UNSUPERVISED SIGNAL RESTORATION USING COPULAS AND PAIRWISE MARKOV CHAINS

N. Brunel and W. Pieczynski

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

ABSTRACT

This work deals with the statistical restoration of hidden discrete signals. The problem we deal with is how to take into account, in recent pairwise and triplet Markov chain context, complex noises that can be non-Gaussian, correlated, and of class-varying nature. We propose to solve this modeling problem using Copulas. The interest of the new modeling is validated by experiments performed in supervised and unsupervised context. In the latter, all parameters are estimated from the only observed data by an original method.

1. INTRODUCTION

Let $X = (X_1, ..., X_n)$ is a stochastic process modeling a hidden discrete signal (each X_i takes its values in a finite set $\Omega = \{1, ..., k\}$), and let $Y = (Y_1, ..., Y_n)$ be a stochastic process modeling the observations (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 the both process X, Y are linked by the mean of some appropriate joint distribution $p(x, y)$. The hidden Markov chain (HMC) model, in which this distribution is written

$$p(x, y) = p(x_1) p(x_2|x_1) ... p(x_n|x_{n-1}) p(y_1|x_1) ... p(y_n|x_n) \quad (1.1)$$

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 (1.1) implies the independence of Y_1, …, Y_n conditionally on X, we will call the model (1.1) HMC with independent noise (HMC-IN).
More recently, this model has been generalized to the so-called “pairwise Markov chain” model, in which the pairwise process $Z = (Z_1, ..., Z_n)$, with $Z_1 = (X_1, Y_1)$, …, $Z_n = (X_n, Y_n)$ is a Markov chain [7]. Thus we have:

$$p(z) = p(z_1) p(z_2|z_1) ... p(z_n|z_{n-1}) \quad (1.2)$$

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 [7]. 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 [7]. However, likely to HMC-IN [1], PMC can be used to estimate $X = x$ from $Y = y$ by different Bayesian methods and first results are encouraging [3].
In fact, considering the same “forward” probability $\alpha(x_i) = p(x_i, y_1, ..., y_i)$, and the new, more general, “backward” probability $\beta(x_i) = p(y_{i+1}, ..., y_n|x_i, y_i)$, we have analogous recursions

$$\alpha(x_1) = p(x_1) p(y_1|x_1), \text{ and}$$
$$\alpha(x_{i+1}) = \sum_{x_i \in \Omega} \alpha(x_i) p(z_{i+1}|z_i) \text{ for } 1 \le i \le N-1 \; ; \quad (1.3)$$
$$\beta(x_N) = 1, \text{ and}$$
$$\beta(x_i) = \sum_{x_{i+1} \in \Omega} \beta(x_{i+1}) p(z_{i+1}|z_i), \text{ for } 1 \le i \le N-1 \quad (1.4)$$

As for each $1 \le i \le n$, $p(x_i|y_1, ..., y_n) \propto \alpha(x_i)\beta(x_i)$, the latter can be calculated and thus the Bayesian MPM method given by $\hat{s}_{MPM}(y_1, ..., y_n) = (\hat{x}_1, ..., \hat{x}_n)$, with

$$\hat{x}_i = \arg\max_{x_i \in \Omega} p(x_i|y_1, ..., y_n) \quad (1.5)$$

can also be calculated.
Otherwise, we have

$$p(x_i, x_{i+1}|y_1, ..., y_n) \propto \alpha(x_i) p(z_{i+1}|z_i) \beta(x_{i+1}), \quad (1.6)$$

which gives $p(x_{i+1}|x_i, y_1, ..., y_n)$ and thus allows one to simulate X according to its posterior distribution $p(x|y)$. Further, these estimations of $X = x$ can be made in an unsupervised manner, where all parameters of the PMC used are estimated from $Y = y$ [3].

2. COPULAS IN HMC-CN, PMC, AND TMC

2.1 Copulas in PMC.

Let us begin by the PMC, which are of a mid generality and which correspond to the experiments described in Section 4. The HMC with correlated noise (HMC-CN) will then appear as a particular case of PMC, and the TMC case will appear an extension of the PMC.

