



Unsupervised signal restoration using hidden Markov chains with copulas

Nicolas Brunel^a, Wojciech Pieczynski^{b,*}

^aUniversité Paris VI, LSTA, 175 rue du Chevaleret, 75013, Paris, France

^bGET/INT Département CITI, CNRS UMR 5157, 9, rue Charles Fourier, 91000 Evry, France

Received 24 June 2004; received in revised form 14 January 2005

Available online 10 May 2005

Abstract

This paper deals with the statistical restoration of hidden discrete signals, extending the classical methodology based on hidden Markov chains. The aim is to take into account the hidden signal and complex relationships between the noises which can be from different parametric models, non-independent, and of class-varying nature. We discuss some possibilities of managing it using copulas. Further, we propose a parameter estimation method and apply resulting unsupervised restoration methods in variety of situations. It is also validated by experiments performed in supervised and unsupervised context.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Hidden Markov chains; Pairwise Markov chains; Triplet Markov chains; Copulas; Parameter estimation; Bayesian restoration; Stochastic expectation-maximization; Statistical image segmentation

1. Introduction

The aim of this paper is to propose original techniques of hidden signal restorations that use simultaneously copulas and hidden Markov chains. Let $Y = (Y_1, \dots, Y_m)$ be a random vector taking its values in R^m . The theory of copulas

allows one to define the distribution $p(y)$ from given marginal distributions $p(y_1), \dots, p(y_m)$ and dependence structure [1]. Firstly introduced by Sklar [2], this theory is widely used in various fields like linguistics [3], biology [4], decision sciences [5], finance and insuring [6], or still psychology [7], among others. We only mentioned recent papers, each of which contain numerous references for these different fields of research. In a more theoretical context, copulas can also be useful to measure the dependence among random variables [2,8,38], and to construct a general model for

*Corresponding author. Tel.: +33 1 60 76 44 25;
fax: +33 1 60 76 44 33.

E-mail addresses: nicolas.brunel@int-evry.fr (N. Brunel),
Wojciech.Pieczynski@int-evry.fr (W. Pieczynski).

stationary Markov process, with non-linear dependence, as it is shown in [9,37], or in [10] for financial applications.

Otherwise, the hidden Markov chains model allows one to recover a hidden random vector $X = (X_1, \dots, X_n)$ from an observed one $Y = (Y_1, \dots, Y_n)$, even for a very large n . Since the seminal papers [11,12], hundreds of papers described different applications in various domains like image processing [13–15], biology [16], finance [17], among others. However, apart [18] there are no results, to our knowledge, concerning a simultaneous use of copulas and hidden Markov chains. Extending the first results presented in [18], the aim of this paper is to describe some possibilities of such a simultaneous use. In particular, the latter allows one to consider complex noise such that (i) its nature can vary with the class, and (ii) it is not independent conditionally on the hidden states. Indeed, instead of assuming that the observed process is the simplest stationary process with fixed marginals conditionally on the state process, we use recent advances in stationary Markov chain modeling based on copulas for specifying processes with prescribed marginals. The first point (i) can be seen as a generalization of ideas in [15], where the noise follows the usual assumption of conditional independence of the classical hidden Markov chain model. Concerning the second point (ii), the approach presented here is a different one, and can be seen as complementary, to the method studied in [19]. The main difference is that the present method does respect the given marginal distributions of the noise, when the previous method does not. We also present some numerical results which attest the interest of the new models, with potential application to the different domains mentioned above. The practical applications are made possible with an original parameter estimation method we propose, based on the general Expectation-Maximization (EM) principle. Finally, some possible extensions of the proposed models and techniques to pairwise and triplet Markov chains [20,21], which generalize hidden Markov chains, are also mentioned.

The paper is organized as follows. Hidden and pairwise Markov chains are recalled in Section 2

and the new models using copulas are described in Section 3. Section 4 is devoted to the parameter estimation problem. Experiments are described in Section 5 and the Section 6 contains some concluding remarks and perspectives.

2. Hidden and pairwise Markov chains

Let $X = (X_1, \dots, X_n)$ denote a stochastic process modeling a hidden discrete signal and let $Y = (Y_1, \dots, Y_n)$ be a stochastic process modeling the observations. We assume that each X_i takes its values in a finite set of classes $\Omega = \{1, \dots, k\}$, which will be assumed known through this paper, and each Y_i takes its values in the set of real numbers R . The problem is then to estimate $X = x$ from $Y = y$. Different Bayesian methods are then very useful tools, once both the processes X, Y are linked by the mean of some appropriate joint distribution $p(x, y)$. The hidden Markov chain with independent noise (HMC-IN) model, in which this distribution is written

$$p(x, y) = p(x_1)p(x_2|x_1) \dots p(x_n|x_{n-1}) \times p(y_1|x_1) \dots p(y_n|x_n) \quad (2.1)$$

is among the most widely used. The name ‘‘HMC’’ is due to the fact that the hidden process X is a Markov one. Further, given that (2.1) implies the independence of Y_1, \dots, Y_n conditionally on X , we will add ‘‘independent noise’’, which gives HMC-IN. However, in real situations the independence of Y_1, \dots, Y_n conditionally on X is often subject to criticism and taking into account possible non-independency undoubtedly improves the adequacy of the model. So, it is possible to propose, especially in the Gaussian noise case, hidden Markov chains (HMC) models in which both X and $Z = (X, Y)$ are Markov chains, but Y_1, \dots, Y_n are no longer independent conditionally on X . Further, the so-called pairwise Markov chains (PMC [21]), whose interest has been shown via experiments in [19], generalize HMC in that the hidden process X is no longer necessarily a Markov chain. Further generalization consists of introducing a latent process $U = (U_1, \dots, U_n)$ and assuming that the triplet $T = (X, U, Y)$ is a

Markov chain; it leads to the so-called triplet Markov chains (TMC [20,22]). It is then possible to show that TMC are strictly more general than PMC, in that the marginal distribution of $Z = (X, Y)$ in a TMC $T = (X, U, Y)$ is not necessarily a Markov chain distribution. So, we have four embedded models called HMC-IN, HMC, PMC, and TMC, and each of them allows us to estimate the hidden process X from the observed process Y .

