Available online at www.sciencedirect.com
SCIENCE DIRECT®

C. R. Acad. Sci. Paris, Ser. I 341 (2005) 189–194

COMPTES RENDUS MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Copules gaussiennes dans les chaînes triplet partiellement de Markov

Wojciech Pieczynski

INT/GET, département CITI, CNRS UMR 5157, 9, rue Charles-Fourier, 91000 Evry, France

Reçu le 3 décembre 2004 ; accepté après révision le 2 juin 2005

Disponible sur Internet le 18 juillet 2005

Présenté par Paul Deheuvels

Résumé

Les chaînes de Markov cachées, largement utilisées dans la problématique de la restauration des données, ont été récemment généralisées aux chaînes couple partiellement de Markov, dans lesquels le processus caché n'est plus nécessairement de Markov et la loi du processus observé conditionnellement au processus caché est quelconque. D'abord, nous montrons l'applicabilité pratique de ces modèles dans le cas gaussien, avec mention particulière pour les bruits à mémoire longue. Ensuite, nous montrons que l'utilisation des copules permet de prendre en compte des marginales quelconques du processus observé conditionnellement au processus caché. Nous terminons en étendant ce dernier modèle aux chaînes triplet partiellement de Markov. ***Pour citer cet article : W. Pieczynski, C. R. Acad. Sci. Paris, Ser. I 341 (2005).***

Abstract

Gaussian copulas in triplet, partially Markov chains. Hidden Markov chains, which are widely used in different data restoration problems, have recently been generalised to pairwise partially Markov chains, in which the hidden chain is no longer necessarily Markovian and the distribution of the observed chain, conditional on the hidden one, is of any form. First, we show the applicability of the models in the Gaussian case, with a particular attention to long range correlation noises. Second, we show that the use of copulas allows one to take into account any other form of marginal distributions of the observed chain, conditionally to the hidden one. We end by extending the latter model to a triplet partially Markov chain case. ***To cite this article: W. Pieczynski, C. R. Acad. Sci. Paris, Ser. I 341 (2005).***

Adresse e-mail : Wojciech.Pieczynski@int-evry.fr (W. Pieczynski).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2005.06.012

Abridged English version

Introduction

Let $X = (X_1, \dots, X_n)$ be a hidden process, and $Y = (Y_1, \dots, Y_n)$ an observed one. Each X_i takes its values in $\Omega = \{\omega_1, \dots, \omega_K\}$, and each Y_i takes its values in R. Different estimations of $X = x$ from Y are workable, once the distribution of X conditional on Y, denoted by $p(x|y)$, is a Markov chain. The simplest model ensuring this property is the hidden Markov chain with independent noise (HMC-IN), in which X is a Markov chain and $p(y|x) = p(y_1|x_1) \cdots p(y_n|x_n)$. This model has since been extended to "pairwise Markov chain" (PMC) models, in which one directly assumes the Markovianity of $Z = (X, Y)$ [5,8]. In PMC both $p(y|x)$ and $p(x|y)$ are Markovian, but X need no longer be a Markov chain. As the noise distribution $p(y|x)$ is Markovian, it is a 'short memory' noise, while 'long memory' noise processes are useful in numerous situations [1,3,4]. The aim of this Note is to show that a 'long memory' noise hidden process obtained from 'partially' Markov chains [9] can be workable and can provide Bayesian estimation of the hidden signal X generalising the classical one.

Markov chains hidden with general Gaussian noise

The pairwise process $Z = (X, Y)$ is a 'partially Markov' chain if (1) holds, with $z^i = (z_1, \dots, z_i)$ and $y^i = (y_1, \dots, y_i)$. As (1) is equivalent to $p(z_1, \dots, z_n) = p(x_1, y_1)p(x_2, y_2|x_1, y^1) \cdots p(x_n, y_n|x_{n-1}, y^{n-1})$, we see that $p(x|y)$ is a Markov chain with transitions given by $p(x_{i+1}|x_i, y) = p(x_{i+1}, y_{i+1}|x_i, y^i)\beta^{i+1}(x_{i+1}, y)/\beta^i(x_i, y)$, where the $\beta^i(x_i, y)$ are computable with $\beta^n(x_n, y) = 1$, and $\beta^i(x_i, y) = \sum_{x_{i+1}} p(x_{i+1}, y_{i+1}|x_i, y^i)\beta^{i+1}(x_{i+1}, y)$ for $1 \leqslant i \leqslant n-1$. Therefore $p(x_{i+1}|x_i, y)$ are computable once $p(x_{i+1}, y_{i+1}|x_i, y^i) = p(x_{i+1}|x_i, y^i)p(y_{i+1}|x_i, x_{i+1}, y^i)$ are computable for each $1 \leqslant i \leqslant n-1$.

