Multivariate Hidden Markov chains for correlated non-Gaussian noise, with applications in segmentation of radar signals

N. Brunel, W. Pieczynsky, F. Barbaresco

20th GRETSI Colloquium at Louvain-la-Neuve, September 6-9, 2005

(This conference proceeding is translated from the original *Chaînes de Markov cachées multivariées à bruit corrélé non Gaussian, avec applications à la segmentation du signal radar*. My translation is in satisfaction of my foreign-language requirement, examined by Dr. Alain Goriely, for the PhD in mathematics at the University of Arizona. Also, I have chosen an article which I can learn something mathematical about. Thus I have, solely for my own reference, appendices on French statistical vocabulary and copulas. The reader is referred to the original article for tables and bibliography, which are not reproduced here. — John Kerl, August 27, 2009.)

Abstract — *We propose a multivariate signal model with temporal and spectral dependence, well-fitted for the modeling of radar signals. The proposed model is a hidden Markov chain for which observations are Spherically Invariant Random Vectors, and temporal variation is described by a copula. It is still possible to estimate the parameters of the SIRV, and we study the robustness of the estimation under different kinds of copulas and various strength of dependence. Finally, we explore the influence of an omitted dependence for statistical segmentation of radar signals based on hidden Markov chains.*

1 Introduction

In this article, we address the modeling and treatment of multivariate signals which are non-Gaussian and correlated temporally and spectrally. The non-Gaussian character of radar signals motivated the introduction of the K-law and its variants for the modeling of intensities [1,7], as well as the "Spherically Invariant Random Vectors" (SIRVs) which are their extension for multidimensional signals [16, 8]. This approach permits a correct statistical representation of effectively measured radar signals [6], and remains sufficiently close to the Gaussian model to permit simple interpretations, all offsetting the problems of the robustness of the estimation and an overly rapid decay of the tails of distributions. Received radar signals are generally dependent on the distance axis, and several approaches have been already proposed to describe that dependence [12]. The objective here is to propose a family of stochastic processes which permit one to simultaneously describe these two characteristics: the marginals pertaining to the SIRV family and whose covariance matrices represent the spectral correlation, and a temporal (or spatial) dependence. We are interested principally in the latter property, which we introduce using copula bias [10,14], a statistical tool still little used in signal processing (although see [3]). We address the influence of this dependence in the context of segmentation, which is often based on hidden Markov models in which the hypothesis of the independence of the observations (conditionally on classes researched) is given. We recall in the following section the definitions and properties of SIRV and copulas. We next show how these allow one to construct the proposed model. We then use the latter to propose a triplet model [15] in order to simulate the observations not respecting the independence hypotheses of "hidden Markov chain" (HMC) models. Varying the type of the copula and the force of the dependence, we test the influence of this dependence on the estimation procedure and the segmentation. We then draw conclusions regarding the robustness and the reliability of non-supervised segmentation procedures in radar signal processing, and more generally for stationary processes.

2 Dependent SIRVs

2.1 SIRV model and notation

A random vector $Z \in \mathbb{R}^M$ is called an SIRV if it may be written as

$$Z = U^{-1/2}\varepsilon \tag{1}$$

where ε is a Gaussian vector distributed as $N(0, \Sigma)$ called "speckle", and U is a positive real-valued random variable, called "texture". The distribution of Z is therefore a continuous mix of centered Gaussians. We extend the definition of SIRVs, such as is usually given in radar, to the case where Z might have non-zero mean. Such distributions are said to be of elliptic contour or simply elliptic, because of the form of their density [9]. In applications, the most-used members of this family are derived from the gamma distribution. When the U variable follows a gamma $\gamma(\nu/2, 2/\nu)$ distribution, Z's distribution is Student's T, parameterized by the triple (m, Σ, ν), and if the variable U^{-1} follows a gamma $\gamma(a, 1/a)$ distribution, it has the K distribution with parameters (m, Σ, a).

