COPULAS IN VECTORIAL HIDDEN MARKOV CHAINS FOR MULTICOMPONENT IMAGE SEGMENTATION

Nicolas Brunel, Wojciech Pieczynski

GET/INT, CITI - CNRS UMR 5157
9 rue Charles Fourier, F91000 Evry

Stéphane Derrode

EGIM, Institut Fresnel (CNRS UMR 6133)
Dom. Univ. de St Jérôme, F13013 Marseille

ABSTRACT

Parametric estimation of non-Gaussian multidimensional probability density function (pdf) is a difficult problem that is required by many applications in signal and image processing. A lot of efforts has been devoted to methods from multivariate analysis such as Principal or Independent Component Analysis (PCA and ICA). In this work, we introduce an alternative solution based on a very general class of multivariate models called 'copulas'. Useful copulas models for image classification are used in the frame of multidimensional mixture estimation arising in the segmentation of multicomponent images, when using a vectorial Hidden Markov Chain (HMC).

1. INTRODUCTION

The aim of this paper is to introduce copulas for multivariate modelling in the framework of statistical image segmentation and to compare this approach with other methods coming from multivariate data analysis.

The main problem of multicomponent images and of their statistical processing is the choice of a relevant statistical model for the relationships between the components. Segmentation procedure of scalar data usually makes the assumption that the expected classes differs from each other by a mean level μ and an inner degree of dispersion σ. The Gaussian hypothesis is an illustration of this *a priori* knowledge on the classes, and the use of Gaussian vectors for multivariate data shows the same implicit assumption for multicomponent images. This assumption can be very restrictive since the difference between the classes can depend on the ways the components are linked (independently of μ and σ). Moreover, if the abundance of univariate parametric models enables to manage deviance to normality, possibilities are far more restricted for multivariate models especially when we want to respect some constraints on the law of each component (due to physical knowledge about the involved phenomena).

Copulas are a statistical tool used for the modelling and estimation of dependence between random variables [1], that gives a general answer to such problem. It enables to widen the ability of multivariate modelling and to keep the same methodology as the one previously used with HMC [2]. It also brings out the influence of dependence between the components for the characterization of the classes.

This paper is organized as follows. In next section, the HMC approach and the estimation methodology are briefly recalled. We then introduce copulas in Section 3, and present some useful models for image analysis. Section 4 discusses estimation methods for multivariate data. In section 5, the methods are illustrated on a multispectral image from a CASI sensor. Conclusions are drawn in section 6.

2. HMC AND UNSUPERVISED ESTIMATION

The estimation principle is recalled here to make the paper self-contained (see [2]). The pixels of the M layers of a multicomponent images are first transformed into 1D chains using a Hilbert-Peano scan on each image. Hence, we get N series of M data, denoted by $\boldsymbol{y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_N)$, where $\boldsymbol{y}_n^t = \left(y_n^1, \dots, y_n^M\right)$, $1 \leq n \leq N$ and ^{t} is transposition. The objective is to classify each $\boldsymbol{y}_n$ into a set of K classes $\Omega = \{\omega_1, \dots, \omega_K\}$ in order to obtain the segmented chain $\boldsymbol{x} = (x_1, \dots, x_N)$. The image is then reconstructed from $\boldsymbol{x}$ by using an inverse Hilbert-Peano scan.

Assuming that $\boldsymbol{y}$ comes from a HMC [3], $\boldsymbol{x}$ is the hidden process which can be recovered by Bayesian estimation procedures. This probabilistic approach is possible if we know the law of the random vector $\boldsymbol{Y}_n^t = \left(\mathrm{Y}_n^1, \dots, \mathrm{Y}_n^M\right)$ conditionally on the state X_n. In the case of unsupervised segmentation, the distribution $P(\boldsymbol{X}, \boldsymbol{Y})$ is unknown. Therefore, we have to estimate the following sets of parameters $(1 \leq k, l \leq K)$: (i) the transition matrix $\boldsymbol{A}$ with entries a_{ω_k, ω_l} ; (ii) the observation densities, i.e. the parameters of the K M-dimensional distributions f_{ω_k}.

The estimation of all the parameters is usually achieved by an iterative search of the Maximum Likelihood Estimator (MLE), like EM algorithm and its variants as Stochastic EM algorithm [4]. We use here another one called Iterative Conditional Expectation (ICE), which uses the conditional expectation of well-suited estimators from the complete data $(\boldsymbol{x}, \boldsymbol{y})$ [5].