Let $Z = (Z_1, ..., Z_n)$ be a PMC verifying (1.2). The distribution $p(z)$ is also defined by $p(z_1, z_2)$, …, $p(z_{n-1}, z_n)$, once for each $2 \le i \le n-1$, $p(z_{i-1}, z_i)$ and $p(z_i, z_{i+1})$ give the same marginal distribution $p(z_i)$. The latter is in particular ensured by the condition $p(z_{i-1}, z_i) = p(z_i, z_{i-1})$ for each $2 \le i \le n$, which will be assumed in the following. Further, assuming that $p(z_1, z_2)$, …, $p(z_{n-1}, z_n)$ are equal, the distribution $p(z)$ of a PMC Z is given by $p(z_1, z_2)$. Writing $p(z_1, z_2) = p(x_1, x_2) p(y_1, y_2 | x_1, x_2)$, we see that the distribution $p(z)$ is defined by a probability $p(x_1, x_2)$ on $\Omega^2 = \{1, ..., k\}^2$, and k^2 probabilities on R^2. To simplify, 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 $1 \le i, j \le k$, the density $p_{ij}(y_1) = p(y_1 | x_1 = i, x_2 = j)$ (which are equal to $p(y_2 | x_2 = i, x_1 = j)$) and the density $p_{ij}(y_2) = p(y_2 | x_1 = i, x_2 = j)$ (which are equal to $p(y_2 | x_2 = j, x_1 = i)$) are known and their correlation ρ_{ij} is given. Of course, in the Gaussian case this gives $p_{ij}(y_1, y_2)$ but in general case numerous different $p_{ij}(y_1, y_2)$ define the same $p_{ij}(y_1)$, $p_{ij}(y_2)$, and ρ_{ij}. So, when using the PMC model in non Gaussian context the following question can arise in real situations: having $p_{ij}(y_1)$, $p_{ij}(y_2)$, and ρ_{ij}, how to define $p_{ij}(y_1, y_2)$? The theory of copulas responds, often in a very neat manner, such questions [4, 8].

The main result of the theory of copulas is the following. Let $h(y_1, y_2)$ be a probability density on R^2, H the associated pdf function, $h_1(y_1)$ and $h_2(y_2)$ the marginal densities, and H_1, H_2 the pdf functions associated with them. Then there exist a function C defined on $[0,1]^2$ such that

$$H(y_1, y_2) = C(H_1(y_1), H_2(y_2)) \tag{2.1}$$

Deriving (2.1) with respect to y_1, y_2 and introducing $c(u,v) = \dfrac{\partial\partial C(u,v)}{\partial u \partial v}$, we have

$$h(y_1, y_2) = h_1(y_1) h_2(y_2) c(H_1(y_1), H_2(y_2)) \tag{2.2}$$

Conversely, having H_1, H_2 and a copula C, one can use (2.1) to define H. So, a given H pdf on R^2 defines a copula C with (2.2), and this copula can also be used to define any another H' from any another H_1', H_2'.

Example 2.1

As an example, let us consider the Gaussian copulas. Let $h(y_1, y_2)$ be Gaussian with correlation ρ and marginal distributions having null means and variances equal to one. (2.1) and (2.2) then define a "Gaussian bivariate copula" by

$$c(u,v) = \frac{h(H_1^{-1}(u), H_2^{-1}(v))}{h_1(H_1^{-1}(u)) h_2(H_2^{-1}(v))} \tag{2.3}$$

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 pdf F_1, F_2, and f_1, f_2, the corresponding densities. Using (2.3), we can define a density for the distribution of (Y_1, Y_2) by

$$f(y_1, y_2) = f_1(y_1) f_2(y_2) \frac{h(H_1^{-1}(F_1(y_1)), H_1^{-1}(F_2(y_2)))}{h_1(H_1^{-1}(F_1(y_1))) h_1(H_1^{-1}(F_2(y_2)))} \tag{2.4}$$

Further, when (Y_1, Y_2) is Gaussian, (2.4) gives again its distribution.

So, in a PMC, we can apply (2.4) k^2 times to $p_{ij}(y_1)$, $p_{ij}(y_2)$, and ρ_{ij} (remember that $p_{ij}(y_1)$ and $p_{ij}(y_2)$ are two different functions). Thus the latter defines all $p_{ij}(y_1, y_2)$, which gives, linked with a distribution $p(x_1, x_2)$ on $\Omega^2 = \{1, ..., k\}^2$, the distribution $p(z_1, z_2) = p(x_1, x_2) p(y_1, y_2 | x_1, x_2)$, which finally gives the distribution of the PMC $Z = (Z_1, ..., Z_n)$.

