractionnaire".

Soit $Y = (Y_n)_{n \in \mathbb{Z}}$ un bruit fractionnaire et $Z = (Z_t)_{t \in \mathbb{R}}$ le processus auto-similaire de paramètre H à accroissements stationnaires associé à Y. Le bruit fractionnaire Y vérifie les propriétés suivantes :

1. Pour tout $n \in \mathbb{Z}$, $\mathbb{E}(Y_n) = m$;

1

2. Soit $\sigma^2 = \mathbb{E}(Z_0^2)$. La famille de covariances de Y est donnée par :

$$\forall k \in \mathbb{Z}, \ \gamma(k) = \frac{\sigma^2}{2} [|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}];$$
(5.1)

3. On a :

(a) Si
$$H = \frac{1}{2}$$
, alors $\gamma(k) = 0$ pour $k \neq 0$;
(b) Si $H < \frac{1}{2}$, alors $\gamma(k) < 0$ pour $k \neq 0$;
(c) Si $H > \frac{1}{2}$, alors $\gamma(k) > 0$ pour $k \neq 0$;
4. Si $H \neq \frac{1}{2}$, alors
 $\gamma(k) \sim_{+\infty} \sigma^2 H(2H - 1) k^{2H - 2}$; (5.2)

5. On a :
(a) Si H = ¹/₂, alors Y est un bruit blanc;
(b) Si H < ¹/₂, alors 2H - 2 < -1 et Y est à dépendance intermédiaire;
(c) Si H > ¹/₂, alors 2H - 2 > -1 et Y est à dépendance longue.

5.1.3 Processus FARIMA

Les "Fractional Autoregressive Integrated Moving Average" (FARIMA) sont, sous certaines conditions, à mémoire longue. Avant d'étudier plus précisement les FARIMA, rappelons quelques notions sur les processus stationnaires du second ordre.

Théorème 5.1.1 (Théorème d'Herglotz). Soit $Y = (Y_n)_{n \in \mathbb{Z}}$ un processus stationnaire du second ordre, à valeurs réelles, de famille de covariances $(\gamma(k))_{k \in \mathbb{Z}}$. Si $(\gamma(k))_{k \in \mathbb{Z}}$ est de type positif (toutes les matrices

($\gamma(0)$	$\gamma(1)$			$\gamma(n)$	
	$\gamma(1)$	$\gamma(0)$	$\gamma(1)$		$\gamma(n-1)$	
	:	:	:	÷	:	
	$\gamma(n)$	$\gamma(n-1)$:	÷	$\gamma(0)$)

ont leurs valeurs propres strictement positives), alors il existe une unique mesure μ sur $]-\pi,\pi[$, appelée "mesure spectrale" du processus Y telle que :

$$\gamma(k) = \int_{]-\pi,\pi[} e^{-ik\lambda} d\mu(\lambda) \text{ pour tout } k \in \mathbb{Z}.$$
(5.3)

De plus, si la famille de covariances est sommable, μ possède une densité f par rapport à la mesure de Lebesgue appelée "densité spectrale" de Y et donnée par :

$$f(\lambda) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma(k) e^{i\lambda k} \text{ pour tout } \lambda \in] -\pi, \pi[.$$
(5.4)

On remarque que la mesure spectrale est la transformée de Fourier de la famille de covariances. Notons que la sommabilité de la famille de covariances est une condition suffisante et non nécessaire d'existence de la densité spectrale. En effet, pour les processus FARIMA, la densité spectrale existe, mais ils peuvent être à dépendance longue, auquel cas la famille de covariances n'est pas sommable.

Exemple 5.1.1 (Bruit blanc). Si Y est un bruit blanc de variance σ^2 , alors sa densité spectrale est :

$$f(\lambda) = \frac{\sigma^2}{2\pi}$$
, pour tout $\lambda \in]-\pi, \pi[$.

Nous introduisons maintenant la définition d'un filtre linéaire. On notera L^2 l'ensemble des variables aléatoires de carré intégrable. On dira qu'une suite $(Z_n)_{n\in\mathbb{N}}$ d'éléments de L^2 converge au sens L^2 vers la variable aléatoire Z si $\lim_{n\to+\infty} \mathbb{E}\left((Z-Z_n)^2\right) = 0$, on rappelle que la limite Z est dans L^2 .

Définition 5.1.6.

1. Un filtre linéaire est une famille de réels $a = (a_k)_{k \in \mathbb{Z}}$;

- 2. Lorsque $\sum_{k \in \mathbb{Z}} |a_k|^p$ converge pour $p \ge 1$, on dit que le filtre est dans l^p et on note $a \in l^p$. Lorsque le filtre est dans l^1 , on dit que le filtre est stable;
- 3. Si pour tout k < 0, $a_k = 0$, le filtre est dit causal;
- 4. On dit que le filtre a admet une transformée en z si la série ∑_{k∈Z} a_kz^k converge dans un voisinage du cercle unité {z ∈ C : |z| = 1}. On notera alors a(z) = ∑_{k∈Z} a_kz^k la transformée en z de a :
- 5. Soient $a = (a_k)_{k \in \mathbb{Z}}$ et $b = (b_k)_{k \in \mathbb{Z}}$ deux familles de réels tels que $a \in l^p$ et $b \in l^q$ avec $\frac{1}{p} + \frac{1}{q} \ge 1$. Le produit de convolution de a et b noté a * b est défini par :

$$(a*b)_n = \sum_{k \in \mathbb{Z}} a_k b_{n-k} ;$$

- 6. Soit $Y^0 = (Y^0_n)_{n \in \mathbb{Z}}$ un processus stationnaire du second ordre et a un filtre.
 - Si a est un filtre stable, alors pour tout n, la série $\sum_{k \in \mathbb{Z}} a_k Y_{n-k}^0$ converge presque sûrement vers une variable aléatoire de carré intégrable et converge au sens L^2 vers la même variable aléatoire. La sortie Y du filtre a est définie comme la limite de la série $\sum_{k \in \mathbb{Z}} a_k Y_{n-k}^0$, que l'on notera $Y = a * Y^0$;
 - si $a \in l^2$ et si la famille de covariances de Y^0 est sommable, alors pour tout n, la série $\sum_{k \in \mathbb{Z}} a_k Y_{n-k}^0$ converge au sens L^2 . La sortie du filtre a est définie comme la limite dans L^2 de la série $\sum_{k \in \mathbb{Z}} a_k Y_{n-k}^0$, on la note également $Y = a * Y^0$.

Remarque : Si le filtre *a* est stable, alors la série $\sum_{k \in \mathbb{Z}} a_k z^k$ converge pour |z| = 1. Mais ceci n'implique pas que *a* possède une transformée en *z*. Lorsque *a* est stable, on notera de même a(z) la somme $\sum_{k \in \mathbb{Z}} a_k z^k$ pour |z| = 1.

La mesure spectrale de la sortie d'un filtre linéaire est donnée par la proposition suivante :

Proposition 5.1.1. Soit $Y^0 = (Y_n^0)_{n \in \mathbb{Z}}$ un processus stationnaire du second ordre de famille de covariances $\gamma_{Y^0} = (\gamma_{Y^0}(k))_{k \in \mathbb{Z}}$ sommable et de mesure spectrale μ_{Y^0} et soit a un filtre linéaire dans l^2 .

Soit $Y = a * Y^0 + m$, où m est un réel, alors on a:

- 1. $\gamma_Y = a * \check{a} * \gamma_{Y^0}, \ o \grave{u} \ \check{a}_k = a_{-k};$
- 2. Si de plus a est stable, la mesure spectrale μ_Y de Y a une densité par rapport à la mesure μ_{Y^0} donnée par :

$$d\mu_Y(\lambda) = |a(e^{i\lambda})|^2 d\mu_{Y^0}(\lambda).$$

Remarque : La stabilité du filtre *a* est une condition suffisante et non nécessaire pour que μ_Y admette une densité par rapport à la mesure μ_{Y^0} . Si le filtre *a* n'est pas stable et si les singularités de a(z) sur le disque unité sont des pôles, il suffit en fait que $\lambda \to |a(e^{i\lambda})|^2$ soit une fonction intégrable par rapport à la mesure μ_{Y^0} pour que μ_Y admette une densité par rapport à la mesure μ_{Y^0} .

On remarque de plus que si le filtre a est dans l^2 , alors la famille de covariances γ_Y est bornée mais n'est pas obligatoirement sommable. Par contre, lorsque le filtre a est stable, la famille de covariances γ_Y est sommable. Ainsi si Y est un processus à dépendance longue, le filtre ane peut pas être stable.

L'exemple le plus connu des processus s'exprimant comme la sortie d'un filtre linéaire est celui des processus "Auto-Regressive Moving Average" (ARMA) :

Définition 5.1.7 (ARMA(p,q)). Soient $p \ge 0$ et $q \ge 0$ deux entiers. Un processus $Y = (Y_n)_{n\in\mathbb{Z}}$ est un ARMA(p,q) de moyenne nulle s'il existe un bruit blanc $B = (B_n)_{n\in\mathbb{Z}}$ de

moyenne nulle, deux polynômes
$$P(X) = 1 + \sum_{k=1}^{k} \alpha_k X^k$$
 et $Q(X) = 1 + \sum_{k=1}^{k} \beta_k X^k$ tels que :

- 1. P et Q n'ont pas de racines en commun;
- 2. P et Q n'ont pas de racines sur le cercle unité ni à l'intérieur du disque unité;

3.
$$Y_n + \sum_{k=1}^p \alpha_k Y_{n-k} = B_n + \sum_{k=1}^q \beta_k B_{n-k}$$
, que l'on note $P * Y = Q * B_n$

Un processus Y est un ARMA de moyenne $m \in \mathbb{R}$ s'il existe Y⁰ un ARMA de moyenne nulle tel que $Y = Y^0 + m$.

Remarque : Un ARMA(p,q) est bien la sortie d'un filtre linéaire. En effet, comme P et Q n'ont pas de racines à l'intérieur du disque unité ni sur le cercle unité, alors la fonction $z \to \frac{Q(z)}{P(z)}$ est holomorphe dans un voisinage du disque unité et donc dans un voisinage du cercle unité. Cette fonction admet un développement de Laurent au voisinage du cercle unité donné par $\frac{Q(z)}{P(z)} = \sum_{k \in \mathbb{N}} a_k z^k$ qui est la transformée en z du filtre causal et stable $a = (a_k)_{k \in \mathbb{N}}$. On montre que Y = a * B ainsi Y est la sortie du filtre linéaire a. Le filtre a est appelé "filtre constructeur".

La proposition suivante montre que les processus ARMA ne sont pas à dépendance longue :

Proposition 5.1.2. Soit Y un processus ARMA(p,q) de filtre constructeur a et de famille de covariances $(\gamma(k))_{k\in\mathbb{Z}}$. Il existe $\rho \in [0, 1]$ et K > 0 tels que :

$$|\gamma(k)| \leq K\rho^k \text{ pour tout } k \geq 0.$$

Les processus FARIMA que nous introduisons ci-dessous généralisent les processus ARMA et peuvent être à dépendance longue.

Soit d un réel non nul, considérons a la fonction de \mathbb{C} dans \mathbb{C} définie par $a(z) = (1-z)^{-d}$.

Cette fonction est holomorphe à l'intérieur du disque unité et admet donc un développement de Laurent pour |z| < 1 donné par :

$$a(z) = \sum_{k=0}^{+\infty} a_k z^k,$$

où $a_k = \frac{\Gamma(d+k)}{\Gamma(d)\Gamma(k+1)}$ (voir Annexe A pour la définition de la fonction Γ). Le filtre causal $a = (a_k)_{k \in \mathbb{N}}$ est stable si d < 0 et dans l^2 si $d < \frac{1}{2}$.

