

Chapitre 4

Chaînes de Markov triplets et segmentation des images

Chapitre rédigé par Wojciech PIECZYNSKI.

Les diverses modélisations markoviennes s'avèrent d'être d'une remarquable efficacité dans différents problèmes de modélisations et de traitements des phénomènes les plus divers. L'utilisation de ces modèles est en très forte progression en économie, finances, génomique, écologie, communications, traitement des signaux et des images, En particulier, les Modèles de Markov cachés (MMC) jouissent d'une grande notoriété liée à leur efficacité dans les traitements du problème de la segmentation, lequel figure parmi les problèmes les plus importants - et les plus difficiles - en traitements d'images. Dans un tel modèle les données cachées, qui modélisent l'image segmentée recherchée, sont considérées comme une réalisation d'un processus de Markov, qui peut être un champ, un arbre, ou une chaîne. Plus généralement, les MMC sont utilisés pour traiter d'autres problèmes inverses en imagerie comme le débruitage ou la détection des contours - voir chapitre 1. La loi du processus caché X , dite loi « a priori », s'interprète généralement comme un outil de « régularisation » de l'image inobservable recherchée. Par ailleurs, la loi de Y conditionnelle à X est dite loi de l'« attache aux données ». Ainsi les diverses extensions des MMC présentées dans ce chapitre constituent autant d'outils plus généraux applicables dans divers problèmes inverses en imagerie et en vision.

L'objet de ce chapitre, qui de par son approche pragmatique s'adresse également aux utilisateurs ne connaissant pas les MMC, est de présenter diverses généralisations de ces derniers proposées depuis 2000. Le premier niveau de généralisation consiste en l'introduction des Modèles de Markov « Couples » (MMCouples), dans lesquels on considère directement la markovianité du couple (processus caché, processus observé). Les MMCouples sont strictement plus généraux que les MMC, et il s'avère que cette propriété se traduit par une plus grande efficacité dans les traitements. Un deuxième niveau consiste en l'introduction d'un processus auxiliaire et en la considération du processus triplet (processus caché, processus auxiliaire, processus observé). On arrive aux Modèles de Markov triplets (MMT) qui présentent un très grand potentiel de généralité. En particulier, ils permettent de traiter des données non stationnaires, ou encore des données « à mémoire longue », modélisées par la semi-markovianité.

Nous nous limitons à l'exposé des processus markoviens les plus simples que sont les chaînes. Certaines des propriétés discutées s'étendent aisément aux champs et arbres de Markov, les possibilités de généralisation des autres présentent d'intéressantes perspectives de recherche.

Tous les traitements sont discutés dans le contexte non supervisé, présentant un intérêt immédiat pour un utilisateur.

4.1. Introduction

Lorsque l'on s'intéresse à la détermination des quantités numériques inaccessibles directement $x = (x_1 \dots, x_n)$ à partir des quantités numériques observables $y =$

(y_1, \dots, y_n) , la théorie des probabilités fournit un cadre rigoureux permettant d'aboutir à des résultats généralement efficaces et parfois spectaculaires. Le couple (x, y) est considéré comme une réalisation de deux processus aléatoires (X, Y) , avec $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_n)$ et les liens entre x et y sont modélisés par la loi de probabilité $p(x, y)$ du couple (X, Y) . La puissance des méthodes probabilistes est due au fait qu'il est possible de proposer, malgré l'absence des liens déterministes entre x et y , des méthodes de recherche de x optimales « en moyenne », ou encore « à long terme », lorsque l'on traite le problème un « grand » nombre de fois. Un certain nombre de ce type de méthodes peuvent être mises en oeuvre dès que l'on est en mesure de calculer les lois des X_i conditionnelles à $Y = y$, qui seront notées $p(x_i|y)$, ce qui sera parmi les principaux objectifs dans différents modèles étudiés dans la suite. Supposons que chaque X_i prend ses valeurs dans un ensemble fini $\Omega = \{\omega_1, \omega_2\}$, et chaque Y_i prend ses valeurs dans \mathbb{R} . Le couple (X, Y) prend alors ses valeurs dans $\Omega^n \times \mathbb{R}^n$. Supposons que sa loi $p(x, y)$ est donnée par $p(x, y) = p(x)p(y|x)$, avec $p(x)$ une loi sur Ω^n et $p(y|x)$ des densités sur \mathbb{R}^n par rapport à la mesure de Lebesgue. Lorsque n n'est pas trop grand, les lois marginales $p(x_i, y)$, qui donnent les lois recherchées $p(x_i|y)$, sont calculables à partir de $p(x, y)$, qui peut être ainsi utilisée dans sa forme la plus générale. Cependant, ce calcul nécessite 2^{n-1} opérations et devient rapidement impossible lorsque n croît. Il est alors nécessaire de se restreindre à des lois $p(x, y)$ de forme particulière, qui permettent le calcul des $p(x_i|y)$, $1 \leq i \leq n$, pour n grand. Dans le modèle classique, que nous appellerons « chaîne de Markov cachée à bruit indépendant » (CMC-BI), $p(x, y)$ est donnée par

$$p(x, y) = p(x_1)p(x_2|x_1) \dots p(x_n|x_{n-1})p(y_1|x_1) \dots p(y_n|x_n) \quad (4.1)$$

Nous garderons un tel type d'écriture, très concis et pratique, en remarquant qu'il contient bien des écritures classiques en traitement du signal et des images.

Par exemple, supposons que $\Omega = \mathbb{R}$, posons $X_0 = \zeta_0$, et considérons les équations suivantes

$$\begin{cases} Y_i = X_i + \varepsilon_i \\ X_i = AX_{i-1} + \zeta_i \end{cases}, \quad (4.2)$$

avec les variables $\varepsilon_1, \dots, \varepsilon_n, \zeta_0, \zeta_1, \dots, \zeta_n$ centrées et indépendantes entre elles. On a un classique processus AR d'ordre 1 caché par du bruit additif, qui fait partie des modèles (4.1).

L'appellation CMC-BI vient du fait que, d'une part, la chaîne cachée X est une chaîne de Markov et, d'autre part, les variables aléatoires Y_1, \dots, Y_n sont indépendantes conditionnellement à $X = x$. Bien que le modèle CMC-BI passe pour une modélisation relativement raffinée, il est parmi les plus simples parmi les modèles permettant de tenir compte, de manière exploitable, de la dépendance mutuelle des différentes variables composantes des deux processus $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_n)$. En effet, la loi markovienne $p(x) = p(x_1)p(x_2|x_1) \dots p(x_n|x_{n-1})$ de X est, dans un sens, la plus simple des lois permettant d'introduire de la dépendance parmi les

variables X_1, \dots, X_n , et la loi $p(y|x) = p(y_1|x_1) \dots p(y_n|x_n)$ est également très simple. Pourtant, ce type de modèles, très largement utilisé, peut donner des résultats spectaculaires dans différents domaines. Son intérêt réside dans la possibilité de calcul, avec le nombre d'opérations élémentaires proportionnel à n , de diverses quantités d'intérêt, dont $p(x_i|y)$. En posant $\alpha_i(x_i) = p(y_1, \dots, y_{i-1}, y_i, x_i)$, que l'on appelle « probabilité progressive », et $\beta_i(x_i) = p(y_{i+1}, \dots, y_n|x_i, y_i)$, que l'on appelle « probabilité rétrograde », on montre classiquement que $p(x_i, y) = \alpha_i(x_i)\beta_i(x_i)$. Par ailleurs, α_i et β_i sont calculables récursivement par

$$\alpha_1(x_1) = p(x_1, y_1), \text{ et}$$

$$\alpha_{i+1}(x_{i+1}) = \sum_{x_i \in \Omega} \alpha_i(x_i)p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1}) \text{ pour } 2 \leq i \leq n-1, \quad (4.3)$$

$$\beta_n(x_n) = 1, \text{ et}$$

$$\beta_i(x_i) = \sum_{x_{i+1} \in \Omega} \beta_{i+1}(x_{i+1})p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1}) \text{ pour } 1 \leq i \leq n-1, \quad (4.4)$$

et le caractère récursif de ces opérations permet le calcul de $p(x_i, y)$ pour n très grand (plusieurs millions...).

Cette possibilité de calcul, ainsi que le fait que $p(x_i|y)$, qui dépend de toutes les composantes y_1, \dots, y_n de y , utilise toute l'information disponible, sont à l'origine du succès, et souvent de sa grande efficacité, du modèle CMC-BI. Ses applications sont innombrables et des milliers d'articles ont été publiés sur le sujet. Citons quelques publications récentes concernant les biosciences [KOS 01, NIC 02, NUE 07], la climatologie [BEL 00], les communications, [CAP 05], l'écologie, [LEB 06], l'économétrie et les finances [GRE 00, THO 02], le traitement de l'écriture [CHE 94], des signaux musicaux [RAP 99], ou encore des images [GIO 97, MAI 01]. Cependant, ce modèle classique demeure relativement rudimentaire et sa simplicité, notamment en ce qui concerne la loi $p(y|x) = p(y_1|x_1) \dots p(y_n|x_n)$, peut s'avérer insuffisante dans un certain nombre de situations. Afin d'y remédier, il est possible d'utiliser des modèles plus généraux, dits Chaînes de Markov couples (CMCouples [DER 04, PIE 03]). Dans ces derniers on suppose directement la markovianité du processus couple $Z = (X, Y)$, ce qui procure les mêmes capacités de recherche du signal caché tout en autorisant des modélisations plus complètes de $p(y|x)$. Enfin, les CMCouples peuvent être généralisées, à leur tour, aux chaînes de Markov triplets (CMT [PIE 02]), dans lesquelles on introduit une troisième chaîne aléatoire $U = (U_1, \dots, U_n)$, qui peut avoir une signification physique ou pas, et l'on considère la markovianité du triplet $T = (X, U, Y)$. Enfin, notons les « chaînes de Markov cachées » (CMC), que nous définissons comme des CMCouples $Z = (X, Y)$ telles que X est une chaîne de Markov (le processus caché est de Markov, d'où l'appellation CMC).

Finalement, on obtient quatre familles de modèles de degré de généralité strictement croissant : les CMC-BI, les CMC, les CMCouples, et les CMT. Notons qu'en

dehors des CMCouples et CMT les CMC-BI ont été généralisées dans différentes directions ; cependant, à notre connaissance, toutes les extensions gardent la markovianité de X et font partie, de ce fait, des CMC. Les CMCouples forment une famille strictement plus grande que les CMC car il existe des CMCouples telles que X n'est pas markovien [PIE 07b]. De même, il existe des CMT $T = (X, U, Y)$ telles que (X, Y) n'est pas une chaîne de Markov. Les CMT présentent, de par la latitude que nous avons de choisir le troisième processus U , de très grandes possibilités de définition - et d'exploitation - des lois $p(x, y)$, et l'objectif de ce chapitre est de décrire les récentes avancées dans l'exploitation de ces modèles. En particulier, nous verrons qu'il est possible d'exploiter des lois $p(x, y)$ dont l'écriture analytique n'est pas possible, ce qui est probablement un des aspects les plus surprenants des CMT. Les différentes modélisations et traitements présentés dans ce chapitre admettent de multiples applications ; en effet, chacun des modèles CMC, CMCouple, et CMT peut s'appliquer à partir du moment où les CMC-BI classiques - ou l'une de leurs extensions classiques - s'appliquent. L'objet de ce chapitre est de décrire ces différentes extensions, préciser leur intérêt dans la solution des différents problèmes pratiques devant lesquels les modèles classiques montrent leurs limites, et de présenter quelques résultats de leur application à la segmentation statistique d'images.

L'organisation du chapitre est la suivante. Les CMCouples sont introduites dans la section suivante. La section 4.3 est consacrée à la description de l'utilisation, récente, des copules dans le cadre des CMCouples [BRU 05a, BRU 05b]. Cette section, dont l'intérêt est de montrer la richesse des possibilités de modélisation du bruit $p(y|x)$, peut éventuellement être omise lors de la première lecture. La quatrième section est consacrée à l'estimation des paramètres à partir des données incomplètes. Nous y présentons la méthode « Estimation conditionnelle itérative » (ECI), généralement utilisée, de par la facilité de sa mise en oeuvre et son bon comportement, dans les CMCouples et les CMT. Cette section est indépendante de la suite et peut également être évitée par un lecteur principalement intéressé par les modèles CMCouples et CMT. Les chaînes de Markov triplets sont introduites dans la section 4.5 et les sections suivantes présentent leurs différentes applications. Dans la sixième nous décrivons l'utilisation des CMT dans le traitement des processus non stationnaires, alors que la section 7 traite de l'utilisation des CMT dans la modélisation de la semi-markovianité du processus caché. La section 8 présente différentes possibilités de modélisation simultanée des divers aspects évoqués dans les sections précédentes par des CMT faisant appel à des chaînes auxiliaires « multi-variées ». Enfin, la neuvième et dernière section contient les conclusions et des perspectives.

