
Segmentation d'images par modèle de mélange conjoint non gaussien

Stéphane Derrode¹, Wojciech Pieczynski²

1. Institut Fresnel (CNRS UMR 6133)
Universités de Marseille et École Centrale Marseille
38, rue Frédéric Joliot-Curie F-13451 Marseille cedex 20
stephane.derrode@centrale-marseille.fr

2. CITI Department (CNRS UMR 5157)
TELECOM SudParis, Institut Télécom
9, Rue Charles Fourier F-91011 Evry cedex
wojciech.pieczynski@it-sudparis.eu

RÉSUMÉ. L'idée à l'origine du modèle de mélange conjoint (MMConjoint) est de classer simultanément deux ensembles d'observations en introduisant un a priori conjoint entre les deux classifications et un lien statistique entre les deux observations. Nous étudions en particulier le cas de mélanges gaussiens et le cas de mélanges paramétriques non gaussiens construits à partir de copules et de marginales non gaussiennes. Nous établissons également un algorithme EM et un algorithme ECI pour l'estimation automatique des paramètres, rendant la méthode de classification non supervisée. Le modèle est illustré à travers la segmentation d'images vectorielles couleur et IRM. Les résultats obtenus améliorent les segmentations individuelles des bandes par des modèles de mélange classiques.

ABSTRACT. The idea behind the Pairwise Mixture Model is to classify simultaneously two sets of observations by introducing a joint prior between the two corresponding classifications and some statistical relations between the two observations. We address both the Gaussian case and non-Gaussian parametric case built with copula-based parametric models and non-Gaussian margins. We also provide EM and ICE algorithms for automatic parameters estimation in order to make classification algorithms unsupervised. The model is illustrated through the segmentation of vectorial images (color and IRM). Results are compared to the segmentations obtained using independent mixture models on individual bands.

MOTS-CLÉS : classification bayésienne, modèle de mélange, copules, segmentation d'images.

KEYWORDS: bayesian classification, probabilistic mixture model, copulas, image segmentation.

DOI:10.3166/TS.29.9-28 © 2012 Lavoisier

Extended abstract

An important problem in signal processing is to restore an unobservable process $\mathbf{x} = \{x_n\}_{n \in \mathbb{N}}$ from an observed one $\mathbf{y} = \{y_n\}_{n \in \mathbb{N}}$. In the classical probabilistic mixture model, observations are assumed mutually independent and the restoration problem consists in classifying data into a finite set of K classes.

It may also be interesting to co-analyze two sets of observations \mathbf{y}^1 and \mathbf{y}^2 to get two classifications denoted by \mathbf{x}^1 and \mathbf{x}^2 , different but dependent. In image segmentation, this situation is encountered in several applications:

- The joint segmentation of co-registered radar and optical satellite images or of Magnetic Resonance Imaging modalities (T1, T2, PD). It is interesting to note that the number of classes can be different in the two images;
- The joint segmentation of two images of the same scene but taken at different dates such that semantic differences between acquisitions can be observed, *e.g.* the lava flow during a volcano eruption in satellite imaging or the evolution of a disease in medical imaging. Another example is given by the emergence of clouds in a temporal set of optical satellite images of the Earth.

In these different contexts, each image provides information for modeling and classification of others. In this work, we develop a particular mixture model that introduces a joint prior between the two corresponding classifications and some statistical relations between the two series of observations. We arrive at a very rich model of multidimensional densities called “Joint Mixture Model” (Joint MM) which is formally defined as follows. We consider two series of observed and hidden data $(\mathbf{y}^1, \mathbf{x}^1)$ and $(\mathbf{y}^2, \mathbf{x}^2)$ with $\forall n \in [1, N]$, $x_n^1 \in \Omega^1 = \{1, \dots, K\}$ and $x_n^2 \in \Omega^2 = \{1, \dots, L\}$. Let us denote $\mathbf{y}_n = (y_n^1, y_n^2)'$ and $\mathbf{x}_n = (x_n^1, x_n^2)'$. In the joint mixture model we assume that data \mathbf{y}_n are realizations of mutually independent random vectors \mathbf{Y}_n with the same distribution characterized by the following finite mixture

$$f(\mathbf{y}_n) = \sum_{k=1}^K \sum_{l=1}^L \gamma_{k,l} f_{k,l}(\mathbf{y}_n),$$

where $\gamma_{k,l} = p(x_n^1 = k, x_n^2 = l)$ is the joint *a priori* probability between components such that $\sum_{k=1}^K \sum_{l=1}^L \gamma_{k,l} = 1$, and $f_{k,l}(\mathbf{y}_n)$ are mixture components.

By extending well-known results for the classical mixture model, Bayesian restoration can be achieved by means of two cost functions, resulting in Maximum A Posteriori (MAP) and Maximum of Posterior Marginals (MPM) classifications:

$$\text{MPM : } \begin{cases} \hat{x}_n^1 = \arg \max_{k \in \Omega^1} p(x_n^1 = k | \mathbf{y}_n) \\ \hat{x}_n^2 = \arg \max_{l \in \Omega^2} p(x_n^2 = l | \mathbf{y}_n) \end{cases}$$

$$\text{MAP : } \hat{\mathbf{x}}_n = \arg \max_{(k,l) \in \Omega^1 \times \Omega^2} p(\mathbf{x}_n = (k, l) | \mathbf{y}_n)$$

Also, in order to make classifications unsupervised, we also provide algorithms for parameters estimation using the EM (*Expectation-Maximization*) principle for Gaussian joint mixtures and the ICE (*Iterative Conditional Estimation*) principle for non-Gaussian ones. As a special case, we consider mixture components build by means of copula functions according to

$$f_{k,l}(\mathbf{y}_n; \tau_{k,l}) = f_{k,l}^{(1)}(y_n^1) f_{k,l}^{(2)}(y_n^2) c_{k,l} \left(F_{k,l}^{(1)}(y_n^1), F_{k,l}^{(2)}(y_n^2); \tau_{k,l} \right),$$

where $F_{k,l}^{(i)}$ is the distribution of the i^e margin, $f_{k,l}^{(i)}$ the corresponding density and $c_{k,l}(\cdot, \cdot; \tau_{k,l})$ the 2D copula density with parameter $\tau_{k,l}$.

The model is illustrated through the segmentation of color and IRM images, using (i) full Gaussian mixtures and (ii) mixtures of densities build upon a Gaussian copula and two gamma margins. Results are compared to the segmentations obtained by applying independent mixture models on individual bands. They clearly show the interest of taking into account cross-observation for classifications.

It is important to note that the Joint MM is not specifically designed for image segmentation but can be used in all places where the classical mixture model is employed. One promising perspective is to enrich the model with a Markovian structure in order to define a mixture of hidden Markov chains.

1. Introduction

Un problème important en traitement du signal consiste à restaurer un processus inobservable $\mathbf{x} = \{x_n\}_{n \in \mathbb{N}}$ à partir d'un processus observé $\mathbf{y} = \{y_n\}_{n \in \mathbb{N}}$. Dans le « modèle de mélange » classique, les observations sont supposées mutuellement indépendantes et le problème de la restauration consiste à classer les données en un nombre fini de K classes.