Let us recall the pairwise Markov chain (PMC) model. Let $Z = (Z_1, \dots, Z_n)$, with $Z_1 = (X_1, Y_1), \dots, Z_n = (X_n, Y_n)$ as defined in the previous section. $Z = (Z_1, \dots, Z_n)$ is called pairwise Markov chain (PMC) simply when it is a Markov chain:

$$p(z) = p(z_1)p(z_2|z_1) \dots p(z_n|z_{n-1}). \tag{2.2}$$

One can then easily see that HMC-IN are PMC (with $p(z_1) = p(x_1)p(y_1|x_1)$ and $p(x_{i+1}, y_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1})$), but PMC are not necessarily HMC-IN. In fact, in PMC we have $p(x_{i+1}, y_{i+1}|x_i, y_i) = p(x_{i+1}|x_i, y_i)p(y_{i+1}|x_{i+1}, x_i, y_i)$ and so we can say that HMC-IN are particular PMC in which $p(x_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)$, $p(y_{i+1}|x_{i+1}, x_i, y_i) = p(y_{i+1}|x_{i+1})$. Otherwise, in PMC the process X is not necessarily a Markov one [21]. However, likely to HMC-IN, PMC can be used to estimate $X = x$ from $Y = y$ by different Bayesian methods and first results are encouraging [19]. In fact, considering the same “forward” probability $\alpha(x_i) = p(x_i, y_1, \dots, y_i)$, and the more general “backward” probability $\beta(x_i) = p(y_{i+1}, \dots, y_n | x_i, y_i)$, we have analogous recursions

$$\begin{aligned} \alpha(x_1) &= p(x_1)p(y_1|x_1), \text{ and} \\ \alpha(x_{i+1}) &= \sum_{x_i \in \Omega} \alpha(x_i)p(z_{i+1}|z_i) \text{ for } 1 \leq i \leq N - 1, \end{aligned} \tag{2.3}$$

and

$$\begin{aligned} \beta(x_N) &= 1, \text{ and} \\ \beta(x_i) &= \sum_{x_{i+1} \in \Omega} \beta(x_{i+1})p(z_{i+1}|z_i), \\ &\text{for } 1 \leq i \leq N - 1. \end{aligned} \tag{2.4}$$

As for each $1 \leq i \leq n$, $p(x_i|y_1, \dots, y_n) \propto \alpha(x_i)\beta(x_i)$, the latter can be calculated and thus the Bayesian MPM method given by $\hat{s}_{MPM}(y_1, \dots, y_n) =$

$(\hat{x}_1, \dots, \hat{x}_n)$, with

$$\hat{x}_i = \arg \max_{x_i \in \Omega} p(x_i|y_1, \dots, y_n) \tag{2.5}$$

can also be calculated.

Otherwise, we have

$$p(x_i, x_{i+1}|y_1, \dots, y_n) \propto \alpha(x_i)p(z_{i+1}|z_i)\beta(x_{i+1}) \tag{2.6}$$

which gives $p(x_{i+1}|x_i, y_1, \dots, y_n)$ and thus allows one to simulate X according to its posterior distribution $p(x|y)$.

Concerning the particular TMC, that we deal within this paper and in which each U_i takes its values in a finite set $A = \{\lambda_1, \dots, \lambda_m\}$, they are not much more complicated than PMC. In fact, as $T = (X, U, Y)$ is a Markov chain, we can put $V = (X, U)$ and say that $T = (V, Y)$ is a PMC (with each v_i taking its values in $\Omega \times A$) and all we said above can be applied. In particular, as $T = (V, Y)$ is a PMC, $p(v_i|y_1, \dots, y_n) \propto \alpha(v_i)\beta(v_i)$ is calculable for each $1 \leq i \leq n$, and thus one can also calculate $p(x_i|y_1, \dots, y_n) = \sum_{j=1}^m p((x_i, u_i = \lambda_j)|y_1, \dots, y_n)$.

The latter makes possible the use of the Bayesian MPM method given by (2.5).

The difference between HMC and PMC, which are not HMC and specifies the interest of the latter with respect to the former, is described in the following proposition.

Proposition 2.1. *Let Z be a PMC verifying: (a) $p(z_i, z_{i+1})$ does not depend on $1 \leq i \leq n - 1$; and (b) $p(z_i, z_{i+1}) = p(z_{i+1}, z_i)$. Then the three following conditions*

- (i) Z is a HMC;
- (ii) for each $2 \leq i \leq n$, $p(y_i|x_i, x_{i-1}) = p(y_i|x_i)$;
- (iii) for each $1 \leq i \leq n$, $p(y_i|x) = p(y_i|x_i)$.

are equivalent.

The proof of the equivalence between (i) and (ii) is given in [21], and the proof of the equivalence between (i) and (iii) is given in [22].

In a similar way, we can specify the conditions under which a TMC is a PMC.

Proposition 2.2. *Let T be a TMC verifying: (a) $p(t_i, t_{i+1})$ does not depend on $1 \leq i \leq n - 1$; and (b)*

$p(t_i, t_{i+1}) = p(t_{i+1}, t_i)$. Then the three following conditions

- (i) T is a PMC;
- (ii) for each $p(t_i, t_{i+1}) = p(t_{i+1}, t_i)$,
 $p(u_i|z_i, z_{i-1}) = p(u_i|z_i)$;
- (iii) for each $1 \leq i \leq n$, $p(u_i|z) = p(u_i|z_i)$.

are equivalent.

The proof, analogous to the proof of the Proposition 2.1, is omitted.

Remark 2.1. PMC are more general than HMC, but they are not necessarily more complex and the same is true concerning the TMC. For example, considering a Markov chain $Y = (Y_1, \dots, Y_n)$ and defining the distribution of $X = (X_1, \dots, X_n)$ conditionally on $Y = (Y_1, \dots, Y_n)$ by $p(x|y) = p(x_1|y_1) \dots p(x_n|y_n)$ we have a PMC which is as simple as a HMC-IN, but which is not a HMC. In an analogous manner, let us consider a TMC, where $U = (U_1, \dots, U_n)$ is a Markov chain and where the distribution of (X, Y) conditionally on $U = (U_1, \dots, U_n)$ is given by $p(x, y|u) = p(x_1|u_1)p(y_1|u_1) \dots p(x_n|u_n)p(y_n|u_n)$. We have a TMC, which is neither PMC nor HMC, and which is as simple as a HMC-IN.