4.2. Chaînes de Markov couples

4.2.1. Modèle général

Considérons $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_n)$ deux processus aléatoires, avec chaque X_i prenant ses valeurs dans un ensemble fini $\Omega = \{\omega_1, \dots, \omega_k\}$ et chaque Y_i prenant ses valeurs dans l'ensemble des nombres réels R . Posons $Z = (Z_1, \dots, Z_n)$, avec $Z_i = (X_i, Y_i)$. Le processus Z sera dit « chaîne de Markov couple » (CMCouple) s'il est de Markov, et il sera dit « chaîne de Markov cachée » (CMC) s'il est une CMCouple telle que X est de Markov. Ainsi Z est une CMCouple si et seulement si sa loi s'écrit (afin de compacter l'écriture, nous utiliserons, à chaque fois que cela sera possible, z_i à la place de (x_i, y_i) et z à la place de (x, y)) :

$$p(z) = p(z_1)p(z_2|z_1) \dots p(z_n|z_{n-1}) \quad (4.5)$$

Sachant que les transitions dans (4.5) peuvent s'écrire

$$p(z_{i+1}|z_i) = p(x_{i+1}|x_i, y_i)p(y_{i+1}|x_{i+1}, x_i, y_i), \quad (4.6)$$

nous remarquons que le modèle classique CMC-BI donné par (4.1) est une CMCouple (4.5) dont les transitions vérifient $p(x_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)$ et $p(y_{i+1}|x_{i+1}, x_i, y_i) = p(y_{i+1}|x_{i+1})$. Par ailleurs, on constate que la première de ces égalités est une condition suffisante pour qu'une CMCouple soit une CMC ; en effet, en intégrant (4.5) par rapport à y_n, \dots, y_1 , on arrive à $p(x) = p(x_1)p(x_2|x_1) \dots p(x_n|x_{n-1})$, ce qui signifie que la chaîne cachée $X = (X_1, \dots, X_n)$ est de Markov. Comme mentionné dans la section précédente, nous nous intéressons aux possibilités de calcul des lois marginales *a posteriori* $p(x_i|y)$. Les probabilités « progressives » α_i et « rétrogrades » β_i de la section précédente sont d'abord étendues aux CMCouples en posant $\alpha_i(x_i) = p(y_1, \dots, y_{i-1}, z_i)$ et $\beta_i(x_i) = p(y_{i+1}, \dots, y_n|z_i)$. De manière analogue aux démonstrations classiques, on montre qu'elles sont calculables par les récursions

$$\alpha_1(x_1) = p(z_1), \text{ et } \alpha_{i+1}(x_{i+1}) = \sum_{x_i \in \Omega} \alpha_i(x_i)p(z_{i+1}|z_i) \text{ pour } 2 \leq i \leq n-1, \quad (4.7)$$

$$\beta_n(x_n) = 1, \text{ et } \beta_i(x_i) = \sum_{x_{i+1} \in \Omega} \beta_{i+1}(x_{i+1})p(z_{i+1}|z_i) \text{ pour } 1 \leq i \leq n-1, \quad (4.8)$$

Comme dans le cas classique des CMC-BI, nous avons alors $p(x_i, y) = \alpha_i(x_i)\beta_i(x_i)$, ce qui donne $p(x_i|y)$. Nous constatons que le modèle CMCouple, dont la plus grande richesse est de nature à mieux modéliser les phénomènes complexes, permet de calculer la loi d'intérêt $p(x_i, y)$ par une démarche de complexité équivalente à celle utilisée dans le cas de CMC-BI.

Nous avons vu ci-dessus comment la plus grande généralité des CMCouples par rapport aux CMC-BI peut être appréhendée, dans le cas général, en comparant les transitions $p(z_{i+1}|z_i)$. Dans le cas des chaînes stationnaires, nous pouvons présenter deux

conditions nécessaires et suffisantes pour qu'une CMCouple soit une CMC. Nous dirons qu'une CMCouple est stationnaire si $p(z_i, z_{i+1})$ ne dépend pas de $1 \leq i \leq n-1$, et si $p(z_i, z_{i+1}) = p(z_{i+1}, z_i)$ pour tout $1 \leq i \leq n-1$. Nous avons le résultat suivant, dont la démonstration peut être consultée dans [PIE 07b] :

Proposition 4.2.1 *Soit $Z = (X, Y)$ une CMCouple stationnaire. Alors les trois conditions*

- (i) X est une chaîne de Markov ;
- (ii) pour tout $2 \leq i \leq n$, $p(y_i|x_i, x_{i-1}) = p(y_i|x_i)$;
- (iii) pour tout $1 \leq i \leq n$, $p(y_i|x) = p(y_i|x_i)$

sont équivalentes.

La condition (ii) permet d'appréhender, dans le contexte des chaînes stationnaires, l'apport des CM Couples par rapports aux CMC localement : dès l'instant où l'on a supposé la markovianité du processus caché X - ce qui est toujours le cas dans les différentes extensions classiques des CMC-BI -, il est impossible de faire apparaître les différences éventuelles des $p(y_i|x_i, x_{i-1})$ lorsque x_{i-1} varie. La condition (iii) permet d'aller encore plus loin dans cette voie : dans une CMC la loi $p(y_i|x)$ ne peut dépendre d'aucun x_j différent de x_i . Ainsi la modélisation du bruit est bien plus complète dans les CM Couples où les lois $p(y_i|x)$ dépendent de toutes les composantes x_1, \dots, x_n de x . Il a également été constaté dans les différentes simulations que cette plus grande richesse du modèle peut se traduire par des gains appréciables de la qualité des segmentations non supervisées des données [DER 04].

Remarque 4.2.1

Lorsque l'on examine les divers traitements utilisant les CMC classiques on réalise que la markovianité de la loi de X n'est jamais utilisée et ne sert qu'indirectement, permettant de montrer la markovianité de la loi de X a posteriori $p(x|y)$, cette dernière étant alors indispensable. Par ailleurs, dans le cadre de la Proposition 4.2.1, supposer la markovianité de X revient à imposer des restrictions aux lois conditionnelles $p(y|x)$. En d'autres termes, en considérant une CM Couple qui n'est pas une CMC, on perd la markovianité X , qui n'intervient pas dans les traitements, et on gagne, de manière non négligeable, en richesse dans la modélisation du bruit. L'apport principal des CM Couples est ainsi de s'affranchir de l'hypothèse de la markovianité de X , qui apparaît comme une contrainte. Notons que cette hypothèse, bien que non indispensable, apparaît dans tous les modèles utilisant la markovianité dans des problèmes de restauration, ou de segmentation, des phénomènes cachés. En effet, qu'il

s'agisse des chaînes, des champs, des arbres, ou des modèles graphiques généraux, la nature markovienne du processus caché apparaît dans l'appellation même « modèles de Markov cachés » (« hidden Markov models » (HMM) en anglais), et donne ainsi l'impression d'être incontournable.

Comme dans le cas des CMC-BI on montre que dans une CMCouple la loi $p(x|y)$ est markovienne avec les transitions données par

$$p(x_{i+1}|x_i, y) = \frac{p(z_{i+1}|z_i)\beta_{i+1}(x_{i+1})}{\beta_i(x_i)} \quad (4.9)$$

et que les lois $p(x_i, x_{i+1}|y)$, qui seront utilisées dans la section 4.4, sont données par

$$p(x_i, x_{i+1}|y) = \alpha_i(x_i)p(z_{i+1}|z_i)\beta_{i+1}(x_{i+1}) \quad (4.10)$$

Par ailleurs, notons que la loi d'une CMCouple stationnaire est définie par $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$. Les transitions peuvent alors s'écrire

$$p(z_{i+1}|z_i) = \frac{p(x_i, x_{i+1})p(y_i, y_{i+1}|x_i, x_{i+1})}{\sum_{\omega \in \Omega} p(x_i, x_{i+1} = \omega)p(y_i|x_i, x_{i+1} = \omega)} \quad (4.11)$$

Finalement, remarquons que la loi $p(z)$ d'une CMCouple Z peut toujours s'écrire

$$\begin{aligned} p(z) &= \frac{p(z_1, z_2) \dots p(z_{n-1}, z_n)}{p(z_2) \dots p(z_n)} \\ &= \frac{p(x_1, x_2) \dots p(x_{n-1}, x_n)}{\underbrace{p(x_2) \dots p(x_{n-1})}_{a(x)}} \\ &\times \frac{p(y_1, y_2|x_1, x_2) \dots p(y_{n-1}, y_n|x_{n-1}, x_n)}{\underbrace{p(y_2|x_2) \dots p(y_{n-1}|x_{n-1})}_{b(x, y)}} \end{aligned} \quad (4.12)$$

Ainsi une CMCouple est une CMC si et seulement si $p(x) = a(x)$ et $p(y|x) = b(x, y)$.

4.2.2. CM Couples gaussiens stationnaires

De manière analogue à ce qui a lieu dans la famille classique des CMC-BI, les modèles « gaussiens » sont parmi les plus simples des CM Couples. Considérons une CM-Couple stationnaire, dont la loi est définie par $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$

telle que les lois conditionnelles $p(y_1, y_2|x_1, x_2)$, égales aux lois conditionnelles $p(y_i, y_{i+1}|x_i, x_{i+1})$ pour tout $i = 1, \dots, n-1$, sont gaussiennes. Si Z n'est pas une CMC on note, en vertu de (4.11) et de la proposition 4.2.1, que les transitions $p(z_{i+1}|z_i)$ ne sont pas des lois gaussiennes (le dénominateur est un mélange de gaussiennes). En conséquence, $p(y|x)$ n'est pas une loi gaussienne. De plus, les lois marginales $p(y_i|x)$, qui ne sont pas gaussiennes, dépendent de tous les x_1, \dots, x_n . Pour montrer cette dernière propriété, on peut utiliser le fait que dans une CMCouple les processus X et Y jouent des rôles symétriques. En inversant X et Y , on peut ainsi introduire $\alpha_i^*(y_i) = p(x_1, \dots, x_{i-1}, z_i)$, $\beta_i^*(y_i) = p(x_{i+1}, \dots, x_n|z_i)$, et montrer que ces quantités vérifient les récursions

$$\alpha_1^*(y_1) = p(z_1), \text{ et}$$

$$\alpha_{i+1}^*(y_{i+1}) = \int_R \alpha_i^*(y_i) p(z_{i+1}|z_i) dy_i \text{ pour } 2 \leq i \leq n-1; \quad (4.13)$$

$$\beta_n^*(y_n) = 1, \text{ et}$$

$$\beta_i^*(y_i) = \int_R \beta_{i+1}^*(y_{i+1}) p(z_{i+1}|z_i) dy_{i+1} \text{ pour } 1 \leq i \leq n-1. \quad (4.14)$$

Cependant, contrairement aux récursions (4.7), (4.8) qui sont calculables car définies par des sommes sur un ensemble de petites taille, les récursions (4.13), (4.14), définies par des intégrales, ne sont pas, sauf pour α_2^* et β_{n-1}^* , calculables analytiquement.

Ainsi dans les CMCouples gaussiens et stationnaires, qui ne sont pas des CMC, les lois du bruit $p(y|x)$ ne sont pas gaussiennes et sont impossibles à calculer. Cependant, ces "difficultés" expriment finalement une richesse et ne nuisent aucunement à la recherche des lois $p(x_i|y)$.

Enfin, on peut également noter que les lois $p(y_i|x_i, x_{i-1})$ et $p(y_i|x_i, x_{i+1})$ sont des gaussiennes, les lois $p(y_i|x_i)$ sont des mélanges de k gaussiennes

$$p(y_i|x_i) = \sum_{x_{i-1} \in \Omega} p(x_{i-1}|x_i) p(y_i|x_i, x_{i-1}) = \sum_{x_{i+1} \in \Omega} p(x_{i+1}|x_i) p(y_i|x_i, x_{i+1}), \text{ et les}$$

densités des lois $p(y_i|x_{i-1}, x_i, x_{i+1})$ sont de la forme

$$p(y_i|x_{i-1}, x_i, x_{i+1}) = \frac{p(y_i|x_{i-1}, x_i) p(y_i|x_i, x_{i+1})}{p(y_i|x_i)}, \text{ qui est un produit de deux densités gaussiennes divisé par un mélange de } k \text{ densités gaussiennes.}$$

4.3. Copules dans les CMCouples

L'intérêt des copules réside dans leur pouvoir de modéliser, de manière particulièrement élégante, les dépendances entre les variables aléatoires. Plus précisément, la dépendance est donnée par une « copule » indépendamment de la forme des lois marginales. La théorie des copules est relativement ancienne [NEL 98]; cependant, leur introduction dans une chaîne de Markov couple, que nous présentons brièvement dans cette section, est relativement récente [BRU 05a, BRU 05b].