Définition 5.1.8. Soient $p \ge 0$ et $q \ge 0$ deux entiers et soit $d \in \left[-\frac{1}{2}, \frac{1}{2}\right[-\{0\}]$. Un processus stationnaire du second ordre $Y = (Y_n)_{n \in \mathbb{Z}}$ est un FARIMA(p,d,q) de moyenne $m \in \mathbb{R}$ s'il existe un ARMA(p,q) de moyenne nulle Y^0 tel que :

$$Y = a * Y^0 + m$$
, où pour tout $k \ge 0$, $a_k = \frac{\Gamma(d+k)}{\Gamma(d)\Gamma(k+1)}$.

On supposera dans la suite que p = q = 0; le processus Y^0 est donc un bruit blanc de moyenne nulle. On notera σ^2 la variance de ce bruit blanc.

Le filtre *a* n'est pas stable si d > 0; cependant, conformément à la remarque de la proposition 5.1.1, la fonction $a(z) = (1-z)^{-d}$ admet 1 pour seule singularité sur le disque unité et cette singularité est un pôle. De plus, $|a(e^{i\lambda})|^2 = \frac{1}{[4\sin^2(\frac{\lambda}{2})]^d}$ définit une fonction intégrable sur $|-\pi,\pi|$. On en déduit qu'un FARIMA possède une densité spectrale par rapport à la mesure

 $]-\pi,\pi[$. On en déduit qu'un FARIMA possède une densité spectrale par rapport à la mesur de Lebesgue donnée par :

$$f(\lambda) = \frac{\sigma^2}{2\pi} \times \frac{1}{\left[4\sin^2\left(\frac{\lambda}{2}\right)\right]^d}.$$
(5.5)

La proposition suivante donne la famille de covariances d'un FARIMA.

Proposition 5.1.3. Soit Y un processus FARIMA(0,d,0) avec $d \in [-\frac{1}{2}, \frac{1}{2}[-\{0\}, sa\ covariance\ est\ donnée\ par\ :$

$$\gamma(k) = \sigma^2 \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(d)\Gamma(1-d)} = \sigma_F^2 \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k-d+1)},$$
(5.6)

оù

$$\sigma_F^2 = \gamma(0) = \sigma^2 \frac{\Gamma(1-2d)}{\left[\Gamma(1-d)\right]^2}$$

Preuve. Voir [19] pages 522-523.

On déduit de cette proposition que :

$$\gamma(k) \sim_{+\infty} ck^{2d-1} \text{ où } c = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} = \sigma_F^2 \frac{\Gamma(1-d)}{\Gamma(d)}.$$
(5.7)

On peut alors donner les conditions de dépendance longue d'un FARIMA(0,d,0):

- si d = 0, le FARIMA(0,0,0) est un bruit blanc;

- si d < 0, le FARIMA(0,d,0) est à dépendance intermédiaire;

- si d > 0, le FARIMA(0,d,0) est à dépendance longue.

Concernant le comportement asymptotique de la famille de covariances lorsque k tend vers l'infini, que ce soit pour les bruits gaussiens fractionnaires ou pour les processus FARIMA, elle décroît en puissance lorsque $k \to +\infty$, soit $\gamma(k) \sim_{+\infty} Ck^{-\alpha}$, où :

 $-\alpha = 2 - 2H$ pour les bruits gaussiens fractionnaires;

 $-\alpha = 1 - 2d$ pour les processus FARIMA.

Ainsi la covariance d'un processus FARIMA de paramètre de dépendance d a le même comportement asymptotique que celle d'un bruit gaussien fractionnaire de paramètre de Hurst $H = d + \frac{1}{2}$.

On peut faire les mêmes remarques concernant la densité spectrale. La densité spectrale d'un bruit gaussien fractionnaire est donnée par la proposition suivante :

Proposition 5.1.4. Soit $Y = (Y_n)_{n \in \mathbb{Z}}$ un bruit gaussien fractionnaire de variance σ^2 et de paramètre de Hurst $H \in [0, 1[$. Sa densité spectrale existe et est donnée par :

$$f(\lambda) = \sigma^2 \frac{2H\Gamma(2H)\sin(\pi H)}{\pi} \times (1 - \cos(\lambda)) \times \sum_{k \in \mathbb{Z}} |2k\pi + \lambda|^{-(2H+1)}.$$
(5.8)

Preuve. Voir [113] pages 28 à 34.

De la formule (5.5), au voisinage de 0, la densité spectrale d'un FARIMA se comporte comme $\frac{\sigma^2}{2\pi}|\lambda|^{-2d}$. De la formule (5.8), la densité spectrale d'un bruit gaussien fractionnaire se comporte comme $\frac{H\Gamma(2H)\sin(\pi H)}{\sigma^2}\sigma^2|\lambda|^{1-2H}$. Ainsi, dans les deux cas, la densité spectrale se comporte comme $C|\lambda|^{\frac{\pi}{-\beta}}$, où $\beta = 2H - 1$ pour les bruits gaussiens fractionnaires et $\beta = 2d$ pour les processus FARIMA.

5.1.4 Estimation des processus à dépendance longue

Nous étudions dans cette sous-section l'estimation des paramètres pour trois types de processus. Le premier est un processus stationnaire du second ordre de moyenne m et dont la famille de covariances est donnée par :

$$\forall k \in \mathbb{Z}, \ \gamma(k) = \sigma^2(|k|+1)^{-\alpha}, \ \text{où } \alpha \in \mathbb{R}^+.$$

Nous étudierons ensuite l'estimation des FARIMA et des bruits gaussiens fractionnaires. L'estimation que nous proposerons sera semi-paramétrique. Nous estimerons la densité spectrale par une méthode non-paramétrique, puis nous estimerons les paramètres à partir d'un échantillon de ce spectre par une méthode paramétrique. Nous choisirons d'étudier d'abord l'estimation des FARIMA, car comme nous le verrons, l'estimation des bruits gaussiens fractionnaires utilise celle des FARIMA.

1. Cas de la covariance $\gamma(k) = \sigma^2(|k|+1)^{-\alpha}$

Les paramètres du modèle sont la moyenne m, la variance σ^2 et le paramètre de dépendance α . Soit $y_{1:N} = (y_1, \ldots, y_N)$ une réalisation du processus à dépendance longue. Les estimateurs sont donnés par :

$$\hat{m} = \frac{1}{N} \sum_{n=1}^{N} y_n,$$
(5.9)

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{m})^2, \qquad (5.10)$$

$$\hat{\alpha} = -\frac{1}{\log(2)} \log\left(\frac{\gamma(\hat{1})}{\hat{\sigma}^2}\right), \tag{5.11}$$

où

$$\gamma(\hat{1}) = \frac{1}{N-1} \sum_{n=1}^{N-1} (y_n - \hat{m}_1) (y_{n+1} - \hat{m}_2)$$
$$\hat{m}_1 = \frac{1}{N-1} \sum_{n=1}^{N-1} y_n,$$
$$\hat{m}_2 = \frac{1}{N-1} \sum_{n=1}^{N-1} y_{n+1}.$$

2. Cas des processus FARIMA

On pourrait estimer $\alpha = 1 - 2d$ (resp. $\alpha = 2 - 2H$) par la méthode précédente, cependant, l'estimateur proposé précédemment donne de mauvais résultats dans des modèles plus spécifiques comme les FARIMA ou les bruits gaussiens fractionnaire. Nous proposons alors une estimation semi-paramétrique inspirée de celle de E. Moulines et P. Soulier [84]. Les paramètres du modèle sont la moyenne m, la variance du bruit blanc filtré σ^2 (resp. la variance du processus FARIMA σ_F^2) et le paramètre de dépendance d.

Etape non paramétrique : estimation du spectre

Soit $y_{1:N} = (y_1, \ldots, y_N)$ une réalisation d'un processus FARIMA. La densité spectrale est alors estimée pour $\lambda \in [-\pi, \pi]$ par :

$$\hat{f}(\lambda) = \frac{1}{2\pi N} \left| \sum_{n=1}^{N} (y_n - \bar{y}) \exp(in\lambda) \right|^2,$$
où $\bar{y} = \frac{1}{N} \sum_{n=1}^{N} y_n.$
(5.12)

Etape paramétrique : estimation de d et de σ

Une fois la densité spectrale estimée, nous disposons d'un échantillon $(\hat{f}(\lambda_1), \ldots, \hat{f}(\lambda_M))$, où

 $\lambda_j = -\pi + \frac{2j\pi}{M}$. D'après (5.5), la densité spectrale d'un FARIMA vérifie :

$$\log(f(\lambda)) = \log\left(\frac{\sigma^2}{2\pi}\right) - d\log\left(4\sin^2\left(\frac{\lambda}{2}\right)\right).$$

Ainsi, posant $\omega_j = \log\left(4\sin^2\left(\frac{\lambda_j}{2}\right)\right), z_j = \log(\hat{f}(\lambda_j))$ et $B = \log\left(\frac{\sigma^2}{2\pi}\right)$, nous devons chercher B et d minimisant :

$$\sum_{j=1}^{M} \left[z_j - (B - d\omega_j) \right]^2.$$

Les estimateurs au sens des moindres carrés de B et d sont alors :

$$\hat{d} = -\frac{\hat{\sigma}_{\omega,z}}{\hat{\sigma}_{\omega}^2},$$

$$\hat{B} = \bar{z} + \hat{d}\bar{\omega},$$
(5.13)

où
$$\bar{\omega} = \frac{1}{M} \sum_{j=1}^{M} \omega_j, \ \bar{z} = \frac{1}{M} \sum_{j=1}^{M} z_j, \ \hat{\sigma}_{\omega}^2 = \frac{1}{M} \sum_{j=1}^{M} (\omega_j - \bar{\omega})^2 \text{ et } \hat{\sigma}_{\omega,z} = \frac{1}{M} \sum_{j=1}^{M} (\omega_j - \bar{\omega})(z_j - \bar{z}).$$

L'estimation de σ est donnée par :

$$\hat{\sigma}^2 = 2\pi \exp(\hat{B}),\tag{5.14}$$

et donc la variance du FARIMA est estimée par :

$$\hat{\sigma}_F^2 = 2\pi \exp(\hat{B}) \times \frac{\Gamma(1-2\hat{d})}{\left[\Gamma(1-\hat{d})\right]^2}.$$

Estimation de la moyenne m

La moyenne *m* ne peut être estimée à partir du spectre. Nous utilisons l'estimateur empirique proposé en (5.9).

3. Cas des bruits gaussiens fractionnaires

Comme la densité spectrale (5.8) d'un bruit gaussien fractionnaire est difficile à exploiter directement, on utilise le fait qu'elle a le même comportement asymptotique lorsque λ tend vers 0 que celle d'un processus FARIMA. Les paramètres à estimer sont la moyenne m, la variance du bruit gaussien fractionnaire σ^2 et le paramètre de Hurst H. Soit $y_{1:N} = (y_1, \ldots, y_N)$ la réalisation d'un bruit gaussien fractionnaire.

Estimation de la moyenne m

De la même façon que précédemment, l'estimateur de la moyenne est celui proposé en (5.9).

Estimation de la variance et du paramètre de Hurst

Soit $(\hat{f}(\lambda_1), \ldots, \hat{f}(\lambda_M))$ un échantillon du spectre estimé par (5.12) tel que les λ_i soient les

plus proches possibles de 0, on prendra dans les expérimentations, $\lambda_j = -10^{-2} + \frac{2j10^{-2}}{M}$ et M = 100. On commence par estimer d et B en utilisant la méthode des moindres carrés décrite précédemment. On pose :

$$\hat{H} = \hat{d} + \frac{1}{2}.$$
(5.15)

Pour l'estimation de la variance, on remarque qu'au voisinage de 0, la densité spectrale d'un FARIMA est équivalente à $\frac{\sigma_0^2}{2\pi} \times |\lambda|^{-2d}$ lorsque σ_0^2 est la variance du bruit blanc filtré et celle d'un bruit gaussien fractionnaire est équivalente à $\frac{H\Gamma(2H)\sin(\pi H)}{\pi}\sigma^2|\lambda|^{1-2H}$, on pose ainsi :

$$\hat{\sigma}^2 = \frac{\pi}{\hat{H}\Gamma(2\hat{H})\sin(\pi\hat{H})}\exp(\hat{B}).$$
(5.16)

L'estimateur du paramètre α dans le cas où la covariance est donnée par $\gamma(k) = \sigma^2(|k| + 1)^{-\alpha}$ sera utilisé dans les deux sections suivantes. Nous utiliserons l'estimation semi-paramétrique de bruit gaussien fractionnaire dans la section 5.4.