3. Copulas in HMC, PMC, and TMC

3.1. Copulas

As said above, copulas allow one to define a joint distribution from a pair formed by the marginal distributions and a dependence structure between the different marginal. The main result of the theory of copulas, which we will state in the case of two variables needed in this paper, is the following. Let $h(y_1, y_2)$ be a probability density on R^2 , H the associated cdf function, $h_1(y_1)$ and $h_2(y_2)$ the marginal densities, and H_1, H_2 the cdf functions associated with them. Then there exists a function C defined on $[0, 1]^2$ such that

$$H(y_1, y_2) = C(H_1(y_1), H_2(y_2)). \tag{3.1}$$

Taking the derivative of (3.1) with respect to y_1, y_2 and introducing $c(u, v) = \partial \partial C(u, v) / \partial u \partial v$,

we have

$$h(y_1, y_2) = h_1(y_1)h_2(y_2)c(H_1(y_1), H_2(y_2)). \tag{3.2}$$

Conversely, having H_1, H_2 and a copula C , one can use (3.1) to define H . So, a given H cdf on R^2 defines a copula C with (3.2), and this copula can also be used to define any another H' from any another H'_1, H'_2 .

Example 3.1. As an example, let us consider the Gaussian copulas. Let $h(y_1, y_2)$ be the density of a Gaussian vector with correlation ρ and marginal distributions having null means and variances equal to one. (3.1) and (3.2) then define a ‘‘Gaussian bivariate copula’’ by

$$c(u, v) = \frac{h(H_1^{-1}(u), H_2^{-1}(v))}{h_1(H_1^{-1}(u))h_2(H_2^{-1}(v))}. \tag{3.3}$$

So, a bivariate Gaussian copula is defined by just one parameter, which is the correlation ρ .

Finally, let Y_1, Y_2 be two real random variables with correlation ρ , and with cdf F_1, F_2 , and f_1, f_2 , the corresponding densities. Using (3.3), we can define a density for the distribution of (Y_1, Y_2) by

$$f(y_1, y_2) = f_1(y_1)f_2(y_2) \times \frac{h(H_1^{-1}(F_1(y_1)), H_2^{-1}(F_2(y_2)))}{h_1(H_1^{-1}(F_1(y_1)))h_2(H_2^{-1}(F_2(y_2)))}. \tag{3.4}$$

The interesting point is that the marginal distributions defined by (3.4) are F_1, F_2 . When (Y_1, Y_2) is Gaussian, (3.4) gives again its distribution.

3.2. Copulas in HMC

Returning to the situation described in the first Section, let $X = (X_1, \dots, X_n)$ denote a stochastic process modeling a hidden discrete signal and let $Y = (Y_1, \dots, Y_n)$ be a stochastic process modeling the observations. We assume that each X_i takes its values in a finite set of classes $\Omega = \{1, \dots, k\}$, and each Y_i takes its values in the set of real numbers R . Finally, let $Z = (Z_1, \dots, Z_n)$ be the pairwise process, with $Z_i = (X_i, Y_i)$.

Let $Z = (Z_1, \dots, Z_n)$ be a PMC verifying the hypotheses (a) and (b) of the Proposition 2.1 above. Furthermore, we will assume that Z is a

HMC which means, according to the Proposition 2.1, that for each $1 \leq i \leq n$, $p(y_i|x) = p(y_i|x_i)$. As the distribution of $Z = (Z_i, Z_{i+1})$ is independent of $1 \leq i \leq n-1$, we can say that the distribution of (Z_1, Z_2) defines the whole distribution of Z . As we have $p(z_1, z_2) = p(x_1, y_1, x_2, y_2) = p(x_1, x_2) \times p(y_1, y_2|x_1, x_2)$, we may say that the whole distribution of Z is defined by a probability $p(x_1, x_2)$ on $\Omega^2 = \{1, \dots, k\}^2$, and k^2 distributions $p(y_1, y_2|x_1, x_2)$ on R^2 . Further, concerning the marginal distributions of $p(y_1, y_2|x_1, x_2)$ we have, according to the Proposition 2.1, $p(y_1|x_1, x_2) = p(y_1|x_1)$ and $p(y_2|x_1, x_2) = p(y_2|x_2)$. Now, let us assume that these k marginal distributions are possibly not all equal; in other words, $p(y_1|x_1 = i)$ can vary with $i \in \Omega$. Further, $p(y_1, y_2|x_1, x_2) \neq p(y_1|x_1)p(y_2|x_2)$, which means that we have a HMC with correlated noise. Finally, we have k marginal distributions $p(y_1|x_1 = i) = p(y_2|x_2 = i)$, and $k(k-1)/2$ correlation coefficients ρ_{ij} and we see that it is possible to use $k(k-1)/2$ copulas to define k^2 distributions of $p(y_1, y_2|x_1, x_2)$. For example, if we take the Gaussian copula for all of them, we have a HMC with correlated noise compatible with the Gaussian HMC in that when all the marginal distributions are Gaussian, the HMC defined with copulas is Gaussian.

Remark 3.1. Let us notice the possible variation of the margin with the class is not a purely mathematical point of view; in fact, it can occur in very practical problems. For instance, in a radar image different classes can produce different shapes of densities [23].