This algorithm faces the problem of estimating the multidimensional pdf arising in the multicomponent HMC model. If Gaussian densities are considered, law parameters can easily be estimated from the first order moments of a M-dimensional sample. However, in a number of image modalities, the noise can not be properly modelled by Gaussians. Moreover, the shape of the distribution of each class in the different layers can vary (e.g. multisensor case). One solution is to use copulas, which are a general tool for modelling multidimensional pdf.

3. MULTIDIMENSIONAL PDF WITH COPULAS

A (bivariate) copula is a cumulative density function (cdf) on the unit square with uniform margins. Such functions enables to give an exhaustive description of the dependence between two real random variables. The clearest way to illustrate it is the Sklar's existence theorem which clarifies the link between marginal laws and the joint law [1] :

Let Y_1, Y_2 be two real random variables, with (respectively) cdf F_1 and F_2. Let F be the corresponding bidimensional cdf on $\mathbb{R}^2$, then F has a copula representation, i.e. there exists a copula $\mathcal{C}$ such that

$$\forall y^1, y^2 \in \mathbb{R}, \quad F(y^1, y^2) = \mathcal{C}\left(F_1(y^1),\ F_2(y^2)\right). \quad (1)$$

Moreover, the copula is unique if F_1 and F_2 are continuous. If the copula is differentiable, we deduce directly from (1) the relationship between the pdfs of the variables (noted with lowercase letters) :

$$f_{1,2}(y^1, y^2) = f_1(y^1)\, f_2(y^2) \partial_{1,2}\mathcal{C}\left(F_1(y^1), F_2(y^2)\right), \quad (2)$$

with $\partial_{1,2}$ denoting the derivation toward the two coordinates. $\partial_{1,2}\mathcal{C} = c$ is called the density of the copula.

These results are also available for a cdf F on $\mathbb{R}^M$ and related marginals $F_1, \ldots, F_M$. If we use a copula for modelling a multivariate variable $\boldsymbol{Y}^t = \left(Y^1, \ldots, Y^M\right)$, we need first to specify the marginal laws of each Y^m. The latter can be known from *a priori* knowledge or by univariate statistical analysis of training data sets. Then, a copula is needed to construct the joint law of $\boldsymbol{Y}$. For illustration purpose and latter use in experiments, three copulas are now presented.

Product copula ($\mathcal{C}_1$) - The product copula is defined by $\mathcal{C}_1(u_1, \ldots, u_M) = u_1 \ldots u_M$, and corresponds to the case of independence between the components since the density of $\boldsymbol{Y}$ is written $\boldsymbol{f_Y}(\boldsymbol{y}) = \prod_{m=1}^M f_m(y^m)$, where f_m is the density of Y^m.

Gaussian copula ($\mathcal{C}_2$) - For a multivariate Gaussian model $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean $\boldsymbol{\mu} \in \mathbb{R}^M$ and variance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{1 \leq i,j \leq M}$, two assumptions are made : (i) the margins are univariate Gaussian variables with law $\mathcal{N}(\mu_i, \sigma_{ii}^2)$; (ii) the dependence structure is such that we have a Gaussian joint law. The Sklar's theorem explains how to get rid of the Gaussian margins. We write $\boldsymbol{\rho}$ as the correlation matrix deduced from $\boldsymbol{\Sigma}$, $\boldsymbol{I}$ the identity matrix of $\mathbb{R}^M$, and $\boldsymbol{\zeta}^t = \left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_M)\right)$ with Φ the cdf of the normalized Gaussian density. If we invert Eq. (2) that gives the relation between the density of the underlying copula and the density of a Gaussian vector, we get

$$c_2(u_1, \ldots, u_M) = |\boldsymbol{\rho}|^{-\frac{1}{2}}\ e^{-\frac{1}{2}\left(\boldsymbol{\zeta}^t(\boldsymbol{\rho}^{-1} - \boldsymbol{I})\boldsymbol{\zeta}\right)} \quad (3)$$

The Gaussian copula is very useful since computations are rather easy and the dependence structure is very intuitive, based on the usual correlation coefficients.