5.2 Chaînes couples partiellement de Markov

Nous définissons ici le modèle de chaînes couples partiellement de Markov. Nous proposerons ensuite un modèle de chaînes couples partiellement de Markov particulier permettant des traitements bayésiens non supervisés dans le cadre des observations à dépendance longue.

5.2.1 Chaînes couples partiellement de Markov : modèle général

Soient $X = (X_n)_{1 \le n \le N}$ et $Y = (Y_n)_{1 \le n \le N}$ deux processus aléatoires tels que chaque X_n prend ses valeurs dans l'ensemble fini $\mathcal{X} = \{\omega_1, \ldots, \omega_K\}$ et chaque Y_n prend ses valeurs dans \mathbb{R} .

Définition 5.2.1 (Chaînes couples partiellement de Markov). Le processus $Z = (X_n, Y_n)_{1 \le n \le N}$ est une chaîne couple partiellement de Markov (CCPM) si sa distribution $p(z_{1:N})$ vérifie :

$$p(z_{1:N}) = p(z_1) \prod_{n=1}^{N-1} p(z_{n+1}|x_n, y_{1:n}),$$
(5.17)

 $o\hat{u} \ y_{1:n} = (y_1, \dots, y_n).$

Remarque : A ce stade, nous disposons des modèles suivants :

- chaînes de Markov cachées à bruit indépendant (CMC-BI);
- chaînes de Markov couples (CMCouple);
- chaînes couples partiellement de Markov (CCPM).

Nous avions vu au chapitre 3 que le modèle CMCouple était plus général que le modèle CMC-BI. Dans le modèle CMCouple, le processus X n'est plus obligatoirement markovien et les variables aléatoires Y_n ne sont plus obligatoirement indépendantes conditionnellement à

X. Quant au modèle CCPM, il généralise le modèle CMCouple. Dans le modèle CCPM, le processus Z n'est plus obligatoirement une chaîne de Markov, ce qui accorde au modèle une plus grande richesse.

En utilisant l'équivalence entre markovianité et factorisation d'une loi vue au chapitre 2, on montre que si $p(x_{1:N}, y_{1:N})$ est la distribution d'une couple partiellement de Markov, alors $p(x_{1:N}|y_{1:N})$ est la distribution d'une chaîne de Markov. La proposition ci-dessous détaille l'algorithme de Baum-Welsh conditionnel dans le cas du modèle CCPM. Celui-ci permet de calculer les lois a posteriori $p(x_n|y_{1:N})$ et $p(x_{n+1}|x_n, y_{1:N})$, à condition toutefois que la distribution $p(z_{n+1}|x_n, y_{1:n})$ soit calculable.

Proposition 5.2.1 (Algorithme de Baum-Welsh). Soit $Z = (X_n, Y_n)_{1 \le n \le N}$ une chaîne couple partiellement de Markov. Les probabilités a posteriori $p(x_n|y_{1:N})$ et $p(x_{n+1}|x_n, y_{1:N})$ sont données par :

$$p(x_n|y_{1:N}) = \tilde{\alpha}_n(x_n)\tilde{\beta}_n(x_n), \qquad (5.18)$$

$$p(x_{n+1}|x_n, y_{1:N}) \propto \frac{\beta_{n+1}(x_{n+1})}{\tilde{\beta}_n(x_n)} p(z_{n+1}|x_n, y_{1:n}),$$
 (5.19)

$$\begin{split} o\hat{u} \quad \tilde{\alpha}_{1}(x_{1}) &= p(x_{1}|y_{1}) \quad et \quad \tilde{\alpha}_{n+1}(x_{n+1}) \propto \sum_{x_{n}} \tilde{\alpha}_{n}(x_{n}) p(z_{n+1}|x_{n}, y_{1:n}) \ pour \ n \geq 1, \\ \tilde{\beta}_{N}(x_{N}) &= 1 \quad et \quad \tilde{\beta}_{n}(x_{n}) = \frac{\sum_{x_{n+1}} \tilde{\beta}_{n+1}(x_{n+1}) p(z_{n+1}|x_{n}, y_{1:n})}{\sum_{x_{n+1}} \sum_{x_{n}} \tilde{\alpha}_{n}(x_{n}) p(z_{n+1}|x_{n}, y_{1:n})} \ pour \ n \leq N-1. \end{split}$$

5.2.2 Observations gaussiennes à dépendance longue

Nous proposons dans cette sous-section un modèle de chaînes couples partiellement de Markov pour lequel la quantité $p(z_{n+1}|x_n, y_{1:n})$ est calculable. Nous verrons par la suite comment ce modèle permet de considérer des observations à dépendance longue. Le modèle particulier est défini par les trois hypothèses suivantes :

- 1. $p(x_{n+1}|x_n, y_{1:n}) = p(x_{n+1}|x_n);$
- 2. $p(y_{n+1}|x_n, x_{n+1}, y_{1:n}) = p(y_{n+1}|x_{n+1}, y_{1:n});$
- 3. Les distributions $p(y_{n+1}|x_{n+1}, y_{1:n})$ sont des lois normales définies par :
 - $p(y_1|x_1)$ est une distribution gaussienne de moyenne m_{x_1} et de variance $\gamma_{x_1}(0)$;
 - $-p(y_{n+1}|x_{n+1}, y_{1:n})$ est une distribution gaussienne de moyenne $\tilde{m}_{x_{n+1}}$ et de variance $\tilde{\gamma}_{x_{n+1}}$ données par :

$$\tilde{m}_{x_{n+1}} = m_{x_{n+1}} + \Gamma_{x_{n+1}}^{2,1} (\Gamma_{x_{n+1}}^n)^{-1} (y_{1:n} - m_{x_{n+1}}^n),$$

$$\tilde{\gamma}_{x_{n+1}} = \gamma_{x_{n+1}} (0) - \Gamma_{x_{n+1}}^{2,1} (\Gamma_{x_{n+1}}^n)^{-1} \Gamma_{x_{n+1}}^{1,2},$$

$$\hat{\operatorname{ou}} m_{x_{n+1}}^{n} = (\underbrace{m_{x_{n+1}}, \dots, m_{x_{n+1}}}_{n \text{ fois}}),$$

$$\Gamma_{x_{n+1}}^{n} = \begin{pmatrix} \gamma_{x_{n+1}}(0) & \gamma_{x_{n+1}}(1) & \dots & \gamma_{x_{n+1}}(n-2) & \gamma_{x_{n+1}}(n-1) \\ \gamma_{x_{n+1}}(1) & \gamma_{x_{n+1}}(0) & \dots & \gamma_{x_{n+1}}(n-3) & \gamma_{x_{n+1}}(n-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{x_{n+1}}(n-1) & \gamma_{x_{n+1}}(n-2) & \dots & \gamma_{x_{n+1}}(1) & \gamma_{x_{n+1}}(0) \end{pmatrix}$$

est une matrice symétrique définie positive et $\Gamma_{x_{n+1}}^{2,1} = (\gamma_{x_{n+1}}(n), \ldots, \gamma_{x_{n+1}}(1)) = (\Gamma_{x_{n+1}}^{1,2})^T$ est un vecteur de taille $n \times 1$;

La première hypothèse implique que X est une chaîne de Markov. Selon le point 3., les distributions $p(z_{n+1}|x_n, y_{1:n})$ sont calculables, il en résulte que les probabilités a posteriori $p(x_n|y_{1:N})$ et $p(x_{n+1}|x_n, y_{1:N})$ sont calculables par la proposition 5.2.1. Lorsque l'une au moins des familles de covariance γ_{ω_j} est à mémoire longue, ce modèle sera appelé "chaîne de Markov cachée à mémoire longue" (CMC-ML).

La proposition suivante résume les propriétés classiques des vecteurs gaussiens.

Proposition 5.2.2.

- Soient W^1 et W^2 deux vecteurs réels gaussiens de dimensions respectives d^1 et d^2 , de moyennes respectives M^1 et M^2 , de matrices de covariance respectives Γ^1 et Γ^2 et tels que le vecteur joint $W = (W^1, W^2)$, de densité $p(w^1, w^2)$, soit un vecteur gaussien de covariance :

$$\Gamma = \left(\begin{array}{cc} \Gamma^1 & \Gamma^{1,2} \\ \Gamma^{2,1} & \Gamma^2 \end{array} \right).$$

Alors la densité $p(w^2|w^1)$ est celle d'une loi normale de moyenne :

$$M^{2|1} = M^2 + \Gamma^{2,1}(\Gamma^1)^{-1}(w^1 - M^1),$$

et de matrice de covariance :

$$\Gamma^{2|1} = \Gamma^2 - \Gamma^{2,1}(\Gamma^1)^{-1}\Gamma^{1,2}.$$

- Réciproquement, soient W^1 et $(W^2|W^1 = w^1)$ deux vecteurs réels gaussiens de dimensions respectives d^1 et d^2 et de densités $p(w^1)$ et $p(w^2|w^1)$. Si W^1 a pour moyenne M^1 et covariance Γ^1 et si $(W^2|W^1 = w^1)$ a pour moyenne $M^{2|1} = Aw^1 + B$ et pour covariance $\Gamma^{2|1}$, alors le vecteur joint $W = (W^1, W^2)$ de densité $p(w^1, w^2)$ est gaussien de moyenne :

$$M = \left(\begin{array}{c} M^1\\ AM^1 + B \end{array}\right),$$

et de matrice de covariance :

$$\Gamma = \left(\begin{array}{cc} \Gamma^1 & \Gamma^1 A^T \\ A\Gamma^1 & \Gamma^{2|1} + A\Gamma^1 A^T \end{array}\right).$$

La proposition 5.2.2 nous permettra également de calculer les lois gaussiennes $p(y_{1:n}|x_{1:n})$, qui seront utiles dans l'estimation des paramètres. Plus précisément :

Proposition 5.2.3. Si $p(y_1|x_1)$ suit la loi normale $\mathcal{N}_{\mathbb{R}}(m_{x_1}, \gamma_{x_1}(0))$ et si $p(y_{n+1}|x_{n+1}, y_{1:n})$ suit la loi normale :

- de moyenne $m_{x_{n+1}} + \Gamma_{x_{n+1}}^{2,1}(\Gamma_{x_{n+1}}^n)^{-1}(y_{1:n} - m_{x_{n+1}}^n);$ - de variance $\gamma_{x_{n+1}}(0) - \Gamma_{x_{n+1}}^{2,1}(\Gamma_{x_{n+1}}^n)^{-1}\Gamma_{x_{n+1}}^{1,2}.$ Alors $p(y_{1:n}|x_{1:n})$ suit la loi normale de \mathbb{R}^n , $\mathcal{N}_{\mathbb{R}^n}(M^{x_{1:n}}, \Gamma^{x_{1:n}})$, où $M^{x_{1:n}}$ et $\Gamma^{x_{1:n}}$ sont calculés par les récursions suivantes :

1. Initialisation :

$$M^{x_1} = m_{x_1} \ et \ \Gamma^{x_1} = \gamma_{x_1}(0) \ ;$$

2. Itération :

$$M^{x_{1:n+1}} = \begin{pmatrix} M^{x_{1:n}} \\ m_{x_{n+1}} + \Gamma^{2,1}_{x_{n+1}} (\Gamma^{n}_{x_{n+1}})^{-1} \left[M^{x_{1:n}} - m^{n}_{x_{n+1}} \right] \end{pmatrix},$$

$$\Gamma^{x_{1:n+1}} = \begin{pmatrix} \Gamma^{x_{1:n}} & \Gamma^{x_{1:n}} (\Gamma^{n}_{x_{n+1}})^{-1} \Gamma^{1,2}_{x_{n+1}} \\ \Gamma^{2,1}_{x_{n+1}} (\Gamma^{n}_{x_{n+1}})^{-1} \Gamma^{x_{1:n}} & \gamma_{x_{n+1}} (0) - \Gamma^{2,1}_{x_{n+1}} (\Gamma^{n}_{x_{n+1}})^{-1} \left[\Gamma^{1,2}_{x_{n+1}} - \Gamma^{x_{1:n}} (\Gamma^{n}_{x_{n+1}})^{-1} \Gamma^{1,2}_{x_{n+1}} \right] \end{pmatrix}.$$
Preuve Voir [66]

Freuve. voir [00].