3.3. Copulas in PMC

Let $Z = (Z_1, \dots, Z_n)$ be a PMC verifying the hypotheses (a) and (b) of the Proposition 2.1, and let us assume that Z is not a HMC. As mentioned above, the distribution of (Z_1, Z_2) , which can be written $p(z_1, z_2) = p(x_1, y_1, x_2, y_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$, defines the whole distribution of Z . So, according to Proposition 2.1, each of the marginal distributions $p(y_1|x_1, x_2)$ and $p(y_2|x_1, x_2)$ of the distribution $p(y_1, y_2|x_1, x_2)$ depends on both x_1 and x_2 . To simplify notations, let

us put $p_{ij}(y_1, y_2) = p(y_1, y_2|x_1 = i, x_2 = j)$. Let us assume that for each $1 \leq i, j \leq k$, the density $p_{ij}(y_1) = p(y_1|x_1 = i, x_2 = j)$ (which is equal to $p(y_2|x_1 = j, x_2 = i)$) and the density $p_{ij}(y_2) = p(y_2|x_1 = i, x_2 = j)$ (which is equal to $p(y_1|x_1 = j, x_2 = i)$) are known and their Gaussian copula is indexed by a known correlation coefficient ρ_{ij} . Of course, in the Gaussian case this gives $p_{ij}(y_1, y_2)$ but in a general case different $p_{ij}(y_1, y_2)$ define the same $p_{ij}(y_1), p_{ij}(y_2)$, and ρ_{ij} .

Finally, by the stationarity (a) and reversibility (b) assumptions, we have k^2 marginal distributions $p_{ij}(y_1)$ (which are the same that the k^2 marginal distributions $p_{ij}(y_2)$, with $p_{ij}(y_2) = p_{ji}(y_1)$) and only $k(k-1)/2$ correlation coefficients ρ_{ij} . As in the HMC above, it is then possible to use $k(k-1)/2$ copulas to define the k^2 distributions $p(y_1, y_2|x_1, x_2)$.

Remark 3.2. We see that PMC do not need much more parameters than HMC; however, they are strictly more general. This greater generality can be seen at the transitions level: as the condition (ii) in Proposition 2.1 is equivalent to $p(x_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)$, the general formulation $p(z_{i+1}|z_i) = p(x_{i+1}, y_{i+1}|x_i, y_i) = p(x_{i+1}|x_i, y_i)p(y_{i+1}|x_{i+1}, x_i, y_i)$ in PMC becomes $p(z_{i+1}|z_i) = p(x_{i+1}|x_i)p(y_{i+1}|x_{i+1}, x_i, y_i)$ in HMC.

Remark 3.3. Let us remark that methods proposed in [19] solve the same problem, which is to take into account the dependence of Y_i, Y_{i+1} conditional on (X_i, X_{i+1}) in Markov chain and non-Gaussian contexts. The main difference is that in the method proposed here the desirable forms of the marginals of the conditional distributions $p(y_i|x_i)$ are controlled, which is not the case in [19]. However, even if the corresponding model is less general than the copula-based one, it furnishes an interesting alternative when we have no knowledge on the shapes of the margins or of the copula. Let us also mention that the method in [19] is inspired by the “multisensor model” proposed in [24] in which Y_i, Y_{i+1} are independent conditionally on (X_i, X_{i+1}) , but there are two (or more) sensors $Y_i = (Y_i^1, Y_i^2)$ and Y_i^1, Y_i^2 are dependent conditionally on X_i . In a similar—though

inverse—manner, the method proposed in this paper can inspire methods to deal with correlated sensors, which is described in [25]. Otherwise, the problem of searching the forms of $p(y_i|x_i)$, which possibly vary with x_i , is processed in [15] in multisensor simple case where Y_i, Y_{i+1} are independent conditionally on (X_i, X_{i+1}) , and Y_i^1, Y_i^2 are independent conditionally on X_i .

3.4. Copulas in TMC

The extension of the copulas in PMC to TMC does not pose particular problems. As $T = (X, U, Y)$ is a Markov chain, putting $V = (X, U)$ we can say that (V, Y) is a PMC and so the previous subsection is applicable. The difference is that each $V_i = (X_i, U_i)$ takes its values in $\Omega \times A$, with $A = \{\lambda_1, \dots, \lambda_m\}$. So, k is replaced with km . We have then two cases: either $V = (X, U)$ is a Markov chain and we apply the results of Section 3.2, or it is not, and we apply the results of Section 3.3.

4. Parameter estimation

The aim of this section is to propose a method of estimation of all the parameters θ from $Y = y$. More precisely, we focus on the construction of a sequential search of a stationary point of the observed log-likelihood in order to compute the Maximum Likelihood Estimator (MLE) of the parameters. The consistency of the MLE has been assessed under general conditions for the HMC-IN model in [26] and only very recently for the switching AR model, somewhat similar to PMC, in [27]. Nevertheless, the main problem in the application of the MLE approach still remains in finding the good root of the log-likelihood. The Expectation-Maximization (EM) algorithm is the main device to reach this purpose (among others [36]), and hundreds of papers have shown its good behavior for mixture models and HMC-IN ones; see [28] and references therein. The latter, having started from an initial value θ^0 , produces a sequence of parameters according to the dynamic $\theta^{q+1} = \arg_{\theta} \max E_{\theta^q}[\text{Log}(p_{\theta}(X, Y)|Y = y)]$. However, in order to reduce the sensibility of EM to

initial conditions (and to avoid cumbersome computations in the M step) it can be interesting to use a “stochastic” EM (SEM [29]), whose principle is

- (i) simulate $X = x^q$ according to $p_{\theta^q}(x|y)$;
- (i) put $\theta^{q+1} = \hat{\theta}(x^q, y)$, where $\hat{\theta} = \hat{\theta}(X, Y)$ is the MLE of θ from complete data (X, Y) .

Fruitful alternatives for parameter estimation can be found by the application of Data Augmentation [30] or more general MCMC algorithms [31,32] that enables to estimate the a posteriori laws of the parameters and could possibly give better results. However, the designing of the suitable Markov chains might become technically more involved due to the apparition of the cdf of the observations, and the presence of copulas that couples the a posteriori laws of the different parameters.

Finally, when considering HMC we have to find, in the case of Gaussian copulas considered, k densities $p_i(y_1)$, $k(k-1)/2$ correlation coefficients ρ_{ij} , and the distribution $p(x_1, x_2)$ on $\Omega^2 = \{1, \dots, k\}^2$. When considering PMC which are not HMC, we have to find, in the case of Gaussian copulas considered, k^2 densities $p_{ij}(y_1)$, $k(k-1)/2$ correlation coefficients ρ_{ij} , and the distribution $p(x_1, x_2)$ on $\Omega^2 = \{1, \dots, k\}^2$.