Let us show that $p(x_{i+1}, y_{i+1}|x_i, y^i)$ are computable for $p(y|x)$. Gaussian and $p(x_{i+1}|x_i, y^i) = p(x_{i+1}|x_i)$. We use the following classical property:

Property (P). *Let $W = (W_1, \dots, W_n)$ be a real Gaussian process, with $M^i = (M_1, \dots, M_i)$ the mean vector and $\Gamma^i = (\gamma_{kl})_{1 \leqslant k,l \leqslant i}$ the variance–covariance matrix (for each $1 \leqslant i \leqslant n$). Then its Gaussian distribution $p(y^n)$ is computable recursively, using the equality $p(y^i) = p(y^{i-1})p(y_i|y^{i-1})$ and the fact that $p(y_i|y^{i-1})$ is Gaussian of mean $M_i + (A^i)^T(\Gamma^{i-1})^{-1}(y^i - M^{i-1})$ and variance $\gamma_{ii} - (A^i)^T(\Gamma^{i-1})^{-1}A^i$, with $A^i = (\gamma_{1,i}, \dots, \gamma_{i-1,i})^T$ (the matrix $(\Gamma^{i-1})^{-1}$ in $p(y_i|y^{i-1})$ is given by $p(y^{i-1})$).*

As $p(x_{i+1}, y_{i+1}|x_i, y^i) = p(x_{i+1}|x_i)p(y_{i+1}|x_i, x_{i+1}, y^i)$, the crucial point is to apply Property (P) to $p(y_{i+1}|x_i, x_{i+1}, y^i)$. As there are K^2 possible (x_i, x_{i+1})(we have K classes), Property (P) is applied K^2 times, which gives, for $j = 1, \dots, K^2$, the corresponding M_j^i mean vector and Γ_j^i variance–covariance matrix.

Remark 1. The marginal posterior distributions $p(x_i|y)$, used in Bayesian 'Maximum Posterior Mode' restoration, can then be computed in three recursions: (i) one computes, using K^2 forward recursions, the transitions $p(x_{i+1}, y_{i+1}|x_i, y^i)$; (ii) one computes, using backward recursions, the quantities $\beta^i(x_i, y)$, which give the transitions $p(x_{i+1}|x_i, y)$ and $p(x_1|y)$; (iii) $p(x_1|y)$ and the transitions $p(x_{i+1}|x_i, y)$ classically give $p(x_i|y)$ with the forward recursions $p(x_{i+1}|y) = \sum_{x_i \in \Omega} p(x_{i+1}|x_i, y)p(x_i|y)$.

Let us now consider a Gaussian 'long memory' chain $Y = (Y_1, \dots, Y_n)$, which means that there exist four constants $\alpha_1 > 0, \alpha_2 > 0, 0.5 < \beta_1 < 1$, and $0.5 < \beta_2 < 1$ such that for each $1 \leqslant k < m, \alpha_1(m-k)^{-\beta_1} \leqslant \mathrm{Cov}(Y_k, Y_m) \leqslant \alpha_2(m-k)^{-\beta_2}$. Having K^2 distributions of such Gaussian long memory chains, we can use them, as above, to define 'partially' Markov chains, in which the hidden process is a Markov chain. One then can see that $p(y|x)$ is still, for each x, a Gaussian long memory chain. This means that it is possible to restore Markov chains hidden with long memory Gaussian noise.