Student copula ($\mathcal{C}_3$) - In the same way, we can compute the Student copula by exploiting the multivariate Student law, with ν degrees of freedom. The computation of the underlying copula is manageable since each margin is a univariate Student law with ν degrees of freedom. The density of the Student copula is then

$$\begin{aligned} c_3(u_1, \ldots, u_M) = |\boldsymbol{\rho}|^{-\frac{1}{2}}\ & \frac{\Gamma\left(\frac{\nu+M}{2}\right)\ \Gamma\left(\frac{\nu}{2}\right)^{M-1}}{\Gamma\left(\frac{\nu+1}{2}\right)^M} \\ & \times \frac{\left(1 + \frac{1}{\nu}\boldsymbol{\xi}^t\boldsymbol{\rho}^{-1}\boldsymbol{\xi}\right)^{-\frac{\nu+M}{2}}}{\prod_{m=1}^M \left(1 + \frac{1}{\nu}\xi_m^2\right)^{-\frac{\nu+1}{2}}} \end{aligned} \quad (4)$$

with $\boldsymbol{\xi}$ the vector with components $\xi_m = T_m^{-1}(u_m)$. Γ is the Euler's Gamma function, T_m is the cdf of a (univariate) Student law with ν degrees of freedom, and $\boldsymbol{\rho}$ is a correlation matrix as for $\mathcal{C}_2$.

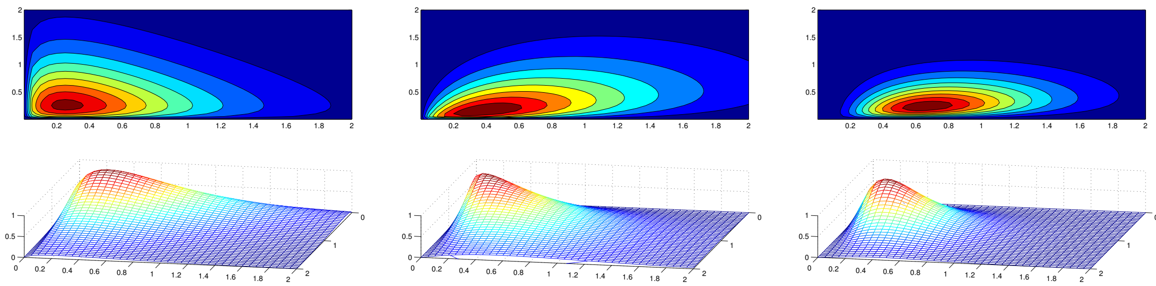
Fig. 1 shows examples of 2D densities obtained with fixed Gamma margins for the three kinds of copulas. The shape and scale parameters of the Gamma laws for the two marginals were set to $(1.5, 0.5)$ and $(0.5, 0.5)$. We can easily observe from the contour plots that the iso-probabilities are quite different, even for $\mathcal{C}_2$ and $\mathcal{C}_3$ copulas.

4. INFERENCE OF MULTIDIMENSIONAL PDF

Suppose we have an i.i.d. sample $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)$ of vectors of dimension M and we want to estimate its distribution $F_{Y^1,\ldots,Y^M}\left(y^1, \ldots, y^M\right)$. Before examining copulas, let us briefly recall the multivariate analysis viewpoint.

4.1. Parametric multivariate analysis

Several strategies are possible, depending on the assumptions made on the links between the components. The most heavy-handed one is independence so that $\boldsymbol{f}$ is written as in $\mathcal{C}_1$ copula. A more sophisticated solution is to consider correlations and consists in applying a PCA algorithm on the data before densities estimation. This can be done by projecting $\boldsymbol{y}$ onto an orthonormal system defined by $\boldsymbol{W}$ so that



(a) Product copula - $\mathcal{C}_1$ (b) Gaussian copula - $\mathcal{C}_2$ (c) Student copula - $\mathcal{C}_3$

Fig. 1. Bivariate densities with (a) independent Gamma margins, (b) correlated Gamma margins using a Gaussian copula ($\rho = 0.5$) and (c) dependent Gamma margins with Student copula ($\rho = 0.5$, $\nu = 10$). Margins are the same for the 3 pdfs.

the new data $\boldsymbol{t}_n = \boldsymbol{W}\,\boldsymbol{y}_n$ are decorrelated. Hence, we get the following estimation :

$$\boldsymbol{f}\left(\boldsymbol{y}_n\right) = \left|\det\left(\boldsymbol{W}\right)\right| \prod_{m=1}^{M} f^m\left(t_n^m\right). \tag{5}$$

However, ideally, data should be independent and this naturally fits an ICA approach. The objective of ICA is to find a linear transformation $\boldsymbol{W'}$ so that the new data $\boldsymbol{s}_n = \boldsymbol{W'}\,\boldsymbol{y}_n$ are mutually independent. A solution can be find under the assumption that the s_n^m components are not Gaussian. This optimization problem requires the optimization of a 'non-Gaussianity' criterion such as the kurtosis or the neguentropy [6]. The density $\boldsymbol{f}$ can be reconstructed using a formula similar to Eq. (5), replacing $\boldsymbol{W}$ by $\boldsymbol{W'}$ and t_n^m by s_n^m.