As in both HMC and PMC cases, X can be simulated according to $p(x|y)$ (see Section 2), all we have to do is to maximize the completed likelihood $\text{Log}(p_{\theta}(X, Y))$. Unfortunately, it is difficult to reach the maximum of likelihoods involving copulas because the parameters of the margins and of the copulas are intricate. The extension of the SEM proposed in this paper lies in the way to reach this maximum. The idea is to consider the estimation of the margins and of the copulas as two separate problems, and to first estimate the parameters of the margins, and then the copula coefficients, following [9].

Finally, the parameters $p(x_1, x_2)$ can be estimated by the empirical estimate:

$$\hat{p}(i, j) = \frac{1_{[x_1=i, x_2=j]} + \dots + 1_{[x_{n-1}=i, x_n=j]}}{n-1} \tag{4.1}$$

Then, we select from the sample $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ the observations $y^{ij} = (y_r)$ such that $x_r = i$ and $x_{r+1} = j$. The estimation of $p_{ij}(y_1)$ is obtained by Maximum Likelihood from the sub-sample y^{ij} .

The parameters ρ_{ij} can be estimated by the empirical correlation coefficient of the transformed data $\Phi^{-1}(F_{ij}(y^{ij}))$, where Φ is the cdf of a standard normal and F_{ij} is the cdf of $p_{ij}(y_1)$, with the newly estimated parameters. This estimator corresponds to the MLE of a Gaussian copula, when the margins are known and equal to F_{ij} . We recover the fact that if X and Y are two random variables with respective cdf F_X and F_Y linked by a Gaussian copula indexed by a parameter ρ then ρ is the covariance of the transformed variables $\Phi^{-1}(F_X(X))$ and $\Phi^{-1}(F_Y(Y))$.

Remark 4.1. The above two-step procedure has its counterpart in the frame of the deterministic EM, where it is called expectation-conditional maximization (ECM introduced by Meng and Rubin, see chapter 5 in [33]). After an E-step, the maximization is performed on smaller groups of parameters, by considering the other parameters fixed (that is Conditional Maximization). As a consequence, simpler, faster and more stable algorithms can be obtained. It was also shown that under wide conditions, convergence to a stationary point is ensured. We could expect to have the same properties for its stochastic version.

5. Experiments

5.1. Hidden Markov chains

We present the case study of HMC based on gamma laws. The aim of these experiments is to show on one hand the influence of wider dependencies in HMC than in HMC-IN, and on the other hand the usefulness of copulas for modeling.

In the first part, we present results on chains and 1D process in order to compare HMC-IN and HMC. In the second part, we show some results on an image, and the ability of the 2 models to mimic a 2D process.

5.1.1. Chains

We use two chains having the same transition matrix A and also the same gamma margins. One is a HMC, with dependence described by a Gaussian copula indexed by a correlation coefficient ρ , so that the classical HMC-IN is a submodel obtained with $\rho = 0$. The chains have two classes, such that $p(y|x = 1)$ is a $\Gamma(1, 0.5)$ and $p(y|x = 2)$ is a $\Gamma(2, 1)$ and the transitions are the following: $P(x_{n+1} = 1|x_n = 1) = P(x_{n+1} = 2|x_n = 2) = 0.8$. We recall that a variable follows a gamma law $\Gamma(a, b)$ if the density is written:

$$f(y; a, b) = (b^a \Gamma(a))^{-1} y^{a-1} \exp(-y/b) 1_{(y \geq 0)} \quad (5.1)$$

where a is a shape parameter and b is a scale one. When a is one, we have the classical exponential law. We stress on the interest of this model, because it is quite difficult to define a bivariate gamma law with gamma margins, although the gamma law is very used for modeling positive random variables. We show in Fig. 1 two examples of such bivariate densities; one can remark very different shapes but the same margins.

We compare first the error rate of the restoration of a HMC by the true model and by the corresponding HMC-IN in a supervised context. We have simulated 100 chains of 1000 observations, and we compute the mean error rates for increasing correlation coefficients. The HMC considered is quite simple: we assume $\rho_{11} = \rho_{22} = \rho$ and $\rho_{12} = \rho_{21} = 0$ (Table 1).

For low correlation (< 0.5) results of restoration are nearly the same, the forgetting of the copula in HMC-IN becomes visible when the correlation is high. The error rate seems to have reached a maximum value for a HMC in this latter case, whereas the error goes on increasing for HMC-IN (by about 2% each time).

We show also in the unsupervised case, that the contribution of the copula is non-negligible or the restoration when we have high correlation (Table 2).

In this case, the superiority of the use of copulas is not as clear as in the supervised case, because we can see that the error rates are quite similar for both models. The HMC-IN performs even better for low correlations, and there is an inversion from 0.5. It is probably due to the presence of more

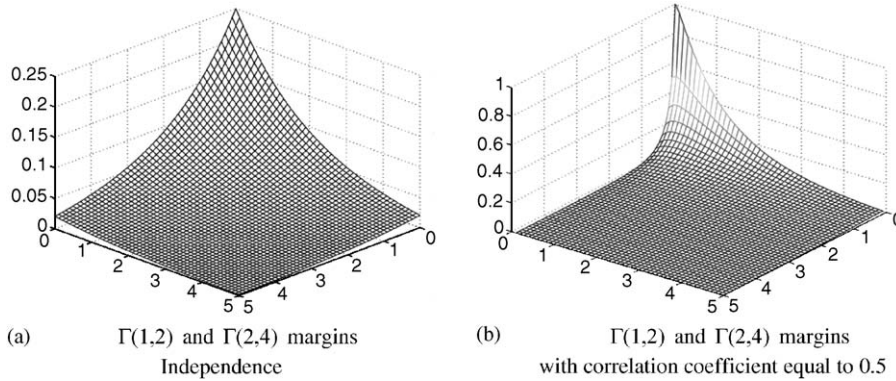


Fig. 1. Bivariate gamma laws with same margins but different copula.

