

---

## Modèles de mélange à copules: simulation, restauration et sélection

Stéphane Derrode\* — Wojciech Pieczynski\*\*

\* Institut Fresnel (CNRS UMR 7249), École Centrale Marseille,  
38, rue Frédéric Joliot-Curie, 13451 Marseille cedex 20  
stephane.derrode@centrale-marseille.fr

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

---

*RÉSUMÉ.* Les modèles de mélanges fournissent un moyen pratique mais formel pour la classification statistique de données. Ce travail introduit des modèles de mélange vectoriels paramétriques pour lesquels les densités conditionnelles aux classes sont construites à l'aide de copules. Après avoir détaillé le principe de la décision bayésienne dans ce contexte, nous abordons les problèmes d'estimation non supervisée des paramètres (méthodes EM, SEM et ECI), en y incluant une méthode de sélection des « meilleurs copules » dans un ensemble fini de copules candidates. Nous traitons enfin des modèles de mélanges multi-échelles, pour lesquels nous explicitons les algorithmes de simulation, restauration et estimation. Les algorithmes sont finalement illustrés par quelques résultats de segmentation d'images vectorielles.

*ABSTRACT.* Probabilistic mixture models provide a convenient yet formal framework for data classification. In this work, we introduce vectorial parametric mixture models for which the class conditional densities are built upon copulas. After giving details about Bayesian decision in this context, we turn to unsupervised parameters estimation (EM, SEM and ICE methods), including a method for selecting the “best copulas” within a finite set of possible copulas. We next deal with multi-scale mixture models, for which we furnish algorithms for simulation, restoration and estimation. Algorithms are finally illustrated through the segmentation of images.

*MOTS-CLÉS :* Classification bayésienne, Modèles de mélange, Copules, Estimation des paramètres (EM/SEM/ECI), Sélection de modèle, Segmentation d'images.

*KEYWORDS:* Bayesian classification, Probabilistic mixture model, Copulas, Parameters estimation (EM/SEM/ICE), Model selection, Image segmentation.

---

2 1<sup>re</sup> soumission à *Cahier Scientifique*

**Table des matières**

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Copules et lois multi-variées : définition et simulation</b>	<b>6</b>
2.1	Synthèse sur les copules . . . . .	6
2.2	Simulation de données 2D selon une méthode d'acceptation-rejet . . .	8
<b>3</b>	<b>Modèle de mélange à copules</b>	<b>10</b>
3.1	Restauration bayésienne . . . . .	10
3.2	Estimation des paramètres . . . . .	12
3.2.1	Algorithme EM . . . . .	12
3.2.2	Algorithmes SEM et ECI . . . . .	14
3.3	Sélection de copules . . . . .	15
3.4	Expériences . . . . .	16
<b>4</b>	<b>Modèle de mélange de mélanges</b>	<b>17</b>
4.1	Restauration bayésienne . . . . .	19
4.2	Estimation des paramètres . . . . .	21
4.2.1	Algorithme EM . . . . .	21
4.2.2	Algorithmes SEM et ECI . . . . .	22
<b>5</b>	<b>Segmentation non supervisée d'images vectorielles</b>	<b>23</b>
5.1	Segmentation d'une image couleur . . . . .	24
5.2	Segmentation de modalités IRM . . . . .	25
<b>6</b>	<b>Conclusion</b>	<b>27</b>
<b>7</b>	<b>Bibliographie</b>	<b>29</b>
<b>A</b>	<b>Liste de copules utilisées dans la partie expérimentale.</b>	<b>33</b>

## 1. Introduction

Notons  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ , un échantillon de  $N$  observations à  $M$  dimensions, indépendantes et issues d'une même distribution de densité de probabilité définie par

$$f(\mathbf{y}; \boldsymbol{\pi}) = \sum_{k=1}^K \pi_k f_k(\mathbf{y}), \quad [1]$$

où  $\pi_k$  représente la  $k^{\text{e}}$  proportion du mélange, ou encore la probabilité pour que l'observation  $\mathbf{y}$  appartienne à la  $k^{\text{e}}$  sous-population de densité correspondante  $f_k$ , appelée  $k^{\text{e}}$  densité du mélange. Ici,  $K$  représente le nombre total de composantes et  $\sum_{k=1}^K \pi_k = 1$ .

Habituellement, les densités du mélange ont une forme paramétrique  $f_k(\mathbf{y}) = f_k(\mathbf{y}; \boldsymbol{\theta}_k)$  connue, excepté le vecteur de paramètres  $\boldsymbol{\theta}_k$ . Ainsi, [1] s'écrit

$$f(\mathbf{y}; \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k f_k(\mathbf{y}; \boldsymbol{\theta}_k). \quad [2]$$

Dans la suite, le modèle  $f(\mathbf{y}; \boldsymbol{\theta})$  sera appelé Modèle de Mélange (MM) fini paramétré par  $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)$ . Nous ferons l'hypothèse que le nombre de composantes  $K$  est connu, de telle manière que seul le vecteur  $\boldsymbol{\theta}$  doit être estimé. Aussi, lorsque nous emploierons le terme de « sélection de modèle », il faudra comprendre la sélection de la forme des composantes, et non la sélection du nombre de composantes du mélange qui ne sera pas abordé dans ce papier ; le lecteur intéressé pourra par exemple consulter (Oliveira-Brochado et Martins, 2005).

C'est en 1874 (Pearson, 1894) qu'apparaît l'un des premiers modèles de mélange gaussien dans le cadre d'une étude concernant la théorie de l'évolution et portant sur des mesures biométriques réalisées sur un échantillon de 1000 crabes. Plus récemment, des mélanges de distributions de Poisson ont été utilisés pour la tomographie par émission de positrons (Vardi *et al.*, 1985), et pour la classification de documents et la recherche d'informations textuelles (Li et Zha, 2006). On peut également citer le mélange de distributions de von Mises-Fisher pour l'analyse de données sur la sphère avec applications à la description de textes et l'expression de gènes (Banerjee *et al.*, 2005). Le lecteur intéressé pourra se référer à (McLachlan et Peel, 2000 ; Titterton *et al.*, 1985) pour un état de l'art complet sur l'histoire et les applications des modèles de mélange finis. On consultera aussi avec intérêt les textes (Frühwirth-Schnatter, 1985 ; Bohning *et al.*, 1998 ; Melnykov et Maitra, 2010) portant sur la théorie, les applications et les développements de ces modèles.

Lorsque l'on considère les MM comme des classifieurs de données, on cherche à attribuer chaque observation de l'échantillon à l'une des  $K$  classes. En accord avec la théorie bayésienne de la décision, cela peut se faire selon le critère du maximum *a posteriori*. C'est à dire qu'une observation  $\mathbf{y}_n$  est associée à la classe qui donne

la plus grande probabilité *a posteriori* d'être issue de ce groupe. Cela est équivalent à rechercher l'index de la classe correspondant au plus grand  $\pi_k f_k(\mathbf{y}_n)$ ,  $k \in \Omega = [1, K]$ .