2.2 Copulas in HMC-CN.

As specified above the HMC-IN are the most frequently used; however, more sophisticated HMC can also be considered. Let $Z = (Z_1, ..., Z_n)$ be a PMC verifying the hypotheses of the previous sub-section. As showed in [7], $X = (X_1, ..., X_n)$ is then a Markov chain if and only if $p(y_i | x_{i-1}, x_i) = p(y_i | x_i)$ (or $p(y_i | x_i, x_{i+1}) = p(y_i | x_i)$, given that $p(y_i | x_{i-1}, x_i) = p(y_i | x_i, x_{i+1})$). This means that we can have a "hidden" Markov chain, in the sense that the hidden process X is a Markov one, in more general situations than (1.1), in which Y_1, .., Y_n are not independent conditionally on X. For example, in the Gaussian case, when the mean and the variance of $p(y_1 | x_1, x_2)$ only depends on x_1 (and thus it is the same for the mean and the variance of $p(y_2 | x_1, x_2)$), the hidden chain X is a Markov one (the PMC is a HMC), even when $p(y_1 | x_1, x_2)$ and $p(y_2 | x_1, x_2)$ are correlated. Of course, in such simpler situations all that has been said above subject to copulas in PMC remain valid.

2.3 Copulas in TMC.

The use of copulas in PMC above can be extended to the so-called "triplet" Markov chains (TMC) model. The latter consists of introducing a third stochastic process $U = (U_1, ..., U_n)$, each U_i taking its values in a finite set $\Lambda = \{\lambda_1, ..., \lambda_m\}$. Thus we have three processes: the hidden process X, the observed process Y, and a latent process U. Assuming that the triplet process $T = (X, U, Y)$ is a Markov chain one can still estimate X from Y [6]. In fact, putting $V_n = (X_n, U_n)$ and V the corresponding process, we see that (V, Y) is a PMC and so different marginal distributions of V conditional on Y can be calculated. For example, $p(v_i, y) \propto \alpha^i(v_i)\beta^i(v_i)$ are calculable, which gives $p(x_i, y) = \sum_{u_i \in \Lambda} p(x_i, u_i, y) = = \sum_{u_i \in \Lambda} p(v_i, y)$, and thus the Bayesian MPM restoration (1.5) is workable. Otherwise, it is possible to show that TMC are strictly more general than PMC (PMC are particular TMC obtained for $U = X$, but for a given TMC $T = (X, U, Y)$, the pairwise process $Z = (X, Y)$ is not necessarily a Markov one [6]).

So, as TMC $T = (X, U, Y)$ also is a PMC (V, Y), copulas can be used as described above. For $\Omega = \{1, ..., k\}$ and the finite set $\Lambda = \{\lambda_1, ..., \lambda_m\}$, there are $(km)^2$ distributions $p_{ij}(y_1, y_2)$ (the set Ω is replaced by $\Omega \times \Lambda$). These distributions can then be defined from their marginal distributions and the correlation coefficients of the latter, with Gaussian copulas, or any other copulas, as described above.

3. PARAMETER ESTIMATION

The aim this Section is to propose a method of estimation of all the parameters θ from $Y = y$. In the classical HMC-IN case the classical Expectation-Maximization (EM [5]) algorithm, which, having started from an initial value θ^0, produces a sequence of parameters according to the principle $\theta^{q+1} = \arg\max_\theta E_{\theta^q}[Log(p_\theta(X, Y) | Y = y]$ and works well in the classical Gaussian HMC-IN case, is difficult to apply in the context considered. We propose an original method, which can be seen as an extension of the "stochastic" EM (SEM [2]), whose principle is:

(i) Simulate $X = x^q$ according to $p_{\theta^q}(x | y)$;

(ii) put $\theta^{q+1} = \hat{\theta}(x^q, y)$, where $\hat{\theta} = \hat{\theta}(X, Y)$ is an estimator of θ from the complete data (X, Y).

So, we have to find, in the case of Gaussian copulas considered, k^2 densities $p_{ij}(y_1)$, k^2 densities $p_{ij}(y_2)$, k^2 correlation coefficients ρ_{ij}, and the distribution $p(x_1, x_2)$ on $\Omega^2 = \{1, ..., k\}^2$. Knowing that X can be simulated according to $p(x | y)$ (see Section 1), all we have to do is to define an estimator $\hat{\theta} = \hat{\theta}(X, Y)$ from the complete data (X, Y).