Table 1
Error rates (%) for restoration of HMC using the true parameters

Correlation	HMC-IN	HMC
0		13.4
0.2	15.8	15.6
0.4	17.7	16.8
0.6	19.5	17.5
0.8	22	17.5

Table 2
Error rates (%) for restoration of HMC using estimated parameters

Correlation	HMC-IN	HMC
0	14.4	14.8
0.2	16.3	17
0.4	18.7	18.9
0.6	20.4	20.4
0.8	23.6	21.8

parameters to estimate in HMC, which make the variance of estimation increase (see Table 3). Consequently, the estimation of the posterior probabilities used for MPM segmentation deteriorates.

The main difference in the unsupervised context is about the estimation of the parameters: the estimators provided by an HMC-IN-based procedure are biased, because it does not use the good

Table 3
Variances of the estimators of the margins for HMC and HMC-IN, while the correlation is increasing

Correlation	HMC-IN		HMC		
	Class	$\sqrt{V(\hat{a})}$	$\sqrt{V(\hat{b})}$	$\sqrt{V(\hat{a})}$	$\sqrt{V(\hat{b})}$
0	1	0.06	0.06	0.06	0.09
	2	0.23	0.07	0.23	0.08
0.2	1	0.08	0.08	0.08	0.11
	2	0.27	0.09	0.37	0.12
0.4	1	0.08	0.07	0.09	0.12
	2	0.38	0.09	0.47	0.14
0.6	1	0.1	0.07	0.12	0.17
	2	0.47	0.09	0.53	0.17
0.8	1	0.14	0.1	0.14	0.3
	2	0.76	0.11	0.67	0.26

likelihood. Hence, if we want to compare the estimators in each case, we have to use the Mean Square Error (MSE) instead of the variance, which is defined by

$$MSE = E[(\hat{\theta} - \theta_0)^2] = E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta_0)^2 \tag{5.2}$$

As for the variances, we can compare the MSE of the estimator of the gamma margins for different correlation in Table 4. For low correlation, the MSE are close to each other in both HMC-IN and HMC cases, but the difference is

Table 4

MSE of the estimators of the margins for HMC and HMC-IN, while the correlation is increasing

Correlation	HMC-IN			HMC	
	Class	MSE(\hat{a})	MSE(\hat{b})	MSE(\hat{a})	MSE(\hat{b})
0	1	0.0045	0.0045	0.0046	0.009
	2	0.051	0.006	0.054	0.007
0.2	1	0.0075	0.007	0.0075	0.013
	2	0.14	0.012	0.15	0.014
0.4	1	0.0134	0.008	0.083	0.015
	2	0.37	0.026	0.25	0.02
0.6	1	0.025	0.014	0.014	0.033
	2	0.64	0.04	0.32	0.03
0.8	1	0.05	0.026	0.02	0.1
	2	1.66	0.073	0.46	0.07

increasing when the correlation grows. We can also remark that the parameters are very badly estimated by HMC-IN. We have the same behavior for the estimation of the transition probabilities: the probabilities $p(x_{n+1} = 1|x_n = 1) = p(x_{n+1} = 2)$ are over-estimated when the correlation is increased.

We wish to stress on the equivalence of the behavior of the two models for estimation and restoration when the correlation is equal to 0. Despite the estimation of a useless parameter, the error rates are the same and mean estimation for the correlations are $\hat{\rho}_{11} = 0.016$ and $\hat{\rho}_{22} = -0.048$ (the other correlations were forced to 0). We are then able to identify a HMC-IN because null correlation is equivalent to independence with a Gaussian copula.

5.1.2. Images

In statistical image analysis, the use of a Peano path enables to transform a 2D process in a 1D, and to use simpler model for segmentation than spatial models as Hidden Markov Fields (HMF) [34]. We show here a simulation of a particular Hidden Markov Field, with non-independent noise. We generate a three-classes isotropic Markov field $X = (X_s)_{s \in S}$ (128×128 pixels), such that $P_X(x) \propto \exp(-\alpha \sum_{(s,t)} \varphi(x_s, x_t))$ with $\alpha = 2$ (φ is -1 if $x_s = x_t$, $+1$ otherwise). We use a neighborhood

Table 5

Estimated parameters for the gamma margins

Class	HMC		HMC-IN	
	\hat{a}	\hat{b}	\hat{a}	\hat{b}
1	0.98	0.48	1.1	0.35
2	1.99	0.97	2.6	0.77
3	3	1.9	4.45	1.46

based on the four nearest neighbors. We generate a correlated Gaussian field $G = (G_s)_{s \in S}$ of the same size of X , such that the law of G_s is centered and with unit variance. Then, by applying the transformation $F_{X_s}^{-1}(\Phi(\cdot))$ to each G_s (with Φ the cdf of the centered normal law with unit variance), we construct Y_s having a cdf F_{X_s} depending on X_s . The same kinds of “cdf transformations” were used in [35] to generate stationary series with prescribed margins. The 2D process $Y = (Y_s)_{s \in S}$ is no more the classical HMF because we do not have the conditional independence assumption. For this experiments, we still use gamma laws: $p(y_s|x_s = 1)$ is a $\Gamma(1, 0.5)$, $p(y_s|x_s = 2)$ is a $\Gamma(2, 1)$, and $p(y_s|x_s = 3)$ is a $\Gamma(3, 2)$.

We obtain the following estimations for the marginal laws (Table 5):

The estimated correlation coefficients for the Gaussian copula are: $\hat{\rho}_{11} = 0.6$, $\hat{\rho}_{12} = 0.45$, $\hat{\rho}_{13} = 0.59$, $\hat{\rho}_{21} = 0.5$, $\hat{\rho}_{22} = 0.5$, $\hat{\rho}_{23} = 0.43$, $\hat{\rho}_{31} = 0.66$, $\hat{\rho}_{32} = 0.58$, $\hat{\rho}_{33} = 0.6$. We remark that the correlation coefficients were not forced to verify equality $\rho_{ij} = \rho_{ji}$ during the estimation; nevertheless the estimations are near, so that the assumption (a) and (b) for the HMC seems to be fair.

Finally, the error rate for the HMC-IN is 22.9%, and only 15.4% for the HMC with copulas (Fig. 2).