Dans le domaine de la reconnaissance de formes et du traitement d'images, les MM sont exploités dans différents contextes : ils sont appliqués soit directement sur les intensités des pixels (mono- ou multi-spectral), soit sur des descripteurs extraits des images tels des coefficients d'ondelettes pour la segmentation des images texturées (Permuter *et al.*, 2006). On peut également citer l'utilisation de MM (gaussien ou plus récemment gaussien généralisé) pour l'estimation du fond dans une séquence vidéo (Stauffer, 1999), ou bien encore pour la poursuite d'objets d'intérêt avec mise à jour adaptative des informations du mélange au fur et à mesure de l'évolution de la scène (mélange gaussien (McKenna *et al.*, 1999), mélange dans le système de Pearson (Ketchantang *et al.*, 2008)).

Ces modèles donnent parfois des résultats très intéressants, même si l'hypothèse d'observations indépendantes est parfois difficile à justifier en traitement d'images où le contexte spatial ou temporel apporte une information primordiale pour la classification. Récemment, pour pallier à ce problème sans entrer dans la complexité des modèles par champs de Markov cachés, des MM spatialement adaptés ou contraints ont été proposés dans la littérature (Sanjay-Gopal et Hebert, 1998 ; Sfikas *et al.*, 2008 ; Nguyen *et al.*, 2012 ; Nguyen et Wu, 2012) ; modèles pour lesquels les poids du mélange dépendent de la position du pixel dans l'image  $\pi_k = \pi_{k,n}$ .

Le modèle de mélange le plus populaire reste celui composé de distributions gaussiennes. Outre la fréquence des données gaussiennes observées, le succès du Modèle de Mélange Gaussien (MMG) s'explique par sa simplicité de mise à oeuvre grâce à la possibilité d'écrire des formules explicites de ré-estimation des paramètres dans le cadre de l'estimation de paramètres par maximum de vraisemblance et de l'algorithme *Expectation-Maximization* (EM) (Dempster *et al.*, 1977 ; Wu, 1983 ; McLachlan et Krishnan, 2008). La maximisation est locale et l'estimation initiale des paramètres est donc primordiale. D'autre part, en dehors du cadre gaussien, il est souvent difficile de déterminer des formules de ré-estimation explicite des paramètres dans la phase de maximisation. Pour pallier à ces difficultés, d'autres algorithmes d'inférence ont été proposés qui reposent en partie sur la possibilité de simuler les données cachées conditionnellement aux données observées :

- l'algorithme SEM (*Stochastic EM*) (Celeux et Diebolt, 1985) qui remplace la phase « Espérance » par une phase « Espérance Stochastique ». Cette phase consiste à simuler des réalisations *a posteriori* sur la base de la valeur courante du paramètre pour approcher l'espérance conditionnelle de la vraisemblance complétée ;

- l'algorithme ECI (Estimation Conditionnelle Itérative) (Pieczynski, 1992) qui ne repose plus sur la recherche du maximum de vraisemblance mais sur le calcul de l'espérance d'un estimateur de chaque paramètre, défini sur les données complètes conditionnellement aux données observées.

Dans un MM, chaque groupe de données  $k$  a sa propre distribution  $f_k(\mathbf{y}; \boldsymbol{\theta}_k)$  et sa propre probabilité d'apparition dans l'échantillon  $\pi_k$ . De nombreux travaux considèrent les MM homogènes, *i.e.* pour lesquels toutes les distributions sont issues d'un même modèle paramétrique, *e.g.* modèle gaussien, gamma, Student't... ; ce qui les diffère étant la valeur des paramètres  $\boldsymbol{\theta}_k$ . Pour prendre en compte des classes de distribution variées au sein de même mélange, des « MM généralisés » ont été proposés dans lequel les formes des densités sont choisies de manière automatique dans un ensemble fini

- soit en minimisant la distance de Kolmogorov entre la fonction de répartition empirique de l'échantillon et les fonctions de répartition de toutes les distributions candidates,

- soit en sélectionnant la « meilleur » famille dans un système de distributions comme celui de Pearson (Delignon *et al.*, 1997 ; Bougarradh *et al.*, 2009).

Dans le cas d'observations vectorielles, le problème est encore accru car l'utilisation de distributions paramétriques multi-variées non gaussiennes est généralement problématique<sup>1</sup>. Récemment, la théorie des copules a suscité un intérêt grandissant dans le domaine du traitement du signal et des images, avec les travaux précurseurs de N. Brunel (Brunel et Pieczynski, 2005) et ceux de P. Varshney (Sundaresan et Varshney, 2011 ; Iyengar *et al.*, 2011), alors qu'elle est utilisée depuis de nombreuses années dans le domaine de la finance. Les copules accroissent les possibilités de modélisations multi-variées en séparant le problème de la recherche d'une forme adéquate pour les marginales du problème de la recherche d'une structure de dépendance pour le vecteur aléatoire (Nelsen, 2006 ; Balakrishnan et Lai, 2009). Il est ainsi possible de croiser différentes marginales classiques et différents modèles de dépendance pour obtenir une large gamme de modèles multi-variés originaux. Ces copules permettent théoriquement de représenter n'importe quelle distribution multi-variée (théorème de Sklar).

Dans ce travail, nous présentons les résultats d'une étude portant sur la possibilité de simuler et de restaurer de manière non supervisée des modèles de mélanges généralisés à base de copules (MMC), c'est-à-dire pour lesquels les ensembles de paramètres  $\boldsymbol{\theta}_k$  comprennent les paramètres des marges (gaussienne, gamma...) et les paramètres spécifiques de la copule (Gaussienne, Student't, Gumbel, Clayton...). Nous étudions en particulier la capacité des algorithmes d'estimation à retrouver la forme de toutes les copules du mélange en sélectionnant les plus adéquates (selon un certain critère proposé par D. Huard (Huard *et al.*, 2006)) dans un ensemble fini de copules candidates. Nous aborderons également le cas des « mélanges de mélanges », encore appelés « mélanges multi-échelles », pour lesquels la distribution de chaque classe  $f_k(\mathbf{y}; \boldsymbol{\theta}_k)$

1. Il existe par exemple de nombreuses formes différentes de lois gamma multi-dimensionnelles (Kotz *et al.*, 2008), toutes d'ailleurs ne donnant pas des marginales gamma !

6 1<sup>re</sup> soumission à *Cahier Scientifique*

est considérée comme un mélange de distributions, permettant en cela de considérer des groupes de formes complexes dont les classes multi-modales.