The matrix Σ is proportional to the variance matrix of Z. The parameters a, ν control the speed of tail decay, and the normal distribution appear as a limiting case of the K and T as they tend to infinity. The densities are written

$$T: \quad f(z \mid \mathbf{m}, \Sigma, \nu) = \frac{\Gamma(\frac{\nu+M}{2})|\Sigma|^{-1/2}}{(\pi\nu)^{M/2}\Gamma(\frac{\nu}{2})}\left(1 + \frac{2p(z)}{\nu}\right)^{-\frac{-\nu+M}{2}} \tag{2}$$

$$K: \quad f(z \mid \mathbf{m}, \Sigma, a) = \frac{2a^a|\Sigma|^{-1/2}}{(2\pi)^{M/2}\Gamma(a)}\left(\sqrt{\frac{p(z)}{a}}\right)^{a-\frac{M}{2}} K_{a-\frac{M}{2}}\left(2\sqrt{a}p(z)\right) \tag{3}$$

where $p(z) = \frac{1}{2}(z-m)'\sigma^{-1}(z-m)$, and $'$ is the transposition operator.

2.2 Copulas

A two-dimensional copula is the cumulative distribution function of two random variables distributed uniformly on the square $[0,1]^2$, [5,10,14]. The interest in copulas is that they allow one to make a link between the joint and marginal distributions. If $\tilde{F}$ is a joint cumulative distribution function of the two random vectors (V_1, V_2), with CDFs F_1, F_2, we may assert thanks to the theorem of Sklar [14] that

$$\tilde{F}(v_1, v_2) = C(F_1(v_1), F_2(v_2)). \tag{4}$$

C is also the CDF of the vector $(F_1(v_1), F_2(v_2))$. For this reason, the derivative function $c(x,y) = \frac{\partial^2}{\partial x \partial y}C(x,y)$ is called the density of the copula C. Copulas generalize in any dimension M as the CDF of random variables with uniform marginals on the hypercube $[0,1]^M$. We propose three different families of parametric copulas which we will use in the simulations of section 4.

Elliptic copulas are derived by by inverting the relation of equation 4 which allows one to obtain the expression of the copula (or of its density) by means of known multivariate families.

The (bivariate) copula of the normal distribution is

$$c(u,v;\rho) = (1-\rho^2)^{-1/2} \exp\left(-\frac{\zeta_1^2+\zeta_2^2-2\rho\zeta_1\zeta_2}{2(1-\rho)^2} + \frac{\zeta_1^2+\zeta_2^2}{2} \right) \tag{5}$$

with $\zeta_1 = \Phi^{-1}(u)$, $\zeta_2 = \Phi^{-1}(v)$, and Φ^{-1} the inverse of the CDF of the centered and reduced (univariate) normal.

The Student copula has a more complex expression:

$$c(u_1,\ldots,u_M) = \frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\sqrt{1-\rho^2}\,\Gamma(\frac{\nu+1}{2})^M} \frac{\left(1+\frac{\zeta_1^2+\zeta_2^2-2\rho\zeta_1\zeta_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}}}{\left((1+\frac{\zeta_1^2}{\nu})(1+\frac{\zeta_2^2}{\nu})\right)^{-\frac{\nu+1}{2}}} \tag{6}$$

with $\zeta_1 = T_\nu^{-1}(u)$, $\zeta_2 = T_\nu^{-1}(v)$, and T_ν^{-1} the inverse of the CDF of the univariate T distribution with ν degrees of freedom. The parameter $\rho \in [-1,1]$ corresponds to the correlation coefficient of the covariance matrix which appears in the definition of the elliptic law. However, if these copulas are used to define the joint distribution of the vector V_1, V_2, ρ no longer corresponds to the (usual) Pearson correlation between V_1, V_2, but rather to the Kendall τ between two random variables. The two measures of dependence are linked by the relation $\tau = 2\arcsin(\rho)/\pi$.

Archimedean copulas constitute another generic family of copulas, defined by the functional form following $C(u,v) = \phi^{-1}(\phi(u)+\phi(v))$, with certain conditions on the function ϕ (among others, it must be positive and decreasing on the interval $[0,1]$). The Clayton copula is constructed with the function $\phi_a(t) = \frac{1}{\alpha}(t^{-\alpha}-1)$, and is

$$C_\alpha(u,v) = \max\left((u^\alpha + v^\alpha - 1)^{-1/\alpha} \right). \tag{7}$$

The parameter α may vary within $[-1,+\infty) \setminus \{0\}$. It is also related to the Kendall τ by the expression $\tau = \alpha/(\alpha+2)$.