4.3.1. Copules

Nous appellerons « copule » toute fonction de répartition C de la loi d'un couple de variables aléatoires prenant ses valeurs dans $[0, 1]^2$ dont les lois marginales sont des lois uniformes sur $[0, 1]$. Soit $h(y_1, y_2)$ une densité de probabilité sur R^2 , que nous supposons continue, H sa fonction de répartition, $h_1(y_1)$, $h_2(y_2)$ les densités des lois marginales, et H_1 , H_2 les fonctions de répartition associées. Selon le théorème de Sklar, il existe alors une copule unique telle que (voir [NEL 98])

$$H(y_1, y_2) = C(H_1(y_1), H_2(y_2)) \quad (4.15)$$

En dérivant (4.15) par rapport aux y_1 , y_2 et en posant $c(u, v) = \frac{\partial \partial C(u, v)}{\partial u \partial v}$, nous avons

$$h(y_1, y_2) = h_1(y_1)h_2(y_2)c(H_1(y_1), H_2(y_2)) \quad (4.16)$$

Réciproquement, en considérant des fonctions de répartition H_1 , H_2 sur R et une copule C , (4.15) définit une fonction de répartition H sur R^2 .

Finalement, nous pouvons dire que se donner une fdr H revient à se donner le triplet (H_1, H_2, C) , où H_1 et H_2 sont les lois marginales de H , et C est la copule associée à H par (4.15). Il en résulte qu'il est possible de construire des lois sur R^2 en considérant diverses marginales - et diverses copules - de manière indépendante. En particulier, il est possible de fabriquer, à partir de (Y_1, Y_2) avec H donnée, des variables aléatoires réelles corrélées de marginales quelconques. En effet, considérons les marginales H_1 , H_2 associées à H , et H'_1 , H'_2 les nouvelles marginales souhaitées. En posant $Y'_1 = (H'_1)^{-1} \circ H_1(Y_1)$ et $Y'_2 = (H'_2)^{-1} \circ H_2(Y_2)$, les variables (Y'_1, Y'_2) sont corrélées et admettent H'_1 , H'_2 pour les fdr de leurs lois. Enfin, notons que la copule C associée à H par (4.15) peut également être vue comme la fdr du couple $(H_1(Y_1), H_2(Y_2))$, dont les lois marginales sont ainsi bien les lois uniformes sur $[0, 1]$.

Exemple 4.3.1

Soit $h(y_1, y_2)$ la densité gaussienne sur R^2 définie par les deux moyennes nulles, les deux variances égales à 1, et par le coefficient de corrélation ρ . Selon (4.15), la copule gaussienne associée est donnée par

$$C(u, v) = H(H_1^{-1}(u), H_2^{-1}(v)), \quad (4.17)$$

ou par sa dérivée

$$c(u, v) = \frac{h(H_1^{-1}(u), H_2^{-1}(v))}{h_1(H_1^{-1}(u))h_2(H_2^{-1}(v))} \quad (4.18)$$

Une copule gaussienne sur R^2 est ainsi définie par un unique paramètre réel ρ . Comme précisé ci-dessus, cette copule peut être utilisée pour définir une loi d'un

couple (Y_1, Y_2) de variables dépendantes, avec les marginales imposées. Considérons deux variables aléatoires réelles Y_1, Y_2 , les fdr marginales F_1, F_2 et f_1, f_2 les densités correspondantes. La densité suivante de la loi de (Y_1, Y_2)

$$f(y_1, y_2) = f_1(y_1)f_2(y_2) \frac{h(H_1^{-1}(F_1(y_1)), H_2^{-1}(F_2(y_2)))}{h_1(H_1^{-1}(F_1(y_1))) h_2(H_2^{-1}(F_2(y_2)))} \quad (4.19)$$

définit alors la loi d'un couple de variables aléatoires (Y_1, Y_2) dépendantes et telles que les lois marginales admettent pour densités f_1 et f_2 . Il est intéressant de noter que lorsque les densités f_1, f_2 sont gaussiennes, alors $f(y_1, y_2)$ donnée par (4.19) est gaussienne. Bien entendu, on peut également considérer f_1, f_2 gaussiennes et une copule différente de la copule gaussienne [BRU 05a].

4.3.2. Copules dans CMCouples

Soit une CMCouple stationnaire $Z = (X, Y)$ dont la loi est définie par $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$. Pour l'ensemble des classes $\Omega = \{\omega_1, \dots, \omega_k\}$, nous avons donc, dans le cas général, $k(k-1)$ densités $p_{ij}(y_1, y_2) = p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$. Chacune de ces densités définit une copule et deux lois marginales sur R^2 . Notons que dans le cas particulier des CMC le nombre des marginales est de k .

Remarquons que la variation de la nature de la marginale $p(y_1|x_1 = \omega_i)$ avec la classe ω_i peut se produire dans des images réelles, comme le montrent différentes études, en particulier celles présentées dans [GIO 97, MAI 01, DEL 02].

4.4. Estimation des paramètres

Le problème de l'estimation des paramètres dans les chaînes de Markov triplet est identique à celui de l'estimation des paramètres dans les CMCouples ; il est ainsi possible de le traiter dès à présent, dans un contexte plus simple. Lorsque l'on aborde ce problème dans le cas des CMC classiques, la méthode la plus souvent utilisée est la méthode dite de « Espérance-Maximisation » (EM [MCL 97]). Cette méthode est fondée sur la maximisation, par une procédure itérative, de la vraisemblance de la chaîne observée et peut donner, lorsque l'initialisation n'est pas trop loin des vrais paramètres, des résultats remarquables. Elle jouit par ailleurs des résultats théoriques partiels concernant son comportement asymptotique ; en particulier, il est possible de montrer qu'elle produit une suite croissante des vraisemblances au cours des itérations. Cependant, en dehors des cas simples, les calculs explicites des différentes valeurs intervenant dans les itérations sont en général impossibles et il est fréquemment fait appel à des approximations, qui sont de nature à annihiler les bonnes propriétés générales du maximum de vraisemblance. Ces difficultés n'étant pas moindres dans les CMCouples, nous nous limitons à la présentation d'une méthode différente, dite

« estimation conditionnelle itérative » (ECI [PIE 92, PIE 07a]). Cette dernière est fondée sur un principe différent, qui utilise la notion de l'espérance conditionnelle et ne fait pas appel à la vraisemblance marginale du processus observé. Notons cependant que l'estimateur du maximum de vraisemblance à partir des données complètes peut intervenir dans l'ECI, garantissant ainsi une « optimalité » analogue à celle offerte par la méthode EM. Il résulte du principe même de l'ECI que sa mise en place est, d'une part, plus simple et, d'autre part, praticable dans un contexte plus général que celle de la méthode EM. Elle a été appliquée dans différentes modélisations complexes comme champs de Markov triplets [BEN 05a, BEN 07a, BEN 05b, BEN 07b], champs de Markov cachés évidentiels [BEN 01], ou chaînes markoviennes ou semi-markoviennes cachées non stationnaires [LAP 06, LAP 07]. Notons que les premiers résultats théoriques concernant le comportement asymptotique de l'ECI, valables dans le cas des observations indépendantes, sont récents [PIE 07a] et les modèles markoviens présentent un très intéressant champ d'investigations théoriques possibles. Enfin, dans certaines paramétrisations des modèles exponentiels ECI et EM peuvent être équivalentes [DEL 97].

Nous présentons deux variantes de l'ECI bien adaptées aux CMCouples et de généralité croissante. La première variante est classique et concerne les modèles paramétriques, la deuxième concerne les mélanges généralisés et, se plaçant dans le contexte des CMCouples avec copules, contient des aspects originaux.

4.4.1. ECI dans un modèle paramétrique

De manière générale, considérons deux processus stochastiques (X, Y) dont la loi dépend d'un paramètre $\theta = (\theta_1, \dots, \theta_s) \in R^s$. Supposons que le problème est celui de l'estimation de θ à partir du processus Y , le seul observable. Il est alors possible d'utiliser l'ECI sous deux conditions suivantes :

- (i) il existe un estimateur $\hat{\theta}(X, Y)$ de θ à partir des données complètes (X, Y) ;
- (ii) il est possible, pour tout θ , de simuler X selon $p(x|y, \theta)$.

Notons le caractère très peu restrictif de ces conditions. En effet, (i) n'en est pas véritablement une car s'il est impossible d'exhiber un estimateur à partir des données complètes, le problème de l'estimation a toutes les chances d'être insoluble à partir des données incomplètes. La condition (ii) est pratiquement toujours vérifiée dans les problèmes de recherche de X à partir de Y par des méthodes bayésiennes ; en effet, ces dernières requièrent généralement la markovianité de $p(x|y, \theta)$.

Le déroulement de l'ECI est le suivant :

- (i) on considère θ^0 valeur initiale de θ ;

(ii) on pose $\theta_r^{q+1} = E[\hat{\theta}_r(X, Y)|Y = y, \theta^q]$ pour les composantes θ_r de θ pour lesquelles cette espérance est calculable ;

(iii) pour les composantes θ_r de θ pour lesquelles l'espérance ci-dessus n'est pas calculable, on simule m valeurs $x^{q,1}, \dots, x^{q,m}$ de X selon $p(x|y, \theta^q)$, et l'on pose $\theta_r^{q+1} = \frac{\hat{\theta}_r(x^{q,1}, y) + \dots + \hat{\theta}_r(x^{q,m}, y)}{m}$.

Nous dirons qu'une CM Couple stationnaire $(X, Y) = (X_1, \dots, X_n, Y_1, \dots, Y_n)$ est « paramétrique » lorsque la loi $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$ dépend d'un ensemble des paramètres θ . Divisons les composantes du vecteur θ en deux sous-vecteurs. Le premier définit la loi $p(x_1, x_2)$ sur Ω^2 , et contient donc k^2 composantes. Le deuxième définit toutes les lois $p(y_1, y_2|x_1, x_2)$ et contient donc $(k-1)k$ groupes des composantes, chaque groupe correspondant à un $(x_1, x_2) = (\omega_i, \omega_j)$. Si toutes ces lois sont gaussiennes, chaque groupe contient cinq composantes réelles : deux moyennes, deux variances, et la covariance. Dans le cas gaussien le vecteur θ admet ainsi $k^2 + 5k(k-1)$ composantes.

Afin de simplifier les notations, posons $p(i, j) = p(x_1 = \omega_i, x_2 = \omega_j)$ et θ_{ij} le groupe des paramètres définissant $p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$. Afin d'appliquer l'ECI, on doit considérer un estimateur $\hat{\theta}(X, Y)$ donné à partir des données complètes (X, Y) . Les composantes $p(i, j)$ peuvent être classiquement estimées à partir de X par les fréquences

$$\hat{p}(i, j) = \frac{1_{[x_1=\omega_i, x_2=\omega_j]} + \dots + 1_{[x_{2n-1}=\omega_i, x_{2n}=\omega_j]}}{n}, \quad (4.20)$$

où 1_A est la fonction indicatrice de l'ensemble A .

L'estimation des groupes θ_{ij} à partir de (X, Y) est possible dès que chaque groupe θ_{ij} peut être estimé à partir d'un échantillon produit par $p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$, ce que nous supposons. En effet, sachant $X = x = (x_1, \dots, x_{2n})$, on définit un sous-échantillon y^{ij} de $y = (y_1, \dots, y_{2n})$ par $(y_{2l}, y_{2l+1}) \in y^{ij}$ si et seulement si $(x_{2l} = \omega_i, x_{2l+1} = \omega_j)$. On dispose ainsi des échantillons des couples (Y_{2l}, Y_{2l+1}) des lois définies par les θ_{ij} , ce qui permet de considérer un estimateur $\hat{\theta}_{ij}$. Nous écrirons $\hat{\theta}_{ij}(x, y)$, sachant que dans cette écriture $\hat{\theta}_{ij}(x, y)$ ne dépend que de y^{ij} , ce dernier étant défini à partir de $x = (x_1, \dots, x_{2n})$.

A titre d'exemple, dans le cas gaussien les deux moyennes sont estimées par les moyennes empiriques, et la matrice de covariance par la matrice de covariance empirique.