As we can see, the estimation of the margins provided by HMC-IN is still biased, and the restored image is less smooth than the original one. When the image is restored by a HMC, the estimated copulas enable to propose a more likely one, with less transitions and small areas within big ones. Moreover the margins are very well estimated.

5.2. Pairwise Markov chains

We compare in this part the general PMC and HMC-IN. This comparison is more difficult than above because the distributions $p_{ij}(y)$ have no equivalent in a HMC-IN model. In fact, the example we deal with will show that PMC can

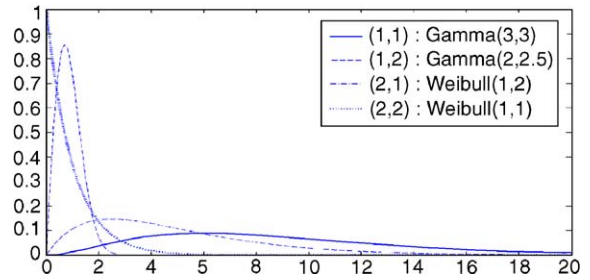


Fig. 3. Distributions $p_{11}(y_1), p_{12}(y_1), p_{22}(y_2), p_{12}(y_2)$ denoted by (1, 1), (1, 2), (2, 1), (2, 2).

Table 6
Error ratios obtained with the Bayesian method MPM

Case	PMC and copulas model			HMC model	
	(a)	(b)	(c)	(b)	(c)
1	5.31%	5.25%	6.44%	7.70%	12.30%
2	2.51%	2.60%	3.37%	7.70%	11.90%

(a): Real parameters, (b): Parameters estimated from (x, y) , (c) : Parameters estimated from y with SEM. Case 1: $\rho_{11} = \rho_{22} = 0.1$, $\rho_{12} = \rho_{21} = 0.5$, Case 2: $\rho_{11} = \rho_{22} = 0.5$, $\rho_{12} = \rho_{21} = 0.8$. Sample size $n = 500$.

describe far more complex phenomenon than a HMC-IN (or HMC).

Let us consider the case of two classes, with the two distributions $p_{11}(y_1), p_{12}(y_1)$ being $\Gamma(3, 3)$ and $\Gamma(2, 2.5)$, and the two distributions $p_{22}(y_2), p_{12}(y_2)$ being Weibull(1,1), and Weibull(2,1), respectively (see their form in Fig. 3). The joint probabilities $p(i, j)$ are $p(1, 1) = 0.4$, $p(1, 2) = p(2, 1) = 0.15$, $p(0, 0) = 0.3$. In HMC-IN, we assume that $p(y_1|1)$ is Gamma and $p(y_1|2)$ is Weibull. According to the results presented in Table 6, we see that when data come from a PMC with Gaussian Copula model, the supervised and unsupervised Bayesian MPM methods based on this model can be much more interesting than the same supervised and unsupervised Bayesian MPM methods based on the classical HMC models. Of course, this is not surprising when the true parameters are used because of the very Bayesian theory; however, it remains when the parameters are estimated, which is interesting for real applications.

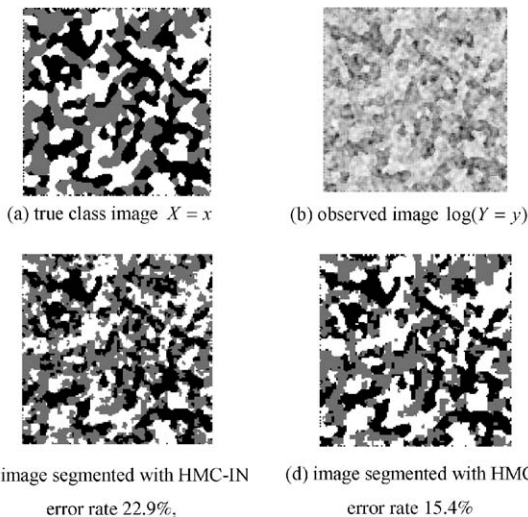


Fig. 2. The true image and the restored ones. Image (b) is the logarithm of true data in order to have a better contrast.

6. Conclusions

Hidden Markov Chains with independent noise processes are widely used in different problems and their success is due to the favorable behavior of different associated Bayesian restoration techniques. This is true even in unsupervised cases, where all parameters have to be estimated from the observations alone. However, these models use a questionable independence assumption, and are often limited by the knowledge of suitable multivariate laws when considering vector variables. When wishing to use dependent and non-Gaussian noise, the theory of Copulas is well suited. The aim of this paper was to introduce copulas in some recent Markov models, and to show how they can be used to describe interactions between hidden stochastic data and observed ones. We presented three possibilities for using copulas in three models of increasing generality: (i) hidden Markov chains with non-independent noise (HMC), (ii) recent pairwise Markov chains (PMC), and (iii) recent triplet Markov chains (TMC). The use of HMC models is particularly direct and workable, which makes it well suited to practical applications. This is all the more true that a simple parameter estimation method proposed in the paper, derived from Stochastic Expectation-Maximization (SEM), works well.

As perspectives, we may envisage an extension of the models and techniques proposed to the multi-sensor case, in which the observations are multivariate. In fact, correlated and non necessarily Gaussian sensors can be modeled using copulas in an analogous way. Another extension could be searching in an automated way, for each class, the very form of the noise distribution. In fact, such studies, performed in [15] in the case of multi-sensor HMC-IN, gave promising results.

Acknowledgments

This work and the PhD of Nicolas Brunel are supported by the Délégation Générale pour l'Armement (DGA/DSP).

The authors thank Alain Hillion, Research Director of “Ecole Nationale Supérieure des

Télécommunications de Bretagne”, for numerous discussions which greatly helped in writing the paper.