The Gaussian copula has the property of making independent the extreme (minimal or maximal) values, in contrast to the Student copula [10]. For the latter, the presence of an extreme value in one of the components leads an extreme value (in the same sense) in the other variable. When the copulas are used for the modeling of the dependence of spatial processes, they allow one to reproduce aggregation phenomena of extreme values. The Clayton copula only correlates minimal values.

2.3 Modeling of dependence by means of copulas

Let $\mathbf{Y} = (Y_n)_{n\geq 1}$ be a stationary process (in the strict sense) with values in $\mathbb{R}^M$ with elliptic marginals. We propose to model the dependence with the aid of the theory of copulas. If $\mathbf{Y}$ were real-valued, it would be possible to describe all the stationary processes with the aid of a copula and the CDF of Y_n, thanks to the theorem of Sklar. However, such a construction poses two difficulties: equation 4 does not extend to random vectors (impossibility theorem, [14]) and the manipulation of a copula of n arguments to model the law of $(Y_1,\ldots,Y_n)$ when n is large poses practical problems. With the goal of proposing a sufficiently large class of stationary processes with SIRV marginals, we introduce the dependence through the bias of a latent Markov scalar process. The SIRV hypothesis allows one to introduce, thanks to equation 1, two processes $\mathbf{U} = (U_n)_{n\geq 0}$, $\boldsymbol{\varepsilon} = (\varepsilon_n)_{n\geq 0}$. We then suppose that $\mathbf{U}$ and $\boldsymbol{\varepsilon}$ are independent processes and that the speckle process $\boldsymbol{\varepsilon}$ is IID, while the texture process $\mathbf{U}$ is a stationary Markov process (hypothesis similar to that taken in [12]). To entirely understand its distribution, it suffices to give the distribution of (U_1,U_2) which we express as a function of their copula C, their marginals being known of the type of SIRV seen above. The interest in the use of a copula to model the process $\mathbf{U}$ is double:

(i) The ease of construction of a Markov chain having stationary distribution with density g (and CDF G). The transition kernel is then written $g(u_{n+1})c(G(u_n), G(u_{n+1}))$;

(ii) The richness of types of dependence which can be envisioned thanks to the numerous families of copulas which exist (elliptic, archimedean copulas).

We have $p(y_n \mid y_{n-1}, \ldots, y_1) \neq p(y_n \mid y_{n-1})$ because the process $(\mathbf{U}, \mathbf{Y})$ is a hidden Markov chain. The estimation of the hidden process $\mathbf{U}$ on the basis of the observations $y_1, \ldots, y_n$ can turn out to be interesting in applications.

Remark: Similar models have already been introduced in econometry for the modeling of financial temporal series in which the variance (or volatility) is itself stochastic [11]. However, most often it is supposed that the logarithm of the volatility has a linear evolution. Copulas allow one to propose non-linear dynamics for the variance process.

3 Estimation of dependent SIRVs

We seek to estimate the parameters $(m, \Sigma, \theta = \nu \text{ or } a)$ of an SIRV based on the realization $y_1, \ldots, y_N$ of the preceding process, which is an identically distributed but not independent sample. We estimate nonetheless (m, Σ, θ) based on equations determining the maximum likelihood estimator (MLE) in the independent case. This avoids having to treat the maximization of the log likelihood of an HMC, which requires procedures which are complex and often costly in terms of calculation time [4].

3.1 Estimation

To calculate the MLE of an SIRV, we propose an EM algorithm exploiting the texture process $\mathbf{U}$ [13]. The complete log-likelihood of a joint process $(\mathbf{U}, \mathbf{Y})$ in the independent case is decomposed into two terms, which allow one to propose a procedure of seeking the maximum likelihood occurring through a succession of maximizations over small intervals, instead of a search over large intervals ($\theta = a$ or ν):

$$\log p\left(y_1^N, u_1^N \mid m, \sigma, \theta\right) = \sum_{i=1}^{N} \log(f(y_i \mid u_i, m, \Sigma)) + \sum_{i=1}^{N} \log(g_\theta(u_i)). \tag{8}$$