Markov chains hidden with non Markovian noise with arbitrary margins

Let $W = (W_1, \ldots, W_n)$ be a real Gaussian process, and let $W^i = (W_1, \ldots, W_i)$ for each $1 \leqslant i \leqslant n$. We will denote by $G^1, \ldots, G^n$ the cumulative distribution functions (cdf) of the distributions of $W^1, \ldots, W^n$, and $g^1, \ldots, g^n$ the corresponding densities. The cdf and the densities of the marginals $W_1, \ldots, W_n$ will be denoted by $G_1, \ldots, G_n$, and $g_1, \ldots, g_n$, respectively. There exist, according to the Sklar theorem, functions $C^1, \ldots, C^n$, called "Gaussian copulas", from $R^1, \ldots, R^n$ to $[0, 1]$ such that for each $1 \leqslant i \leqslant n$, $G^i(y^i) = C^i(G_1(y_1), \ldots, G_i(y_i))$ (see [7]).

Otherwise, let $Y = (Y_1, \ldots, Y_n)$ be a real process, with $Y^i = (Y_1, \ldots, Y_i)$ for each $1 \leqslant i \leqslant n$ as above. The cdf functions of $Y_1, \ldots, Y_n$ will be denoted by $H_1, \ldots, H_n$, and their densities by $h_1, \ldots, h_n$. One can then show that $F^i(y^i) = C^i(H_1(y_1), \ldots, H_i(y_i))$ is a cdf admitting $H_1, \ldots, H_i$ as marginal cdf. This leads to $p(y^i) = h_1(y_1) \cdots h_i(y_i) c^i(H_1(y_1), \ldots, H_i(y_i))$, where c^i is the derivative of C^i. One then notices that this also is the distribution of $Y^i = (Y_1, \ldots, Y_i)$, with $Y_q = H_q^{-1}(G_q(W_q))$ for each $1 \leqslant q \leqslant i$. Putting $\varphi_j(y_j) = H_j^{-1}(G_j(y_j))$, we get $p(y^i) = \frac{g^i(\varphi_1^{-1}(y_1), \ldots, \varphi_i^{-1}(y_i))}{g_1(\varphi_1^{-1}(y_1)) \cdots g_i(\varphi_i^{-1}(y_i))} h_1(y_1) \cdots h_i(y_i)$. Knowing that $p(y_i | y^{i-1}) = \frac{p(y^i)}{p(y^{i-1})}$, we have (2) (in which "g" is used to denote Gaussian densities). As the Gaussian densities $g(y_i | y^i)$ are computable according to the previous section, $p(y_i | y^{i-1})$ given by (2) is computable once $h_i(y_i)$, $\varphi_i^{-1}(y_i)$, and $g_i(\varphi_i^{-1}(y_i))$ are computable. Even if the analytical formulas do not exist for all of them, approximations can be used, as has been done in [2] in a simpler model.

Finally, using the results of the previous section we can say that it is possible to recover the hidden signal X in the following case: (i) X is a Markov chain; (ii) $Z = (X, Y)$ is a partially Markov chain; and (iii) the cdf corresponding to the densities $p(y|x)$ are of the form $F^n(y^n) = C^n(H_1(y_1), \ldots, H_n(y_n))$ ($F^n(y^n)$ depends on x which is omitted), where C^n are Gaussian copulas, and $H_1, \ldots, H_n$ are cdf of marginals $p(y_1|x), \ldots, p(y_n|x)$, which can be chosen of any form.

Triplet partially Markov chains

The partially Markov chains given by the transitions (1) can be extended to 'triplet' partially Markov chains (TPMC), in which one introduces a third auxiliary random chain $U = (U_1, \ldots, U_n)$, where each U_i takes its values in a finite set Λ. Then one assumes that (1) is verified using $X^* = (U, X)$ instead of X. The restorations of X from Y are still possible; for instance, having computed $p(x_i^* | y) = p(x_i, u_i | y)$, one can compute $p(x_i | y)$ by summing with respect to u_i. Such a model is then a strictly more general model; in particular, if $p(x_{i+1}^* | x_i^*, y^i) = p(x_{i+1}^* | x_i^*)$ then $X^* = (U, X)$ is a Markov chain, but X is not necessarily a Markov chain.

Let us notice that TPMC can be used to manage the possible non stationarities of (X, Y); the study of a simpler case presented in [6] appears to indicate that managing non stationarities may be a promising application of TPMC.