La suite du papier se décline de la façon suivante. Dans le chapitre 2, nous commençons par effectuer une brève présentation de la théorie des copules, et par décrire un algorithme permettant de simuler un échantillon de données selon une loi bi-dimensionnelle spécifiée par copule. Dans le chapitre 3, nous décrivons le modèle de mélange à copules : restauration bayésienne, estimation de paramètres (EM, SEM et ECI) et sélection de copules. Dans le chapitre 4, nous présentons une extension du MM –les modèles de mélange multi-échelles ou MMM–, pour lesquels deux critères de décision bayésienne et trois algorithmes d’estimation de paramètres adaptés au modèle sont décrits. Enfin, le chapitre 5 illustre les méthodes de classification en segmentation non supervisée d’images vectorielles. Nous terminons en proposant dans la conclusion plusieurs perspectives d’enrichissement et d’approfondissement.

## 2. Copules et lois multi-variées : définition et simulation

Nous commençons par faire une brève synthèse sur la théorie des copules, puis nous exposons un algorithme d’acceptation-rejet pour simuler des données selon une loi bidimensionnelle spécifiée par copule.

### 2.1. Synthèse sur les copules

Les fonctions copules sont des distributions multi-variées  $C$  définie sur l’hypercube unité  $[0, 1]^M$  avec des distributions marginales uniformes sur l’intervalle  $[0, 1]$ . Plus précisément,

$$C(u_1, \dots, u_M) = Pr(U_1 \leq u_1, \dots, U_M \leq u_M), \quad [3]$$

avec  $U_m \rightsquigarrow U_{[0,1]}$  pour  $m = 1, 2, \dots, M$ . De plus, si le vecteur  $(Y_1, \dots, Y_M) \in \mathbb{R}^M$  a une distribution multi-variée continue  $F(y_1, \dots, y_M) = Pr(Y_1 \leq y_1, \dots, Y_M \leq y_M)$  alors, étant donné le théorème de Sklar, cf. (Nelsen, 2006), il existe une unique fonction copule  $C : [0, 1]^M \rightarrow [0, 1]$  de  $F$  telle que

$$F(y_1, \dots, y_M) = C(F_1(y_1), \dots, F_M(y_M)), \quad [4]$$

pour tout  $(y_1, \dots, y_M) \in \mathbb{R}^M$  et pour toutes distributions marginales continues  $F(y_m) = Pr(Y_M \leq y_m)$ ,  $y_m \in \mathbb{R}$ .

En plus si  $C$  est une copule sur  $[0, 1]^M$  et  $F_1(y_1), \dots, F_M(y_M)$  sont des fonctions de répartition continues sur  $\mathbb{R}$ , alors  $F(y_1, \dots, y_M) = C(F_1(y_1), \dots, F_M(y_M))$  est une fonction de répartition sur  $\mathbb{R}^M$  avec comme distributions marginales uni-variées  $F_1(y_1), \dots, F_M(y_M)$ .

Ainsi, il est possible de construire des distributions multi-variées avec des distributions marginales pré-spécifiées et une fonction copule qui représente la struc-

Tableau 1 – Paramètres des marges utilisées dans la figure 2 et dans la partie expérimentale.

Marges gaussiennes $\mathcal{N}(\mu, \sigma^2)$ $p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$	$f_1 \rightsquigarrow \mathcal{N}(0.00, 1.00)$ $f_2 \rightsquigarrow \mathcal{N}(1.00, 2.00)$
Marges gamma $\mathcal{G}(\mu, \alpha, \theta)$ $p(y) = \frac{\theta^\alpha}{\Gamma(\alpha)} (y - \mu)^{\alpha-1} e^{-\theta(y-\mu)}$	$f_1 \rightsquigarrow \mathcal{G}(1.50, 8.00, 0.13)$ $f_2 \rightsquigarrow \mathcal{G}(2.00, 2.67, 1.12)$

ture de dépendance désirée. En dérivant [4] par rapport à  $y_1, \dots, y_M$  et en notant  $c(u_1, \dots, u_M) = \frac{\partial C(u_1, \dots, u_M)}{\partial u_1 \dots \partial u_M}$ , nous obtenons

$$f(y_1, \dots, y_M) = f_1(y_1) \dots f_M(y_M) c(F_1(y_1), \dots, F_M(y_M)). \quad [5]$$

Le lecteur intéressé par de plus amples détails sur les copules trouvera dans les notes de cours de T. Roncalli (Roncalli, 2002) une synthèse pédagogique et complète.

Dans la partie expérimentale, nous nous intéresserons aux copules bi-dimensionnelle  $c_0, \dots, c_8$  décrites dans l'annexe A. Le tableau donne des détails concernant leur fonction de répartition, leur densité de probabilité, l'intervalle de validité de leur paramètre  $\theta$  et celui du tau de Kendall  $\tau^2$  d'expression

$$\tau = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1. \quad [6]$$

Ces fonctions copules ont été choisies car elles sont toutes à un seul paramètre  $\theta^3$  et que celui-ci peut s'exprimer simplement en fonction de  $\tau$ . Ainsi les deux écritures  $c_p(u_1, u_2; \theta)$  et  $c_p(u_1, u_2; \tau)$  seront interchangeables. Il est important de noter que les valeurs admissibles pour  $\tau$  ne sont pas les mêmes pour toutes les copules : certaines d'entre-elles n'admettent pas la valeur  $\tau = 0$ , alors que d'autres n'admettent pas de corrélations négatives. Dans cet ensemble, seules la copule gaussienne et la copule de Student permettent de couvrir l'ensemble des valeurs admissibles pour  $\tau \in [-1, 1]$ .

Dans la figure 1 sont représentées les densités 2D d'une copule gaussienne ( $c_1$ ) et d'une copule A12 ( $c_7$ ) avec le même tau de Kendall ( $\tau = 0.5$ ). La figure 2 représente des densités de lois 2D construites à partir de ces deux copules et de deux

2. Le tau de Kendall est aussi appelé corrélation de rang. D'autres mesures de dépendance existent comme le  $\rho$  de Spearman ou l'indice de Gini ou encore, la corrélation médiane de Blomquist, cf. (Mercier, 2005).

3. exceptés (i)  $c_2$  la copule de Student pour laquelle le degré de liberté  $\nu$  est supposé connu, et (ii)  $c_0$  la copule produit qui ne possède pas de paramètre et qui permet de modéliser l'indépendance.