4.2. Estimation of copulas

Following the Sklar's decomposition, we use a parametric model assuming for instance that the margins are Gamma laws, and that the copula is gaussian.

$$\begin{aligned} F_{Y^1,\ldots,Y^M}\left(y^1,\ldots,y^M\,;\,\eta,(\theta_m)_{1\leq m\leq M}\right) = \\ \mathcal{C}\left(F_{Y^1}(y^1|\theta_1),\ldots,F_{Y^M}(y^M|\theta_M)\,;\,\eta\right) \end{aligned} \tag{6}$$

Parameters θ_m characterize the margins and η defines the copula. We propose to use the following two-step estimation methodology ('Inference For Margins' - IFM [7]) :

1. compute the MLE $\hat{\theta}_m$ of each margin from the M samples $(y_n^m)_{1\leq n\leq N}$ and then create the new data set $\forall n, m,\ u_n^m = F(y_n^m|\hat{\theta}_m)$.
2. compute the MLE $\hat{\eta}$ from the new sample $(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)$.

Gaussian copula ($\mathcal{C}_2$) Whatever the margins used, $\boldsymbol{\rho}$ can be estimated by the following matrix

$$\hat{\boldsymbol{\rho}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\zeta}_n\,\boldsymbol{\zeta}_n^t. \tag{7}$$

Student copula ($\mathcal{C}_3$) The MLE estimation of $\boldsymbol{\rho}$ can be computed by solving a fix point equation, deduced from the first order condition of the maximization of the log-likelihood. The Gaussian estimator given by Eq. (7) can be used for the initialization of the iterative search of the root.

The IFM method furnishes only an approximation of the MLE since the global maximization of the likelihood is replaced by two successive (and easier) maximizations. Nevertheless, the procedure gives good estimations and is easy to implement. For its use in HMC, it is only required to adapt the general frame of the ICE algorithm to this two-steps estimation.

Remark 1 : One can then ask for the difference between the use of a PCA analysis and of a Gaussian copula for the estimation of the multidimensional pdf. In the first case, we estimate the laws of the principal components of the signal. For instance, we make the assumption that the laws of the principal components belong to an *a priori* parametric model, such as the Pearson's system of distributions. It can be seen from Eq. (5) that the densities of the margins are hard to recover since we have to integrate over $M-1$ variates, whereas we directly identify the laws of the margins with copulas and then the dependence structure (characterized here by a matrix $\boldsymbol{\rho}$). So, in general, PCA does not furnish an idea on the induced parametric shape for margins.

Remark 2 : We have to choose among numerous families of copulas for the multidimensional reconstruction of the law, whereas it was straightforward by ICA.

5. SEGMENTATION OF MULTISPECTRAL IMAGES

This section is intended to illustrate the multidimensional pdf modelling capability of copulas. Hence, vectorial HMC-based classifications have been applied on an airborne

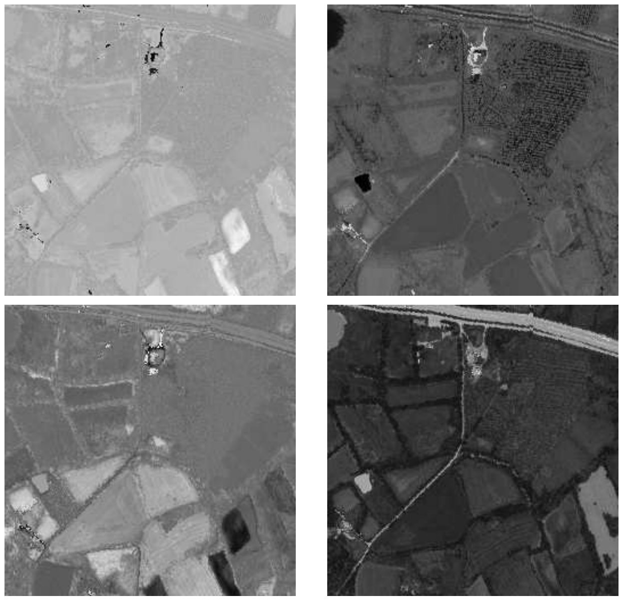


Fig. 2. Four bands of a multicomponent image. 256×256.