L'estimateur $\hat{\theta}(X, Y)$ étant défini par (4.20) et les estimateurs $\hat{\theta}_{ij}(X, Y)$, l'application de l'ECI se déroule de la manière suivante. Le calcul de l'espérance $E[\hat{\theta}_r(X, Y)|Y = y, \theta^q]$ donnant θ_r^{q+1} est possible pour les composantes du premier

sous-vecteur de θ donné par les $p(i, j)$. En effet, sachant que l'espérance d'une fonction indicatrice est la probabilité de l'ensemble correspondant, en prenant l'espérance conditionnelle de (4.20) on obtient

$$p^{q+1}(i, j) = \frac{p(x_1 = \omega_i, x_2 = \omega_j | y, \theta^q) + \cdots + p(x_{2n-1} = \omega_i, x_{2n} = \omega_j | y, \theta^q)}{n} \quad (4.21)$$

avec les probabilités $p(x_{2n-1} = \omega_i, x_{2n} = \omega_j | y, \theta^q)$ calculables par (4.10), section 4.2. Concernant les groupes des paramètres θ_{ij} , l'espérance conditionnelle n'est pas calculable et l'on utilise le point (iii) de l'ECI, utilisant les simulations. Notons que la loi de X conditionnelle à $Y = y$ étant une chaîne de Markov avec les transitions calculables par (4.9), section 4.2, les simulations ne posent pas de problème particulier. Dans la pratique, on effectue souvent un seul tirage x^q et on pose $\theta_{ij}^{q+1} = \hat{\theta}_{ij}(x^q, y)$.

4.4.2. ECI dans un modèle de mélange généralisé avec copules

Considérons une CMCouple stationnaire de loi définie par $p(z_1, z_2) = p(x_1, x_2)p(y_1, y_2 | x_1, x_2)$. Supposons que chacune des $(k-1)k$ densités marginales $p_{ij}(y_1) = p(y_1 | x_1 = \omega_i, x_2 = \omega_j)$ est d'une forme appartenant à un ensemble des formes admissibles, qui sera noté Φ^{ij} . A titre d'exemple on peut imaginer, pour deux classes $\Omega = \{\omega_1, \omega_2\}$, que Φ^{11} contient deux éléments : « lois gaussiennes » et « lois gamma », $\Phi^{12} = \Phi^{21}$ contient trois éléments : « lois gaussiennes », « lois gamma », et « lois beta », et Φ^{22} contient deux éléments : « lois gaussiennes » et « lois K ». Le problème est alors double : identifier la forme de chacune de ces densités et estimer les paramètres fixant la densité dans la famille identifiée. Ce type de problème, traité dans [GIO 97] dans le cas des CMC-BI, est dit « estimation de mélange généralisé ». Nous décrivons ci-après une démarche de ce type étendue aux CMCouples. D'une part, elle généralise la démarche présentée dans [GIO 97] et, d'autre part, elle généralise, par l'utilisation des copules, la démarche de la sous-section précédente.

Nous avons vu à la section 4.3 que chacune des $(k-1)k$ lois $p_{ij}(y_1, y_2) = p(y_1, y_2 | x_1 = \omega_i, x_2 = \omega_j)$ est définie par $(k-1)k$ densités marginales et $(k-1)k$ copules. Supposons que pour chaque (i, j) la forme de la copule associée à $p_{ij}(y_1, y_2)$ appartient à l'ensemble \mathcal{X}^{ij} des formes admissible. Pour chaque (i, j) nous devons donc :

(i) choisir dans Φ^{ij} la forme de la marginale $p_{ij}(y_1)$ et estimer les paramètres correspondants ;

(ii) choisir dans \mathcal{X}^{ij} la forme de la copule donnée par $p_{ij}(y_1, y_2)$ et estimer les paramètres correspondants.

Considérons l'estimation à partir des données complètes et supposons $X = x = (x_1, \dots, x_{2n})$ donné. Comme précédemment, on définit, pour chaque (i, j) , un sous-échantillon y^{ij} de $y = (y_1, \dots, y_{2n})$ par $(y_{2l}, y_{2l+1}) \in y^{ij}$ si et seulement si $(x_{2l} = \omega_i, x_{2l+1} = \omega_j)$. On dispose ainsi, pour chaque (i, j) , d'un échantillon des couples (Y_{2l}, Y_{2l+1}) dont les lois marginales appartiennent à une des familles $F_1^{ij}, \dots, F_{r(i,j)}^{ij}$ dans $\Phi^{ij} = \{F_1^{ij}, \dots, F_{r(i,j)}^{ij}\}$. En supposant que ces familles sont paramétrées respectivement par $\theta_1^{ij}, \dots, \theta_{r(i,j)}^{ij}$ et qu'il existe des estimateurs adéquats, notons $\hat{\theta}_1^{ij}, \dots, \hat{\theta}_{r(i,j)}^{ij}$ les valeurs estimées. Ces valeurs donnent $r(i, j)$ densités « candidates » à la bonne loi, et l'on choisit celle parmi ces candidates qui optimise un certain critère qui peut être, comme dans [GIO 97], la minimisation de la distance de Kolmogorov à l'histogramme. Notons H_1^{ij}, H_2^{ij} les fdr des lois marginales ainsi estimées ; alors la loi du couple $(H_1^{ij}(Y_1), H_2^{ij}(Y_2))$ est la copule recherchée. En supposant qu'elle appartient à une des familles $C_1^{ij}, \dots, C_{m(i,j)}^{ij}$ dans $\mathcal{X}^{ij} = \{C_1^{ij}, \dots, C_{m(i,j)}^{ij}\}$, on peut alors estimer $m(i, j)$ « candidates » et choisir, de manière analogue à celle utilisée pour rechercher les lois marginales, celle qui minimise un certain critère.

L'ensemble de cette procédure est ensuite intégrée dans l'ECI où, à chaque itération, x est remplacé par x^q simulé.

Remarque 4.4.1

Nous avons noté dans la Remarque 4.2.1 que l'habitude prise de considérer systématiquement le processus caché comme markovien contribuait probablement à la persistance de l'utilisation des « modèles de Markov cachés » (MMC), avec modélisations simples des bruits. Notons également que les difficultés à étudier, au plan théorique, l'estimation des paramètres, constituent probablement une deuxième raison de la persistance de ces modèles. A titre d'exemple, les CM Couples sont brièvement mentionnées dans [CAP 05] mais les auteurs renoncent à leur étude car les résultats théoriques concernant l'estimation par le maximum de vraisemblance qu'ils considèrent par la suite sont intimement liés à la markovianité du processus caché (page 5).

4.5. Chaînes de Markov triplets

Le problème demeure identique à celui considéré dans les sections précédentes : retrouver la réalisation cachée de la chaîne X à partir de la réalisation observée de la chaîne Y . Considérer une chaîne de Markov triplet consiste en l'introduction d'une troisième chaîne aléatoire $U = (U_1, \dots, U_n)$, qui sera ici supposée à valeurs finies (chaque U_n prend ses valeurs dans un ensemble fini $\Lambda = \{\lambda_1, \dots, \lambda_m\}$), et supposer que le triplet $T = (X, U, Y)$ est une chaîne de Markov. Notons que U peut avoir une signification physique ou non ; notons également que le modèle CMT est strictement plus général que le modèle CM Couple ; en effet, dans une CMT $T = (X, U, Y)$

la chaîne $Z = (X, Y)$ n'est pas nécessairement de Markov. Cependant, en posant $V = (X, U)$, la chaîne $T = (V, Y)$ est une CM couple et il est possible d'utiliser tous les résultats des sections précédentes. En particulier, la méthode d'estimation des paramètres (on dit aussi « d'apprentissage ») ECI est applicable, y compris dans le contexte des copules.

En restauration bayésienne MPM, l'intérêt des CMT réside dans le fait que lorsque le cardinal de Λ n'est pas trop élevé les marginales *a posteriori* $p(x_i|y)$ sont calculables. En effet, $p(v_i|y) = p(x_i, u_i|y)$ le sont car $T = (V, Y)$ est une CM couple, et alors les lois $p(x_i|y)$ sont données simplement par

$$p(x_i|y) = \sum_{u_i \in \Lambda} p(x_i, u_i|y) \quad (4.22)$$

Nous obtenons alors un modèle très riche, pouvant donner lieu à un grand nombre de modèles particuliers. Nous étudions dans les sections suivantes différentes situations où les différentes CMT sont appliquées, éventuellement de manière non supervisée, à la restauration des données cachées. Nous nous limitons dans cette section à l'énoncé d'un résultat généralisant les différents cas particuliers, très pratique pour mettre en lumière les différentes spécificités des CMT particulières. Ensuite, nous présentons deux exemples de CMT, illustrant leurs richesses et leurs différences avec les modèles classiques.

Le résultat suivant, qui est une généralisation de la Proposition 4.2.1, permet de construire différentes CMT particulières.

Soit $W = (G, H) = (G_1, H_1, \dots, G_n, H_n)$ une chaîne aléatoire, avec chaque $W_i = (G_i, H_i)$ à valeurs dans un ensemble produit $\Gamma \times \mathcal{H}$. Soit γ une mesure σ -additive sur Γ , et soit η une mesure σ -additive sur \mathcal{H} . Supposons W de Markov et notons avec la même lettre p les différentes densités par rapport aux différentes mesures liées aux mesures γ et η . En pratique, chacune des mesures γ et η est soit la mesure de comptage, soit la mesure de Lebesgue ; cependant, des mesures mixtes, faisant intervenir des masses de Dirac et la mesure de Lebesgue, peuvent également être envisagées, comme dans [SAL 07].

Nous avons le résultat suivant, dont la démonstration suit le schéma général de la démonstration de la Proposition 4.2.1.

Proposition 4.5.1

Soit $W = (G, H)$ une chaîne de Markov vérifiant

(a) $p(w_i, w_{i+1})$ ne dépend pas de $1 \leq i \leq n - 1$;

(b) $p(w_i = a, w_{i+1} = b) = p(w_i = b, w_{i+1} = a)$ pour chaque $1 \leq i \leq n - 1$, a , et b .

Alors les trois conditions suivantes

(i) G est une chaîne de Markov ;

(ii) pour tout $2 \leq i \leq n$, $p(h_i|g_i, g_{i-1}) = p(h_i|g_i)$;

(iii) pour tout $1 \leq i \leq n$, $p(h_i|g) = p(h_i|g_i)$,

sont équivalentes.

Nous présentons ci-dessous deux exemples. Le premier illustre le fait que certaines TMC peuvent être de complexité comparable à celle du modèle classique CMC-BI, et le deuxième généralise les modèles classiques de mélange.

Exemple 4.5.1

Soit $T = (X, U, Y)$ une CMT dont la loi est donnée par $p(x_1, u_1, y_1)$ et les transitions $p(t_{i+1}|t_i)$ de la forme

$$p(t_{i+1}|t_i) = p(u_{i+1}|u_i)p(x_{i+1}|u_{i+1})p(y_{i+1}|u_{i+1}) \quad (4.23)$$

Une telle CMT est relativement simple tout en étant très différente des CMC-BI classiques. En particulier, on peut montrer les propriétés suivantes : U est une chaîne de Markov, aucune des chaînes X , Y , (X, Y) n'est de Markov, les chaînes X , Y sont indépendantes conditionnellement à U .

Cet exemple montre qu'il est possible de construire très simplement des modèles exploitables aussi simples que les CMC-BI et très différents de ces derniers. En particulier, aucune des trois propriétés suivantes définissant les CMC-BI : (i) X est de Markov ; (ii) les variables aléatoires Y_1, \dots, Y_n sont indépendantes conditionnellement à X ; (iii) $p(y_i|x) = p(y_i|x_i)$ pour tout $i = 1, \dots, n$, n'est nécessairement vérifiée.