1. Introduction

Soient $X = (X_1, \ldots, X_n, \ldots)$, $Y = (Y_1, \ldots, Y_n, \ldots)$ deux processus, X étant non observable et Y observé. Chaque X_i est à valeurs dans un ensemble fini $\Omega = \{\omega_1, \ldots, \omega_K\}$, et chaque Y_i est à valeurs dans R. Il est possible d'estimer $X = x$ à partir de Y lorsque la loi de X conditionnelle à Y, notée $p(x|y)$, est une loi de Markov. Cette dernière propriété est obtenue classiquement en supposant la markovianité de X et en prenant $p(y|x)$ (la loi de Y conditionnelle à X) suffisamment simple. En notant, pour simplifier, $X = (X_1, \ldots, X_n)$ et $Y = (Y_1, \ldots, Y_n)$, le modèle le plus classique, que nous appellerons « chaîne de Markov cachée avec bruit indépendant » (CMC-BI, avec « BI » pour « bruit indépendant », ce qui traduit l'indépendance des $Y_1, \ldots, Y_n$ conditionnellement à X), consiste à prendre $p(y|x) = p(y_1|x_1) \cdots p(y_n|x_n)$. Dans de nombreux cas il s'agit là d'une simplification relativement brutale de la réalité physique, qui peut s'avérer préjudiciable à la qualité de l'estimation de $X = x$. Afin d'y remédier, les modèles plus généraux, dits chaînes de Markov « Couple » (CMCouple) ont été proposés [8] et leur intérêt pratique en traitement du signal et des images s'est avéré non négligeable [5]. Dans ces derniers, on suppose directement la markovianité du couple $Z = (X, Y)$. Cela implique la markovianité de $p(y|x)$, ce qui est plus général et permet donc de modéliser cette loi de manière plus satisfaisante, et la markovianité de $p(x|y)$, ce qui permet l'estimation de $X = x$. Il est à noter que CMCouple n'est pas nécessairement une « chaîne de Markov cachée » (CMC) dans la mesure où le processus caché X n'est pas nécessairement de Markov. Cependant, la markovianité de $p(y|x)$ comporte des restrictions comme la décroissance exponentielle de la covariance (processus à « mémoire courte »), alors même que les processus « à mémoire longue » (la décroissance de la covariance plus lente) présentent un intérêt dans différentes situations [1,3,4]. Les chaînes « partiellement de Markov » permettent de lever cette restriction ; en effet, dans ces modèles $p(x|y)$ est de Markov, ce qui permet l'estimation de $X = x$, et $p(y|x)$ est quelconque [9].

L'objet de cette note est de montrer l'applicabilité de deux modèles, importants pour la pratique, et présenter leur généralisation aux chaînes de Markov triplet. Dans un premier temps, nous montrons que les transitions de la chaîne de Markov a posteriori $p(x|y)$ sont calculables dans une chaîne partiellement de Markov avec $p(y|x)$ gaussien quelconque, ce qui permet de traiter, en particulier, le cas des $p(y|x)$ à mémoire longue. Dans un deuxième temps, nous utilisons les copules gaussiennes afin de généraliser le modèle gaussien aux modèles à marginales quelconques. Enfin, ce dernier modèle est étendu aux chaînes triplet partiellement de Markov.

2. Chaînes de Markov cachées par bruit gaussien quelconque

Les chaînes couple partiellement de Markov (CCouplePM) $Z = (Z_1, \ldots, Z_n)$ vérifient

$$p(z_{i+1}|z^i) = p(z_{i+1}|x_i, y^i) = p(x_{i+1}|x_i, y^i)p(y_{i+1}|x_i, x_{i+1}, y^i) \tag{1}$$