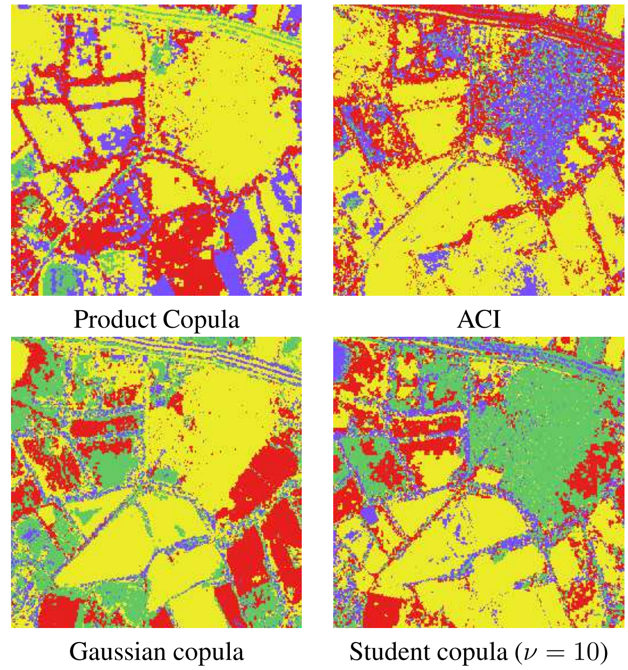


Fig. 3. Segmentation results. The same colors have been used for most similar classes.

hyperspectral CASI image, reduced to 4 bands through an adapted projection pursuit method [8], see Fig. 2. The orignal image contains 17 spectral bands from 450 to 950 nm and ground resolution is two meters.

Results from ACI (non-Gaussianity criterion from Hyvärinen [6]), and for product, Gaussian and Student copulas are presented in Fig 3. In all the experiments, segmentations have been achieved with 4 classes that should correspond to forest, fields, roads and wasteland. Hence, the parameters to be estimated by ICE correspond to a mixture of 4 classes in a 4D space. Gamma laws have been chosen for independent components in ACI, as well as for margins in copulas.

It is obvious that the classifications obtained are all different and this difference is only due to the way multivariate densities are estimated. However, it is quite difficult, and not the purpose of this work, to determine which result is the more appropriate and should depend on the application considered.

6. CONCLUSION

Copulas are a statistical concept representing the dependence between random variables in a very general way. In this work, copulas have been used to model multidimensional pdf arising in the segmentation of multicomponent images, when using a vectorial HMC-model. Three types of copulas have been illustrated according to the segmentation of a four spectral band images, and visually compared to the segmentation obtained from ACI analysis. Such copulas are not limited to the context of HMC and will be further investigated in a number of problems and models where multidimensional pdf estimation is concerned.

7. REFERENCES

[1] R. B. Nelsen, *An introduction to copulas*, vol. 139 of *Lectures Notes in Statistics*, Springer-Verlag, 1998.

[2] S. Derrode, G. Mercier, and W. Pieczynski, "Unsupervised multicomponent image segmentation combining a vectorial HMC model and ICA," in *IEEE ICIP*, Barcelona, Spain, 2003.

[3] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. of IEEE*, vol. 77, no. 2, pp. 257–286, February 1989.

[4] G. J. McLachlan and T. Khrishnan, *The EM algorithm and extensions*, Wiley Interscience, 1996.

[5] W. Pieczynski, "Statistical image segmentation," *Mach. Graph. and Vis.*, vol. 1, no. 1/2, pp. 261–268, 1992.

[6] A. Hyvärinen and E. Oja, "ICA : algorithms and applications," *Neur. Net.*, vol. 13, pp. 411–430, 2000.

[7] H. Joe, *Multivariate models and dependence concepts*, vol. 73 of *Monographs on Stat. and App. Prob.*, Chapman and Hall, 1997.

[8] G. Mercier, S. Derrode, and M. Lennon, "Hyperspectral image segmentation with Markov chain model," in *IGARSS*, Toulouse, France, 2003.