Notons que le modèle est relativement complexe car $p(y_n|x_{1:n})$ dépend de tous les x_k pour $k \leq n$. Nous reviendrons sur cette remarque lors de l'estimation des paramètres.

5.2.3Estimation des paramètres

Nous détaillons dans cette sous-section l'algorithme ICE dans le cas d'un modèle CMC-ML gaussien. Lorsque chaque X_n prend ses valeurs dans l'ensemble fini $\mathcal{X} = \{\omega_1, \ldots, \omega_K\},\$ les paramètres à estimer sont les K moyennes m_{ω_j} , les K variances $\sigma_{\omega_j}^2 = \gamma_{\omega_j}(0)$ et les K paramètres de dépendance qui selon le modèle sont :

- α_{ω_j} lorsque $\gamma_{\omega_j}(k) = \sigma_{\omega_j}^2 (k+1)^{-\alpha_{\omega_j}}$;

- H_{ω_j} lorsque $(\gamma_{\omega_j}(k))_{k\in\mathbb{N}}$ est la famille de covariances d'un bruit gaussien fractionnaire; $-d_{\omega_i}$ lorsque $(\gamma_{\omega_i}(k))_{k\in\mathbb{N}}$ est la famille de covariances d'un processus FARIMA;

et les K^2 paramètres $p(x_n = \omega_i, x_{n+1} = \omega_j)$. On suppose que la chaîne X est une chaîne de Markov stationnaire et réversible.

Afin d'utiliser l'algorithme ICE, nous devons nous donner un estimateur à partir des données complètes. Concernant les paramètres $p_{i,j}$, nous considérerons l'estimateur classique :

$$\hat{p}_{i,j}(x_{1:N}, y_{1:N}) = \frac{1}{2(N-1)} \sum_{n=1}^{N-1} \left[I(x_n = \omega_i, x_{n+1} = \omega_j) + I(x_n = \omega_j, x_{n+1} = \omega_i) \right].$$
(5.20)

L'espérance conditionnelle $\mathbb{E}(\hat{p}_{i,j}(X, y_{1:N})|y_{1:N}; \theta_q)$ est calculable et la re-estimation $p_{i,j}^{q+1}$ de $p_{i,j}$ donne :

$$p_{i,j}^{q+1} = \frac{1}{2(N-1)} \sum_{n=1}^{N-1} \left[p(x_n = \omega_i, x_{n+1} = \omega_j | y_{1:N}; \theta_q) + p(x_n = \omega_j, x_{n+1} = \omega_i | y_{1:N}; \theta_q) \right].$$
(5.21)

L'estimation des paramètres m_{ω_j} , $\sigma_{\omega_j}^2$ et des paramètres de dépendance est plus délicate car si $x_{n+1} \neq x_n$, la covariance de $p(y_{1:n+1}|x_{1:n+1})$ n'est pas égale à $\Gamma_{x_{n+1}}^{n+1}$.

Pour voir les difficultés liées à l'estimation de ces paramètres, considérons l'exemple suivant :

Exemple 5.2.1. Supposons K = 2 et N = 10 et considérons le modèle où la covariance est donnée par $\gamma(k) = \sigma^2 (k+1)^{-\alpha}$. On observe un échantillon $(x_{1:10}, y_{1:10})$ et on cherche à estimer les moyennes m_{ω_1} et m_{ω_2} , les variances $\sigma_{\omega_1}^2$ et $\sigma_{\omega_2}^2$ et les paramètres de dépendance α_{ω_1} et α_{ω_2} .

Supposons que l'on ait $x_{1:10} = (\omega_1, \omega_1, \omega_2, \omega_2, \omega_1, \omega_1, \omega_2, \omega_2, \omega_2, \omega_2)$. S'il s'agissait du modèle classique de chaînes de Markov cachées à bruit corrélé, nous aurions $p(y_n|x_{1:10}) = p(y_n|x_n)$. Sous une telle hypothèse, les estimateurs des moyennes et des covariances seraient donnés par :

$$\hat{m}_{\omega_{1}}(x_{1:10}, y_{1:10}) = \frac{y_{1} + y_{2} + y_{5} + y_{6}}{4},
\hat{m}_{\omega_{2}}(x_{1:10}, y_{1:10}) = \frac{y_{3} + y_{4} + y_{7} + y_{8} + y_{9} + y_{10}}{6},
\hat{\gamma}_{\omega_{1}}(1) = \frac{(y_{1} - \hat{m}_{\omega_{1}}, y_{2} - \hat{m}_{\omega_{1}}) \left(\begin{array}{c} y_{1} - \hat{m}_{\omega_{1}} \\ y_{2} - \hat{m}_{\omega_{1}} \end{array}\right) + (y_{5} - \hat{m}_{\omega_{1}}, y_{6} - \hat{m}_{\omega_{1}}) \left(\begin{array}{c} y_{5} - \hat{m}_{\omega_{1}} \\ y_{6} - \hat{m}_{\omega_{1}} \end{array}\right)}{2},$$

et une formule similaire pour $\hat{\gamma}_{\omega_2}(1)$. L'utilisation de la dernière formule est possible dans le cas du modèle classique de chaînes de Markov cachées car $p(y_n, y_{n+1}|x_{1:n+1}) = p(y_n, y_{n+1}|x_n, x_{n+1})$. Cependant, cette égalité n'est plus vraie dans le cas des chaînes de Markov cachées dont les observations sont à dépendance longue. Considérons les échantillons $x_{1:4} = (\omega_1, \omega_1, \omega_2, \omega_2)$ et $y_{1:4}$ extraits de $x_{1:10}$ et de $y_{1:10}$. La matrice

$$\Gamma_{\omega_1}^2 = \begin{pmatrix} \gamma_{\omega_1}(0) & \gamma_{\omega_1}(1) \\ \gamma_{\omega_1}(1) & \gamma_{\omega_1}(0) \end{pmatrix}$$
(5.22)

est bien la matrice de covariance de $p(y_1, y_2 | x_1, x_2)$ tandis que la matrice

$$\Gamma_{\omega_2}^2 = \begin{pmatrix} \gamma_{\omega_2}(0) & \gamma_{\omega_2}(1) \\ \gamma_{\omega_2}(1) & \gamma_{\omega_2}(0) \end{pmatrix}$$
(5.23)

n'est plus la matrice de covariance de $p(y_3, y_4|x_{1:4})$.

Il en est de même si le modèle considéré est un modèle FARIMA ou bruit gaussien fractionnaire. La densité spectrale de $p(y_1, y_2|x_1, x_2)$ est bien la transformée de Fourier de la famille de covariances définie par (5.22) mais celle de $p(y_3, y_4|x_{1:4})$ n'est pas la transformée de Fourier de la famille de covariances définie par (5.23).

Revenons au problème général. Notons $(m_{\omega_j}^{(q)})_{1 \le j \le K}$, $(\Gamma_{\omega_j}^{n,(q)})_{1 \le j \le K}$ les moyennes et les matrices de covariance correspondant aux paramètres estimés lors de l'étape q de ICE. D'après ce qui précède, si $x_{n_1} = \ldots = x_{n_1+n_2}$, sous le paramètre θ_q , la distribution $p(y_{n_1:n_1+n_2}|x_{1:n_1+n_2})$ n'est pas la distribution normale de moyenne $m_{x_{n_1}}^{n_2+1,(q)} = (\underbrace{m_{x_{n_1}}^{(q)}, \ldots, m_{x_{n_1}}^{(q)}}_{n_2+1 \text{ fois}})$ et de covariance

 $\Gamma_{x_{n_1}}^{n_2+1,(q)}$. En fait, la distribution $p(y_{n_1:n_1+n_2}|x_{1:n_1+n_2})$ est la distribution marginale de la loi normale $p(y_{1:n_1+n_2}|x_{1:n_1+n_2})$ de moyenne $M^{x_{1:n_1+n_2},(q)}$ et de covariance $\Gamma^{x_{1:n_1+n_2},(q)}$ calculées

par la procédure décrite dans la proposition 5.2.3 avec $m_{\omega_j} = m_{\omega_j}^{(q)}$ et $\Gamma_{\omega_j}^n = \Gamma_{\omega_j}^{n,(q)}$. Notons $M^{x_{n_1:n_1+n_2},(q)}$ et $\Gamma^{x_{n_1:n_1+n_2},(q)}$ la moyenne et la covariance de la distribution $p(y_{n_1:n_1+n_2}|x_{1:n_1+n_2})$ sous θ_q . Afin de pouvoir re-estimer les paramètres par l'algorithme ICE, nous devons effectuer une transformation de l'échantillon, soit $\tilde{y}_{n_1:n_1+n_2} = Ay_{n_1:n_1+n_2}$, de façon à ce que $p(\tilde{y}_{n_1:n_1+n_2}|x_{1:n_1+n_2})$ suive une loi normale de moyenne $m_{x_{n_1}}^{n_2+1,(q)}$ et de covariance $\Gamma_{x_{n_1}}^{n_2+1,(q)}$. Notons $\Gamma^{x_{n_1:n_1+n_2},(q)} = CC^T$ et $\Gamma_{x_{n_1}}^{n_2+1,(q)} = DD^T$ les transformées de Choleski des matrices $\Gamma^{x_{n_1:n_1+n_2},(q)}$ et $\Gamma_{x_{n_1}}^{n_2+1,(q)}$. Alors la transformation qui convient est :

$$\tilde{y}_{n_1:n_1+n_2} = DC^{-1} \left(y_{n_1:n_1+n_2} - M^{x_{n_1:n_1+n_2},(q)} \right) (C^T)^{-1} D^T + m_{x_{n_1}}^{n_2+1,(q)}.$$
(5.24)

Finalement, l'algorithme ICE proposé fonctionne de la manière suivante :

- 1. Sous θ_q , calcular les distributions $p(y_{1:n}|x_{1:n})$ en utilisant la proposition 5.2.3 avec $m_{\omega_i} = m_{\omega_i}^{(q)}$ et $\Gamma_{\omega_i}^n = \Gamma_{\omega_i}^{n,(q)}$ pour tout $1 \le i \le K$;
- 2. considérer $J(i) = \{n_1, \ldots, n_r\}$ sous-ensemble de $\{1, \ldots, N\}$, tel que pour tout $1 \leq j \leq r, x_{n_j-1} \neq x_{n_j}$ et il existe $m_j \geq 0$ tel que $x_{n_j} = x_{n_j+1} = \ldots = x_{n_j+m_j} = \omega_i$ et $x_{n_j} \neq x_{n_j+m_j+1}$. Considérer les r échantillons correspondants $(y_{n_1}, \ldots, y_{n_1+m_1}), \ldots, (y_{n_r}, \ldots, y_{n_r+m_r});$
- 3. soient, pour tout $1 \leq j \leq r$, $M^{x_{n_j:n_j+m_j},(q)}$ et $\Gamma^{x_{n_j:n_j+m_j},(q)}$ les moyennes et variances de $p(y_{n_j}, \ldots, y_{n_j+m_j}|x_{1:n_j+m_j})$ calculés à l'étape 1. Calculer les transformées de Choleski $\Gamma^{x_{n_j:n_j+m_j},(q)} = C_j C_j^T$ et $\Gamma^{m_j+1,(q)}_{\omega_i} = D_j D_j^T$ et poser :

$$\begin{pmatrix} \tilde{y}_{n_j} \\ \tilde{y}_{n_j+1} \\ \vdots \\ \tilde{y}_{n_j+m_j} \end{pmatrix} = D_j C_j^{-1} \left(\begin{pmatrix} y_{n_j} \\ y_{n_j+1} \\ \vdots \\ y_{n_j+m_j} \end{pmatrix} - M^{x_{n_j:n_j+m_j},(q)} \right) (C_j^T)^{-1} D_j^T + m_{\omega_i}^{m_j+1,q} ;$$

4. estimer m_i , $\gamma_i(0)$ et le paramètre de dépendance correspondant à la classe *i* à partir des échantillons $(\tilde{y}_{n_i}, \ldots, \tilde{y}_{n_i+m_i})$ par les estimateurs de la sous-section 5.1.4.