Il peut toutefois être intéressant de co-analyser deux séries d'observations \mathbf{y}^1 et \mathbf{y}^2 pour obtenir deux classifications, notées \mathbf{x}^1 et \mathbf{x}^2 , distinctes mais dépendantes. Dans le domaine de la segmentation d'images, ce type de modélisation peut être d'un grand intérêt dans plusieurs situations :

- segmentation conjointe de deux images satellites optique et radar recalées. Le nombre de classes peut être différent dans les deux images car les principes physiques d'observation de la Terre sont très différents pour ces deux modalités. Ce type de problème se rencontre également dans le domaine de l'imagerie médicale, avec en particulier la segmentation conjointe d'images T1, T2 et Densité de Protons en imagerie par résonance magnétique (IRM) ;

- segmentation de deux images recalées de même modalité mais dont l'une présente des données cachées. Ces données manquantes peuvent être dues à des changements intervenus entre les dates d'acquisition : impact d'une catastrophe naturelle en imagerie satellitaire ou évolution temporelle d'une maladie en imagerie médicale. Elles peuvent également apparaître en imagerie optique de la Terre suite à la présence de nuages occultant la scène observée.

Dans ces différents contextes, chaque image apporte une information pour la modélisation et la classification des autres. Rogers *et al.* (2008) ont développé un modèle de mélange permettant la co-analyse de séquences transcriptomique et protéomique par l'introduction d'un *a priori* commun sur les composantes du mélange. Dans ce travail nous étendons ces résultats en incluant un lien statistique entre les observations elle-mêmes, aboutissant à un modèle très riche de densités multidimensionnelles appelé « Modèle de Mélange Conjoint » (MMConjoint). Nous abordons le problème de l'estimation non supervisée des paramètres de ce modèle avec un algorithme de type *Expectation-Maximization* (EM) (Dempster *et al.*, 1977) dans le cas d'un mélange gaussien (Rogers *et al.*, 2008), et d'un algorithme de type Estimation Conditionnelle Itérative (ECI) (Pieczynski, 1992) dans le cas d'un mélange paramétrique non gaussien défini par des densités à copules (Derrode, Pieczynski, 2011). Notons que différentes comparaisons entre EM et ECI ont été effectuées dans le cadre de mélanges gaussiens classiques et il s'avère que, dans ce contexte, l'efficacité de deux méthodes est comparable (Peng, Pieczynski, 1995). Cependant, lorsque l'on considère des copules, le maniement de la vraisemblance peut être compliqué et l'application du EM difficile ; nous considérons en conséquence la méthode ECI, qui est plus facile à mettre en œuvre.

Cet article aborde la restauration bayésienne non supervisée d'un MMConjoint par extension des travaux concernant les MM classiques. Il s'organise de la manière suivante. La section 2 est consacrée à la présentation du modèle MMConjoint et propose deux critères bayésiens pour la restauration de données selon ce modèle. Le modèle est présenté dans le cas particulier de deux séries d'observations ; il se généralise cependant sans difficulté à un nombre quelconque (mais limité) de séries d'observations, comme l'illustrent les résultats expérimentaux impliquant 3 images. La section 3 aborde le problème de l'estimation non supervisée des paramètres du modèle, d'abord dans le cas gaussien à l'aide d'un algorithme EM, puis dans le cas d'un mélange de densités à copule à l'aide d'un algorithme de type ECI. Les résultats de tests comparatifs réalisés sur des données simulées sont également présentés. La section 4 présente l'application du modèle en segmentation d'images couleur et d'images IRM de modalités différentes. La dernière section conclut sur le modèle MMConjoint et propose quelques pistes pour l'enrichir.

2. Modèle de mélange conjoint

Soit $\mathbf{y} = \{y_1, \dots, y_N\}$ un ensemble de N données observées ($y_n \in \mathbb{R}$). Soit également $\mathbf{x} = \{x_1, \dots, x_N\}$ un ensemble de N données cachées dont chaque élément x_n appartient à $\Omega = \{1, \dots, K\}$. Dans le modèle classique de mélange probabiliste (Titterton *et al.*, 1985), les données y_n sont considérées comme des réalisations de variables aléatoires réelles Y_n mutuellement indépendantes et de même distribution définie par le mélange

$$f_{\text{MM}}(y_n) = \sum_{k=1}^K \pi_k f_k(y_n), \quad (1)$$

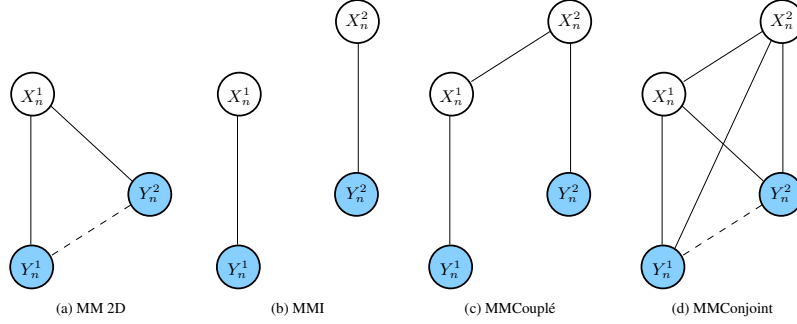


Figure 1. Graphes de dépendance pour les modèles (a) 2D, (b) produit, (c) couplé et (d) conjoint

où les densités $f_k(y_n) = p(y_n | x_n = k)$ sont parfois appelées lois d'« attache aux données » et $\pi_k = p(x_n = k)$ sont les probabilités *a priori*. Ce modèle s'interprète en considérant que les données sont distribuées selon K classes. La variable $x_n \in \Omega$ est un élément essentiel du problème mais elle ne peut pas être observée dans la pratique (x_n est une donnée cachée ou manquante).

2.1. Le modèle MMConjoint

On considère maintenant deux couples de données observées et cachées $(\mathbf{y}^1, \mathbf{x}^1)$ et $(\mathbf{y}^2, \mathbf{x}^2)$, de même dimension N , avec $\forall n \in [1, N]$, $x_n^1 \in \Omega^1 = \{1, \dots, K\}$ et $x_n^2 \in \Omega^2 = \{1, \dots, L\}$. Nous noterons $\mathbf{y}_n = (y_n^1, y_n^2)'$ et $\mathbf{x}_n = (x_n^1, x_n^2)'$. Dans le modèle de mélange conjoint, nous considérons que les couples de données \mathbf{y}_n sont des réalisations de vecteurs aléatoires \mathbf{Y}_n mutuellement indépendants ayant la même distribution caractérisée par le mélange fini suivant

$$f_{\text{MMConjoint}}(y_n^1, y_n^2) = \sum_{k=1}^K \sum_{l=1}^L \gamma_{k,l} f_{k,l}(y_n^1, y_n^2), \quad (2)$$

où $\gamma_{k,l} = p(x_n^1 = k, x_n^2 = l)$ est la probabilité *a priori* conjointe entre les composantes. Le graphe de dépendance de ce modèle est représenté dans la figure 1d. Ce modèle est différent d'un modèle classique bi-dimensionnelle (aussi appelé modèle de mélange vectoriel) dans lequel le vecteur d'observations \mathbf{y}_n est utilisé pour estimer une seule classification x_n (MM 2D), cf. figure 1a.

REMARQUE 1. — Pour simplifier l'exposé, nous décrivons le MMConjoint en nous limitant à deux séries d'observations même si le modèle s'écrit aisément pour un nombre fini de séries d'observations.