Exemple 4.5.2

Considérons le cas classique CMC-BI donnée par (4.1) et supposons que les densités $p(y_i|x_i)$ sont des mélanges $p(y_i|x_i = \omega_j) = \sum_{l=1}^m \alpha_{lj} f_{lj}(y_i)$. De tels modèles sont dits « modèles de mélange » et peuvent être utilisés, notamment, lorsque la forme des lois $p(y_i|x_i)$ n'est pas connue et ces dernières sont approchées par des mélanges, par exemple des lois gaussiennes [DIZ 07]. Un tel modèle est une CMT dont la loi est donnée par $p(x_1, u_1, y_1)$ et les transitions $p(t_{i+1}|t_i)$ de la forme

$$p(t_{i+1}|t_i) = p(x_{i+1}|x_i)p(u_{i+1}|x_{i+1})p(y_{i+1}|u_{i+1}, x_{i+1}) \quad (4.24)$$

avec $p(u_i = \lambda_l | x_i = \omega_j) = \alpha_{lj}$ et $p(y_i | u_i = \lambda_l, x_i = \omega_j) = f_{lj}(y_i)$. Notons que dans cet exemple nous avons

$$p(y_i | x) = p(y_i | x_i) \quad (4.25)$$

Considérons alors une CMT stationnaire et plus générale que celle donnée par (4.24), dont la loi s'écrit

$$p(t_{i+1} | t_i) = p(x_{i+1}, u_{i+1} | x_i, u_i) p(y_{i+1} | u_{i+1}, x_{i+1}) \quad (4.26)$$

et qui est telle que $p(x_{i+1} | u_{i+1}, u_i) \neq p(x_{i+1} | u_{i+1})$ et $p(u_{i+1} | x_{i+1}, x_i) \neq p(u_{i+1} | x_{i+1})$. La chaîne $V = (X, U)$ est alors de Markov et $p(y_i | u, x) = p(y_i | u_i, x_i)$: ainsi (V, Y) est la très classique CMC-BI. Cependant, en appliquant la Proposition 5.1 à $G = X$ et $H = U$, on note que la condition (ii) n'est pas vérifiée et donc X n'est pas une chaîne de Markov. On peut montrer de manière analogue que (X, Y) n'est pas de Markov non plus et, de plus, les variables aléatoires Y_1, \dots, Y_n ne sont pas nécessairement indépendantes conditionnellement à X . Enfin, montrons que $p(y_i | x)$ dépend de tous les x_1, \dots, x_n . Nous avons $p(y_i | x) = \sum_{u_i \in \Lambda} p(y_i, u_i | x) = \sum_{u_i \in \Lambda} p(y_i | x, u_i) p(u_i | x) = \sum_{u_i \in \Lambda} p(y_i | x_i, u_i) p(u_i | x)$, ce qui montre que $p(y_i | x)$ est un mélange dont les coefficients $p(u_i | x)$ dépendent de tous les x_1, \dots, x_n . On peut également noter que les lois $p(y_i | x_i)$ sont des mélanges extrêmement riches : pour k valeurs possibles pour chaque x_n et m valeurs possibles pour chaque u_n , le nombre des composantes dans le mélange $p(y_i | x_i)$ est de mk^{n-1} . En effet, en posant $x^* = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, nous avons $p(y_i | x_i) = \sum_{(x^*, u_i)} p(x^*, u_i, y_i | x_i) = \sum_{(x^*, u_i)} p(y_i | x_i, x^*, u_i) p(x^*, u_i | x_i) = \sum_{(x^*, u_i)} p(y_i | x_i, u_i) p(u_i | x) p(x^* | x_i)$ et mk^{n-1} est le nombre des (x^*, u_i) possibles.

Ainsi, ce modèle généralise de manière conséquente, qui pourrait s'avérer intéressante dans les applications, les modèles de mélange classiques, dans la mesure que $p(y_i | x_i)$ est un mélange de mk^{n-1} composantes au lieu de m .

4.6. Chaînes de Markov triplets et non stationnarité

Les CMT $T = (X, U, Y)$ peuvent être utilisées pour traiter du problème de la non stationnarité de la chaîne $Z = (X, Y)$. La troisième chaîne U admet alors une signification physique et chaque état dans $\Lambda = \{\lambda_1, \dots, \lambda_m\}$ modélise une stationnarité particulière. Plus précisément, nous dirons qu'une chaîne aléatoire $Z = (X, Y)$ est une « CMT m stationnaire » s'il existe une chaîne aléatoire U à valeurs dans un ensemble de m éléments telle que $T = (X, U, Y)$ soit une CMT stationnaire. En effet, U modélise alors le fait qu'il y ait m stationnarités différentes, présentes à

des instants aléatoires, dans la chaîne $Z = (X, Y)$. Il est alors possible de s'intéresser aux différents types de restauration. On peut, classiquement, rechercher $X = x$, comme décrit dans les sections précédentes. On peut également s'intéresser au couple $(X, U) = (x, u)$, ou même à la seule chaîne $U = u$. La souplesse des démarches bayésiennes permet alors, via différentes fonctions de perte, de proposer différentes méthodes de segmentation, éventuellement non supervisée, fondées sur la markovianité de la loi conditionnelle $p(x, u|y)$.

L'intérêt de l'utilisation des CMT dans le cadre des données comportant un nombre fini de stationnarités a été validé dans le contexte de la segmentation d'images par différentes expérimentations présentées dans [LAN 06, LAN 04, LAN 05], dont nous présentons un exemple ci-dessous.

Notons que l'introduction d'une troisième chaîne modélisant les différentes stationnarités n'est pas originale car différents modèles « à sauts », où le processus des sauts remplit ce rôle, ont été proposés et étudiés. Cependant, dans ces différentes modélisations le processus des sauts est toujours supposé markovien, ce qui est, en vertu de la Proposition 5.1, est un cas particulier relativement contraignant du cadre des CMT considéré dans cette section.

Exemple 4.6.1

Considérons l'image binaire du zèbre, de taille 256x256, représentée à la Figure 6.1. L'image est donnée par une grille carrée de pixels, chaque pixel étant noir ou blanc. La grille est transformée en une suite selon le « parcours de Hilbert-Peano », comme proposé dans [GIO 97]. On obtient ainsi une chaîne $X = x$, l'image observée devenant une chaîne $Y = y$. Le bruitage synthétique est gaussien et spatialement indépendant, de variance un pour les deux classes, et des moyennes, respectivement, de 0 et de 2. En dehors du problème de la stationnarité le bruitage correspond ainsi au modèle classique CMC-BI. L'image observée est alors segmentée de deux manières différentes : en utilisant le modèle classique CMC-BI, et en utilisant le modèle triplet avec la loi donnée par

$$p(x_1, u_1, y_1) = p(x_1, u_1)p(y_1|x_1),$$

$$p(t_{i+1}|t_i) = p(x_{i+1}, u_{i+1}|x_i, u_i)p(y_{i+1}|x_{i+1}) \quad (4.27)$$

Notons que la CMT utilisée est relativement rudimentaire ; en particulier, U et Y sont indépendantes conditionnellement à X . La méthode classique donne $\hat{X}^1 = \hat{x}^1$, avec un taux d'erreur de 7.2%, alors que la méthode utilisant la CMT donne $\hat{X}^2 = \hat{x}^2$, avec un taux d'erreur de 3.5%. Dans cet exemple on considère qu'il existe dans l'image trois stationnarités différentes ($m = 3$). La recherche de $U = u$ donne alors $\hat{U} = \hat{u}$ également présentée sur la Figure 4.1 et l'on constate que les trois zones de stationnarités différentes, que sont le fond (stationnarité « noir »), les bandes relativement

larges présentes sur le corps de l'animal (stationnarité « gris »), et les bandes relativement étroites présentes sur le cou et les pattes de l'animal (stationnarité « blanc ») sont bien retrouvées.

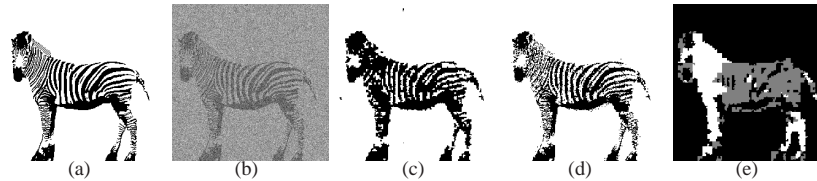


Figure 4.1. (a) Image $X = x$ d'un zèbre, (b) sa version bruitée $Y = y$ par du bruit gaussien indépendant, (c) segmentation non supervisée utilisant les CMC-BI $\tau = 7.2\%$, (d) segmentation non supervisée utilisant les CMT : $\tau = 3.5\%$, (e) segmentation $\hat{U} = \hat{u}$ en trois stationnarités différentes utilisant les CMT. τ désigne le taux d'erreur.

Des études similaires ont été menées dans le cadre de la segmentation d'images par des méthodes fondées sur les champs de Markov triplets [BEN 05a, BEN 05b, BEN 07b]. En particulier, de telles modélisations et méthodes de traitement associées semblent très prometteuses dans le traitement de l'important problème de la classification des textures [BEN 07a, BLA 07].

Enfin, mentionnons que les chaînes et champs de Markov triplets permettent l'utilisation de la théorie de l'évidence de Dempster-Shafer dans le contexte markovien ce qui, en particulier, permet également l'étude des données non stationnaires [LAN 05, PIE 06].

4.7. Chaînes semi-markoviennes cachées et chaînes de Markov triplet

Dans de nombreux domaines d'application, les CMC-BI se sont avérées insuffisantes et il a été fait appel à des modèles plus généraux, dits « chaînes semi-markoviennes cachées » (CSMC). On peut citer les applications en segmentations d'images médicales [FAI 05], traitement de la parole [MOO 04], ou poursuite [YU 03] (notons également les modèles hybrides, mélangeant les aspects « HMC » et « HSMC » [GUE 05]). Sachant qu'une distribution semi-markovienne peut-être vue comme la distribution marginale d'une distribution markovienne, il s'avère que les CSMC sont des CMT particulières. Cette situation permet de proposer, dans un cadre unifié, diverses extensions des CSMC classiques que nous décrivons succinctement dans la suite de cette section.

4.7.1. Chaînes semi-markoviennes cachées comme CMT particuliers

Considérons un processus $X = (X_1, \dots, X_n, \dots)$, avec chaque X_i à valeurs dans $\Omega = \{\omega_1, \dots, \omega_k\}$. X est une chaînes « semi-markovienne », si sa loi est donnée par la loi de X_1 , notée $p(x_1)$, la suite des matrices de transition $(p(x_i|x_{i-1}))_{i \geq 2}$ vérifiant $p(x_i|x_{i-1}) = 0$ pour $x_i = x_{i-1}$, et k suites de lois sur \mathbb{N}^* . Pour chaque $i = 1, \dots, k$, la suite des lois correspondantes sera notée $p^*(\cdot|x_n = \omega_i)$. Une réalisation de la chaîne semi-markovienne $X = (X_1, \dots, X_n, \dots)$ est obtenue par : (i) $X_1 = x_1$ est simulé selon $p(x_1)$; (ii) on simule un entier naturel $N_1 = n_1$ selon $p^*(\cdot|x_1)$; (iii) pour $1 \leq i \leq n_1$, on pose $x_i = x_1$; (iv) on simule $X_{n_1+1} = x_{n_1+1}$ selon $p(x_{n_1+1}|x_{n_1})$; on simule un entier naturel $N_2 = n_2$ selon $p^*(\cdot|x_{n_1})$;

Une chaîne de Markov est alors une chaîne semi-markovienne particulière dans laquelle $p^*(\cdot|x_n = \omega_i)$, qui est la loi du temps de séjour dans ω_i sachant que l'on y arrive à l'instant n , sont des lois exponentielles. L'origine de l'introduction des CSMC vient de la constatation selon laquelle dans un certain nombre de phénomènes réels, la loi de la durée des séjours dans un état n'est pas exponentielle.

En considérant un processus réel observé $Y = (Y_1, \dots, Y_n, \dots)$, la loi d'une CSMC est obtenue en définissant $p(y|x)$ de la même manière que dans les CMC : pour chaque n et $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_n)$, on pose $p(y|x) = p(y_1|x_1) \dots p(y_n|x_n)$. On montre alors que les CSMC sont des TMC particuliers et offrent donc, au prix d'une plus grande complexité des calculs, mêmes possibilités de traitement que les CMC. La chaîne U est telle que pour chaque n , U_n modélise le temps restant de séjour de X_n dans x_n . Si ce temps est supérieur à 1 on a ainsi $U_{n+1} = U_n - 1$. Nous avons donc le résultat suivant :

Proposition 4.7.1

Une chaîne semi-markovienne cachée est une chaîne de Markov triplet $T = (X, U, Y)$, avec U_n à valeurs dans \mathbb{N}^* , définie par $p(x_1, u_1, y_1) = p(x_1)p(u_1|x_1)p(y_1|x_1)$ et les transitions

$$\begin{aligned} p(x_{n+1}, u_{n+1}, y_{n+1}|x_n, u_n, y_n) &= p(x_{n+1}|x_n, u_n, y_n) \times \\ p(u_{n+1}|x_n, u_n, y_n, x_{n+1}) &p(y_{n+1}|x_n, u_n, y_n, x_{n+1}, u_{n+1}) \end{aligned} \quad (4.28)$$

données par (δ désignant la masse de Dirac) :

$$p(x_{n+1}|x_n, u_n, y_n) = p(x_{n+1}|x_n, u_n) = \begin{cases} \delta(x_n) & \text{si } u_n > 1 \\ p(x_{n+1}|x_n) & \text{si } u_n = 1 \end{cases} \quad (4.29)$$

$$\begin{aligned} p(u_{n+1}|x_n, x_{n+1}, u_n, y_n) &= \\ = p(u_{n+1}|x_{n+1}, u_n) &= \begin{cases} \delta(u_n - 1) & \text{si } u_n > 1 \\ p(u_{n+1}|x_{n+1}) & \text{si } u_n = 1 \end{cases} \end{aligned} \quad (4.30)$$

$$p(y_{n+1}|x_n, u_n, y_n, u_{n+1}, x_{n+1}) = p(y_{n+1}|x_{n+1}) \quad (4.31)$$

Preuve 4.7.1 La preuve est immédiate.