avec $z^i = (z_1, \ldots, z_i)$ et $y^i = (y_1, \ldots, y_i)$. On retrouve le modèle classique CMC-BI pour $p(x_{i+1}|x_i, y^i) = p(x_{i+1}|x_i)$, $p(y_{i+1}|x_i, x_{i+1}, y^i) = p(y_{i+1}|x_{i+1})$, et le modèle CMCouple pour $p(x_{i+1}|x_i, y^i) = p(x_{i+1}|x_i, y_i)$, $p(y_{i+1}|x_i, x_{i+1}, y^i) = p(y_{i+1}|x_i, x_{i+1}, y_i)$. Par ailleurs, (1) est équivalent à la factorisation suivante de la loi de Z : $p(z_1, \ldots, z_n) = p(x_1, y_1)p(x_2, y_2|x_1, y^1) \cdots p(x_n, y_n|x_{n-1}, y^{n-1})$. Cette dernière factorisation permet de montrer classiquement que $p(x|y)$ est une chaîne de Markov dont les transitions sont données par $p(x_{i+1}|x_i, y) = p(x_{i+1}, y_{i+1}|x_i, y^i)\beta^{i+1}(x_{i+1}, y)/\beta^i(x_i, y)$, avec les $\beta^i(x_i, y)$ vérifiant les récusions rétrogrades $\beta^n(x_n, y) = 1$, et $\beta^i(x_i, y) = \sum_{x_{i+1}} p(x_{i+1}, y_{i+1}|x_i, y^i)\beta^{i+1}(x_{i+1}, y)$ pour $1 \leqslant i \leqslant n-1$. Ainsi ces transitions sont calculables dès que les transitions $p(x_{i+1}, y_{i+1}|x_i, y^i) = p(x_{i+1}|x_i, y^i)p(y_{i+1}|x_i, x_{i+1}, y^i)$ le sont pour tout $1 \leqslant i \leqslant n-1$. Montrons que c'est le cas dans une classe générale des modèles où $p(y|x)$ est gaussienne. Nous allons utiliser la propriété classique suivante :

Propriété (P). *Soit $W = (W_1, \ldots, W_n)$ un processus réel gaussien, avec, pour tout $1 \leqslant i \leqslant n$, $M^i = (M_1, \ldots, M_i)$ vecteur moyen et $\Gamma^i = (\gamma_{kl})_{1 \leqslant k,l \leqslant i}$ matrice de covariance de $W^i = (W_1, \ldots, W_i)$. Alors il est possible de calculer, pour tout i, la densité gaussienne $p(y^i)$ de paramètres M^i, Γ^i. Classiquement, on procède de proche en proche en utilisant l'identité $p(y^i) = p(y^{i-1})p(y_i|y^{i-1})$ et le fait que $p(y_i|y^{i-1})$ est gaussienne de moyenne $M_i + (A^i)^T(\Gamma^{i-1})^{-1}(y^{i-1} - M^{i-1})$ et de variance $\gamma_{ii} - (A^i)^T(\Gamma^{i-1})^{-1}A^i$, avec $A^i = (\gamma_{1,i}, \ldots, \gamma_{i-1,i})^T$ (la matrice $(\Gamma^{i-1})^{-1}$ figurant dans $p(y_i|y^{i-1})$ est donnée par la densité gaussienne $p(y^{i-1})$).*

Pour assurer la gaussianité de $p(y|x)$, considérons le cas particulier des transitions $p(x_{i+1}, y_{i+1}|x_i, y^i)$ de la forme $p(x_{i+1}, y_{i+1}|x_i, y^i) = p(x_{i+1}|x_i)p(y_{i+1}|x_i, x_{i+1}, y^i)$ (on peut remarquer que X est alors de Markov). La calculabilité de ces transitions est alors assurée par l'application de la Propriété (P) aux transitions $p(y_{i+1}|x_i, x_{i+1}, y^i)$, cela pour tous les couples (x_i, x_{i+1}).

Définition 2.1. Considérons K classes $\Omega = \{\omega_1, \ldots, \omega_K\}$, numérotons les K^2 couples de classes par $j = 1, \ldots, K^2$. Pour tout $j = 1, \ldots, K^2$ et tout $1 \leqslant i \leqslant n$, considérons M_j^i et Γ_j^i comme dans la Propriété (P). On appelle chaîne de Markov cachée gaussienne (CMCG) toute CCouplePM $Z = (X, Y)$ telle que $p(y|x)$ est gaussienne et dont la loi est donnée par $p(x_1, x_2, y_1, y_2)$, et, pour tout $i \geqslant 2$, les transitions $p(x_{i+1}, y_{i+1}|x_i, y^i) = p(x_{i+1}|x_i)p(y_{i+1}|x_i, x_{i+1}, y^i)$ avec pour $(x_i, x_{i+1}) = j$, la loi gaussienne $p(y_{i+1}|x_i, x_{i+1}, y^i)$ donnée par M_j^{i+1} et Γ_j^{i+1} (de la même manière que les transitions $p(y_i|y^{i-1})$ sont données par M^i et Γ^i dans la Propriété (P)).