The parameters $p(x_1, x_2)$ can then be estimated by the classical "empirical" estimate

$$\hat{p}(i, j) = \frac{1_{[x_1 = i, x_2 = j]} + ... + 1_{[x_{n-1} = i, x_n = j]}}{n - 1} \quad (3.1)$$

and the correlation coefficients ρ_{ij} can be estimated by the classical "empirical" correlation coefficients

$$\hat{\rho}_{ij}(x, y) = \frac{(y_1 - \hat{\mu}_{ij}(x, y))^2 1_{[x_1 = i]} + ... + (y_n - \hat{\mu}_{ij}(x, y))^2 1_{[x_n = i]}}{n} \quad (3.2)$$

$$\hat{\mu}_{ij}(x, y) = \frac{y_1 1_{[x_1 = i]} + ... + y_n 1_{[x_n = i]}}{n} \quad (3.3)$$

Concerning the k^2 densities $p_{ij}(y_1)$, let us assume that each of them is of a known form, and each of them depends on some parameters, which can be estimated by some estimator. Selecting from (x_1, y_1), (x_2, y_2), …, (x_n, y_n) the observations $y^{ij} = (y_r)$ such that $x_r = i$ and $x_{r+1} = j$, we know that all y_r are produced according to $p_{ij}(y_1)$, so the sample y^{ij} can be used to estimate $p_{ij}(y_1)$.

4. EXPERIMENTS

Let us consider the case of two classes, with the two distributions $p_{11}(y_1)$, $p_{12}(y_1)$ being Gamma(3,3), 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 Figure 1). The probabilities $p(i, j)$ are $p(1,1) = 0.4$, $p(1,2) = p(2,1) = 0.15$, $p(0,0) = 0.3$. In the classical HMC verifying (1), we assume that $p(y_1|1)$ is Gamma and $p(y_1|2)$ is Weibull (the parameters are estimated with SEM). According to the results presented in Table 1, 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.

	PMC and Copulas model			HMC model	
Case	(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%

Table 1. Error ratios obtained with the Bayesian method MPM. (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 $n = 500$.

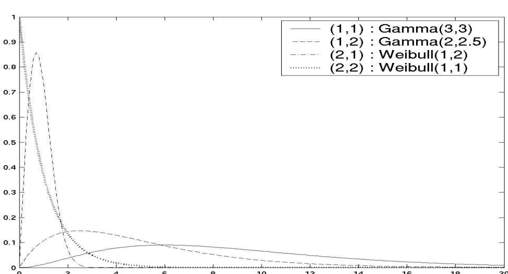

Figure 1. Distributions $p_{11}(y_1)$, $p_{12}(y_1)$, $p_{22}(y_2)$, $p_{12}(y_2)$ denoted by $(1,1)$, $(1,2)$, $(2,1)$, $(2,2)$.

5. CONCLUSIONS

Hidden Markov chains with independent noise (HMC-IN) are widely used in different problems and this success is generally due to the very good behavior of different associated Bayesian unsupervised restorations. However, these models are mainly used in the Gaussian noise context and the noise independence is difficult to clearly justify in numerous real situations. When wishing to use correlated and non-necessarily Gaussian noise, the theory of Copulas is a quite well suited one. The aim of this paper was to introduce copulas in recent Pairwise Markov Chains (PMC) models, which generalize the classical HMC-IN. We presented three possibilities of the use of copulas in three models of increasing generality: (i) hidden Markov chains with correlated noise (HMC-CN), (ii) PMC, and (iii) "Triplet Markov Chain" (TMC). Furthermore, an original parameter estimation method is proposed and some experiments in the PMC context are described. In particular, the latter show the interest of the Gaussian Copulas in the unsupervised Bayesian Maximum of the Posterior Mode (MPM) restoration of hidden data.

6. REFERENCES

[1] L. E. Baum, T. Petrie, G. Soules, N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *An. Math. Statistic.*, 41, pp. 164-171, 1970.

[2] G. Celeux and J. Diebolt, The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, *Comput. Statist. Quater*. 2, pp. 73-82, 1985.

[3] S. Derrode and W. Pieczynski, SAR image segmentation using generalized Pairwise Markov Chains, SPIE's International Symposium on Remote Sensing, September 22-27, Crete, Greece, 2002.

[4] P. Deheuvels, La fonction de dépendance empirique et ses propriétés, Académie Royale de Belgique, Bulletin de la Classe des Sciences, 5ème série, No. 65, pp. 274-292, 1979.

[5] G. J. McLachlan and T. Krishnan, EM Algorithm and Extensions, Wiley, 1997.

[6] W. Pieczynski, Chaînes de Markov Triplet, *C. R.. Acad. Sci. – Mathématiques*, Paris, Ser. I 335, pp. 275-278, 2002.

[7] W. Pieczynski, Pairwise Markov chains, *IEEE Trans. on PAMI*, Vol. 25, No. 5, pp. 634-639, 2003.

[8] R. B. Nelsen, An introduction to copulas, *Lecture Notes in Statistics*, Vol. 139, Springer, New York, 1998.