Le modèle de mélange couplé (MMCouplé) étudié dans (Rogers *et al.*, 2008) est un cas particulier du MMConjoint. En effet, si

$$\forall (k, l) \in \Omega^1 \times \Omega^2, f_{k,l}(y_n^1, y_n^2) = f_{k,l}^{(1)}(y_n^1) f_{k,l}^{(2)}(y_n^2)$$

avec $\forall l \in \Omega^2, f_{k,l}^{(1)}(\cdot) = f_k^{(1)}(\cdot)$ et $\forall k \in \Omega^1, f_{k,l}^{(2)}(\cdot) = f_l^{(2)}(\cdot)$, nous aboutissons au modèle couplé

$$f_{\text{MMCouplé}}(y_n^1, y_n^2) = \sum_{k=1}^K p(k) f_k^{(1)}(y_n^1) \sum_{l=1}^L p(l|k) f_l^{(2)}(y_n^2). \quad (3)$$

dont le graphe de dépendance est représenté dans la figure 1c.

Sous l'hypothèse supplémentaire $p(l|k) = p(l)$, le modèle se simplifie en un produit de deux mélanges indépendants (MMI)

$$f_{\text{MMI}}(y_n^1, y_n^2) = \sum_{k=1}^K \pi_k^{(1)} f_k^{(1)}(y_n^1) \sum_{l=1}^L \pi_l^{(2)} f_l^{(2)}(y_n^2) \quad (4)$$

dont le graphe de dépendance est décrit dans la figure 1b.

2.2. Paramètres du modèle

Le modèle fait intervenir KL densités de probabilités (ddp) bi-dimensionnelles modélisant les couples (y_n^1, y_n^2) conditionnellement aux composantes (k, l) du mélange. Dans un MMConjoint gaussien, toutes les lois d'attache aux données sont considérées gaussiennes. Ainsi, l'ensemble des paramètres caractérisant un tel modèle est décrit par les probabilités conjointes *a priori*, les vecteurs moyens et les matrices de covariance.

Il est possible d'étendre le cas gaussien en considérant la théorie des copules (Nelsen, 1999 ; Roncalli, 2002) qui enrichit les capacités de modélisation multivariées en séparant les problèmes de la recherche d'une forme adéquate pour les marginales et d'une structure de dépendance, appelée copule, entre les composantes du vecteur aléatoire. Dans ce cadre, chaque loi d'attache aux données du mélange conjoint s'écrit sous la forme

$$f_{k,l}(y_1, y_2; \tau_{k,l}) = f_{k,l}^{(1)}(y_1) f_{k,l}^{(2)}(y_2) c_{k,l} \left(F_{k,l}^{(1)}(y_1), F_{k,l}^{(2)}(y_2); \tau_{k,l} \right), \quad (5)$$

où $F_{k,l}^{(i)}$ représente la fonction de répartition de la i^{e} marge, $f_{k,l}^{(i)}$ la densité correspondante et $c_{k,l}(\cdot, \cdot; \tau_{k,l})$ la copule 2D de paramètres $\tau_{k,l}$. Rappelons qu'une copule bi-dimensionnelle est une ddp sur $[0, 1] \times [0, 1]$ dont les marges sont uniformes. L'ensemble des paramètres Θ d'un MMConjoint est donc constitué ainsi

$$\Theta = \{ \Theta_{k,l} \}_{(k,l) \in \Omega^1 \times \Omega^2}, \quad (6)$$

avec

$$\forall (k, l) \in \Omega^1 \times \Omega^2, \Theta_{k,l} = \left(\gamma_{k,l}, \tau_{k,l}, \theta_{k,l}^{(1)}, \theta_{k,l}^{(2)} \right),$$

et où $\theta_{k,l}^{(1)}$ et $\theta_{k,l}^{(2)}$ désignent les paramètres des deux marges $f_{k,l}^{(1)}$ et $f_{k,l}^{(2)}$. De très nombreuses copules ont été étudiées (Nelsen, 1999 ; Roncalli, 2002), parmi lesquelles les copules à un seul paramètre comme les copules gaussienne, de Clayton et de Gumbel-Hougaard. On trouvera les expressions des FdR, des ddp et du tau de Kendall de ces copules dans (Derrode, 2011).

Tableau 1. Paramètres utilisés pour représenter les mélanges gaussiens de la figure 2 et pour les simulations de la section 3.3

	(k, l)	$\gamma_{k,l}$	$\mu_{k,l}^{(1)}$	$\sigma_{k,l}^{2(1)}$	$\mu_{k,l}^{(2)}$	$\sigma_{k,l}^{2(2)}$	$\rho_{k,l}$
PMM	(1, 1)	0,30	0	2	1	2	0,8
	(1, 2)	0,20	0,4	2	-0,6	2	0,3
	(2, 1)	0,15	1,6	2	0,7	2	0,3
	(2, 2)	0,35	2	2	-1	2	0,7
CMM	(1, 1)	0,30	0	2	1	2	0
	(1, 2)	0,20	0	2	-1	2	0
	(2, 1)	0,15	2	2	1	2	0
	(2, 2)	0,35	2	2	-1	2	0

Il est ainsi possible de combiner des distributions marginales « classiques » (lois gaussienne, gamma, bêta de 1^{re} espèce...) avec différents modèles de dépendance. Par exemple, la copule gaussienne peut être combinée avec des marges gaussiennes, aboutissant à une gaussienne bi-dimensionnelle. De cette manière le modèle MMConjoint gaussien évoqué précédemment apparaît comme un cas particulier de MMConjoint avec copules où $\theta_{k,l}^{(i)} = \left\{ \mu_{k,l}^{(i)}, \sigma_{k,l}^{(i)} \right\}$ et la copule est paramétrée par la corrélation $\tau_{k,l} = \rho_{k,l}$. Il est aussi possible d'adjoindre à la copule gaussienne des marges gamma, de manière à construire un nouveau modèle paramétrique dont la dépendance est caractérisée par la corrélation et dont les fluctuations des données prises marginalement sont caractérisées par des lois gamma.

En utilisant ce potentiel de combinaisons dans l'équation (5), nous aboutissons à une modélisation d'une grande richesse, capable de prendre en compte des situations très variées sans augmenter significativement la complexité.

La figure 2 montre un exemple de mélange de $K = L = 2$ composantes pour les modèles de mélange couplé et conjoint gaussien. Les paramètres utilisés sont reportés dans le tableau 1. Un exemple de MMConjoint faisant intervenir des copules gaussiennes et des marges gamma est présenté dans (Derrode, Pieczynski, 2011, cf. figure 1). Voyons maintenant comment appliquer la décision bayésienne dans ce contexte.