Proposition 2.2. *Dans une CMCG $Z = (X, Y)$ les transitions $p(x_{i+1}, y_{i+1}|x_i, y^i) = p(x_{i+1}|x_i)p(y_{i+1}|x_i, x_{i+1}, y^i)$ sont calculables. En conséquence, les transitions $p(x_{i+1}|x_i, y)$ et les marginales $p(x_i|y)$ le sont également.*

Démonstration. Les transitions $p(x_{i+1}|x_i)$ étant données, il reste à montrer la calculabilité des transitions $p(y_{i+1}|x_i, x_{i+1}, y^i)$. Pour$(x_i, x_{i+1}) = j$, cette dernière s'obtient en utilisant la Propriété (P), ce qui termine la preuve. □

Remarque 1. La méthode bayésienne « Maximum des Marginales a Posteriori » (MPM) d'estimation de $X = x$ à partir de $Y = y$ requière le calcul des $p(x_i|y)$. En vertu de ce qui précède, ce calcul peut être fait dans le cas de CMCG en trois phases suivantes : (i) on calcule, par K^2 récursions progressives, les transitions $p(x_{i+1}, y_{i+1}|x_i, y^i)$ grâce à la Propriété (P) et la Proposition 2.2 ; (ii) on calcule, par des récursions rétrogrades, les $\beta^i(x_i, y)$ et on en déduit les transitions $p(x_{i+1}|x_i, y)$ et $p(x_1|y)$; (iii) ces dernières donnent classiquement $p(x_i|y)$ par les récursions progressives $p(x_{i+1}|y) = \sum_{x_i \in \Omega} p(x_{i+1}|x_i, y)p(x_i|y)$. Notons que le calcul (i), qui est une nouveauté par rapport aux modèles classiques, est adaptatif dans le sens où lorsque l'on souhaite tenir compte d'une nouvelle observation y_{n+1}, les résultats du calcul des K^2 récursions progressives jusqu'à n ne sont pas perdus et servent au calcul de la seule « $n + 1$ » ième récursion.

Définition 2.3. Un processus gaussien $Y = (Y_1, \ldots, Y_n, \ldots)$ sera dit « à mémoire longue » s'il existe des constantes $\alpha_1 > 0$, $\alpha_2 > 0$, $0{,}5 < \beta_1 < 1$, et $0{,}5 < \beta_2 < 1$ telles que pour tous $1 \leqslant k < m$, $\alpha_1(m-k)^{-\beta_1} \leqslant \mathrm{Cov}(Y_k, Y_m) \leqslant \alpha_2(m-k)^{-\beta_2}$.

Proposition 2.4. *Considérons K classes $\Omega = \{\omega_1, \ldots, \omega_K\}$, et K^2 suites $(\Gamma_j^i)_{1 \leqslant j \leqslant K^2}$ de matrices de covariance comme dans la Définition* 2.1. *Si toutes les Γ_j^i sont à mémoire longue, alors $p(y|x)$ qu'elles définissent l'est également.*

Démonstration. En suivant la Remarque 1, la covariance de (Y_k, Y_m) conditionnelle à x dépend de $j = (x_{m-1}, x_m)$ et est précisément le coefficient $\lambda_{km,j}$ de la matrice $\Gamma_j^m = (\lambda_{kl,j})_{k \leqslant m, l \leqslant m}$. Sachant que $\alpha_1^j(m-k)^{-\beta_1^j} \leqslant \lambda_{km,j}^m \leqslant \alpha_2^j(m-k)^{-\beta_2^j}$ et en prenant $\alpha_1 = \inf_j \alpha_1^j$, $\alpha_2 = \sup_j \alpha_2^j$, $\beta_1 = \sup_j \beta_1^j$, et $\beta_2 = \inf_j \beta_2^j$, on a $\alpha_1(m-k)^{-\beta_1} \leqslant \lambda_{km,j}^m \leqslant \alpha_2(m-k)^{-\beta_2}$, ce qui termine la preuve. □

Ce qui précède permet donc d'introduire, en particulier, un modèle de chaîne de Markov cachée par du « bruit » à mémoire longue, et de préciser une méthode d'estimation de cette chaîne.