4.7.2. Extensions de chaînes semi-markoviennes cachées

Le résultat de la Proposition 4.7.1 permet de proposer, en utilisant divers résultats récents portant sur les CMT, deux séries de généralisations des CSMC.

Dans la première série, nous restons dans la modélisation générale 4.28, en envisageant diverses extensions des identités 4.29-4.31.

Une première généralisation est obtenue en remplaçant dans 4.29 $p(x_{n+1}|x_n)$ par $p(x_{n+1}|x_n, y_n)$ (pour $u_n = 1$). Afin de justifier l'intérêt de cette extension imaginons que les indices des divers processus sont les pixels d'une ligne dans une image numérique représentant une scène de trois classes « eau », « désert » et « forêt ». Chaque X_n prend ses valeurs dans $\Omega = \{\omega_1, \omega_2, \omega_3\}$, chaque Y_n prend ses valeurs dans un sous-ensemble de \mathbb{R} des valeurs numériques possibles, et la loi $p(y_n|x_n)$ modélise - entre autres - la « variabilité naturelle » (sable de diverses natures, verdure de la forêt, couleur de l'eau, ...). Pour $x_n = \omega_3$ et $u_n = 1$ on sait que la forêt s'arrête à n et la probabilité $p(x_{n+1}|x_n = \omega_3)$ est celle d'avoir la classe « eau » ou « désert » à $n+1$. On peut alors imaginer que la probabilité d'avoir « eau » à $n+1$ dépend de l'aspect de la forêt à n : plus cette dernière est verte, plus l'on a des chances d'avoir de l'eau à $n+1$. Cela signifie que remplacer $p(x_{n+1}|x_n)$ par $p(x_{n+1}|x_n, y_n)$ peut avoir une justification physique. De plus, il est maintenant avéré que ce type d'extension peut améliorer les traitements de manière significative. En effet, appliquée dans le contexte plus simple où remplacer $p(x_{n+1}|x_n)$ par $p(x_{n+1}|x_n, y_n)$ conduit au passage des chaînes de Markov cachées aux chaînes de Markov couples, elle a permis d'améliorer notablement la qualité de la segmentation statistique non supervisée de certaines images [DER 04]. Même type de raisonnement aboutit à une justification de la généralisation de l'équation 4.31 qui consiste en le remplacement de $p(y_{n+1}|x_{n+1})$ par $p(y_{n+1}|x_{n+1}, x_n)$, ce qui constitue une deuxième extension du modèle 4.29-4.31. Sachant que u_{n+1} représente le temps de séjour résiduel de X_{n+1}, X_{n+2}, \dots dans x_n , on peut imaginer, en reprenant l'exemple ci-dessus, que l'aspect y_{n+1} de la classe x_{n+1} (un arbre isolé présente des caractéristiques moyennes différentes d'un arbre situé dans une forêt), que $p(y_{n+1}|x_n, x_{n+1}, u_{n+1}) \neq p(y_{n+1}|x_n, x_{n+1})$. De plus, on peut imaginer que l'aspect moyen (donc la loi de Y_{n+1}) de la classe x_{n+1} est différent selon qu'il y a changement de classe entre n et $n+1$ (ce qui équivaut à $u_n = 1$), ou pas (ce qui équivaut à $u_n > 1$). Cela signifie que $p(y_{n+1}|x_n, u_n, x_{n+1}, u_{n+1}) \neq p(y_{n+1}|x_n, x_{n+1}, u_{n+1})$. Enfin, on pourrait également imaginer que $p(u_{n+1}|x_{n+1})$ dans 4.30 dépend aussi de y_n .

Lorsque l'on considère simultanément ces différentes extensions, les formules 4.29-4.31 deviennent :

$$p(x_{n+1}|x_n, u_n, y_n) = p(x_{n+1}|x_n, u_n) = \delta(x_n) \text{ si } u_n > 1, \\ \text{et } p(x_{n+1}|x_n, y_n) \text{ si } u_n = 1. \quad (4.32)$$

$$p(u_{n+1}|x_n, x_{n+1}, u_n, y_n) = p(u_{n+1}|u_n, x_{n+1}) = \delta(u_n - 1) \text{ si } u_n > 1, \\ \text{et } p(u_{n+1}|x_{n+1}, y_n) \text{ si } u_n = 1. \quad (4.33)$$

$$p(y_{n+1}|x_n, u_n, y_n, u_{n+1}, x_{n+1}) = p(y_{n+1}|x_n, u_n, u_{n+1}, x_{n+1}) \quad (4.34)$$

Les extensions 4.32-4.34 des transitions 4.29-4.31 admettent ainsi des justifications concrètes ; en d'autres termes, 4.29-4.31 sont des simplifications des situations réelles pouvant s'avérer préjudiciables à la qualité des traitements.

La deuxième série d'extensions est obtenue en exploitant le fait que la CMT $T = (X, U, Y)$ de la proposition de la section précédente est également un CMCouple $T = (V, Y)$, avec $V = (X, U)$. On peut alors appliquer le passage classique d'une chaîne couple à une chaîne triplet en considérant un processus latent

$W = (W_1, \dots, W_n, \dots)$, chaque W_i prenant ses valeurs dans $\Lambda = \{\lambda_1, \dots, \lambda_m\}$, et une nouvelle chaîne triplet $T' = (V, W, Y)$. Tous les cas considérés dans la première série ci-dessus peuvent alors être étendus à T' en remplaçant, dans toutes les transitions figurant dans chacun des modèles, les (x, u) par les (x, u, w) . Les modélisations de la première série permettent d'estimer (x, u) et donc celles de la deuxième série permettent d'estimer (x, u, w) . Ces diverses estimations donnent x recherché.

4.8. Chaînes auxiliaires multi-variées

Les différentes propriétés des chaînes aléatoires cachées modélisées par des CMT discutées dans les sections précédentes peuvent être considérées simultanément, ce qui aboutit à un grand nombre de modèles particuliers. Cela revient à considérer une chaîne auxiliaire U sous forme multi-variée $U = (U^1, \dots, U^d)$, où chaque U^i modélise une certaine propriété, éventuellement parmi celles mentionnées dans les sections précédentes. On arrive rapidement à des modèles assez complexes dont la mise en oeuvre, notamment dans les cas des méthodes de traitement non supervisé, devient rapidement relativement lourde. Il se pose alors, au plan de l'utilité pratique de tels modèles, au moins deux questions :

- les CMC-BI classiques ayant fait preuve de leur grande robustesse, le gain dû à la grande généralité de tels modèles, même dans l'hypothèse de l'utilisation des paramètres optimaux, est-il suffisant pour justifier leur utilisation ?
- la quantité des paramètres croissant rapidement dans les modèles faisant appel aux chaînes auxiliaires multi-variées, les méthodes de leur estimation permettant des traitements non supervisés ne risquent-elles pas, notamment dans des cas très bruités, d'être mises en défaut ?

Il est très difficile de répondre à de telles questions de manière générale ; de plus, l'introduction de ces modèles est très récente et l'on dispose de relativement peu d'études pratiques. Cependant, les résultats de certaines de ces études sont très encourageants. Nous présentons ci-dessous des résultats concernant les CMT modélisant simultanément la semi-Markoviannité de la chaîne cachée et sa non stationnarité [LAP 06]. Considérons une CMT $T = (X, U, Y)$, avec $U = (U^1, U^2)$, $X = (X_1, \dots, X_n)$, $U^1 = (U_1^1, \dots, U_n^1)$, $U^2 = (U_1^2, \dots, U_n^2)$ et $Y = (Y_1, \dots, Y_n)$. Les variables X_i , U_i^1 , U_i^2 et Y_i prennent leurs valeurs respectivement dans $\Omega = \{\omega_1, \dots, \omega_k\}$, $\Lambda_1 = \mathbb{N}^*$,

$\Lambda_2 = \{1, \dots, m\}$ et \mathbb{R} . Le processus U^1 modélise la semi-Markoviannité de X et le processus U^2 modélise sa non stationnarité. La loi de T est donnée par $p(t_1)$ et les transitions $p(t_{i+1}|t_i)$ qui peuvent présenter un grand nombre d'écritures particulières. Considérons d'emblée le cas particulier suivant

$$p(t_{i+1}|t_i) = p(u_{i+1}^2|x_i, u_i^1, u_i^2) \times \\ \times p(x_{i+1}|u_{i+1}^2, x_i, u_i^1, u_i^2) p(u_{i+1}^1|x_{i+1}, u_{i+1}^2, x_i, u_i^1, u_i^2) p(y_{i+1}|x_{i+1}) \quad (4.35)$$

avec

$$p(u_{i+1}^2|x_i, u_i^1, u_i^2) = \delta_{u_i^2}(u_{i+1}^2) \text{ si } u_i^1 > 1, \\ \text{et } p(u_{i+1}^2|u_i^2) \text{ si } u_i^1 = 1; \quad (4.36)$$

$$p(x_{i+1}|u_{i+1}^2, x_i, u_i^1, u_i^2) = \delta_{x_i}(x_{i+1}) \text{ si } u_i^1 > 1, \\ \text{et } p(x_{i+1}|u_{i+1}^2, x_i) \text{ si } u_i^1 = 1; \quad (4.37)$$

$$p(u_{i+1}^1|x_i, x_{i+1}, u_i^1, u_i^2, u_{i+1}^2) = \delta_{u_i^1-1}(u_{i+1}^1) \text{ si } u_i^1 > 1, \\ \text{et } p(u_{i+1}^1|x_{i+1}, u_{i+1}^2) \text{ si } u_i^1 = 1; \quad (4.38)$$

On suppose par ailleurs que pour $u_i^1 = 1$ dans 4.37 on a $p(x_{i+1} = x_i|u_{i+1}^2, x_i) = 0$; la réalisation $U_i^1 = u_i^1$ représente alors le temps exact restant du séjour de la chaîne X dans l'état x_i .

Ce type de modèles a été étudié dans [LAP 06] et les premiers résultats numériques sont encourageants. Ces résultats concernent un modèle légèrement différent du modèle précédent, la différence consistant dans le fait que l'on considère $\Lambda_1 = \{1, \dots, l\}$ fini et, pour $u_i^1 = 1$ dans 4.37, on suppose que $p(x_{i+1} = x_i|u_{i+1}^2, x_i)$ peut être non nul; la réalisation $U_i^1 = u_i^1$ représente le temps minimal restant du séjour de la chaîne X dans l'état x_i . Un tel modèle s'apparente à un modèle de chaîne de Markov cachée très classique. En effet, la chaîne $V = (X, U^1, U^2)$ est une chaîne discrète finie, chaque V_i prenant ses valeurs dans $\Omega \times \Lambda_1 \times \Lambda_2 = \{\omega_1, \dots, \omega_k\} \times \{1, \dots, l\} \times \{1, \dots, m\}$, et elle est classiquement bruitée par $p(y_i|v_i) = p(y_i|x_i)$. En considérant T stationnaire, sa loi est définie par $p(t_1, t_2) = p(v_1, v_2)p(y_1|x_1)p(y_2|x_2)$. Dans le cas gaussien, nous avons ainsi, dans le cas le plus général, $(klm)^2$ paramètres définissant $p(v_1, v_2)$, ainsi que k moyennes, et k variances définissant les lois $p(y_i|x_i)$.

Notons que l'hypothèse $p(y_i|v_i) = p(y_i|x_i)$ implique que les chaînes Y et (U^1, U^2) sont indépendantes conditionnellement à la chaîne X ce qui pourrait, a priori, préjuger de la difficulté de l'estimation de (U^1, U^2) . Cependant, une telle difficulté n'apparaît pas dans la pratique.

Nous présentons ci-après deux séries de résultats.

Dans la première, nous simulons une chaîne de Markov triplet $T = (X, U, Y)$ vérifiant 4.35-4.38, avec 4.37 modifié comme précisé ci-dessus. Ensuite, la réalisation $Y = y$ est segmentée par trois méthodes MPM, fondées respectivement sur

- les chaînes de Markov cachées avec bruit indépendant (CMC-BI) classiques;
- les chaînes semi-markoviennes cachées avec bruit indépendant (CSMC-BI);

– et le vrai modèle CMT ; qui est une chaîne semi-markovienne cachée non stationnaire (CSMCNS-BI).

L'objet de cette étude est de tester si les deux premières méthodes sont capables d'approcher, en qualité, la solution optimale donnée par la troisième.