8 1<sup>re</sup> soumission à *Cahier Scientifique*

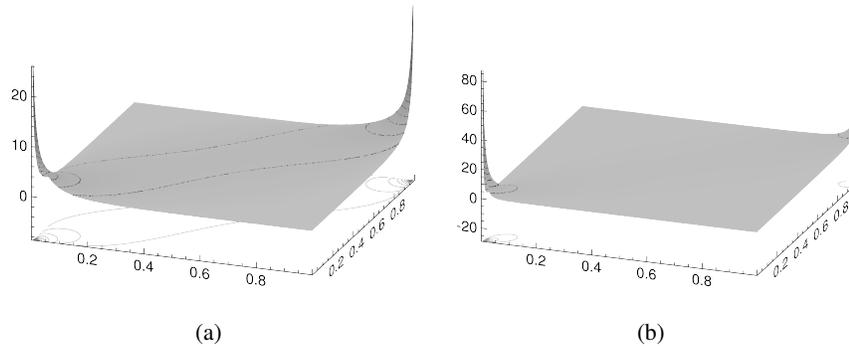


Figure 1 – Densité d’une copule gaussienne (a) et d’une copule A12 (b) avec  $\tau = 0.5$ .

marges gaussiennes ou de deux marges gamma. Ces marges ont, deux à deux, mêmes moyennes et mêmes variances, cf. tableau 1.

## 2.2. Simulation de données 2D selon une méthode d’acceptation-rejet

Soient  $Y_1$  et  $Y_2$  deux variables réelles ayant pour densité de probabilité  $f^{(1)}(\cdot)$  et  $f^{(2)}(\cdot)$ , et pour fonction de répartition  $F^{(1)}(\cdot)$  and  $F^{(2)}(\cdot)$ . En considérant une représentation par copule pour  $f(y_1, y_2)$  selon [5], le tirage de couples de données  $(y_1, y_2)$  se fait d’abord en tirant  $y_1$  avec la loi donnée par  $f^{(1)}$ <sup>4</sup>, puis en tirant  $y_2$  avec la loi donnée par

$$p(Y_2 = y_2 | y_1) = f^{(2)}(y_2) c\left(F^{(1)}(y_1), F^{(2)}(y_2)\right) = f_{y_1}(y_2). \quad [7]$$

Pour cela, nous proposons d’appliquer le principe de la méthode d’acceptation-rejet (Devroye, 1986), qui a l’avantage d’être indépendante de la forme de la copule, même si elle procure généralement un taux de rejet important.

Soit un nombre réel  $R$  et une densité de probabilité  $g$  telle que  $\forall x \in \mathbb{R}, f_{y_1}(x) \leq R g(x)$ , la simulation de  $Y_2$  conditionnellement à  $Y_1 = y_1$  peut être réalisé par l’algorithme suivant :

- 1) Échantillonne  $Y_2 = y_2$  selon  $g$  ;
- 2) Échantillonne  $V = v$  selon  $\mathcal{U}([0, 1])$ , la loi uniforme sur  $[0, 1]$  ;
- 3) Accepte  $y_2$  comme échantillon si

$$v \leq \frac{f_{y_1}(y_2)}{R g(y_2)}. \quad [8]$$

4. ce qui se fait aisément pour des marges classiques telles les lois Gaussienne, Gamma, Beta. . .

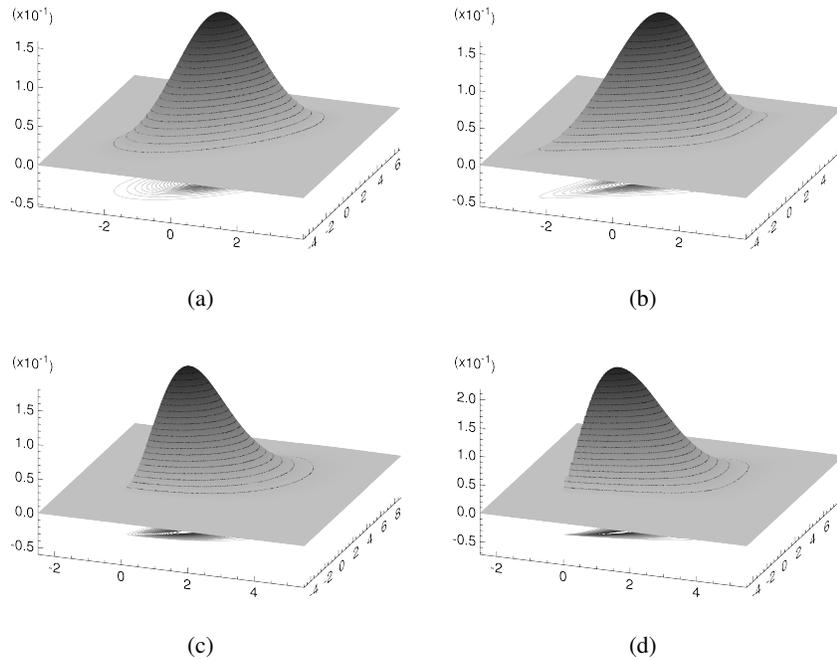


Figure 2 – Exemple de densités construites à l’aide d’une copule gaussienne (colonne de gauche) et d’une copule A12 (colonne de droite), pour des marges gaussiennes (première ligne) et des marges gamma (deuxième ligne). Les paramètres des marges sont donnés dans le tableau 1.

4) Sinon, retourne à l’étape 1.

En choisissant  $R = \max_{u_2 \in [0,1]} c(u_1, u_2)$  et  $g = f^{(2)}$ , le test de l’eq. [8] s’écrit maintenant

$$v \leq \frac{c(u_1, F^{(2)}(y_2))}{\max_{u_2 \in [0,1]} c(u_1, u_2)}$$

où  $u_1 = F^{(1)}(y_1)$ .

Ainsi, il est possible de générer des échantillons de  $Y_2|Y_1$  quelque soit la forme de la copule, une fois que l’on connaît un moyen de simuler un échantillon selon la loi de  $Y_2$ . L’efficacité de l’algorithme, *i.e.* le nombre de réjections avant une acceptation de l’échantillon  $y_2$ , dépend de la valeur de  $R$  et ainsi de la forme de la copule. Le coefficient  $R$  a une forme analytique pour certaines copules utilisées dans ce travail ; pour les autres, une méthode numérique peut facilement être utilisée.

10 1<sup>re</sup> soumission à *Cahier Scientifique*

### 3. Modèle de mélange à copules

Dans ce chapitre, nous considérons donc le modèle de mélange suivant

$$f(\mathbf{y}; \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k f_k(\mathbf{y}; \boldsymbol{\theta}_k), \quad [9]$$

où les densités  $f_k$  sont spécifiées par des copules (acronyme : MMC). Ainsi nous noterons  $\boldsymbol{\theta} = \{\pi_1, \dots, \pi_K, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K\}$  avec  $\forall k \in \Omega = [1, K], \boldsymbol{\theta}_k = \{\tau_k, \psi_k^1, \dots, \psi_k^M\}$  et  $\psi_k^i$  les paramètres propres à la  $i^e$  marginale de la classe  $k$ . On peut remarquer que si toutes les copules sont gaussiennes et que toutes les marges le sont également, alors nous retrouvons un modèle de mélange multi-dimensionnel gaussien classique.