3. Chaînes de Markov cachées par bruit non markovien à marginales quelconques

L'utilisation des copules permet d'étendre le résultat de la Proposition 2.2 au cas où les marginales ne sont pas nécessairement gaussiennes [7]. Soit $W = (W_1, \ldots, W_n)$ un processus réel gaussien, avec, pour tout $1 \leqslant i \leqslant n$, $M^i = (M_1, \ldots, M_i)$ vecteur moyen et $\Gamma^i = (\gamma_{kl})_{1 \leqslant k,l \leqslant i}$ matrice de covariance de $W^i = (W_1, \ldots, W_i)$. On notera $G^1, \ldots, G^n$ les fonctions de répartition des loi des $W^1, \ldots, W^n$, et $g^1, \ldots, g^n$ leurs densités. Les fonctions de répartition et les densités des lois des $W_1, \ldots, W_n$ seront notées respectivement $G_1, \ldots, G_n$, et $g_1, \ldots, g_n$. Il existe alors, selon le théorème de Sklar, des applications $C^1, \ldots, C^n$, appelées « copules gaussiennes », de $R^1, \ldots, R^n$ dans $[0, 1]$ telles que pour tout $1 \leqslant i \leqslant n$, $G^i(y^i) = C^i(G_1(y_1), \ldots, G_i(y_i))$. Par ailleurs, soient $Y^i = (Y_1, \ldots, Y_i)$ des variables aléatoires à valeurs dans R^i des lois marginales données par les fonctions de répartition $H_1, \ldots, H_i$, ou encore par les densités $h_1, \ldots, h_i$. On montre alors que $F^i(y^i) = C^i(H_1(y_1), \ldots, H_i(y_i))$ est la fonction de répartition d'une loi respectant les marginales définies par $H_1, \ldots, H_i$. En dérivant cette dernière équation, nous obtenons $p(y^i) = h_1(y_1) \cdots h_i(y_i)c^i(H_1(y_1), \ldots, H_i(y_i))$, où c^i est la dérivée de C^i. On note alors que c'est également la densité de la loi de $Y^i = (Y_1, \ldots, Y_i)$, avec $Y_q = H_q^{-1}(G_q(W_q))$ pour tout$1 \leqslant q \leqslant i$.

En posant $\varphi_j(y_j) = H_j^{-1}(G_j(y_j))$, nous avons alors $p(y^i) = \frac{g^i(\varphi_1^{-1}(y_1),\ldots,\varphi_i^{-1}(y_i))}{g_1(\varphi_1^{-1}(y_1))\cdots g_i(\varphi_i^{-1}(y_i))} h_1(y_1)\cdots h_i(y_i)$. Sachant que $p(y_i|y^{i-1}) = \frac{p(y^i)}{p(y^{i-1})}$, cela implique (avec « g » pour densité gaussienne) :

$$p(y_i|y^{i-1}) = g\big(\varphi_i^{-1}(y_i)|\varphi_1^{-1}(y_1),\ldots,\varphi_{i-1}^{-1}(y_{i-1})\big)\frac{h_i(y_i)}{g_i(\varphi_i^{-1}(y_i))}. \tag{2}$$

Sachant que les densités gaussiennes $g(y_i|y^i)$ sont calculables en vertu de la Proposition 2.2, $p(y_i|y^{i-1})$ donné par (2) est calculable dès que $h_i(y_i), \varphi_i^{-1}(y_i)$, et $g_i(\varphi_i^{-1}(y_i))$ le sont. Même si les formules analytiques de ces fonction ne peuvent être données, cette question peut être considérée, du point de vue des applications pratiques, comme peu gênante ; en effet, il est possible d'utiliser des approximations rapides sur ordinateur, comme cela a été fait dans [2] dans le cadre d'un modèle plus simple.