Le modèle CMT considéré est le suivant. Soient $k = 2$, $l = 5$ et $m = 2$. Les deux gaussiennes modélisant le bruit ont la même variance commune égale à 1, et

les moyennes de 1 et 1.5. Par ailleurs, on prend $p(u_{i+1}^2 | u_i^2) = \begin{pmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{pmatrix}$

dans 4.36,

$p(x_{i+1} | u_{i+1}^2 = 0, x_i) = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}$ et $p(x_{i+1} | u_{i+1}^2 = 1, x_i) = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$

dans 4.37, et $p(u_{i+1}^1 | x_{i+1}, u_{i+1}^2) = 0.1$ dans 4.38.

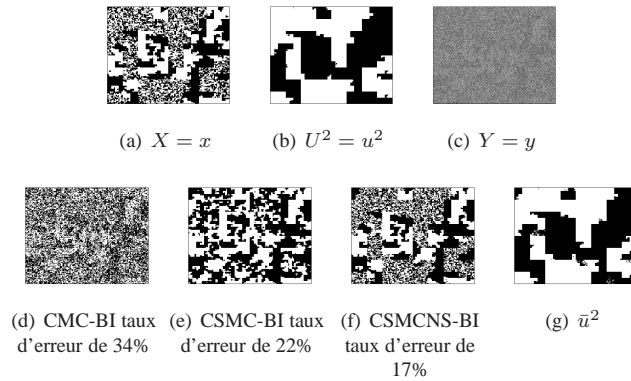


Figure 4.2. Réalisation de (X, U^2, Y) et segmentations non supervisées avec CMC-BI, CSMC-BI et CSMCNS. \hat{u}^2 est l'estimée de $U^2 = u^2$ avec CSMCNS-BI.

Les paramètres définissant les lois respectives de X nécessaires à la mise en place des segmentations fondées sur les CMC-BI et CSMC-BI sont estimés à partir de X . Les résultats présentés à la Figure 4.2 montrent que la méthode fondée sur le modèle classique CMC-BI donne des résultats relativement médiocres, aussi bien du point de vue visuel que quantitatif, avec un taux d'erreur de 34%. Ces résultats sont améliorés par la méthode fondée sur le modèle CSMC-BI ; visuellement, on commence à distinguer les deux stationnarités, par ailleurs le taux d'erreur passe à 22%. Cette première comparaison montre que la semi-markoviannité des données cachées ne peut pas être négligée, malgré la très bonne robustesse des CMC-BI classiques. Il est également intéressant de noter que cette première comparaison est faite dans un contexte des

données non stationnaires, qui ne correspondent à aucun des deux modèles comparés. Enfin, la segmentation optimale utilisant le modèle CSMCNS-BI donne de très bons résultats visuels, le taux d'erreur devenant 17%. Cet exemple met ainsi en lumière l'existence d'une situation simple où les CMC-BI classiques sont inopérants ; de plus, l'utilisation simultanée des deux chaînes auxiliaires permet d'améliorer les résultats obtenus avec l'utilisation d'une seule d'entre elles.

Dans la deuxième série nous considérons des données qui ne correspondent à aucun des trois modèles ci-dessus et l'objet en est de montrer l'existence des situations dans lesquelles l'utilisation de la CMT, soit la modélisation la plus complète et la plus complexe des trois, présente un intérêt. Les paramètres nécessaires à la mise en oeuvre des différentes méthodes sont estimés avec l'ECI. Ainsi, l'objectif de cette série de résultats est d'apporter des éléments de réponse à la deuxième question ci-dessus : malgré une plus grande complexité impliquant un plus grand nombre de paramètres à estimer, les traitements fondés sur le modèle CSMCNS-BI sont les plus efficaces, malgré un niveau très élevé de bruit.

Nous considérons une image dessinée à la main, représentée à la Figure 4.3. On distingue deux types de stationnarités - on dit aussi deux « textures » qui sont considérées comme une réalisation u^2 de la chaîne auxiliaire U^2 , et également représentées à la Figure 4.3. L'image des classes $X = x$ est ensuite bruitée par deux gaussiennes de moyennes 1 et 1.5, et de variance 1, donnant l'image observée $Y = y$. Comme ci-dessus, cette dernière est ensuite segmentée par les trois méthodes bayésiennes MPM non supervisées, les paramètres étant estimés par des méthodes de type ECI, fondées sur les trois modèles CMC-BI, CSMC-BI et CSMCNS-BI. Nous constatons que la tendance de la série précédente persiste : l'utilisation du modèle CMC-BI donne des résultats médiocres, avec un taux d'erreur de 35%, CSMC-BI les améliore avec un taux d'erreur de 23%, et CSMCNS-BI donne des meilleurs résultats avec un taux d'erreur de 14%. Par ailleurs, les trois méthodes d'estimation des paramètres du bruit donnent des résultats comparables, présentés dans le Tableau 1. Il en résulte que malgré la généralité croissante des modèles, impliquant le nombre croissant des paramètres à estimer et donc la difficulté croissante à le faire convenablement, la qualité de l'estimation obtenue dans les trois modèles est comparable.

Ainsi cette deuxième série d'expérimentations nous permet d'affirmer l'existence de situations dans lesquelles les questions (i),(ii) posées au début de la section admettent des réponses. En effet, il apparaît que les traitements non supervisés fondés sur les CMC-BI peuvent être inefficaces et les CSMCNS-BI assurent des solutions de qualité tout-à-fait intéressante. Il est également intéressant de noter que les résultats fondés sur les CSMC-BI sont, comme dans la première série de résultats ci-dessus, de qualité intermédiaire, ce qui montre que l'ajout des chaînes auxiliaires successives améliore les résultats obtenus sur des données ne suivant, a priori, aucun modèle. Concernant le comportement de l'ECI, nous notons la bonne estimation des paramètres du bruit. Enfin, il est à noter que le parcours de Peano appliqué à l'ensemble des pixels produit une chaîne non stationnaire, même dans le cas des images stationnaires « visuellement », ce qui renforce l'intérêt des résultats.

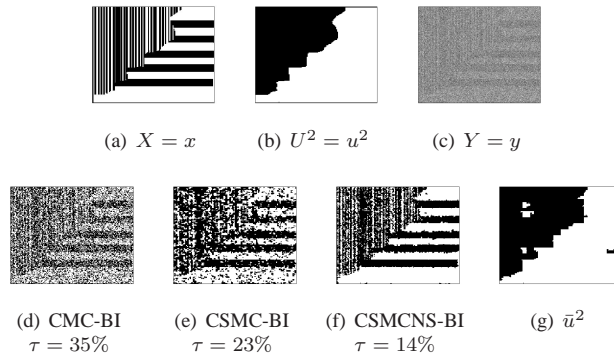


Figure 4.3. Réalisation de (X, U^2, Y) et segmentations non supervisées avec CMC-BI, CSMC-BI et CSMCNS. τ représente le taux d'erreur et \bar{u}^2 est l'estimée de $U^2 = u^2$ avec CSMCNS-BI.

Classe	CMC-BI		CSMC-BI		CSMCNS-BI	
	Moyenne	Ecart-type	Moyenne	Ecart-type	Moyenne	Ecart-type
ω_1	0.84	0.91	1.09	1.04	0.9	0.94
ω_2	1.65	0.89	1.46	1.02	1.49	0.99
τ	35%		23%		14%	

Tableau 1 Estimation des paramètres de bruit, correspondant aux données de la Figure 4.8, par la méthode ECI dans les trois modèles CMC-BI, CSMC-BI et CSMCNS-BI. τ est le taux d'erreur obtenu en appliquant la méthode bayésienne MPM utilisant les paramètres estimés.

4.9. Conclusions et perspectives

Nous avons exposé dans ce chapitre certains résultats récents concernant les modèles « chaînes de Markov triplet » (CMT) $T = (X, U, Y)$, où le processus observé Y est à valeurs continues, alors que le processus caché X , ainsi que le processus auxiliaire U sont à valeurs dans des ensembles finis. Un tel modèle offre une très grande variété de modèles particuliers. Certains d'entre eux sont des extensions des modélisations classiques ; d'autres permettent de traiter des cas importants pour la pratique, comme celui des processus aléatoires non stationnaires. Des CMT de complexité croissante permettent également d'aborder, dans un cadre unifié, des problématiques réunissant plusieurs aspects classiques. En particulier, des résultats de segmentation des chaînes semi-markoviennes cachées non stationnaires ont été présentés. Enfin, la théorie des copules, récemment introduite dans le cadre des modèles de Markov couples [BRU 05a, BRU 05b], permet des modélisations très complètes du

bruit. Dans tous les cas de figure la méthode générale d'apprentissage « Estimation conditionnelle itérative » (ECI) permet d'estimer les paramètres des divers modèles et autorise ainsi la conception des méthodes bayésiennes non supervisées de segmentation.

Notons que des études similaires du modèle où les trois chaînes X , U et Y sont continues, sont en cours et un certain nombre de résultats encourageants ont été publiés [FQU 06, DES 03b, DES 03a]. L'adjonction de U permet de considérer les problèmes classiques de filtrage ou de lissage dans des modèles plus généraux, de nature à mieux correspondre aux données réelles. En particulier, on montre que tous les traitements, qu'ils soient exacts de par la nature gaussienne des processus considérés [FQU 06, DES 03a], ou approchés par des techniques de filtrage ou lissage particulière [DES 03b], s'étendent aisément aux nouveaux modèles. Comme dans le cas considéré dans le présent chapitre, l'introduction de la chaîne auxiliaire permet ainsi de proposer des modèles très généraux, dans lesquels (X, Y) n'est plus nécessairement markovienne [DES 03a].

D'autres possibilités de mixer les chaînes à valeurs continues avec celles à valeurs discrètes existent et permettent la conception des modélisations et des traitements associés originaux, susceptible de mieux correspondre à un certain nombre de situations et ainsi améliorer les résultats obtenus avec des démarches classiques. Citons l'exemple traité dans [BEN 07a], où la chaîne X est discrète alors que les chaînes U et Y sont continues. Enfin, le choix entre « discret » et « continu » n'est pas exhaustif et il est possible de considérer des chaînes composées des variables admettant pour lois des lois « mixtes », comportant simultanément une partie continue et une partie discrète. Les premiers modèles de ce type ont été proposés dans le contexte des données indépendantes dans [CAI 93] ; ensuite, différents modèles de Markov cachés à « états mixtes » ont été introduits (champs [PIE 94], chaînes [GER 02], et arbres [MON 03]). Ces différents modèles ont été généralement utilisés à des fins de segmentation floue [CAR 06, RUA 00, RUA 02, SAL 06] ; notons également une interprétation de la partie continue de la mesure mixte plus générale, récemment utilisée pour modéliser les textures dans [BOU 06]. Tous ces modèles peuvent être étendus aux modèles triplets et la première étude de ce type, traitant de la non stationnarité, est très encourageante [SAL 07].

Les modèles de type « chaîne de Markov » étudiés dans ce chapitre s'étendent relativement directement aux modèles de type « arbres de Markov » qui présentent également de multiples possibilités d'applications, notamment en imagerie multirésolution. Les premières modélisations de type « Markov couple » [DES 05, DES 06, MON 03], aisément généralisable aux Markov « triplet » sont ainsi susceptibles de bénéficier des différentes généralisations étudiées, ou simplement suggérées dans le cadre des chaînes. Par ailleurs, un certain nombre de considérations exposées dans ce chapitre dans le cadre des chaînes de Markov s'adaptent aisément au contexte des champs de Markov. Les champs de Markov triplet [BEN 05a] permettent ainsi des modélisations et des traitements des champs non stationnaires [BEN 05b], ou des traitements

de l'important problème de classification des textures [BEN 05b, BLA 07]. Notons qu'un lien avec la « théorie de l'évidence », non exposé dans ce chapitre, a également été établi et étudié dans [BEN 05a, PIE 06]. D'autres modélisations décrites dans ce chapitre dans le cadre des chaînes de Markov, en particulier celles faisant intervenir des champs auxiliaires multivariés ou des mesures de probabilité mixtes, font ainsi partie des perspectives les plus prometteuses. Notons que ces différentes considérations peuvent être étendues aux modèles graphiques markoviens généraux. De plus, ces extensions peuvent être envisagées dans le cadre des modèles « partiellement de Markov », comme proposé dans le cas des chaînes dans [PIE 05]. Les premières études de tels modèles, autorisant l'introduction des bruits « à mémoire longue », donnent des résultats encourageants [LAP 07, LAN 08].

Enfin, notons la possibilité de considérer des observations multivariées, y compris de nature différente comme dans [BRI 06, BRI 08], ce qui démultiplie les possibilités de conception des modèles de Markov triplets originaux.