Remarque : Lorsque la covariance est de la forme $\gamma(k) = \sigma^2 (k+1)^{-\alpha}$, l'estimation de α utilise seulement les estimations de $\gamma(0)$ et $\gamma(1)$, on peut donc se contenter d'échantillons $(\tilde{y}_{n_j}, \tilde{y}_{n_j+1})$ de taille $m_j + 1 = 2$. L'algorithme ICE s'écrit alors :

- 1. Sous θ_q , calcular les distributions $p(y_{1:n}|x_{1:n})$ en utilisant la proposition 5.2.3 avec $m_{\omega_i} = m_{\omega_i}^{(q)}$ et $\Gamma_{\omega_i}^n = \Gamma_{\omega_i}^{n,(q)}$ pour tout $1 \le i \le K$;
- 2. considérer $J(i) = \{n_1, \ldots, n_r\}$ sous-ensemble de $\{1, \ldots, N\}$, tel que pour tout $1 \le j \le r, x_{n_j} = x_{n_j+1} = \omega_i$. Considérer les r échantillons correspondants $(y_{n_1}, y_{n_1+1}), \ldots, (y_{n_r}, y_{n_r+1});$
- 3. soient, pour tout $1 \leq j \leq r$, $M^{x_{n_j:n_j+1},(q)}$ et $\Gamma^{x_{n_j:n_j+1},(q)}$ les moyennes et variances de $p(y_{n_j}, y_{n_j+1}|x_{1:n_j+1})$ calculées à l'étape 1. Calculer les transformées de Choleski $\Gamma^{x_{n_j:n_j+1},(q)} = C_j C_j^T$ et $\Gamma^{2,(q)}_{\omega_i} = D_j D_j^T$ et poser :

$$\begin{pmatrix} \tilde{y}_{n_j} \\ \tilde{y}_{n_j+1} \end{pmatrix} = D_j C_j^{-1} \left(\begin{pmatrix} y_{n_j} \\ y_{n_j+1} \end{pmatrix} - M^{x_{n_j:n_j+1},(q)} \right) (C_j^T)^{-1} D_j^T + m_{\omega_i}^{2,q} ;$$

4. estimer les paramètres m_{ω_i} , $\sigma_{\omega_i}^2$ et α_{ω_i} par :

$$m_{\omega_i}^{q+1} = \frac{1}{2} \left(m_{\omega_i,1}^{q+1} + m_{\omega_i,2}^{q+1} \right) ;$$

$$(\sigma_{\omega_i}^{q+1})^2 = \frac{1}{2} \left(\hat{\gamma}_1(0) + \hat{\gamma}_2(0) \right) ;$$

$$\alpha_{\omega_i}^{q+1} = -\frac{\log\left(\frac{\hat{\gamma}(1)}{(\sigma_{\omega_i}^{q+1})^2}\right)}{\log(2)},$$

où

$$\begin{pmatrix} m_{\omega_i,1}^{q+1} \\ m_{\omega_i,2}^{q+1} \end{pmatrix} = \frac{1}{r} \sum_{j=1}^r \begin{pmatrix} \tilde{y}_{n_j} \\ \tilde{y}_{n_j+1} \end{pmatrix},$$

 et

$$\begin{split} \Gamma_{\omega_{i}}^{2,(q+1)} &= \frac{1}{r} \sum_{j=1}^{r} \left(\begin{array}{c} \tilde{y}_{n_{j}} - m_{\omega_{i},1}^{q+1} \\ \tilde{y}_{n_{j}+1} - m_{\omega_{i},2}^{q+1} \end{array} \right) \left(\tilde{y}_{n_{j}} - m_{\omega_{i},1}^{q+1}, \tilde{y}_{n_{j}+1} - m_{\omega_{i},2}^{q+1} \right) \\ &= \left(\begin{array}{c} \hat{\gamma}_{1}(0) & \hat{\gamma}(1) \\ \hat{\gamma}(1) & \hat{\gamma}_{2}(0) \end{array} \right). \end{split}$$

5.2.4 Expérimentations

Nous proposons trois séries d'expériences. Pour les trois séries, la chaîne de Markov X est stationnaire et réversible à valeurs dans $\mathcal{X} = \{\omega_1, \omega_2\}$ avec $p(x_1 = \omega_1, x_2 = \omega_1) = p(x_1 = \omega_2, x_2 = \omega_2) = 0.495$ et $p(x_1 = \omega_1, x_2 = \omega_2) = p(x_1 = \omega_2, x_2 = \omega_1) = 0.005$. La covariance est de la forme $\gamma(k) = \sigma^2 (k+1)^{-\alpha}$, la taille des échantillons est N = 1000. Les simulations sont présentées dans la figure 5.1. Dans la première expérience, on considère une chaîne de Markov cachée à bruit indépendant dont les moyennes sont égales respectivement à 1 et 2 et dont les variances sont égales à 1. La réalisation observée $y_{1:N}$ est ensuite segmentée en utilisant trois méthodes. La première utilise les vrais paramètres du modèle. La seconde méthode utilise ICE pour estimer les paramètres et nous supposons que les données observées sont issues du vrai modèle de chaînes de Markov cachées à bruit indépendant (CMC-BI). Quant à la dernière méthode, on suppose que les données sont issues du modèle de chaînes de Markov cachées à mémoire longue (CMC-ML), les paramètres étant estimés par ICE proposé ci-dessus. Pour les trois méthodes, les états cachés sont estimés par MPM.



FIG. 5.1 – (a) Chaîne de Markov à deux états, (b) Version bruitée avec bruit indépendant, (c) Bruit à dépendance longue, mêmes moyennes, mêmes variances, $\alpha_{\omega_1} = 0.1$ et $\alpha_{\omega_2} = 1$, (d) Bruit à dépendance longue, $m_{\omega_1} = 1$ et $m_{\omega_2} = 2$, variance commune égale à 1, $\alpha_{\omega_1} = 0.1$ et $\alpha_{\omega_2} = 0.9$.

Paramètres	CMC-BI		CMC	C-ML	Vraies valeurs		
	$\omega_1 \omega_2$		ω_1	ω_2	ω_1	ω_2	
m	0.92	1.99	0.89	1.96	1	2	
σ^2	1 1		0.98 1.05		1	1	
α	-	-	> 100	> 100	-	-	
Taux d'erreur	5.2%		5.2%		4.1%		

TAB. 5.1 – Estimation de la loi d'observation en utilisant les modèles CMC-BI et CMC-ML. La réalisation $y_{1:N}$ est issue du modèle CMC-BI.

104



FIG. 5.2 – Segmentation de $y_{1:N}$ issue du modèle CMC-BI : (a) Fondée sur les vrais paramètres, 4.1% d'erreur, (b) Modèle CMC-BI : 5.2% d'erreur, (c) Modèle CMC-ML : 5.2% d'erreur.

D'après les figures 5.2,(b) et 5.2,(c), les résultats obtenus avec CMC-BI et CMC-ML sont comparables, ce qui montre la bonne robustesse, dans le cadre de l'expérience, du modèle à dépendance longue. On peut remarquer également que les moyennes et variances sont bien estimées et le paramètre de dépendance α est supérieur à 100 (voir tableau 5.1), ainsi le modèle CMC-ML est capable de reconnaître des situations où la covariance décroît rapidement.



FIG. 5.3 – Segmentation de $y_{1:N}$ issue du modèle CMC-ML (deuxième expérience) : (a) Fondée sur les vrais paramètres, 2.1% d'erreur, (b) Modèle CMC-BI : 48% d'erreur, (c) Modèle CMC-ML : 1.9% d'erreur.

La deuxième expérience est complémentaire de la précédente. Les deux moyennes sont égales à 0, les deux variances à 1 tandis que les paramètres de dépendance valent respectivement 0.1 et 1 (figure 5.1,(c)). Les données simulées ne peuvent alors être issues du modèle CMC-BI. Il est alors intéressant de savoir si le modèle CMC-BI peut tout de même bien estimer les états cachés.

106

Paramètres	CMC-BI		CMC	C-ML	Vraies valeurs		
	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2	
m	0.61	-0.41	0.24	0.22	0	0	
σ^2	0.35	0.26	0.37	0.81	1	1	
α	-	-	0.28	1.1	0.1	1	
Taux d'erreur	48%		1.9%		2.1%		

TAB. 5.2 – Estimation de la loi d'observation en utilisant les modèles CMC-BI et CMC-ML. La réalisation $y_{1:N}$ est issue du modèle CMC-ML (deuxième expérience).

On peut remarquer sur la figure 5.3 que les résultats obtenus en utilisant le modèle CMC-ML sont très bons tandis que les résultats obtenus en utilisant le modèle CMC-BI sont très médiocres. Il en résulte qu'il existe des situations dans lesquelles une CMC-ML ne peut être approchée par une CMC-BI. Concernant l'estimation des paramètres présentée dans le tableau 5.2, nous voyons que les paramètres sont mal estimés lorsque l'on utilise le modèle CMC-BI. Cependant, les paramètres ne sont pas non plus très bien estimés lorsque l'on utilise le modèle CMC-BI. CMC-ML.

Dans le dernier exemple, les moyennes des deux classes sont respectivement égales à 1 et 2, les paramètres de corrélation sont égaux à 0.1 et 0.9, la variance est égale à 1.



FIG. 5.4 – Segmentation de $y_{1:N}$ lorsque (X, Y) est une chaîne de Markov cachée à dépendance longue (troisième expérience) : (a) Basée sur les vrais paramètres, 4.9% d'erreur, (b) Modèle CMC-BI : 32.4% d'erreur, (c) Modèle CMC-ML : 4.1% d'erreur.

Paramètres	CMC-BI		CMC-ML		Vraies valeurs	
	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2
m	1.06	2.25	1.33	1.73	1	2
σ^2	0.42	0.67	0.53	1.17	1	1
α	-	-	0.22	0.66	0.1	0.9
Taux d'erreur	32.4%		4.1%		4.9%	

TAB. 5.3 – Estimation de la loi d'observation en utilisant les modèles CMC-BI et CMC-ML. La réalisation $y_{1:N}$ est issue du modèle CMC-ML (troisième expérience).

Selon les résultats de la figure 5.4, nous constatons que l'estimation des états cachés par le modèle CMC-ML est nettement meilleure que celle par le modèle CMC-BI. Cependant, on peut également constater que l'estimation des paramètres du modèle (Tableau 5.3) n'est pas très bonne en utilisant le modèle de chaînes de Markov cachées à dépendance longue. De plus, on peut remarquer que même avec des paramètres mal estimés, les états cachés restent bien estimés par le modèle CMC-ML. Ces différents résultats montrent que le modèle CMC-ML est très différent du modèle CMC-BI. Par ailleurs, la méthode ICE proposée semble bien adaptée à la segmentation non supervisée.