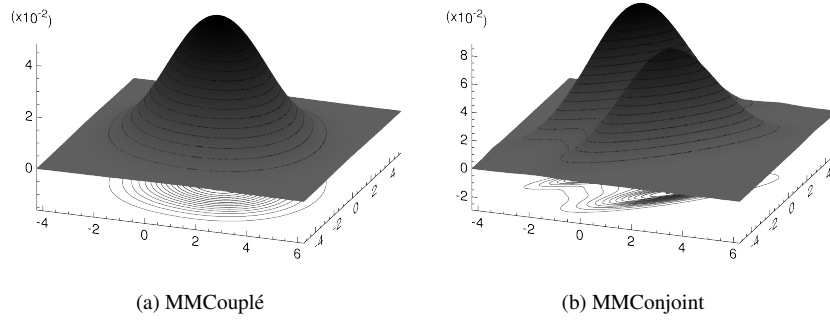


Figure 2. Exemples de mélanges couplé et conjoint gaussien utilisant les paramètres du tableau 1

2.3. Classification bayésienne

L'estimation de \mathbf{x}_n à partir de \mathbf{y}_n se fait par une règle de décision s . Une telle règle est caractérisée par une fonction de coût \mathcal{L} qui mesure l'erreur entre la vraie valeur \mathbf{x}_n et son estimée $\tilde{\mathbf{x}}_n$. L'estimateur bayésien $\hat{\mathbf{x}}_n = s(\mathbf{y}_n)$ est celui qui minimise le coût moyen, aboutissant à

$$\hat{\mathbf{x}}_n = \arg \min_{\tilde{\mathbf{x}}_n \in \Omega^1 \times \Omega^2} \sum_{\mathbf{x}_n \in \Omega^1 \times \Omega^2} \mathcal{L}(\mathbf{x}_n, \tilde{\mathbf{x}}_n) p(\mathbf{x}_n | \mathbf{y}_n).$$

Nous pouvons considérer les deux fonctions de coût suivantes :

$$\begin{aligned} \mathcal{L}_1(\mathbf{x}_n, \tilde{\mathbf{x}}_n) &= \mathbb{1}_{x_n^1 \neq \tilde{x}_n^1} + \mathbb{1}_{x_n^2 \neq \tilde{x}_n^2} \\ \mathcal{L}_2(\mathbf{x}_n, \tilde{\mathbf{x}}_n) &= \mathbb{1}_{\mathbf{x}_n \neq \tilde{\mathbf{x}}_n} = \mathbb{1}_{x_n^1 \neq \tilde{x}_n^1 \text{ ou } x_n^2 \neq \tilde{x}_n^2} \end{aligned}$$

où $\mathbb{1}_c = 1$ si la condition c est vraie, 0 sinon. \mathcal{L}_1 est appelée « Maximum a Posteriori des Marges » (MPM) alors que \mathcal{L}_2 est appelée « Maximum A Posteriori » (MAP). Après calculs, nous aboutissons à

$$MPM : \begin{cases} \hat{x}_n^1 = \arg \max_{k \in \Omega^1} p(x_n^1 = k | \mathbf{y}_n) \\ \hat{x}_n^2 = \arg \max_{l \in \Omega^2} p(x_n^2 = l | \mathbf{y}_n) \end{cases} \quad (7)$$

$$MAP : \hat{\mathbf{x}}_n = \arg \max_{(k,l) \in \Omega^1 \times \Omega^2} p(\mathbf{x}_n = (k,l) | \mathbf{y}_n) \quad (8)$$

où

$$p(x_n^1 = k | \mathbf{y}_n) = \sum_{l=1}^L p(\mathbf{x}_n = (k, l) | \mathbf{y}_n)$$

$$p(x_n^2 = l | \mathbf{y}_n) = \sum_{k=1}^K p(\mathbf{x}_n = (k, l) | \mathbf{y}_n).$$

Dans le modèle classique, le MAP et le MPM aboutissent à la même méthode.

Dans beaucoup d'applications, les paramètres ne sont pas connus ; il est donc intéressant de développer des méthodes adéquates d'apprentissage non supervisé.

3. Estimation non supervisée des paramètres

Nous abordons maintenant la question de l'estimation non supervisée des paramètres intrinsèques au mélange et des paramètres des lois d'attache aux données d'un MMConjoint, cf. équation (6). Nous distinguons le cas gaussien pour lequel un algorithme de type EM est proposé, du cas plus général avec copules où un algorithme de type ECI est présenté.

3.1. MMConjoint gaussien : EM

En notant δ le paramètre à estimer et $\delta^{(\ell)}$ une estimation de δ à l'itération ℓ , le principe de EM (McLachlan, Khrishnan, 2008) consiste à calculer

$$\delta^{(\ell)} = \arg \max_{\delta} Q(\delta^{(\ell-1)}, \delta) \quad (9)$$

de manière récursive jusqu'à obtenir la convergence de tous les paramètres estimés. Dans le cas d'un MMConjoint, en notant H la vraisemblance conjointe observations-états, Q s'écrit

$$Q(\delta^{(\ell-1)}, \delta) = \mathbb{E} \left[\underbrace{\ln H(\mathbf{x}^1, \mathbf{x}^2, \mathbf{y}^1, \mathbf{y}^2)}_{\mathcal{H}} \middle| \mathbf{y}^1, \mathbf{y}^2; \delta^{(\ell-1)} \right]$$

avec

$$\begin{aligned} \mathcal{H} &= \sum_{n=1}^N \ln \gamma_{k,l} f_{k,l}(y_n^1, y_n^2) \\ &= \sum_{n=1}^N \ln \sum_{k=1}^K \sum_{l=1}^L \gamma_{k,l} f_{k,l}(y_n^1, y_n^2) \mathbb{1}_{\mathbf{x}_n=(k,l)} \\ &= \sum_{n=1}^N \sum_{k=1}^K \sum_{l=1}^L \ln(\gamma_{k,l} f_{k,l}(y_n^1, y_n^2)) \mathbb{1}_{\mathbf{x}_n=(k,l)} \end{aligned}$$

En calculant l'espérance de \mathcal{H} , nous pouvons écrire $\xi_n^{(\ell-1)}(k, l) = \mathbb{E} [\mathbb{1}_{\mathbf{x}_n=(k,l)} | \mathbf{y}^1, \mathbf{y}^2; \delta^{(\ell-1)}]$ de la manière suivante

$$\begin{aligned} \xi_n^{(\ell-1)}(k, l) &= \sum_{i=1}^K \sum_{j=1}^L \mathbb{1}_{\mathbf{x}_n=(k,l)} p(\mathbf{x}_n = (i, j) | \mathbf{y}^1, \mathbf{y}^2; \delta^{(\ell-1)}) \\ &= p(\mathbf{x}_n = (k, l) | \mathbf{y}^1, \mathbf{y}^2; \delta^{(\ell-1)}) \\ &= p(\mathbf{x}_n = (k, l) | y_n^1, y_n^2; \delta^{(\ell-1)}) \\ &= \frac{p(\mathbf{x}_n = (k, l), \mathbf{y}_n; \delta^{(\ell-1)})}{p(\mathbf{y}_n; \delta^{(\ell-1)})} \end{aligned}$$

de telle manière que nous pouvons finalement écrire les probabilités conjointes *a posteriori* ainsi

$$\xi_n^{(\ell-1)}(k, l) = \frac{\gamma_{k,l}^{(\ell-1)} f_{k,l}^{(\ell-1)}(y_n^1, y_n^2)}{\sum_{i=1}^K \sum_{j=1}^L \gamma_{i,j}^{(\ell-1)} f_{i,j}^{(\ell-1)}(y_n^1, y_n^2)}. \quad (10)$$

Finalement, nous obtenons

$$Q(\delta^{(\ell-1)}, \delta) = \sum_{n=1}^N \sum_{k=1}^K \sum_{l=1}^L \ln \left(\gamma_{k,l}^{(\ell-1)} f_{k,l}^{(\ell-1)}(y_n^1, y_n^2) \right) \xi_n^{(\ell-1)}(k, l).$$

Selon l'équation (9), l'estimation à l'itération suivante est obtenue en maximisant Q . En utilisant des calculs classiques et en suivant l'approche développée pour le modèle original (Titterington *et al.*, 1985), nous obtenons les formules de ré-estimation des paramètres du MMConjoint gaussien :

$$\gamma_{k,l}^{(\ell)} = \frac{1}{N} \sum_{n=1}^N \xi_n^{(\ell-1)}(k, l) \quad (11)$$

$$\mu_{k,l}^{(\ell)} = \frac{\sum_{n=1}^N \xi_n^{(\ell-1)}(k, l) \mathbf{y}_n}{\sum_{n=1}^N \xi_n^{(\ell-1)}(k, l)} \quad (12)$$

$$\mathbf{\Gamma}_{k,l}^{(\ell)} = \frac{\sum_{n=1}^N \xi_n^{(\ell-1)}(k, l) (\mathbf{y}_n - \mu_{k,l}^{(\ell)}) (\mathbf{y}_n - \mu_{k,l}^{(\ell)})'}{\sum_{n=1}^N \xi_n^{(\ell-1)}(k, l)} \quad (13)$$

L'algorithme de classification non supervisé de séries de données selon le modèle MMConjoint gaussien est représenté en annexe A.