Nous présentons dans un premier temps la restauration bayésienne de données selon ce modèle, puis les algorithmes les plus usités pour l'estimation des paramètres et, enfin, une méthode de sélection de la meilleur copule dans une ensemble fini de modèles candidats. Nous construisons ainsi un algorithme de restauration non supervisées très général où seuls le nombre de classes et la forme des marginales doivent être précisés. La dernière section rapporte les résultats d'expériences systématiques sur des données simulées qui illustrent le bon comportement de la méthode.

#### 3.1. Restauration bayésienne

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

$$\hat{x}_n = \arg \min_{\tilde{x}_n \in \Omega} \sum_{x_n \in \Omega} L(x_n, \tilde{x}_n) p(x_n | \mathbf{y}_n).$$

Sans information *a priori*, on considère généralement la fonction de perte « 0-1 » définie par  $L_{0-1}(x_n, \tilde{x}_n) = \mathbb{1}_{x_n \neq \tilde{x}_n}$  où  $\mathbb{1}_c = 1$  si la condition  $c$  est vraie, 0 sinon. Nous aboutissons au résultat bien connu suivant

$$\hat{x}_n = \arg \max_{k \in \Omega} p(X_n = k | \mathbf{y}_n) \quad [10]$$

qui consiste à sélectionner le classe qui maximise la probabilité *a posteriori*  $p(X_n = k | \mathbf{y}_n) = \gamma_n(k) \propto \pi_k f_k(\mathbf{y}; \boldsymbol{\theta}_k)$ .

Dans l'objectif de mesurer l'influence numérique des formes des copules sur le mélange, nous avons conduit l'expérience suivante

1) Simulation de  $N$  données complètes  $(x_n, \mathbf{y}_n)$  selon un MMC à  $M = 2$  dimensions et  $K = 2$  classes avec deux poids  $\pi_k$ , deux copules de même forme et de même tau de Kendall  $\tau$  et quatre marges données. La simulation se fait classiquement

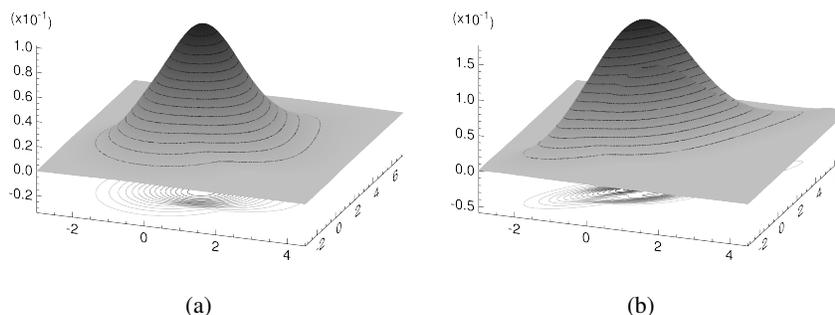


Figure 3 – Exemples de mélanges de copules de Gumbel avec des paramètres identiques excepté le tau de Kendall.

en deux étapes. D’abord le tirage de  $x_n$  est effectué selon les poids  $\pi_k$ , puis le tirage du vecteur d’observations  $\mathbf{y}_n$  est effectué selon la densité  $f_k(\mathbf{y}; \boldsymbol{\theta}_k)$ , en utilisant la méthode d’acceptation-rejet présentée dans la section 2.2.

2) Ensuite, restauration des observations simulées en utilisant tous les paramètres utilisés pour la simulation excepté la forme de la copule que l’on fait varier parmi les copules de la liste de l’annexe A.

Les expériences présentées ci-dessous ont été conduites avec  $N = 2000$  données, avec les poids suivants :  $\pi_1 = 0.6$  et  $\pi_2 = 0.4$ . Les marges de la première et de la deuxième copule sont respectivement données par les deux lois gaussiennes et les deux lois gamma du tableau 1. Nous avons considéré deux valeurs du tau de Kendall :

–  $\tau_1 = 0.16$  ce qui correspond à une corrélation linéaire de 0.25. L’ensemble des copules éligibles est noté  $\Pi_1 = \{c^1, c^2, c^3, c^4, c^5, c^6\}$ .

–  $\tau_2 = 0.70$  ce qui correspond à une corrélation linéaire de 0.89. L’ensemble des copules éligibles est noté  $\Pi_2 = \{c^1, c^2, c^3, c^6, c^7, c^8\}$ .

Les taux d’erreur moyen de classification reportés dans les tableaux 2(a) (pour  $\tau_1$ ) et 2(b) (pour  $\tau_2$ ) sont des résultats moyens de 300 expériences indépendantes. Les résultats soulignés en gras représentent le minimum de chaque colonne. Des exemples de mélanges utilisés dans l’expérience sont dessinés dans la figure 3.

Les commentaires sur les résultats peuvent se résumer à travers ces quelques points :

– Quelque soit la valeur du tau de Kendall, la restauration avec la copule adéquate donne toujours le taux moyen d’erreur le plus faible (à une exception près donnée par la copule de Student  $c_1$  qui est confondue avec la copule FGM  $c_4$  lorsque  $\tau = 0.16$ ) ;

– Lorsque  $\tau$  est bas, les taux d’erreur moyens peuvent être très proches, montrant que des confusions peuvent apparaître dans les expériences si la corrélation est faible.

Tableau 2 – Taux d’erreur de classification moyen pour l’expérience de la section 3.1 avec  $\tau_1 = 0.16$  et  $\tau_2 = 0.70$ . La variance des taux d’erreur moyens sur les 300 expériences est donnée entre parenthèses. Les tableaux doivent être lus par colonne.

	$c^0$	$c^1$	$c^2$	$c^3$	$c^4$	$c^5$	$c^6$
$c^0$	<b>20.72 (.7)</b>	22.61 (.7)	21.50 (.7)	22.28 (.8)	22.94 (.7)	23.10 (.7)	22.23 (.7)
$c^1$	21.06 (.7)	22.21 (.7)	21.28 (.7)	21.94 (.8)	22.18 (.7)	22.37 (.7)	21.98 (.7)
$c^2$	20.93 (.7)	22.35 (.7)	<b>21.10 (.7)</b>	21.93 (.8)	22.41 (.7)	22.66 (.7)	22.06 (.7)
$c^3$	21.05 (.7)	22.28 (.7)	21.21 (.7)	<b>21.90 (.8)</b>	22.27 (.7)	22.48 (.7)	22.09 (.7)
$c^4$	21.39 (.7)	<b>22.18 (.7)</b>	21.37 (.7)	21.96 (.8)	<b>22.01 (.7)</b>	22.20 (.7)	22.03 (.7)
$c^5$	21.50 (.8)	22.24 (.7)	21.52 (.7)	22.05 (.8)	22.06 (.7)	<b>22.20 (.7)</b>	22.09 (.7)
$c^6$	20.98 (.8)	22.27 (.7)	21.31 (.7)	22.06 (.8)	22.34 (.7)	22.50 (.7)	<b>21.91 (.7)</b>