5.3 Chaînes semi-markoviennes cachées à dépendance longue

Nous proposons dans cette section un autre modèle partiellement de Markov permettant de modéliser des observations à dépendance longue. Dans ce nouveau modèle, chaque y_n ne dépend plus de tous les y_k tels que $k \leq n$ mais seulement des y_{n_1}, \ldots, y_n tels que $x_{n_1} =$ $\ldots = x_n$ et $x_{n_1-1} \neq x_{n_1}$. Ce nouveau modèle nous permettra de traiter des processus de taille plus grande. En effet, dans le modèle précédent, le calcul de la transformée de Choleski $\Gamma^{x_{n_1:n_1+n_2}} = CC^T$ nécessite le stockage des matrices $\Gamma^{x_{1:n}}$, ce qui rend l'algorithme d'estimation ICE inopérant lorsque le processus est de grande taille ($N \simeq 10^4$), en raison de débordement de mémoire. En effet, avec un ordinateur disposant d'une mémoire de 3 giga-octets, le logiciel MATLAB ne peut gérer des matrices de plus de 25×10^7 éléments réels. Cependant, le calcul de la transformée de Choleski nécessite également le stockage de C, ainsi les matrices ne peuvent avoir plus de 12.5×10^7 éléments réels et donc les processus ne peuvent être de taille plus grande que N = 11025. Par ailleurs, ce nouveau modèle est étendu au cas où la chaîne cachée est semi-markovienne.

5.3.1 Chaînes triplets partiellement de Markov et dépendance longue

On considère un processus triplet $Z = (X_n, U_n, Y_n)_{1 \le n \le N}$ tel que chaque X_n prend ses valeurs dans $\mathcal{X} = \{\omega_1, \ldots, \omega_K\}$, chaque U_n prend ses valeurs dans un ensemble fini $\Lambda = \{0, 1, \ldots, L\}$ et chaque Y_n prend ses valeurs dans \mathbb{R} .

Définition 5.3.1 (Chaînes triplets partiellement de Markov).

Le processus $Z = (X_n, U_n, Y_n)_{1 \le n \le N}$ est une chaîne triplet partiellement de Markov si le processus couple (V, Y) où V = (X, U) est une chaîne couple partiellement de Markov.

Afin de modéliser des observations à dépendance longue, nous proposons la distribution $p(x_{1:N}, u_{1:N}, y_{1:N})$ suivante :

$$p(x_{1:N}, u_{1:N}, y_{1:N}) = p(x_1)\delta_0(u_1)p(y_1|x_1)\prod_{n=1}^{N-1} p(x_{n+1}|x_n)p(u_{n+1}|x_n, x_{n+1}, u_n)p(y_{n+1}|x_{n+1}, u_{n+1}, y_{1:n})$$
(5.25)

où

$$p(u_{n+1}|x_n, x_{n+1}, u_n) = \begin{cases} \delta_{u_n+1}(u_{n+1}) \text{ si } x_n = x_{n+1} \text{ et } u_n < L; \\ \delta_0(u_{n+1}) \text{ si } x_n \neq x_{n+1} \text{ ou } u_n = L; \end{cases}$$
(5.26)

$$p(y_{n+1}|x_{n+1}, u_{n+1}, y_{1:n}) = \begin{cases} p(y_{n+1}|x_{n+1}) & \text{if } u_{n+1} = 0; \\ p(y_{n+1}|x_{n+1}, y_{n-u_{n+1}+1:n}) & \text{if } u_{n+1} > 0; \end{cases}$$
(5.27)

et $p(y_{n+1}|x_{n+1}, y_{n-u_{n+1}+1:n})$ est tel que $p(y_{n-u_{n+1}+1:n+1}|x_{n+1})$ suive une loi normale de moyenne

$$m_{x_{n+1}}^{u_{n+1}+1} = (\underbrace{m_{x_{n+1}}, \dots, m_{x_{n+1}}}_{u_{n+1}+1 \text{ fois}}),$$

et de matrice de covariance

$$\Gamma_{x_{n+1}}^{u_{n+1}+1} = \begin{pmatrix} \gamma_{x_{n+1}}(0) & \gamma_{x_{n+1}}(1) & \dots & \gamma_{x_{n+1}}(u_{n+1}-1) & \gamma_{x_{n+1}}(u_{n+1}) \\ \gamma_{x_{n+1}}(1) & \gamma_{x_{n+1}}(0) & \dots & \gamma_{x_{n+1}}(u_{n+1}-2) & \gamma_{x_{n+1}}(u_{n+1}-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{x_{n+1}}(u_{n+1}) & \gamma_{x_{n+1}}(u_{n+1}-1) & \dots & \gamma_{x_{n+1}}(1) & \gamma_{x_{n+1}}(0) \end{pmatrix}.$$

On remarque que X est une chaîne de Markov, de plus u_n représente le temps de séjour minimal écoulé dans l'état x_n depuis le changement d'état. Si L = N, u_n est le temps de séjour exact écoulé depuis le changement d'état. Ce modèle ressemble à celui des chaînes semi-markoviennes, il en est cependant très différent. Dans le modèle des chaînes semimarkoviennes, le processus U permet de définir la loi de X, tandis que dans notre modèle triplet partiellement de Markov, la loi de X est déjà définie comme celle d'une chaîne de Markov et le processus U est défini à partir de X.

Comme pour le modèle couple partiellement de Markov décrit dans la section précédente, l'algorithme de Baum-Welsh est utilisable si les quantités $p(y_{n+1}|x_{n+1}, u_{n+1}, y_{1:n})$ sont calculables. En effet, lorsque $u_{n+1} = 0$, $p(y_{n+1}|x_{n+1}, u_{n+1}, y_{1:n}) = p(y_{n+1}|x_{n+1})$ est une loi normale dans \mathbb{R} de moyenne $m_{x_{n+1}}$ et de variance $\gamma_{x_{n+1}}(0)$; et lorsque $u_{n+1} > 0$, $p(y_{n+1}|x_{n+1}, u_{n+1}, y_{1:n}) = p(y_{n+1}|x_{n+1}, y_{n-u_{n+1}+1:n})$ est une loi normale de moyenne

$$M_{x_{n+1}} = m_{x_{n+1}} + \left(\gamma_{x_{n+1}}(u_{n+1}), \dots, \gamma_{x_{n+1}}(1)\right) \left(\Gamma_{x_{n+1}}^{u_{n+1}}\right)^{-1} \left(y_{n-u_{n+1}+1:n} - m_{x_{n+1}}^{u_{n+1}}\right),$$

et de variance

$$\tilde{\gamma}_{x_{n+1}}(0) = \gamma_{x_{n+1}}(0) - (\gamma_{x_{n+1}}(u_{n+1}), \dots, \gamma_{x_{n+1}}(1)) \left(\Gamma_{x_{n+1}}^{u_{n+1}}\right)^{-1} \begin{pmatrix} \gamma_{x_{n+1}}(u_{n+1}) \\ \vdots \\ \gamma_{x_{n+1}}(1) \end{pmatrix}.$$

Pour calculer les quantités $p(y_{n+1}|x_{n+1}, y_{n-u_{n+1}+1:n})$, il n'est pas nécessaire de stocker les matrices $\Gamma_{x_{n+1}}^{u_{n+1}}$, celles-ci pouvant être de grande taille dans le cas des applications en segmentation d'image. En effet, les matrices $\Gamma_{x_{n+1}}^{u_{n+1}}$ sont de Toeplitz (la valeur de l'élément (i, j) de la matrice ne dépend que de la différence |i - j|) et de plus le vecteur $(\gamma_{x_{n+1}}(u_{n+1}), \ldots, \gamma_{x_{n+1}}(1))$ est extrait de la matrice $\Gamma_{x_{n+1}}^{u_{n+1}+1}$. Ainsi nous pouvons utiliser l'algorithme de Durbin-Levinson décrit dans la sous-section 6.3.1 du chapitre 6 pour calculer $(\gamma_{x_{n+1}}(u_{n+1}), \ldots, \gamma_{x_{n+1}}(1)) (\Gamma_{x_{n+1}}^{u_{n+1}})^{-1}$.

5.3.2 Le modèle semi-markovien

Le modèle précédent est enrichi en introduisant un autre processus auxiliaire modélisant la semi-markovianité. Considérons le processus $Z = (X_n, U_n^1, U_n^2, Y_n)_{1 \le n \le N}$, où chaque X_n prend ses valeurs dans $\mathcal{X} = \{\omega_1, \ldots, \omega_K\}$, chaque U_n^1 prend ses valeurs dans $\Lambda^1 = \{1, \ldots, L_1\}$, chaque U_n^2 prend ses valeurs dans $\Lambda^2 = \{0, \ldots, L_2\}$ et chaque Y_n prend ses valeurs dans \mathbb{R} . Dans ce modèle, on considère que X est une chaîne semi-markovienne telle que la loi de (X, U^1) soit définie par les formules (3.15), (3.16) et (3.17). La loi de (X, U^1, U^2, Y) est alors définie par :

$$p(x_{1:N}, u_{1:N}^1, u_{1:N}^2, y_{1:N}) = p(x_1)p(u_1^1|x_1)\delta_0(u_1^2)p(y_1|x_1)$$

$$\times \prod_{n=1}^{N-1} p(x_{n+1}|x_n, u_n^1)p(u_{n+1}^1|x_{n+1}, u_n^1)p(u_{n+1}^2|x_n, x_{n+1}, u_n^2)p(y_{n+1}|x_{n+1}, u_{n+1}^2, y_{1:n}), \quad (5.28)$$

avec

$$p(u_1^1|x_1) = \bar{d}(x_1, u_1^1); \tag{5.29}$$

$$p(x_{n+1}|x_n, u_n^1) = \begin{cases} \delta_{x_n}(x_{n+1}) \text{ si } u_n^1 > 1; \\ r(x_{n+1}|x_n) \text{ si } u_n^1 = 1; \end{cases}$$
(5.30)

$$p(u_{n+1}^1|x_{n+1}, u_n^1) = \begin{cases} \delta_{u_n^1 - 1}(u_{n+1}^1) & \text{is } u_n^1 > 1; \\ \bar{d}(x_{n+1}, u_{n+1}^1) & \text{is } u_n^1 = 1; \end{cases}$$
(5.31)

$$p(u_{n+1}^2|x_n, x_{n+1}, u_n^2) = \begin{cases} \delta_{u_n^2 + 1}(u_{n+1}^2) & \text{si } x_n = x_{n+1} \text{ et } u_n < L_2; \\ \delta_0(u_{n+1}^2) & \text{si } x_n \neq x_{n+1} \text{ ou } u_n = L_2; \end{cases}$$
(5.32)

$$p(y_{n+1}|x_{n+1}, u_{n+1}^2, y_{1:n}) = \begin{cases} p(y_{n+1}|x_{n+1}) & \text{if } u_{n+1}^2 = 0; \\ p(y_{n+1}|x_{n+1}, y_{n-u_{n+1}^2+1:n}) & \text{if } u_{n+1}^2 > 0; \end{cases}$$
(5.33)

où pour tout ω_j , $\bar{d}(\omega_j, .)$ est une densité de probabilité sur Λ^1 et r(.|.) est un noyau de transition sur \mathcal{X}^2 conformément aux notations de la sous-section 3.3.3 du chapitre 3.

Dans la suite, nous appelerons ce modèle "chaînes semi-markoviennes cachées à mémoire longue" (CSMC-ML). Lorsque le processus X est une chaîne de Markov, on l'appelera "chaînes triplets partiellement de Markov à mémoire longue" (CTPM-ML) pour le différencier du modèle étudié dans la section 5.2.