The complete log-likelihood of the HMC $(\mathbf{U}, \mathbf{Y})$ only differs by the presence of the term

$$\sum_{i=1}^{N-1} \log(c(G_\theta(u_i), G_\theta(u_{i+1})))$$

representing the dependence between observations. Thus the estimate used under the (false) hypothesis of independence returns to modify the EM algorithm corresponding to the true model during the update of the parameter θ. The re-estimation formulas in the independent case are:

$$m^{(n+1)} = \frac{\sum_{i=1}^N w_i^{(n)} z_i}{\sum_{i=1}^N w_i^{(n)}} \quad \text{and} \quad \Sigma^{(n+1)} = \frac{1}{N} \sum_{i=1}^{N} w_i^{(n)} (z_i - \mathbf{m}^{(n+1)})(z_i - \mathbf{m}^{(n+1)})' \tag{9}$$

with $w_i^{(n)} = E[U_i \mid z_i, m^{(n)}, \sigma^{(n)}, \theta^{(n)}]$. The texture parameter is obtained by the solution of a unidimensional non-linear equation depending on the density g_θ which may be written in the form

$$\theta^{(n+1)} = \psi(\theta^{(n)}, (w_i^{(n)})_{1 \leq i \leq N}). \tag{10}$$

3.2 Influence of dependence

The maximization of the likelihood (8) always gives consistent estimators, but ones whose variance is larger than that of the MLE. The EM algorithm, with samples of small size, often gives a biased estimator, due to being trapped in local maxima of the likelihood. In the case of a sufficiently large sample ($N = 500$), this problem is attenuated and tables 1 and 2 allow one to evaluate the effect of the dependence on the estimation of the tail parameter (estimated by Monte Carlo over 200 trials). We use the square root of the mean quadratic diffence (RMQD) of the estimator in order to give an indication of the fluctuation of the estimator, and also of its bias.

The parameters (m, Σ) are correctly estimated, and the variance of the estimators remains stable and close to the independent case. The difference becomes notable at $\tau = 0.8$. The tail parameter is the one most affected by the dependence, that which returns to misestimate the multiplicative factor of the covariance of an SIRV.

When the distribution tails are thicker (e.g. the case of the T distribution with regard to the K distribution considered) the influence of the dependence of the texture is more important. The correlation of extremes by the Student copula does not deteriorate the quality of the estimate with regard to the Gaussian copula. It is, on the other hand, the type of the dependence structure which modifies the segmentation performance: with the same τ of Kendall, the Clayton copula more strongly reduces the variance of the estimators.

4 Segmentation

4.1 Conditional dependence in the HMC

In contrast to the homogeneous case treated in section 2, we suppose that we have several zones of different characteristics, and that these are representable by a hidden Markov process $\mathbf{X} = (X_n)_{n\geq 1}$ with K classes, such that the distribution of Y_n conditioned on $X_n = k$ is elliptic with parameters $(m_k, \Sigma_k, \theta_k)$. We seek to estimate the process $\mathbf{X}$ in a non-supervised manner, using the Maximum a Posteriori Marginals (MPM) Bayesian estimator, supposing that we have a hidden Markov chain.

We introduce a dependence between the observations, conditioned on the states through the bias of the texture process $\mathbf{U}$ which allows one to evaluate the robustness (by means of the differences in the hypotheses of the model) of the non-supervised segmentation procedures in the case where the hypothesis of conditional independence of the observations is placed under suspicion. This supplementary spatial correlation is introduced by a copula c modeling the dependence of (U_n, U_{n+1}), such that the process $(\mathbf{X}, \mathbf{U}, \mathbf{Y})$ is a homogeneous stationary Markov chain (this is a particular case of triplet Markov chains). This new model also generalizes the univariate model [13], as well as the multivariate model [2].