References

- [1] R.B. Nelsen, An introduction to copulas, Lecture Notes in Statistics, vol. 139, Springer, New York, 1998.
- [2] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publications de L'institut de Statistique de l'Université de Paris 8 (1959) 229–231.
- [3] T. Cameron-Faulkner, E. Lieven, M. Tomasello, A construction based analysis of child directed speech, *Cognitive Sci.* 27 (6) (2003) 843–873.
- [4] M.A. Fishman, L. Stone, Indiscriminate polyandry and male parental effort, *Bull. Math. Biol.* 66 (1) (2004) 47–63.
- [5] R.M.S. Accioly, F.Y. Chiyoshi, Modeling dependence with copulas: a useful tool for field development decision process, *J. Petroleum Sci. Eng.* 44 (1–2) (2004) 83–91.
- [6] B. Vaz de Melo Mendes, R. Martins de Souza, Measuring financial risks with copulas, *Internat. Rev. Financial Anal.* 13 (1) (2004) 27–45.
- [7] F. Tuerlinckx, A multivariate counting process with Weibull-distributed first-arrival times, *J. Math. Psychol.* 48 (1) (2004) 65–79.
- [8] P. Deheuvels, Multivariate tests of dependence, *Analytical Methods in Probability* (Oberwolfach, 1980), *Lectures Notes in Mathematics*, vol. 861, Springer, Berlin, 1981, pp. 42–50.
- [9] H. Joe, J.J. Xu, The estimation method of inference functions for margins for multivariate models, Department of Statistics, University of Columbia, Technical Report, 166, 1996, <http://hajek.stat.ubc.ca/~harry/ifm.pdf>
- [10] E. Bouye, N. Gaussel, M. Salmon, Investigating dynamic dependence using copulae, Research Report, September 2001, [http://www.cass.city.ac.uk/facfin/papers/Investigating Dynamic Dependence Using Copulae.pdf](http://www.cass.city.ac.uk/facfin/papers/Investigating%20Dynamic%20Dependence%20Using%20Copulae.pdf)
- [11] L.E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Statist.* 41 (1970) 164–171.
- [12] G.D. Fornay, The Viterbi algorithm, *Proc. IEEE* 61 (3) (1973) 268–277.
- [13] K. Aas, L. Eikvil, R. Bang Huseby, Applications of hidden Markov chains in image analysis, *Pattern Recognition* 32 (4) (1999) 703–713.
- [14] C. Collet, J. Zallat, Y. Takakura, Clustering of Mueller matrix images for skeletonized structure detection, *Optics-Express* (2004) 1271–1280.
- [15] N. Giordana, W. Pieczynski, Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation, *IEEE Trans. Pattern Anal. Machine Intel.* 19 (5) (1997) 465–475.

- [16] Y. Guedon, Estimating hidden semi-Markov chains from discrete sequences, *J. Comput. Graph. Statist.* 12 (3) (2003) 604–639.
- [17] L.C. Thomas, D.E. Allen, N. Morkel-Kingsbury, A hidden Markov chain model for the term structure of bond credit risk spreads, *Internat. Rev. Financial Anal.* 11 (3) (2002) 311–329.
- [18] 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.
- [19] S. Derrode, W. Pieczynski, Signal and image restoration using Pairwise Markov Chains, *IEEE Trans. Signal Processing* 52 (9) (2004) 2477–2489.
- [20] W. Pieczynski, C. Hulard, T. Veit, Triplet Markov Chains in hidden signal restoration, *SPIE's International Symposium on Remote Sensing*, 22–27 September, Crete, Greece, 2002.
- [21] W. Pieczynski, Pairwise Markov chains, *IEEE Trans. Pattern Anal. Machine Intel.* 25 (5) (2003) 634–639.
- [22] W. Pieczynski, Triplet Markov Chains and Theory of Evidence, *Internat. J. Approx. Reason.* (2003) submitted.
- [23] Y. Delignon, W. Pieczynski, Modeling non Rayleigh Speckle Distribution in SAR Images, *IEEE Trans. Geosci. Remote Sens.* 40 (6) (2002) 1430–1435.
- [24] W. Pieczynski, J. Bouvrais, C. Michel, Estimation of Generalized Mixture in the Case of Correlated Sensors, *IEEE Trans. Image Process.* 09 (2) (2000) 308–311.
- [25] N. Brunel, W. Pieczynski, S. Derrode, Copulas in vectorial hidden Markov chains for multicomponent image segmentation, *ICASSP'05*, Philadelphia, USA, March 2005.
- [26] R. Douc, C. Matias, Asymptotics of the Maximum Likelihood Estimator for general Hidden Markov Models, *Bernoulli* 7 (3) (2001) 381–420.
- [27] R. Douc, E. Moulines, T. Ryden, Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, *Ann. Statist.* 32 (5) (2004) 2254–2304.
- [28] G.J. McLachlan, D. Peel, *Finite Mixture Models*, Wiley series in applied probability, 2000.
- [29] G. Celeux, J. Diebolt, The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, *Comput. Statist. Quater.* 2 (1985) 73–82.
- [30] M.A. Tanner, *Tools for Statistical Inference: Methods for the exploration of posterior distributions and likelihood functions*, third ed., Springer series in Statistics, Springer, New-York, 1996.
- [31] W.R. Gilks, S. Richardson, D.J. Spiegelhalter (Eds.) (Eds.), *Markov Chain Monte Carlo in Practice*, *Interdisciplinary Statistics*, Chapman & Hall/CRC, London/Boca Ration, 1996.
- [32] C.P. Robert, G. Casella, *Monte-Carlo Statistical Methods*, Springer texts in Statistics, Springer, New-York, 1999.
- [33] G.J. McLachlan, T. Krishnan, *EM Algorithm and Extensions*, Wiley, New York, 1997.
- [34] F. Salzenstein, W. Pieczynski, Sur le choix de méthode de segmentation statistique d'images, *Traitement du Signal* 15 (2) (1998) 119–128.
- [35] M. Pitt, S. Walker, Construction of stationary Time series with application to volatility models, Working Paper, November 1999.
- [36] G.E.B. Archer, D.M. Titterton, Parameter estimation for hidden Markov chains, *J. Statist. Plann. Inference* 108 (1–2) (1 November 2002) 365–390.
- [37] W.F. Darsow, B. Nguyen, E.T. Olsen, Copulas and Markov processes, *Illinois J. Math.* 36 (1992) 600–642.
- [38] B. Fernández Fernández, J.M. González-Barrios, Multi-dimensional dependency measures, *J. Multivar. Anal.* 89 (2) (2004) 351–370.