3.2. MMConjoint avec copules : ECI

Avec EM, il n'est pas aisé de déterminer les formules de ré-estimation des paramètres des lois d'attaches aux données dès lors que celles-ci ne sont plus gaussiennes. La méthode itérative ECI (Pieczynski, 1992) offre une solution alternative à ce problème. C'est une méthode générale d'estimation de paramètres dont le principe ne repose pas sur la maximisation de la vraisemblance comme EM mais sur le calcul de l'espérance d'un estimateur $\hat{\delta}(\mathbf{x}, \mathbf{y})$ du paramètre δ , défini sur les données complètes, conditionnellement aux données observées. Dans le cas d'un MMConjoint, la ré-estimation de δ à l'itération ℓ s'écrit

$$\delta^{(\ell)} = \mathbb{E} \left[\hat{\delta}(\mathbf{x}^1, \mathbf{x}^2, \mathbf{y}^1, \mathbf{y}^2) \mid \mathbf{y}^1, \mathbf{y}^2; \delta^{(\ell-1)} \right]. \quad (14)$$

Ce principe conduit à deux situations différentes selon que l'on cherche à estimer les probabilités jointes *a priori* ou les paramètres des lois d'attache aux données.

En ce qui concerne les probabilités jointes *a priori* $\gamma_{k,l}$, la ré-estimation ECI conduit à la même formule que dans le cas EM, cf. équation (11). Cette dernière repose sur le calcul des probabilités jointes *a posteriori* qui sont elles-mêmes inchangées, cf. équation (10).

Les paramètres des lois d'attache aux données, quant à eux, peuvent se calculer à partir d'estimateurs empiriques sur les données complétées $(\hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2, \mathbf{y}^1, \mathbf{y}^2)$, où $\hat{\mathbf{x}}^1$ et $\hat{\mathbf{x}}^2$ sont obtenus, à l'itération ℓ , par simulation, en effectuant des tirages selon les lois suivantes

$$p(x_n^1 = k \mid \mathbf{y}_n; \Theta^{(\ell-1)}) = \sum_{l=1}^L \xi_n^{(\ell-1)}(k, l), \quad (15)$$

$$p(x_n^2 = l \mid x_n^1 = k, \mathbf{y}_n; \Theta^{(\ell-1)}) = \frac{\xi_n^{(\ell-1)}(k, l)}{p(x_n^1 = k \mid \mathbf{y}_n; \Theta^{(\ell-1)})}. \quad (16)$$

Il est en général très aisé de trouver des estimateurs à partir des données complètes pour les paramètres des lois marginales et pour les paramètres de copules. En ce qui concerne les copules à un seul paramètre que nous avons évoquées dans la section précédente, les paramètres peuvent s'exprimer à l'aide du tau de Kendall $\tau_{k,l}$ (aussi appelé corrélation de rang) dont on trouve aisément, dans la littérature, un estimateur à partir des données complètes basé sur les taux de concordance et discordance.

L'algorithme de classification non supervisé de séries de données selon le modèle MMConjoint avec copules est représenté en annexe B.

3.3. Résultats expérimentaux d'estimation - Cas gaussien :

L'objectif de l'expérience est d'évaluer le comportement de l'algorithme d'estimation gaussien sur des données simulées selon la procédure suivante. Étant donné tous les paramètres du modèle, la simulation des deux couples de données (y^1, y^2) et (x^1, x^2) vérifiant l'équation (2) est réalisée ainsi

- tirage de x^1 , selon la loi $p(x^1) = \sum_{l=1}^L \gamma_{x^1, l}$;
- tirage de x^2 , selon la loi $p(x^2 | x^1) = \frac{\gamma_{x^1, x^2}}{p(x^1)}$;
- tirage de y^1 , selon la loi marginale gaussienne $f_{x^1, x^2}^{(1)}(y^1) = p(y^1 | x^1, x^2)$;
- tirage de y^2 , selon la loi conditionnelle gaussienne $p(y^2 | x^1, x^2, y^1)$.

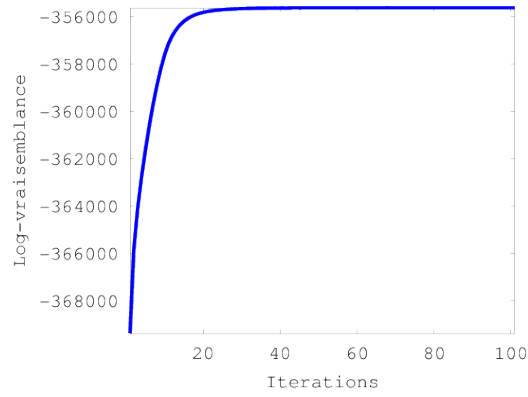
En itérant N fois le procédé de simulation, nous obtenons deux séries d'observations \mathbf{y}^1 et \mathbf{y}^2 , et leurs classifications correspondantes \mathbf{x}^1 et \mathbf{x}^2 . Ces dernières sont utilisées pour réaliser des mesures de taux d'erreur avec les classifications estimées. Pour les expériences présentées ci-après, nous avons simulé $N = 10000$ couples de données en utilisant les paramètres du tableau 1 (première ligne).

Tableau 2. Paramètres estimés par l'algorithme EM/MMConjoint (exp. section 3.3)

	(k, l)	$\gamma_{k, l}$	$\mu_{k, l}^{(1)}$	$\sigma_{k, l}^{2(1)}$	$\mu_{k, l}^{(2)}$	$\sigma_{k, l}^{2(2)}$	$\rho_{k, l}$
PMM	(1, 1)	0,30	0,01	2,07	1,02	2,08	0,96
	(1, 2)	0,18	0,29	1,86	-0,76	1,52	0,38
	(2, 1)	0,15	1,49	1,67	0,93	1,64	0,44
	(2, 2)	0,37	2,00	2,02	-0,99	2,04	0,89

En utilisant l'algorithme décrit en annexe A avec un nombre d'itérations EM fixé à $I = 100$, nous avons obtenu les paramètres reportés dans le tableau 2. Nous avons également reproduit dans la figure 3a l'évolution de la vraisemblance, et dans la figure 3b l'évolution des probabilités *a priori* $\gamma_{k, l}$ au cours des itérations EM. Dans cette expérience, EM converge rapidement et les paramètres estimés sont proches des paramètres utilisés pour simuler les données. Ce bon comportement est confirmé par les taux d'erreur de classification reportés dans le tableau 3, que ce soit pour le MAP (8) ou pour le MPM (7). En effet, on peut constater que les taux d'erreur dans le cas non supervisé sont proches des taux obtenus dans le cas supervisé. Pour confirmer expérimentalement la plus grande richesse du modèle conjoint par rapport au modèle couplé (équation (3)), nous avons également reporté les taux d'erreur obtenus grâce au MMCouplé (dans le cas supervisé) sur ce jeu de données. On peut constater une augmentation significative des taux d'erreur comparé au modèle conjoint. Comme attendu le modèle couplé n'est pas capable de capturer la complexité des données.