 (a)  $\tau_1 = 0.16$ 

	$c^0$	$c^1$	$c^2$	$c^3$	$c^6$	$c^7$	$c^8$
$c^0$	<b>2.65 (.7)</b>	24.73 (.6)	24.57 (.7)	24.61 (.7)	25.29 (.7)	24.61 (.6)	24.56 (.7)
$c^1$	27.45 (.8)	<b>12.93 (.7)</b>	11.53 (.6)	11.60 (.7)	11.83 (.6)	12.18 (.7)	11.93 (.7)
$c^2$	26.31 (.8)	13.07 (.7)	<b>11.32 (.6)</b>	11.41 (.7)	12.19 (.6)	11.99 (.7)	11.71 (.7)
$c^3$	26.61 (.8)	13.25 (.7)	11.60 (.7)	<b>11.13 (.7)</b>	13.25 (.7)	12.38 (.7)	11.71 (.7)
$c^6$	28.43 (1.)	14.15 (.7)	12.90 (.6)	14.15 (.6)	<b>11.24 (.5)</b>	13.08 (.6)	13.75 (.7)
$c^7$	26.20 (.8)	13.11 (.7)	11.36 (.6)	11.59 (.7)	11.89 (.6)	<b>11.98 (.7)</b>	11.78 (.7)
$c^8$	26.30 (.8)	13.15 (.7)	11.39 (.7)	11.24 (.7)	12.54 (.6)	12.10 (.7)	<b>11.65 (.7)</b>

 (b)  $\tau_2 = 0.70$ 

Lorsque  $\tau$  est élevé, les différences entre les taux d’erreur moyens sont plus marqués ;

– Les copules présentant des formes proches pour une certaine valeur de  $\tau$  montrent le même comportement et des taux d’erreur très proches.

Dans de nombreuses applications, les paramètres ne sont pas connus ; il est donc intéressant de développer des méthodes adéquates d’apprentissage non supervisé, objet de la prochaine section.

### 3.2. Estimation des paramètres

Nous commençons par présenter la méthode EM, qui est particulièrement bien adaptée au cas gaussien, puis les méthodes SEM et ECI qui peuvent aussi être utilisées dans le cas de densités non gaussiennes.

#### 3.2.1. Algorithme EM

L’algorithme *Expectation-Maximization* (EM) (Dempster *et al.*, 1977) est une méthode générale pour trouver l’estimée du maximum de vraisemblance d’un ensemble de paramètres  $\theta$  d’une distribution donnée à partir d’un échantillon. On sait en effet que cette stratégie d’estimation conduit à des estimateurs asymptotiquement efficace, c’est-à-dire « optimaux » lorsque le nombre de données observées devient important.

Le problème est que la fonction de vraisemblance  $\mathcal{L}$  prend ici la forme relativement complexe suivante :

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{y}) = p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{n=1}^N \sum_{k=1}^K \pi_k f_k(\mathbf{y}_n; \boldsymbol{\theta}_k), \quad [11]$$

où  $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n, \dots, \mathbf{y}_N\}$  désigne l'ensemble des  $N$  vecteurs d'observations dont on dispose pour estimer les paramètres du modèle, et  $\boldsymbol{\theta}$  regroupe les paramètres du modèle à estimer.

La maximisation de la fonction de vraisemblance par rapport à  $\boldsymbol{\theta}$  ne peut pas se faire directement, ni même à l'aide d'algorithme d'optimisation classique car elle est non convexe. L'idée à la base de l'algorithme EM consiste à raisonner sur les données complètes (données observées et données cachées) tout en prenant en compte le fait que l'information disponible sur les données cachées ne peut venir que des données observées. Le principe de l'algorithme EM est de compenser les données cachées en les remplaçant par leur moyenne. Plus précisément, chaque itération  $\ell$  se décompose en deux étapes (Dempster *et al.*, 1977) :

⇒ **Expectation** Calcul de la fonction auxiliaire  $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(\ell)}) = \mathbb{E} \left[ \ln p(\mathbf{y}, \mathbf{x} | \boldsymbol{\theta}) | \mathbf{y}, \boldsymbol{\theta}^{(\ell)} \right]$  où

- $\boldsymbol{\theta}$  désigne les vrais paramètres,
- $\boldsymbol{\theta}^{(\ell)}$  désigne la valeur estimée des paramètres du modèle à la  $\ell^e$  itération de l'algorithme,
- $\mathbf{y}$  et  $\mathbf{x}$  désignent l'ensemble des données observées et cachées associées,
- $p(\mathbf{y}, \mathbf{x} | \boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$  désigne la vraisemblance conjointe des données observées et cachées.

⇒ **Maximisation** de la fonction auxiliaire permettant d'obtenir une nouvelle estimée :  $\boldsymbol{\theta}^{(\ell+1)} = \arg \max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(\ell)})$ .

Des calculs classiques conduisent à l'expression suivante pour la fonction auxiliaire

$$\mathcal{Q}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\ell)}) = \sum_{n=1}^N \sum_{k=1}^K \gamma_n^{(\ell)}(k) \ln(\pi_k f_k(\mathbf{y}_n; \boldsymbol{\theta}_k)). \quad [12]$$

Dans un cas purement gaussien où toutes les densités sont celles de lois gaussiennes de vecteur moyen  $\boldsymbol{\mu}_k$  et de matrice de covariance  $\Gamma_k$ , la maximisation de  $\mathcal{Q}$  conduit aux formules de ré-estimation suivantes :

### ⇔ Vecteurs moyens et matrices de covariance

$$\boldsymbol{\mu}_k^{(\ell+1)} = \frac{\sum_{n=1}^N \gamma_n^{(\ell)}(k) \mathbf{y}_n}{\sum_{n=1}^N \gamma_n^{(\ell)}(k)}, \quad [13]$$

$$\Gamma_k^{(\ell+1)} = \frac{\sum_{n=1}^N \gamma_n^{(\ell)}(k) \left( \mathbf{y}_n - \boldsymbol{\mu}_k^{(\ell+1)} \right) \left( \mathbf{y}_n - \boldsymbol{\mu}_k^{(\ell+1)} \right)^T}{\sum_{n=1}^N \gamma_n^{(\ell)}(k)}. \quad [14]$$

### ⇔ Poids du mélange

$$\pi_k^{(\ell+1)} = \frac{1}{N} \sum_{n=1}^N \gamma_n^{(\ell)}(k). \quad [15]$$

Lorsque le mélange n'est plus gaussien, la maximisation de la fonction auxiliaire peut devenir délicate, et il devient alors impossible d'écrire des formules de ré-estimation explicites. Les algorithmes SEM et ECI permettent de répondre à ce problème, par le biais de simulation *a posteriori*.