4.2 Non-supervised segmentation with HMC-IN

The segmentations are obtained under the hypothesis of conditional independence of observations, for a model of 3 classes in dimension 2. In the example considered, the means are $m_1 = (0, 0)'$, $m_2 = (1.5, 1.5)'$, $m_3 = (3, 3)'$, and the variances are all normalized with distinct correlation coefficients $\rho_1 = 0.4$, $\rho_2 = 0.2$, $\rho_3 = 0.5$. Finally, the tail parameters are $\nu_1 = 5$, $\nu_2 = 10$, $\nu_3 = 15$ for the T distribution, and $a_1 = 2.5, a_2 = 5, a_3 = 7.5$ for the K distribution. The error rates reduce slowly with comparison to the independent case (Gaussian case with $\tau = 0$), and show a strong spread only in the case of the Kendall tau of 0.8. The deviations of the error rates remain stable but become larger in the independent case for $\tau \geq 0.59$, indicating a greater variablity in the quality of non-supervised segmentations, which is particularly clean in the case of the Clayton copula (for the T distribution and in lesser measure for the K distribution).

5 Conclusion

The statistical model proposed allows one to describe different dependence structures in the SIRV process, often used in radar. The process $(\mathbf{U}, \mathbf{Y})$ is an HMC whose originality is the use of copulas for the modeling of the dependence of the hidden process, and we have presented several copulas inducing different types of dependence. A short experimental study shows that dependence reduces the estimates of the distribution tails, moreso if the tails are large and the sample size is small. In the frame of non-supervised segmentation, we have extended the model of the couple chain $(\mathbf{X}, \mathbf{U})$ introduced in [3] to the triplet $(\mathbf{X}, \mathbf{U}, \mathbf{Y})$ in order to introduce a conditional independence in the observations. The influence of the latter remains weak when the classes are well separated (notably by distinct means), but it is notable when the dependence increases strongly. A deeper study needs to be taken when the classes have close characteristics (among others, zero means) and to identify the families of copulas likely to represent the dependence of real radar data. The development of estimation methods based on the true likelihood of the model also constitute an axis of development, in order to apply non-supervised segmentation methods in the context of Markov triplet models [15].

A Selected vocabulary items

French	**English**
Fonction de répartition (FDR)	Cumulative distribution function (CDF)
Éstimateur de maximum vraisemblance (EMV)	Maximum likelihood estimator (MLE)
Taux d'erreur	Error rates
Écart-type	Deviation

B Copulas and Sklar's theorem

(The following is from the *Oxford Dictionary of Statistics.*)

A copula is a function that relates a joint CDF to marginal CDFs of the individual variables. If the marginals are known but the joint is unknown, then a copula can be used to suggest a suitable form for the joint distribution.

Let F be the multivariate distribution function for the random variables $X_1, \ldots, X_n$ and let the CDF of X_j be F_j for all j. Define random variables $U_j = F_j(X_j)$, so that the marginal distribution of each U_j is uniformly distributed on the unit interval. Assume that for each value u_j there is a unique value $x_j = F^{-1}(u_j)$ (i.e. assume that the marginal CDFs are invertible) and let the joint CDF of $U_1, \ldots, U_n$ be C. Then

$$C(u_1, \ldots, u_n) = P(U_j < u_j \text{ for all } j) = F\{F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\},$$

for all $u_1, \ldots, u_n \in (0, 1)$ since $U_j < u_j$ if and only if $X_j < F_j^{-1}(u_j)$. The function C is called the *copula.* An equivalent relation to the above is

$$C\{F_1(x_1), \ldots, F_n(x_n)\} = F(x_1, \ldots, x_n),$$

for all $x_1, \ldots, x_n$ where $u_j = F_j(x_j)$ for each j. *Sklar's theorem* states that, for a given F, there is a unique C such that this equation holds.

Note that it may well not be possible to express the inverse functions F^{-1} in closed form (cf. the multivariate normal distribution).

Assuming that the copula and the marginal CDFs are differentiable, the corresponding result for PDFs is

$$f(x_1, \ldots, x_n) = c\{F_1(x_1), \ldots, F_n(x_n)\} f_1(x_1), \ldots, f_n(x_n).$$

If $X_1, \ldots, X_n$ are independent, then the left-hand side factors and we have

$$c\{F_1(x_1), \ldots, F_n(x_n)\} \equiv 1.$$

Thus the copula encapsulates the interdependencies between the X variables and is therefore also known as the *dependence function*. The joint PDFof $U_1, \ldots, U_n$ is

$$c(u_1, \ldots, u_n) = f(x_1, \ldots, x_n)/\{f_1(x_1) \cdots f_n(x_n)\},$$

where $x_j = F^{-1}(u_j)$ for each j.