Ces résultats classiques permettent alors d'étendre les CCMG de la Proposition 2.2 aux modèles dans lesquels les marginales de $p(y|x)$ sont quelconques :

Proposition 3.1. *Soit une chaîne partiellement de Markov Z. Supposons que les transitions données par* (1) *sont telles que $p(x_{i+1}|x_i, y^i) = p(x_{i+1}|x_i)$ et, en notant $j = (x_i, x_{i+1})$, $p(y_{i+1}|x_i, x_{i+1}, y^i) = p(y_{i+1}|j, y^i) = g_j((\varphi_j^i)^{-1}(y^i)|(\varphi_j^{i-1})^{-1}(y^{i-1}))\frac{h_{j,i}(y_i)}{g_{j,i}(\varphi_{j,i}^{-1}(y_i))}$ (à $j = (x_i, x_{i+1})$ fixé, la transition $p(y_{i+1}|x_i, x_{i+1}, y^i)$ est de la forme* (2)*). On suppose que $h_{j,i}(y_i)$, $\varphi_{j,i}^{-1}(y_i)$, et $g_{j,i}(\varphi_{j,i}^{-1}(y_i))$ sont calculables ou approchables numériquement.*

Alors les transitions $p(y_{i+1}|x_i, x_{i+1}, y^i)$ sont calculables. En conséquence, il est possible d'estimer $X = x$ à partir de Y par différentes méthodes bayésiennes.

4. Chaînes triplet partiellement de Markov

Les chaînes partiellement de Markov données par les transitions (1) s'étendent aisément aux chaînes « triplet », dans lesquelles on introduit un processus latent $U = (U_1, \ldots, U_n, \ldots)$, chaque U_i prenant ses valeurs dans un ensemble fini Λ, et l'on suppose que (1) est vérifié avec $X^* = (U, X)$ à la place de X. La calculabilité de $p(x_i^*|y) = p(x_i, u_i|y)$ entraîne alors la calculabilité de $p(x_i|y)$ par la sommation sur u_i. Notons que le modèle ainsi obtenu est strictement plus général ; en particulier, si $p(x_{i+1}^*|x_i^*, y^i) = p(x_{i+1}^*|x_i^*)$ alors $X^* = (U, X)$ est une chaîne de Markov, mais X ne l'est pas nécessairement. Notons également que la gestion de la non stationnarité de (X, Y), dont l'intérêt est montré dans le cas simple où (X, Y) est une CMC-BI dans [6], semble être une application particulièrement prometteuse des chaînes triplet.

Références

[1] J. Beran, M.S. Taqqu, Statistics for Long-Memory Processes, Monographs Statist. Appl. Probab., Chapman and Hall, New York, 1994.
[2] N. Brunel, W. Pieczynski, Unsupervised signal restoration using copulas and pairwise Markov chains, IEEE Workshop on Statistical Signal Processing (SSP 2003), Saint Louis, Missouri, September 28–October 1, 2003.
[3] O. Cappé, E. Moulines, J.C. Pesquet, A. Petropulu, X. Yang, Long-range dependence and heavy-tail modeling for teletraffic data, IEEE Signal Processing Magazine 19 (3) (2002) 14–27.
[4] F. Chapeau-Blondeau, M. Guglielmi, Modèles de signaux à longue dépendance statistique, in : M. Guglielmi (Ed.), Signaux Aléatoires : Modélisation, Estimation, Détection, Traité IC2, Hermes, Paris, 2004, Chapitre 4.
[5] S. Derrode, W. Pieczynski, Signal and image segmentation using pairwise Markov chains, IEEE Trans. Signal Processing 52 (9) (2004) 2477–2489.
[6] P. Lanchantin, W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chain using evidential priors, IEEE Trans. Signal Processing 53 (8) (2005).
[7] R.B. Nelsen, An Introduction to Copulas, Lecture Notes in Statist., vol. 139, Springer, New York, 1998.
[8] W. Pieczynski, Pairwise Markov chains, IEEE Trans. Pattern Analysis and Machine Intelligence 25 (5) (2003) 634–639.
[9] W. Pieczynski, Triplet partially Markov chains and tree, in: 2nd International Symposium on Image/Video Communications over Fixed and Mobile Networks (ISIVC'04), Brest, France, July 2004.