#### 3.2.2. Algorithmes SEM et ECI

L'algorithme SEM introduit par G. Celeux (Celeux et Diebolt, 1985) inclut une phase de simulation entre la phase du calcul de l'espérance de la fonction auxiliaire et celle de sa maximisation. L'estimation de tous les paramètres est alors réalisée à partir d'estimateurs empiriques sur les données complétées  $(\hat{\mathbf{x}}^{(\ell)}, \mathbf{y})$ , où  $\hat{\mathbf{x}}^{(\ell)} = \{x_1^{(\ell)}, \dots, x_N^{(\ell)}\}$  est obtenu par simulation, en effectuant des tirages selon  $\gamma_n^{(\ell)}(k)$ .

La méthode itérative ECI (Pieczynski, 1992) offre également 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 et SEM, mais sur le calcul de l'espérance d'un estimateur  $\hat{\delta}(\mathbf{x}, \mathbf{y})$  du paramètre  $\delta$ , défini sur les données complètes, conditionnellement aux données observées. Dans le cas d'un modèle de mélange, la ré-estimation de  $\delta$  à l'itération  $\ell + 1$  s'écrit

$$\delta^{(\ell+1)} = \mathbb{E} \left[ \hat{\delta}(\mathbf{x}, \mathbf{y}) \mid \mathbf{y}; \delta^{(\ell)} \right]. \quad [16]$$

Ce principe conduit à deux situations différentes selon que l'on cherche à estimer les poids du mélange ou les paramètres des lois d'attache aux données. En ce qui concerne les poids du mélange  $\pi_k$ , la ré-estimation ECI conduit aux mêmes formules

que dans le cas EM, cf eq. [15]. Les paramètres des lois d'attache aux données, quand à eux, peuvent se calculer à partir d'estimateurs empiriques sur les données complétées  $(\hat{\mathbf{x}}^{(\ell)}, \mathbf{y})$  par simulation, comme dans le cas SEM.

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 considérerons pour les simulations, le tau de Kendall peut être estimé à partir du taux de concordance  $c$  et du taux de discordance  $d$  selon  $\hat{\tau} = \frac{c-d}{c+d}$ .

### 3.3. Sélection de copules

Une des propriétés très intéressantes des modèles de mélange est la possibilité d'estimer les paramètres à partir des seules observations. Mais, une question intéressante est de savoir s'il est possible de retrouver automatiquement les formes exactes des copules utilisées lors de la simulation des données. Pour vérifier cela, nous avons inclus la méthode bayésienne de sélection de copule introduite par D. Huard *et al* (Huard *et al.*, 2006), permettant de sélectionner la « meilleur forme » pour les  $K$  copules à chaque itération de l'algorithme SEM ou ECI.

L'identification de lois multi-variées, lorsqu'elles sont définies avec des copules, est le sujet de nombreux papiers récents, parmi lesquels (Nikoloulopoulos et Karlis, 2008 ; Genest *et al.*, 2009b ; Genest *et al.*, 2009a ; Xiaomei *et al.*, 2010 ; Noh *et al.*, 2010). Dans ce travail, nous utiliserons la méthode de « sélection bayésienne de copules » (Huard *et al.*, 2006) pour sa simplicité et sa faible complexité. Soit  $\mathbf{y} = \{\mathbf{y}^1, \mathbf{y}^2\}$  un ensemble d'observations 2D avec  $\mathbf{y}^1 = \{y_1^1, \dots, y_N^1\}$  and  $\mathbf{y}^2 = \{y_1^2, \dots, y_N^2\}$ , la « meilleure copule »  $c^s$  d'un ensemble fini de copules  $\Pi = \{c^1, \dots, c^R\}$  est désignée selon le critère suivant

$$s = \arg \max_{r \in [1, R]} \frac{1}{\tau_M^r - \tau_m^r} \int_{\tau_m^r}^{\tau_M^r} \prod_{n=1}^N c^r (F^1(y_n^1), F^2(y_n^2); \tau) d\tau, \quad [17]$$

où  $F^1$  et  $F^2$  sont les fonctions de répartition marginales des données  $\mathbf{y}^1$  et  $\mathbf{y}^2$  (leur forme est supposée connue). Les coefficients  $\tau_m^r$  et  $\tau_M^r$  de la copule  $c^r$  représentent les valeurs minimales et maximales admissibles du tau de Kendall. Ces valeurs sont reportées dans l'annexe A pour les copules que nous considérons dans ce travail. Une particularité intéressante de la méthode est qu'elle ne repose pas sur l'estimation de  $\tau$ . En traitement d'images, elle a déjà été appliquée avec succès dans (Sakji-Nsibi et Benazza-Benyahia, 2009) pour la l'indexation d'images vectorielles à l'aide d'on-delettes.

En résumé, à chaque itération de l'algorithme d'estimation des paramètres, que ce soit SEM ou ECI, nous incorporons une étape de sélection de la meilleur forme de copule pour les  $K$  copules.

### 3.4. Expériences

Pour évaluer la combinaison de ECI avec l'algorithme de sélection de copules, nous avons conduit les expériences suivantes, toujours sur données simulées.

Tout d'abord, nous avons simulé un échantillon de  $N = 2000$  observations selon le mélange d'une copule Gaussienne ( $\tau_1 = 0.7$ ) et d'une copule de Gumbel ( $\tau_2 = 0.8$ ) en proportion  $\pi_1 = 0.6$  et  $\pi_2 = 0.4$  avec des marges spécifiées par le tableau 1. Nous avons ensuite restauré les données en laissant ECI sélectionner les formes des copules dans l'ensemble  $\Pi_1 = \{c^1, c^3, c^6\}$  (resp. Gaussienne, Gumbel, Clayton)<sup>5</sup>.

La figure 4 dresse l'évolution des paramètres du mélange, à partir d'une initialisation des paramètres par un algorithme des k-moyennes. Excepté les poids, les paramètres initiaux sont très éloignés des vrais paramètres (représentés par des traits horizontaux sur la figure). Ce sont les copules de Clayton et de Gumbel qui sont sélectionnées à la première itération de l'algorithme avec des taux de Kendall respectifs de  $\tau_1 = 0.48$  et  $\tau_2 = 0.54$ . Sur cette simulation la convergence vers les paramètres attendus est atteinte après une centaine d'itérations, sachant que les bonnes formes des copules (Gaussienne et Gumbel) sont sélectionnées dès l'itération 7 et ne changent plus jusqu'à convergence.