5.3.3 Estimation des paramètres

Nous ne traiterons ici que de l'estimation de la loi d'observation, l'estimation de la loi $p(x_{1:N}, u_{1:N}^1, u_{1:N}^2)$ ayant été traitée au chapitre 3. Si θ_q désigne le vecteur paramètre obtenu à l'étape q de ICE, on procède de la manière suivante :

1. simuler un échantillon $(x_{1:N}, u_{1:N}^1, u_{1:N}^2)$ selon la loi a posteriori $p(x_{1:N}, u_{1:N}^1, u_{1:N}^2|y_{1:N}; \theta_q)$;

- 2. pour chaque ω_j , considérer l'échantillon $(y_m^{\omega_j})_{1 \le m \le N_j}$, où pour tout $m, x_m = \omega_j$ et N_j est le nombre de x_n égaux à ω_j ;
- 3. estimer les paramètres correspondant à l'état ω_j à partir de $(y_m^{\omega_j})_{1 \le m \le N_j}$ en utilisant les estimateurs présentés à la sous-section 5.1.4.

5.3.4 Expérimentations

Le modèle de chaînes semi-markoviennes cachées à dépendance longue généralise à la fois le modèle de chaînes semi-markoviennes cachées (CSMC) et le nouveau modèle de chaînes triplets partiellement de Markov à dépendance longue (CTPM-ML). Le but de cette sous-section est de regarder les améliorations apportées par ces deux généralisations. Les expérimentations seront faites sur des processus de taille $N = 128 \times 128$. Les réalisations de ces processus seront représentées comme des images bi-dimensionnelles grâce au parcours d'Hilbert-Peano.

Nous présentons trois séries d'expériences. Dans la première série, les données sont issues du modèle CSMC et la question est de savoir si le modèle CSMC-ML, plus complexe, donne des résultats similaires à ceux obtenus par le modèle CSMC. Dans la seconde série, les données sont issues du modèle CSMC-ML et nous estimons les paramètres et les états cachés en utilisant les modèles CSMC, CTPM-ML et CSMC-ML. Nous conclurons cette sous-section par la segmentation d'une image réelle, les données n'étant alors probablement issues d'aucun des 3 modèles. Dans toutes les expériences proposées dans cette sous-section, la covariance est de la forme $\gamma(k) = \sigma^2 (k+1)^{-\alpha}$.

Dans la première expérience, on considère une chaîne semi-markovienne cachée (CSMC) $(X, U^1, Y) = (X_n, U_n^1, Y_n)_{1 \le n \le N}$ telle que chaque X_n prend ses valeurs dans $\mathcal{X} = \{\omega_1, \omega_2\}$ et chaque U_n^1 prend ses valeurs dans $\Lambda^1 = \{1, \ldots, 10\}$. La loi $p(x_{1:N}, u_{1:N}^1)$ est définie par (3.15), (3.16) et (3.17) et la loi $p(y_{1:N}|x_{1:N}, u_{1:N}^1)$ est donnée par :

$$p(y_{1:N}|x_{1:N}, u_{1:N}^1) = \prod_{n=1}^N p(y_n|x_n).$$

Lorsque $x_n = \omega_1$, $p(y_n|x_n)$ est une loi normale de moyenne 1 et de variance 20 et lorsque $x_n = \omega_2$, c'est une loi normale de moyenne 2 et de variance 20. La distribution $\bar{d}(\omega_j, .)$ est celle d'une loi uniforme sur Λ^1 pour tout $j \in \{1, 2\}$. De plus, $r(x_{n+1}|x_n) = 0.999$ si $x_n = x_{n+1}$ et $r(x_{n+1}|x_n) = 0.001$ si $x_n \neq x_{n+1}$. La réalisation $y_{1:N}$ est ensuite segmentée en utilisant trois méthodes. La première suppose que les données sont issues du modèle CSMC et utilise les vrais paramètres. La seconde méthode suppose que les données sont issues du modèle CSMC et utilise les données sont issues du modèle CSMC et utilise les données sont issues du modèle CSMC. La première par ICE. Quant à la troisième méthode, on suppose que les données sont issues du modèle CSMC et utilise les données sont issues du modèle CSMC. Les résultats sont présentés par ICE. Dans le modèle CSMC-ML, on supposer que $L_2 = 50$. Les résultats sont présentés dans la figure 5.5 et dans le tableau 5.4.



FIG. 5.5 – Segmentation de $y_{1:N}$ lorsque (X, Y) est une chaîne semi-markovienne cachée à bruit indépendant : (a) Réalisation de X, (b) Réalisation de Y, (c) Modèle CSMC avec les vrais paramètres, 3.74% d'erreur, (d) Modèle CSMC en non supervisé : 4.55% d'erreur, (e) Modèle CSMC-ML en non supervisé : 4.57% d'erreur.

Paramètres	CSMC		CSM	C-ML	Vraies valeurs		
	$\omega_1 \qquad \omega_2$		ω_1	$\omega_1 \qquad \omega_2$		ω_2	
m	1.06	2.04	0.98	1.97	1	2	
σ^2	19.81	20.71	19.84	20.46	20	20	
α	-	-	15.28	5.95	-	-	
Taux d'erreur	4.55%		4.5	7%	3.74%		

TAB. 5.4 – Estimation des paramètres à partir de données issues d'une chaîne semimarkovienne à bruit indépendant.

Nous constatons que les résultats sont similaires en utilisant les modèles CSMC et CSMC-ML, ce qui montre la capacité du modèle CSMC-ML à traiter des données issues du modèle CSMC. De plus l'estimation du paramètre α prouve que le modèle CSMC-ML est capable de reconnaître les situations où la covariance décroît rapidement.

Dans la seconde expérience, nous segmentons des données issues du modèle CSMC-ML. Le but de cette expérience est de savoir lequel des deux modèles CSMC ou CTPM-ML donne de meilleurs résultats. La chaîne semi-markovienne (X, U^1) considérée suit la même loi que dans l'expérience précédente. Concernant le processus U^2 , on prendra $L_2 = 50$. Les lois $p(y_n|x_n)$ seront respectivement la loi normale de moyenne 1 et de variance 1 lorsque $x_n = \omega_1$ et la loi normale de moyenne 2 et de variance 1 lorsque $x_n = \omega_2$. Le paramètre de dépendance est égal pour les deux classes à 0.5.



FIG. 5.6 – Segmentation de $y_{1:N}$ lorsque (X, Y) est une chaîne semi-markovienne cachée à mémoire longue : (a) Réalisation de X, (b) Réalisation de Y, (c) Modèle CSMC en non supervisé : 31.15% d'erreur, (d) Modèle CTPM-ML en non supervisé : 21.53% d'erreur, (e) Modèle CSMC-ML en non supervisé : 3.16% d'erreur.

Paramètres	CSMC		CTPM-ML		CSMC-ML		Vraies valeurs	
	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2
m	0.97	2.44	1.08	2.22	0.98	1.97	1	2
σ^2	0.59	0.56	0.83	0.78	0.96	0.93	1	1
α	-	-	0.69	0.72	0.62	0.61	0.5	0.5
Taux d'erreur	31.15%		21.53%		3.16%		2.85%	

TAB. 5.5 – Estimation des paramètres à partir de données issues de chaînes semi-markoviennes cachées à dépendance longue.

Nous constatons d'après la figure 5.6 et le tableau 5.5, que négliger la dépendance longue donne de plus mauvais résultats que négliger la semi-markovianité. Cependant, d'après la figure 5.6,(e), la considération de la semi-markovianité est importante et améliore nettement les résultats comparés à ceux obtenus à la figure 5.6,(d). Ces remarques sont encore justifiées dans l'estimation des paramètres présentée dans le tableau 5.5. Cependant les qualités d'estimation des paramètres par CTPM-ML et CSMC-ML sont quasi similaires. Globalement, l'expérience montre que le modèle CSMC-ML est plus riche, de manière significative, que chacun des deux modèles CTMP-ML et CSMC.

Dans la dernière expérience, on choisit de segmenter à l'aide des trois modèles CSMC, CTPM-ML et CSMC-ML une image dessinée. De l'image réelle représentant des cercles concentriques, on obtient la réalisation $x_{1:N}$ par le parcours d'Hilbert-Peano. La réalisation $x_{1:N}$ est ensuite bruitée à l'aide du modèle défini par les formules (5.26) et (5.27). Dans cette expérience, on prendra $L_2 = 50$, les moyennes des deux classes seront respectivement égales à 1 et 2, la variance sera égale à 1, et le paramètre de dépendance sera égal à 0.9.



FIG. 5.7 – Segmentation d'une image réelle bruitée avec de la dépendance longue : (a) Réalisation de X, (b) Réalisation de Y, (c) Modèle CSMC en non supervisé : 24.70% d'erreur, (d) Modèle CTPM-ML en non supervisé : 18.27% d'erreur, (e) Modèle CSMC-ML en non supervisé : 6.31% d'erreur.

Paramètres	CSMC		CTPM-ML		CSMC-ML		Vraies valeurs	
	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2	ω_1	ω_2
m	0.75	2.26	0.91	2.04	0.99	1.99	1	2
σ^2	0.69	0.66	0.92	0.93	1.01	1.02	1	1
α	-	-	1.07	1.07	0.93	0.92	0.9	0.9
Taux d'erreur	24.7	70%	18.	27%	6.3	1%	5.9	5%

TAB. 5.6 – Estimation des paramètres à partir de données issues d'une image réelle

Des résultats présentés dans la figure 5.7 et dans le tableau 5.6, nous voyons que le modèle semi-markovien est suffisamment général pour prendre en compte des propriétés statistiques de l'image que le modèle markovien ne prend pas en compte. Ainsi, on peut constater une meilleure segmentation en utilisant le modèle CSMC-ML (Figure 5.7,(e)) qu'en utisant le modèle CTPM-ML (Figure 5.7,(d)). Cependant, si on néglige la dépendance longue, les paramètres de la loi d'observation sont mal estimés (Tableau 5.6), ce qui entraîne une mauvaise estimation des états cachés (Figure 5.7,(c)).

5.4 Observations non gaussiennes à dépendance longue

Le dernier modèle que nous présentons dans ce chapitre permet de modéliser des observations non gaussiennes. Nous avons vu au chapitre 4 qu'il est possible d'écrire la loi jointe d'un vecteur aléatoire à partir de ses lois marginales et d'une fonction d'agrégation appelée "copule". Dans ce chapitre, les copules vont nous permettre ainsi de modéliser des observations non gaussiennes à dépendance longue.