En appliquant ces algorithmes à la segmentation conjointe d'images, il est possible de vérifier visuellement l'apport du modèle conjoint par rapport à une segmentation indépendante des images avec un modèle de mélange classique.



(a) Vraisemblance

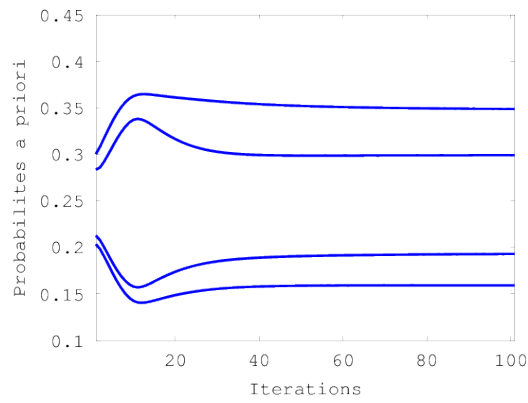
(b) $\gamma_{k,l}$

Figure 3. (a) Évolution de la vraisemblance. (b) Évolution de $\gamma_{k,l}$ pour $(k, l) \in \Omega^1 \times \Omega^2$. D'après le tableau 1, les valeurs attendues sont 0, 15, 0, 20, 0, 30 et 0, 35

Tableau 3. Taux d'erreur pour l'expérience de la section 3.3. S signifie « Supervisé » et NS « Non Supervisé »

	MPM		MAP	
	\mathbf{x}^1	\mathbf{x}^2	\mathbf{x}^1	\mathbf{x}^2
MMConjoint (NS)	15,70	14,06	16,25	14,21
MMConjoint (S)	15,59	14,11	15,85	14,19
MMCouplé (S)	22,78	21,50	21,96	20,73

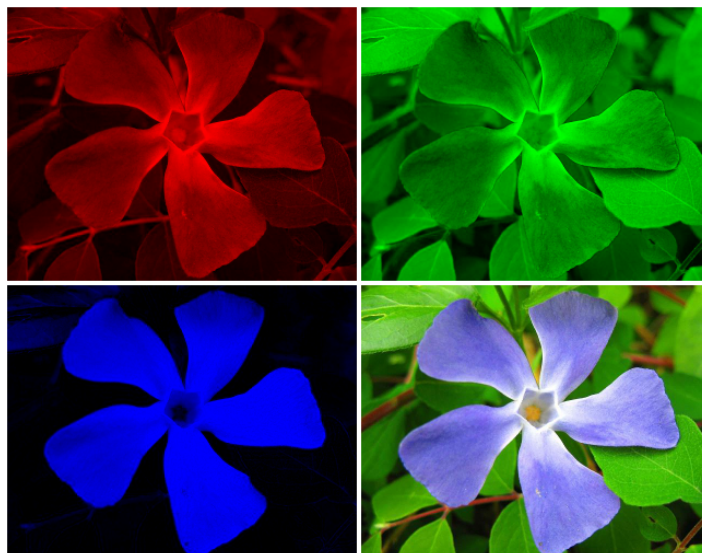


Figure 4. Une image couleur de fleur (bas droit), et sa décomposition R (haut gauche), V (haut droit) et B (bas gauche). Dimensions : 320×250

4. Segmentation non supervisée d'images

Cette section rapporte les résultats de segmentations obtenues avec le modèle conjoint pour différentes expériences portant sur des images multibandes (image couleur et images IRM de modalités différentes). Pour ces images, nous n'avons pas de « vérités-terrain » à notre disposition ou d'expert pour interpréter les résultats. Aussi la qualité et l'intérêt des résultats sont appréhendés par comparaison avec ceux obtenus en appliquant un modèle de mélange classique sur chaque bande indépendamment. Les deux expériences présentées ci-après font intervenir 3 images, ce qui illustre la capacité des algorithmes à traiter un nombre de bandes supérieur à deux.

4.1. Segmentation d'une image couleur

La première expérience concerne la segmentation d'une image couleur, cf. figure 4. Nous avons segmenté les 3 bandes colorimétriques avec, d'une part, des mélanges classiques gaussiens sur chaque bande (indépendamment les uns par rapport aux autres) et, d'autre part, avec le modèle de mélange conjoint gaussien. Les résultats sont reportés dans les figures 5 et 6 respectivement, en considérant à chaque fois 2 classes par bande. Pour un meilleur rendu des plans segmentés, nous avons remplacé le numéro de classe par la valeur moyenne des pixels attribués à cette classe. Nous avons également reconstruit l'image couleur par recombinaisons des plans colorimétriques segmentés, de manière à pouvoir comparer globalement les segmentations avec

l'image originale. Comme les bandes sont segmentées en 2 classes, les segmentations couleur sont formées de 8 classes.

Les segmentations couleur des deux modèles laissent apparaître des différences notables que l'on peut interpréter en analysant les plans colorimétriques segmentés. Ainsi, en comparant les segmentations des plans verts, il apparaît que le modèle classique a de grandes difficultés à dissocier la fleur des feuilles qui l'entourent. Les autres plans montrent également des différences notables qui se retrouvent dans les images segmentées couleur. Visuellement et comparativement à l'image originale, le résultat de segmentation couleur avec le modèle MMConjoint gaussien est plus satisfaisant et plus riche en détails que son équivalente pour le modèle de mélange indépendant. On peut pointer en exemple les fines branches marron et le cœur de la fleur.

4.2. Segmentation de modalités IRM

Cette seconde expérience porte sur la segmentation conjointe de trois modalités IRM d'une coupe de cerveau. Les images, reproduites dans la figure 7, ont été générées par le simulateur « BrainWeb » (www.bic.mni.mcgill.ca/brainweb/) (Kwan *et al.*, 1999). L'image couleur représente un rendu colorimétrique des trois coupes selon la correspondance suivante : T1→rouge, T2→vert et DP→bleue. Les résultats présentés ci-après sont extraits de (Derrode, Pieczynski, 2011).

Les résultats de segmentation par des modèles de mélange gaussien indépendants à $K = 2$ classes sont reportés dans la figure 8 ; l'image couleur est le composé colorimétrique des 3 classifications et comporte donc 8 classes. Les résultats obtenus en segmentant conjointement les 3 coupes selon le modèle MMConjoint sont reportés dans la figure 9a lorsque l'on considère des copules gaussiennes et des marges gaussiennes. La différence entre les segmentations indépendantes et les segmentations conjointes est particulièrement notable. Le modèle conjoint est capable de prendre en considération les corrélations entre les informations fournies par chaque bande pour améliorer la classification individuelle de chacune d'elles grâce aux copules gaussiennes.