4.10. Bibliographie

- [BEL 00] BELLONE E., HUGUES J., GUTTORP P., « A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts », *Climate research*, vol. 15, n°1, p. 1-15, 2000.
- [BEN 01] BENDJEBBOUR A., DELIGNON Y., FOUQUE L., SAMSON V., PIECZYNSKI W., « Multisensor images segmentation using Dempster-Shafer fusion in Markov fields context », *Trans. on Geoscience and Remote Sensing*, vol. 39, n°8, p. 1789-1798, 2001.
- [BEN 05a] BENBOUDJEMA D., Champs de Markov triplets et segmentation bayésienne non supervisée d'images, PhD thesis, thèse de l'Institut National des Télécommunications, Evry, 2005, soutenue le 12 décembre 2005.
- [BEN 05b] BENBOUDJEMA D., PIECZYNSKI W., « Unsupervised image segmentation using triplet Markov fields », *Computer Vision and Image Understanding*, vol. 99, n°3, p. 476-498, 2005.
- [BEN 07a] BEN MABROUK M., PIECZYNSKI W., « Unsupervised segmentation of random discrete data using triplet Markov chains », Chania, Greece, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA), May 2007.
- [BEN 07b] BENBOUDJEMA D., PIECZYNSKI W., « Unsupervised statistical segmentation of non stationary images using triplet Markov fields », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, n°8, p. 1367-1378, 2007.
- [BLA 07] BLANCHET J., Modèles Markoviens et extensions pour la classification de données complètes, PhD thesis, thèse de l'Université Joseph Fourier, Grenoble, 2007, soutenue le 10 octobre 2007.
- [BOU 06] BOUTHEMY P., HARDOUIN C., PIRIOU G., YAO J. -F., « Mixed-state auto-models and motion texture modeling », *Journal of Mathematical Imaging and Vision*, vol. 25, n°3, p. 387-402, 2006.

- [BRI 06] BRICQ S., COLLET C., ARMSPACH J.-P., « Triplet Markov chains for 3D MRI brain segmentation using a probabilistic atlas », Arlington, Virginia, USA, Proceedings of IEEE International Symposium on Biomedical Imaging, ISBI'06, April 6-9, 2006.
- [BRI 08] BRICQ S., COLLET C., ARMSPACH J.-P., « Unifying framework for multimodal brain MRI segmentation based on Hidden Markov Chains », *Medical Image Analysis. Elsevier*, vol. doi :10.1016/j.media.2008.03.001, 2008.
- [BRU 05a] BRUNEL N., Sur quelques extensions des chaînes de Markov cachées et couples, Application à la segmentation non supervisée des signaux radar, PhD thesis, thèse de l'Université Paris VI, 2005, soutenue le 5 décembre 2005.
- [BRU 05b] BRUNEL N., PIECZYNSKI W., « Unsupervised signal restoration using hidden Markov chains with copulas », *Signal Processing*, vol. 85, n°12, p. 2304-2315, 2005.
- [CAI 93] CAILLOL H., HILLION A., PIECZYNSKI W., « Fuzzy random fields and unsupervised image segmentation », *IEEE Trans. on Geoscience and Remote Sensing*, vol. 34, n°4, p. 801-910, 1993.
- [CAP 05] CAPPÉ O., MOULINES E., RYDEN T., *Inference in hidden Markov models*, Springer, Series in Statistics, 2005.
- [CAR 06] CARINCOTTE C., DERRODE S., BOURENNANE S., « Unsupervised change detection on SAR images using fuzzy hidden Markov chains », *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, n°2, p. 432-441, 2006.
- [CHE 94] CHEN M., KUNDU A., ZHOU J., « Off-Line Handwritten Work Recognition Using a Hidden Markov Model Type Stochastic Network », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, n°5, 1994.
- [DEL 97] DELMAS J.-P., « An equivalence of the EM and ICE algorithm for exponential family », *IEEE Trans. on Signal Processing*, vol. 45, n°10, p. 2613-2615, 1997.
- [DEL 02] DELIGNON Y., PIECZYNSKI W., « Modeling non-Rayleigh speckle distribution in SAR images », *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, n°6, p. 1430-1435, 2002.
- [DER 04] DERRODE S., PIECZYNSKI W., « Signal and Image segmentation using pairwise Markov chains », *IEEE Trans. on Signal Processing*, vol. 52, n°9, p. 2477-2489, 2004.
- [DES 03a] DESBOUVRIES F., PIECZYNSKI W., « Modèles de Markov Triplet et Filtrage de Kalman », *Comptes Rendus de l'Académie des Sciences-Mathématiques*, vol. 336-8, p. 667-70, 2003.
- [DES 03b] DESBOUVRIES F., PIECZYNSKI W., « Particle Filtering in pairwise and triplet Markov chains », Grado-Gorizia, Italy, Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP 2003), June 8-11 2003.
- [DES 05] DESBOUVRIES F., LECOMTE J., « Multiscale Bayesian restoration in pairwise Markov trees », *IEEE Trans. on Automatic Control*, vol. 50, n°8, p. 1185-1190, 2005.
- [DES 06] DESBOUVRIES F., LECOMTE J., PIECZYNSKI W., « Kalman filtering in pairwise Markov trees », *Signal Processing*, vol. 86, n°5, p. 1049-54, 2006.

- [DIZ 07] DI ZIO M., GUAMERA U., ROCCI R., « A mixture of mixture models for a classification problem : the unity measure error », *Computational Statistics & Data Analysis*, vol. 51, n°5, p. 2573-2585, 2007.
- [FAI 05] FAISAN S., THORAVAL L., ARMSPACH J., METZ-LUTZ M.-N., HEITZ F., « Un-supervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models », *IEEE Trans. on Medical Imaging*, vol. 24, n°2, p. 263-276, 2005.
- [FQU 06] AIT-EL FQUIH B., DESBOUVRIES F., « Kalman filtering in triplet Markov chains », *IEEE Trans. on Signal Processing*, vol. 54, n°8, p. 2957-63, 2006.
- [GER 02] GERMAIN M., VOORONS M., BOUCHER J. M., BENIE G. B., « Fuzzy statistical classification method for multiband image fusion », *Proceedings of the Fifth International Conference on Information Fusion*, vol. 1, p. 178-184, 2002.
- [GIO 97] GIORDANA N., PIECZYNSKI W., « Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, n°5, p. 465-475, 1997.
- [GRE 00] GREGOIR S., LENGART F., « Measuring the probability of a business cycle turning point by using a multivariate qualitative hidden Markov model », *Journal of forecasting*, vol. 19, n°2, page 81, 2000.
- [GUE 05] GUEDON Y., « Hidden hybrid Markov/semi-Markov chains », *Computational Statistics & Data Analysis*, vol. 49, n°3, p. 663-688, 2005.
- [KOS 01] KOSKI T., *Hidden Markov models for bioinformatics*, Kluwer Academic Publishers, 2001.
- [LAN 04] LANCHANTIN P., PIECZYNSKI W., « Unsupervised non stationary image segmentation using triplet Markov chains », Brussels, Belgium, Advanced Concepts for Intelligent Vision Systems (ACIVS 04), Aug. 31-Sept. 3, 2004.
- [LAN 05] LANCHANTIN P., PIECZYNSKI W., « Unsupervised restoration of hidden non stationary Markov chain using evidential priors », *IEEE Trans. on Signal Processing*, vol. 53, n°8, p. 3091-3098, 2005.
- [LAN 06] LANCHANTIN P., Chaînes de Markov Triplets et segmentation non supervisée des signaux, PhD thesis, thèse de l'Institut National des Télécommunications, Evry, 2006, soutenue le 5 décembre 2006.
- [LAN 08] LANCHANTIN P., LAPUYADE-LAHORGUE J., PIECZYNSKI W., « Unsupervised Segmentation of triplet Markov chains hidden with long-memory noise », *Signal Processing*, vol. 88, n°5, p. 1134-1151, 2008.
- [LAP 06] LAPUYADE-LAHORGUE J., PIECZYNSKI W., « Unsupervised Segmentation of hidden semi-Markov non stationary chains », Paris, France, Twenty sixth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, MaxEnt2006, July 8-13, 2006.
- [LAP 07] LAPUYADE-LAHORGUE J., PIECZYNSKI W., « Partially Markov models and unsupervised segmentation of semi-Markov chains with long-dependent noise », Chania, Greece, International Symposium on Applied Stochastic Models and Data Analysis,

(ASMDA 2007), May 2007.

- [LEB 06] LE BER F., BENOÎT M., SCOTT C., MARI J.-F., MIGNOLET C., « Studying crop sequences with CarrotAge, a HMM-based data mixing software », *Ecological Modelling*, vol. 191, n°1, p. 170-185, 2006.
- [MAI 01] MAÎTRE H., *Traitement des images RSO*, Hermes, Collection IC2, 2001.
- [MCL 97] MCLACHLAN G., KRISHNAN T., *EM algorithm and extensions*, Wiley, Series in Probabilities and Statistics, 1997.
- [MON 03] MONFRINI E., LECOMTE J., DESBOUVRIES F., PIECZYNSKI W., « Image and Signal Restoration using Pairwise Markov Trees », Saint Louis, Missouri, IEEE Workshop on Statistical Signal Processing (SSP 2003), September 28-October 1, 2003.
- [MOO 04] MOORE M., SAVIC M., « Speech reconstruction using a generalized HSMM (GHSMM) », *Digital Signal Processing*, vol. 14, n°1, p. 37-53, 2004.
- [NEL 98] NELSEN R., *An introduction to Copulas*, Springer-Verlag, 1998, Number 139 in Lecture notes in Statistics.
- [NIC 02] NICOLAS P., BIZE L., MURI-MAJOUBE F., HOEBEKE M., RODOLPHE F., DUSKO EHRlich S., PRUM B., BESSIÈRES P., « Mining bacillus subtilis chromosome heterogeneities using hidden Markov models », *Nucleic Acid Research*, vol. 30, 2002.
- [NUE 07] NUEL G., PRUM B., *Analyse statistique des séquences biologiques : modélisation markovienne, alignements et motifs*, Hermes, Collection Bioinformatique, 2007.
- [PIE 92] PIECZYNSKI W., « Statistical image segmentation », *Machine Graphics and Vision*, vol. 1, n°1/2, p. 261-268, 1992.
- [PIE 94] PIECZYNSKI W., CAHEN J. M., « Champs de Markov flous cachés et segmentation d'images », *Revue de Statistique Appliquée*, vol. 42, n°3, p. 13-31, 1994.
- [PIE 02] PIECZYNSKI W., HULARD C., VEIT T., « Triplet Markov Chains in hidden signal restoration », Crete, Greece, SPIE's International Symposium on Remote Sensing, September 22-27, 2002.
- [PIE 03] PIECZYNSKI W., « Pairwise Markov chains », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, n°5, p. 634-639, 2003.
- [PIE 05] PIECZYNSKI W., « Copules gaussiennes dans les chaînes triplet partiellement de Markov », *Comptes Rendus de l'Académie des Sciences-Mathématiques*, vol. 341, n°3, p. 189-194, 2005.
- [PIE 06] PIECZYNSKI W., BENBOUDJEMA D., « Multisensor triplet Markov fields and theory of evidence », *Image and Vision Computing*, vol. 24, n°1, p. 61-69, 2006.
- [PIE 07a] PIECZYNSKI W., « Convergence of the iterative conditional estimation and application to mixture proportion identification », Madison, Wisconsin, USA, IEEE Workshop on Statistical Signal Processing (SSP 2007), August 26-29, 2007.
- [PIE 07b] PIECZYNSKI W., « Multisensor triplet Markov chains and theory of evidence », *International Journal of Approximate Reasoning*, vol. 45, n°1, p. 1-16, 2007.

- [RAP 99] RAPHAEL C., « Automatic segmentation of acoustic musical signals using hidden Markov models », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, n°4, page360, 1999.
- [RUA 00] RUAN S., JAGGI C., XUE J., FADILI J., BLOYET D., « Brain tissue classification of magnetic resonance images using partial volume modeling », *IEEE Trans. on Medical Imaging*, vol. 19, n°12, p. 1179-1187, 2000.
- [RUA 02] RUAN S., MORETTI B., FADILI J., BLOYET D., « Fuzzy Markovian segmentation in application of magnetic resonance images », *Computer Vision and Image Understanding*, vol. 85, p. 54-69, 2002.
- [SAL 06] SALZENSTEIN F., COLLET C., « Fuzzy Markov random fields versus chains for multispectral image segmentation », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, n°11, p. 1753-1767, 2006.
- [SAL 07] SALZENSTEIN F., COLLET C., LECAM S., HATT M., « Non stationary fuzzy Markov chains », *Pattern Recognition Letters*, vol. 28, n°16, p. 2201-2208, 2007.
- [THO 02] THOMAS L., ALLEN D., MORKEL-KINGSBURY N., « A hidden Markov chain model for the term structure of bond credit risk spreads », *International Review of Financial Analysis*, vol. 11, n°3, p. 311-329, 2002.
- [YU 03] YU S.-Z., KOBAYASHI H., « A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking », *Signal Processing*, vol. 83, n°2, p. 235-250, 2003.