5.4.1 Dépendance longue et copules

On considère un processus triplet $(X, U, Y) = (X_n, U_n, Y_n)_{1 \le n \le N}$ tel que chaque X_n prend ses valeurs dans un ensemble fini $\mathcal{X} = \{\omega_1, \ldots, \omega_K\}$, chaque U_n prend ses valeurs dans l'ensemble fini $\Lambda = \{0, \ldots, L\}$ et chaque Y_n prend ses valeurs dans \mathbb{R} . La distribution $p(x_{1:N}, u_{1:N}, y_{1:N})$ est celle d'une chaîne triplet partiellement de Markov définie par (5.25), (5.26) et (5.27). A la différence avec l'ancien modèle où la distribution $p(y_{n-u_{n+1}+1:n+1}|x_{n+1})$ était celle d'une loi normale, dans le modèle proposé, elle s'écrit :

$$p(y_{n-u_{n+1}+1:n+1}|x_{n+1}) = \prod_{k=n-u_{n+1}+1}^{n+1} p(y_k|x_{n+1}) \\ \times c_{x_{n+1}}^{u_{n+1}+1} \left(F_{x_{n+1}}(y_{n-u_{n+1}+1}), \dots, F_{x_{n+1}}(y_{n+1}) \right),$$
(5.34)

où pour chaque $\omega_j \in \mathcal{X}$, F_{ω_j} est une fonction de répartition marginale, $c_{\omega_j}^{u_{n+1}+1}$ est la copule gaussienne de matrice de corrélation :

$$R_{\omega_{j}}^{u_{n+1}+1} = \begin{pmatrix} 1 & \gamma_{\omega_{j}}(1) & \dots & \gamma_{\omega_{j}}(u_{n+1}-1) & \gamma_{\omega_{j}}(u_{n+1}) \\ \gamma_{\omega_{j}}(1) & 1 & \gamma_{\omega_{j}}(1) & \dots & \gamma_{\omega_{j}}(u_{n+1}-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{\omega_{j}}(u_{n+1}) & \gamma_{\omega_{j}}(u_{n+1}-1) & \dots & \gamma_{\omega_{j}}(1) & 1 \end{pmatrix},$$

et chaque $p(y_k|x_{n+1} = \omega_j)$ est une distribution de fonction de répartition F_{ω_j} . Ainsi, si $x_n = x_{n+1} = \ldots = x_{n+k}$, alors (y_n, \ldots, y_{n+k}) est la réalisation d'un vecteur aléatoire dont les lois marginales sont de fonction de répartition F_{x_n} et dont la copule est une copule gaussienne. Posons $z_n = \phi^{-1} \circ F_{x_n}(y_n)$, où ϕ est la fonction de répartition de la loi normale centrée et réduite. La distribution $p(x_{1:N}, u_{1:N}, z_{1:N})$ est également définie par (5.25), (5.26) et (5.27). Dans ce cas il s'agit du modèle triplet partiellement de Markov tel que $p(z_{n-u_{n+1}+1:n+1}|x_{n+1})$ soit la distribution d'une loi normale à marginales centrées et réduites et de matrice de corrélation égale à $R_{x_{n+1}}^{u_{n+1}+1}$. Si pour tout $1 \leq j \leq K$, les familles de covariance $(\gamma_{\omega_j}(k))_{k \in \mathbb{N}}$ satisfont les propriétés de dépendance longue, l'échantillon $z_{1:N}$ est alors à dépendance longue conditionnellement aux classes; en d'autres termes, si $x_n = x_{n+1} = \ldots = x_{n+k}$, l'échantillon (z_n, \ldots, z_{n+k}) est la réalisation d'un processus à dépendance longue. Dans cette définition, nous introduirons également la notion de stationnarité du second ordre en copule.

Définition 5.4.1. Soit $Y = (Y_n)_{n \in \mathbb{Z}}$ un processus réel tel que pour tout $n \in \mathbb{Z}$, Y_n ait pour fonction de répartition F.

- On dit que le processus Y est stationnaire du second ordre en copule si le processus $U = (F(Y_n))_{n \in \mathbb{Z}}$ est stationnaire du second ordre;
- Lorsque Y est stationnaire du second ordre en copule, on définit les corrélations de Spearman $(\rho_S(k))_{k\in\mathbb{N}}$ par :

$$\rho_S(k) = \frac{\mathbb{E}\left[(F(Y_n) - \mathbb{E}(F(Y_n)))(F(Y_{n-k}) - \mathbb{E}(F(Y_{n-k}))) \right]}{\sqrt{\mathbb{E}\left((F(Y_n) - \mathbb{E}(F(Y_n)))^2 \right) \mathbb{E}\left((F(Y_{n-k}) - \mathbb{E}(F(Y_{n-k})))^2 \right)}}$$

- On dit que le processus Y est à dépendance longue si :

$$\rho_S(k) \sim_{+\infty} Ck^{-\alpha} \text{ où } \alpha \in]0,1] \text{ et } C \in \mathbb{R}.$$

Remarque : D'après ce qu'on a précisé au chapitre 4, les coefficients de mélangeance ne dépendent que de la copule. Ainsi si f est un C^1 -difféomorphisme croissant de \mathbb{R} dans \mathbb{R} , les processus $(Y_n)_{n\in\mathbb{Z}}$ et $(f(Y_n))_{n\in\mathbb{Z}}$ ont la même mélangeance. La dépendance longue peut se définir de manière plus générale : un processus est à dépendance longue s'il est ρ -mélangeant et son coefficient de ρ -mélangeance converge vers 0 en $n^{-\alpha}$ où $\alpha \in [0, 1]$.

5.4.2 Estimation des paramètres

Soit ϕ la fonction de répartition d'une loi normale centrée-réduite. Nous détaillons ici l'estimation de la loi d'observation $p(y_{1:N}|x_{1:N})$ par l'algorithme ICE. Si θ_q désigne le vecteur paramètre obtenu à l'étape q, on procède de la manière suivante :

- 1. simuler un échantillon $(x_{1:N}, u_{1:N})$ selon la loi a posteriori $p(x_{1:N}, u_{1:N}|y_{1:N}; \theta_q)$;
- 2. pour chaque ω_j , considérer l'échantillon $(y_m^{\omega_j})_{1 \le m \le N_j}$, où pour tout $m, x_m = \omega_j$ et N_j est le nombre de x_n égaux à ω_j ;
- 3. estimer la loi marginale $p(y_n|x_n = \omega_j)$ à partir de l'échantillon $(y_m^{\omega_j})_{1 \le m \le N_j}$. Soit $F_{\omega_j}^{q+1}$ la fonction de répartition correspondant aux paramètres re-estimés;
- 4. estimer les paramètres de corrélation à partir de l'échantillon $(\phi^{-1} \circ F_{\omega_j}^{q+1}(y_m^{\omega_j}))_{1 \le m \le N_j}$ avec les estimateurs présentés à la sous-section 5.1.4.

5.4.3 Expérimentations

Nous proposons dans cette sous-section deux expériences. Les expériences présentées permettent de savoir si le modèle à dépendance longue gaussien peut être utilisé dans le cas où les données ne sont pas issues d'un modèle gaussien. Elles permettent également de tester la robustesse lorsque l'on utilise différent modèle de bruit. Dans les deux expériences, la taille des échantillons est N = 1000, la chaîne de Markov X à valeurs dans $\mathcal{X} = \{\omega_1, \omega_2\}$ est stationnaire réversible de loi donnée par $p(x_n = \omega_1, x_{n+1} = \omega_1) = p(x_n = \omega_2, x_{n+1} = \omega_2) = 0.495$ et $p(x_n = \omega_1, x_{n+1} = \omega_2) = p(x_n = \omega_2, x_{n+1} = \omega_1) = 0.005$ et la famille de covariance $(\gamma_{\omega_j}(k))_{k \in \mathbb{N}}$ est celle d'un bruit gaussien fractionnaire. Les paramètres de Hurst pour les deux expériences sont respectivement égaux à $H_{\omega_1} = 0.5$ et $H_{\omega_2} = 0.99$. On définit pour la suite la loi Γ^- dite "Gamma signée" comme la loi de la variable aléatoire ϵW , où ϵ suit une loi uniforme sur $\{-1, 1\}$ et W suit une loi Γ de paramètres a et b.

Dans la première expérience, les lois marginales sont gaussiennes de moyenne nulle et de variance égale à 1. Les données sont ensuite segmentées par trois méthodes. La première utilise les vrais paramètres du modèle et utilise MPM pour estimer les états cachés. La deuxième suppose que les données sont issues du modèle triplet partiellement de Markov à observations gaussiennes à dépendance longue. La troisième méthode suppose que les données sont à dépendance longue mais les lois marginales sont des lois Γ^- . Le but de cette expérience est de savoir si choisir un autre modèle que le modèle gaussien peut dégrader les résultats si les données sont réellement gaussienne.



FIG. 5.8 – Segmentation de données issues du modèle triplet partiellement de Markov à bruit gaussien à dépendance longue : (a) Réalisation de X, (b) Réalisation de Y, (c) Avec les vrais paramètres, 1.3% d'erreur, (d) Modèle gaussien à dépendance longue, 9.3% d'erreur, (e) Modèle Γ^- à dépendance longue, 26.1% d'erreur.

Dans la seconde expérience, nous considérons le problème inverse. Les données sont simulées selon le modèle à dépendance longue dont les lois marginales sont des lois Γ^- de paramètres $a_{\omega_1} = a_{\omega_2} = 1$ et $b_{\omega_1} = b_{\omega_2} = 1$. Le but de cette expérience est de savoir si le modèle gaussien peut être suffisamment robuste pour être utilisé même si les données ne sont probablement pas gaussiennes. Notons que les copules sont gaussiennes dans les duex modèles, qui ne se différencient que par les lois marginales.



FIG. 5.9 – Segmentation de données issues du modèle triplet partiellement de Markov à bruit Γ^- à dépendance longue : (a) Réalisation de X, (b) Réalisation de Y, (c) Avec les vrais paramètres, 0.8% d'erreur, (d) Modèle gaussien à dépendance longue, 6.5% d'erreur, (e) Modèle Γ^- à dépendance longue, 1.9% d'erreur.

Ces expériences montrent qu'il est important de bien choisir la loi d'observation. En effet, si nous segmentons des données en utilisant une loi différente de celle du vrai modèle, les résultats peuvent se dégrader de manière non négligeable.

Conclusion

Nous avons proposé dans ce chapitre plusieurs modèles dans lesquels la chaîne inobservable est cachée par du bruit à dépendance longue. Dans tous les modèles proposés, les probabilités marginales a posteriori sont calculables, ce qui permet l'estimation de la chaîne inobservable par la méthode bayésienne MPM. Dans le premier modèle "chaînes de Markov cachée avec du bruit à mémoire longue gaussien" (CMC-ML), la loi de chacune des marginales Y_n (conditionnellement aux classes) dépend de toutes les marginales Y_k passées; cependant, les calculs se font grâce au caractère gaussien du bruit. Afin d'estimer les paramètres de ce modèle, nous avons proposé un algorithme ICE original et nous avons montré son bon comportement au travers des simulations. Nous avons également montré la bonne robustesse du modèle CMC-ML lorsque les données sont issues du modèle classique des chaînes de Markov cachées avec du bruit indépendant (CMC-BI). Le second modèle que nous avons proposé, qui est une "chaîne triplet partiellement de Markov cachée par du bruit à mémoire longue" (CTPM-ML), permet de palier aux difficultés algorithmiques du premier modèle et peut ainsi être utilisé en traitement d'image où les données traitées sont, en général, très volumineuses. Ce modèle a été ensuite étendu, en introduisant un processus auxiliaire, de façon à considérer la semi-markovianité éventuelle du processus caché. Nous avons montré aux travers des expérimentations que ni le modèle de chaînes semi-markoviennes cachées à bruit indépendant, ni le modèle CTPM-ML dans le lequel le processus caché est une chaîne de Markov, ne parvient à donner d'aussi bons résultats que le modèle général, où la chaîne cachée est semi-markovienne (les données sont issues de ce dernier). Enfin, tous les modèles précédents peuvent être généralisés avec l'introduction des lois marginales non nécessairement gaussiennes. En effet, en gardant les copules gaussiennes il est possible d'étendre les calculs faisables dans le cas gaussien aux cas où les marginales du bruit sont quelconques. Finalement, en utilisant tous les résultats exposés jusqu'à présent, il est possible de proposer un modèle général dans lequel la chaîne cachée est semi-markovienne éventuellement non stationnaire. et le bruit est à mémoire longue et à marginales quelconques. Un tel modèle peut être utilisé à des fins de segmentations non supervisées, les paramètres pouvant être estimés par une méthode ICE adéquate. En guise de perspectives, il serait intéressant d'étudier le cas de lois marginales à queues lourdes telles que les lois stables de Lévy. Ces lois sont couramment utilisées pour modéliser les phénomènes atypiques dans les données financières. Il devrait ainsi être envisageable de segmenter des données financières en considérant simultanément deux propriétés de celles-ci : la dépendance longue et l'apparition de phénomènes atypiques tels que les "cracks'".