Par contre, on ne constate que très peu de différences en comparant les résultats de segmentation du modèle MMConjoint où la seule différence réside dans la forme des marginales (gaussienne vs. gamma), cf. figure 9b. Dans ce cas, nous avons en effet pu constater que les lois gamma estimées par le MMC étaient très proches en forme des lois gaussiennes, ce qui est cohérent avec la nature gaussienne des bruits utilisés par le simulateur « BrainWeb ». On peut cependant légitimement se poser la question du choix, pour une application donnée, de la « meilleure » forme pour les lois marginales. Pour cela il est par exemple possible de calculer, à chaque itération ECI, la distance de Kolmogorov entre les fonctions de répartition (FdR) empiriques et les FdR de modèles paramétriques candidates – méthode introduite et utilisée avec succès dans un autre contexte dans (Giordana, Pieczynski, 1997) –, ou bien encore en considérant les lois candidates dans le système de Pearson (Derrode, 2009).

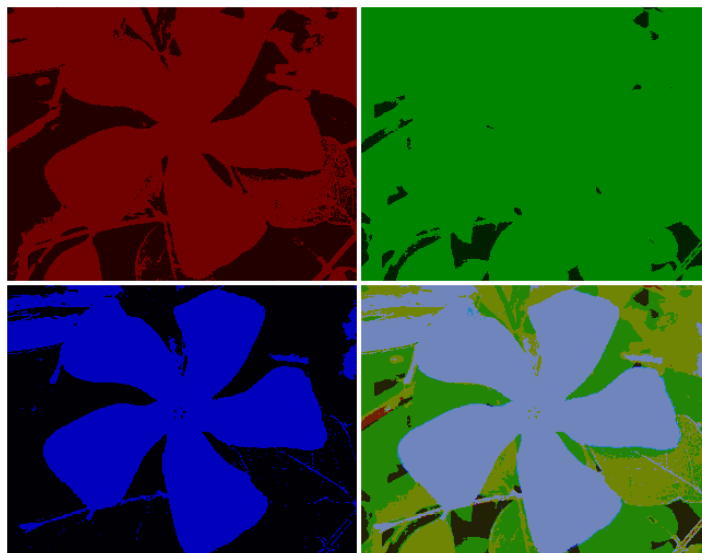


Figure 5. Segmentations individuelles des bandes en 2 classes chacune par des modèles de mélange indépendants, et recombinaison couleur (bas droit)

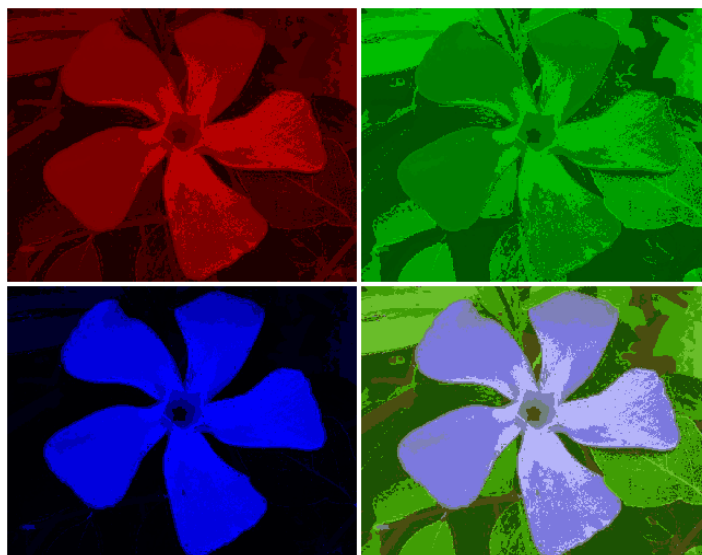


Figure 6. Segmentations des bandes en 2 classes chacune par le modèle de mélange conjoint gaussien, et recombinaison couleur (bas droit)

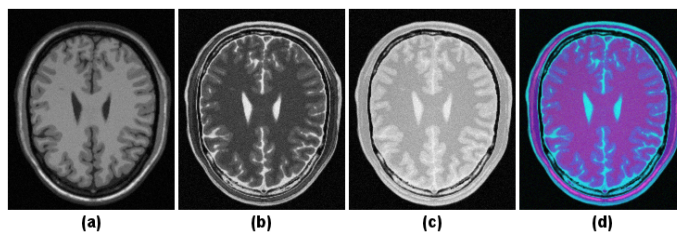


Figure 7. Une coupe IRM du cerveau en modalités (a) T1, (b) T2, (c) DP (Densité de Protons) et (d) rendu coloré des 3 images. Dimensions des images : 180×216

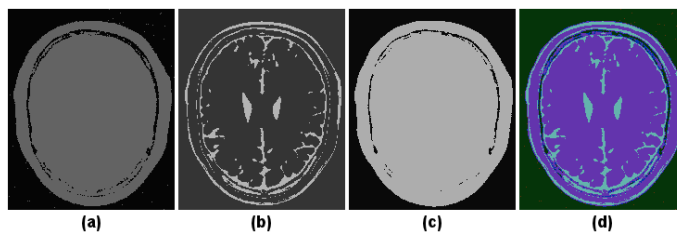
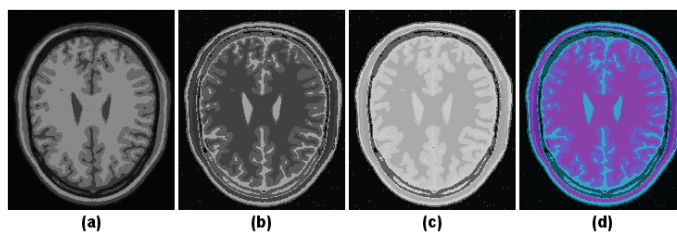
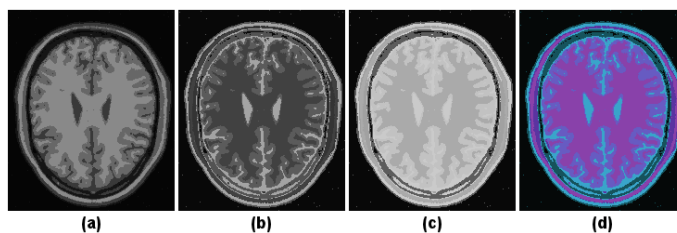


Figure 8. Résultats de segmentation (MPM) des 3 modalités par des modèles de mélange gaussien indépendants, avec 2 classes pour chaque coupe



(a) Marges gaussiennes



(b) Marges gamma

Figure 9. Résultats de segmentation (MPM) des 3 modalités IRM par le MMConjoint, avec des copules gaussiennes et (a) des marges gaussiennes, (b) des marges gammas, avec 2 classes pour chaque coupe

Lorsque l'on ne connaît pas *a priori* la forme de la copule à utiliser, il est possible d'intégrer dans l'estimation le choix automatique de la meilleur copule parmi un ensemble fini de candidates. On pourra par exemple utiliser des distances entre FdR ou bien encore la sélection bayésienne de copules proposée dans (Huard *et al.*, 2006).

5. Conclusion

Ce travail décrit un algorithme pour la classification conjointe de deux (ou plus) séries d'observations avec un modèle de mélange multidimensionnel appelé « Modèle de Mélange Conjoint ». Ce modèle prend en compte l'interdépendance entre les observations et entre les états cachés, résultant en un mélange très riche de composantes. Après avoir présenté deux critères bayésiens de classification adaptés au modèle, nous avons proposé deux algorithmes d'estimation non supervisée des paramètres du modèle, l'un de type EM pour le cas d'un mélange purement gaussien, l'autre de type ECI pour le cas de mélanges intégrant des densités à copule.

Les applications de ce type de modèle sont fort nombreuses. En particulier, nous avons illustré l'algorithme dans deux situations différentes : la segmentation d'images couleur et d'images multimodales en IRM. D'autres applications dans le domaine de la segmentation mériteraient d'être évaluées comme la segmentation vidéo, en considérant les images consécutives du flux vidéo comme entrées de l'algorithme MMConjoint. Il est aussi important de remarquer que le modèle n'est pas spécifique à la segmentation d'images mais peut être utilisé dans toute application nécessitant de classer conjointement plusieurs séries de données recalées (bio-informatique, finance...).

En ce qui concerne la modélisation proposée, la perspective la plus prometteuse consistera à enrichir le modèle de mélange conjoint en intégrant une structure de chaîne de Markov entre les séries d'états cachés. Ce dernier modèle pourra être vu comme une extension du modèle de chaîne de Markov cachée vectorielle, proposé dans (Brunel *et al.*, 2005).

Bibliographie

- Brunel N., Pieczynski W., Derrode S. (2005, 19-23 mars). Copulas in vectorial hidden Markov chains for multicomponent image segmentation. In *Proc. of the int. conf on acoustics, speech and signal processing (icassp'05)*. Philadelphie, USA.
- Dempster A., Laird N., Rubin D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. of the Royal Statistical Society, series B*, vol. 39, p. 1-38.
- Derrode S. (2009). *Description du système de Pearson pour son implémentation en langage C à l'aide de la librairie GSL*. Rapport de recherche, www.fresnel.fr/perso/derrode/HMC_Ext/Doc/Pearson.pdf. École Centrale Marseille & Institut Fresnel.
- Derrode S. (2011). *Some details about a few copulas*. Rapport de recherche, www.fresnel.fr/perso/derrode/HMC_Ext/Doc/ListCopulas.pdf. Institut Fresnel.
- Derrode S., Pieczynski W. (2011). Segmentation conjointe d'images et copules. In *Ateliers taima'11*. Hammamet, Tunisie.

- Giordana N., Pieczynski W. (1997). Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, n° 5, p. 465-475.
- Huard D., Évin G., Favre A.-C. (2006). Bayesian copula selection. *Computational Statistics & Data Analysis*, vol. 51, p. 809-822.
- Kwan R.-S., Evans A., Pike G. (1999). MRI-simulation based evaluation of image processing and classification methods. *IEEE trans. on Medical Imaging*, vol. 18, n° 11, p. 1085-1097.
- McLachlan G., Krishnan T. (2008). *The EM algorithm and extensions*. New York, John Wiley & Sons, Inc.
- Nelsen R. B. (1999). *An introduction to copulas*. New York, USA, Lecture Notes in Statistics.
- Peng A., Pieczynski W. (1995). Adaptive mixture estimation and unsupervised local Bayesian image segmentation. *Graphical Models and Image Processing*, vol. 57, n° 5, p. 389-399.
- Pieczynski W. (1992). Statistical image segmentation. *Machine Graphics and Vision*, vol. 1, n° 1/2, p. 261-268.
- Rogers S., Girolami M., Kolch W., Waters K., et al. (2008). Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. *Bioinformatics*, vol. 24, n° 24, p. 2894-2900.
- Roncalli T. (2002). *Gestion des risques multiples ou copules et aspects multidimensionnels du risque*. Rapport de recherche, www.thierry-roncalli.com/download/Lecture-Notes-Copula-Ensa.pdf. Groupe de Recherche Opérationnelle - Crédit Lyonnais.
- Titterton D. M., Smith A. F. M., Makov U. E. (1985). *Statistical analysis of finite mixture distributions*. John Wiley.

Annexe A. Algorithme d'estimation non supervisée des paramètres par EM (cas gaussien)

L'algorithme requiert en entrée $\mathbf{y}^1, \mathbf{y}^2, K$ et L .

1. Initialisation des paramètres à l'itération $\ell = 0$:

Classer \mathbf{y}^1 and \mathbf{y}^2 séparément en utilisant une méthode classique telle que l'algorithme des k-moyennes.

Utiliser des estimateurs empirique classiques pour obtenir $\Theta^{(0)}$.

2. Estimation selon EM.

for $\ell = 1$ to I **do**

Calculer les probabilités *a posteriori* $\xi_n^{(\ell-1)}(\cdot, \cdot)$ à partir de (10).

Calculer *a priori* probabilités $\gamma_{\cdot, \cdot}^{(\ell)}$ à partir de (11).

Calculer les paramètres des lois d'attache aux données $\mu_{\cdot, \cdot}^{(\ell)}$ et $\Gamma_{\cdot, \cdot}^{(\ell)}$ à partir de (12) et (13).

end for

3. Classification à partir de $\Theta^{(I)}$

Calculer les critères MPM and MAP en utilisant (7) et (8).

Annexe B. Algorithme d'estimation non supervisée des paramètres par ECI (cas avec copules)

L'algorithme requiert en entrée \mathbf{y}^1 , \mathbf{y}^2 , K et L .

1. Initialisation des paramètres à l'itération $\ell = 0$:

Classer \mathbf{y}^1 and \mathbf{y}^2 séparément en utilisant une méthode classique telle que l'algorithme des k-moyennes.

Utiliser des estimateurs empirique classiques pour obtenir $\Theta^{(0)}$.

2. Estimation selon EM.

for $\ell = 1$ to I **do**

Calculer les probabilités *a posteriori* $\xi_n^{(\ell-1)}(\cdot, \cdot)$ à partir de (10).

Calculer *a priori* probabilités $\gamma_{\cdot, \cdot}^{(\ell)}$ à partir de (11).

Simuler une réalisation *a posteriori* de \mathbf{x}^1 et \mathbf{x}^2 en utilisant (15) et (16).

Estimer les paramètres des lois marginales à partir de $\hat{\mathbf{x}}^1$ et $\hat{\mathbf{x}}^2$.

Estimer les taux de Kendall empiriques des copules à partir de $\hat{\mathbf{x}}^1$ et $\hat{\mathbf{x}}^2$.

end for

3. Classification à partir de $\Theta^{(I)}$

Calculer les critères MPM and MAP en utilisant (7) et (8).

Stéphane Derrode a obtenu sa thèse de 3^e cycle de l'université de Rennes en 1999, et une HDR d'Aix-Marseille Université en 2008. Depuis septembre 2001, il est maître de conférences à l'École Centrale Marseille dans le département Mathématiques et Informatique et chercheur au sein du Groupe Signaux Multidimensionnels de l'Institut Fresnel (CNRS UMR 7249). Ses centres d'intérêt portent sur les modèles markoviens non gaussiens pour la classification et le filtrage de séries temporelles de données. Les applications concernent plus particulièrement la segmentation d'images satellitaires et la biométrie dynamique.

Wojciech Pieczynski a reçu un doctorat d'Etat en Statistique Mathématique de l'Université Paris VI en 1986. Après avoir enseigné au Centre Universitaire de Tiaret, Algérie (1981-83), à l'Université de Brazzaville, Congo (1983-87), et à l'École Nationale Supérieure des Télécommunications de Bretagne (1987-90), il est professeur à Télécom SudParis depuis 1990, et responsable du département Communications, Images, et Traitement de l'Information (CITI) depuis septembre 2004. Ses recherches portent sur les modèles de Markov cachés et les traitements statistiques du signal et des images.