Le cas présenté en détails est un cas favorable dans la mesure où l'algorithme d'estimation est capable de retrouver les paramètres à l'origine du mélange. Ce n'est pas toujours le cas. D'autres simulations, non reportées ici, nous ont permis de faire les conclusions suivantes :

- Pour un échantillon de petite taille comme celui de cette expérience, ce bon comportement n'est plus observé lorsque les taux de Kendall sont faibles. Les algorithmes (ECI comme SEM) convergent vers une solution du mélange différente de celle à l'origine de la simulation. Outre la taille de l'échantillon, il faut noter la grande difficulté des mélanges testés qui sont similaires à ceux utilisés dans le tableau 2(b), et pour lequel le taux d'erreur moyen est de l'ordre de 22%. La seconde raison provient de l'inadéquation de l'algorithme d'initialisation (k-moyennes) qui donne des estimées initiales trop éloignées des paramètres attendus. Des expériences complémentaires ont permis de vérifier qu'une initialisation plus fiable garantissait la convergence, même en cas de corrélation faible et d'échantillon de relativement petite taille.

- Avec la méthode de sélection de copule présentée ci-dessus, il nous est impossible de traiter des échantillons de taille plus importante suite à des problèmes de dépassement de capacité des calculs. C'est la raison pour laquelle il nous sera impossible de sélectionner automatiquement les copules en segmentation d'images étant donné le nombre généralement important de pixels à traiter. Une solution pour pallier à ce problème consistera à remplacer cette méthode par une méthode concurrente comme celle présentée dans (dos Santos Silva et Lopes, 2008).

5. Les formes des marges sont imposées, par contre leurs paramètres sont estimés par ECI.

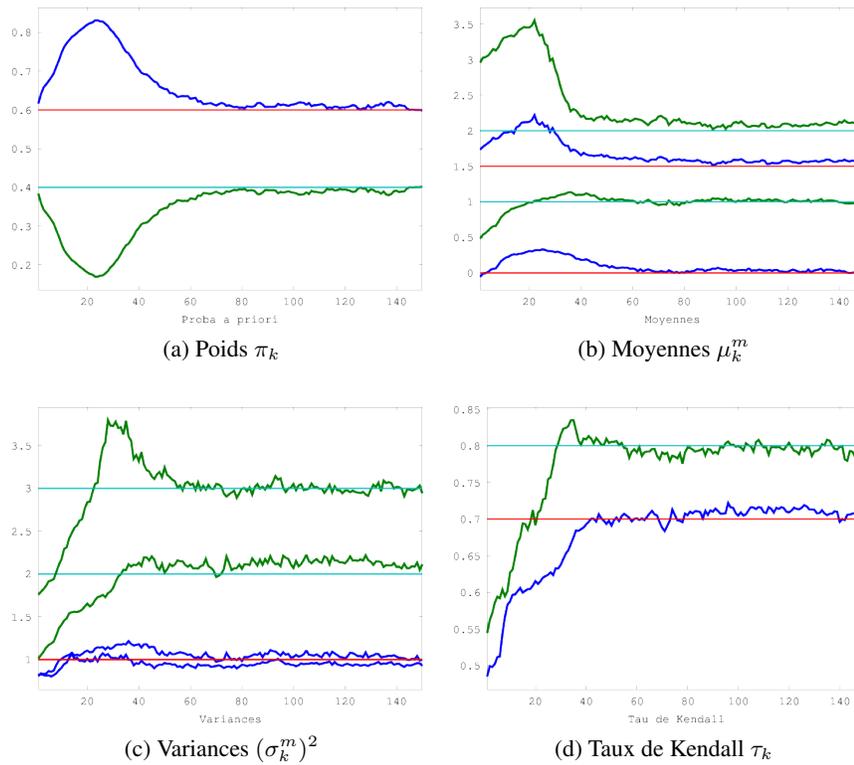


Figure 4 – Tracés de l’évolution des paramètres estimés d’un MMC par ECI. Les traits horizontaux représentent les valeurs de simulation.

Voyons maintenant comment il est possible d’étendre ce modèle en considérant que chaque densité du mélange est elle-même le résultat d’un mélange de densités.

#### 4. Modèle de mélange de mélanges

Ce chapitre reprend pour l’essentiel le travail présenté dans (Derrode et Pieczynski, 2012), mais propose de nouvelles interprétations et de nouvelles illustrations.

Une extension intéressante aux modèles de mélange est de considérer que chaque classe est elle-même le résultat d'un mélange de (sous-)classes, ce que l'on peut écrire de la manière suivante

$$f_k(\mathbf{y}; \boldsymbol{\theta}_k) = \sum_{l=1}^L \pi_{l|k} f_{l|k}(\mathbf{y}; \boldsymbol{\theta}_{k,l}), \quad [18]$$

où chaque  $f_{k,l}$  est une densité  $M$ -dimensionnelle que l'on peut représenter à l'aide de copules (dès lors que  $M > 2$ ). D'après [2], le mélange « complet » s'écrit sous la forme suivante

$$f(\mathbf{y}; \boldsymbol{\theta}) = \sum_{k=1}^K \sum_{l=1}^L \pi_{k,l} f_{k,l}(\mathbf{y}; \boldsymbol{\theta}_{k,l}) \quad [19]$$

avec  $\boldsymbol{\theta} = \{\pi_{1,1}, \pi_{1,2}, \dots, \pi_{K,L}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K\}$ ,  $\forall k \in \Omega$ ,  $\boldsymbol{\theta}_k = \{\boldsymbol{\theta}_{k,1}, \dots, \boldsymbol{\theta}_{k,L}\}$  et où  $\sum_{k=1}^K \sum_{l=1}^L \pi_{k,l} = 1$ . Ce modèle est appelé « mélange de mélanges » (Di Zio *et al.*, 2007), « mélange multi-échelles », « mélange hiérarchique » (Vermunt et Magidson, 2005) ou encore « mélange multi-couches » (Li, 2005 ; Jörnsten et Keleş, 2008) (acronyme utilisé ici MMM). Si  $L = 1$ , nous retrouvons le modèle de mélange classique. Il est également intéressant de noter que

- le modèle peut s'étendre à un nombre d'échelles supérieur à deux ;
- chaque classe  $k$  à l'échelle 1 peut être composée, non pas d'un nombre fixe  $L$ , mais d'un nombre variable  $L(k)$  de classes ;

Nous éviterons ces possibilités extensions dans la mesure où les écritures s'en trouvent alourdies et nous nous limiterons au modèle de l'éq. [19]. Il est important de noter que les MMM proposent une modélisation très riche<sup>6</sup>, permettant notamment de prendre en compte des classes multi-modales et offrant en cela une alternative intéressante aux modèles de mélange non paramétriques (Li *et al.*, 2007 ; Mallapragada *et al.*, 2010).

Les observations  $\mathbf{y}_n$  ont un classement à chaque échelle :

- nous noterons  $x_n^1 \in \Omega^1 = [1, K]$ , la classification à la première échelle, et
- conditionnellement à la classe affectée à la première échelle, nous noterons  $x_n^2 | x_n^1 \in \Omega^2 = [1, L]$ , la classification à la seconde échelle,

et finalement  $\mathbf{x}_n = (x_n^1, x_n^2) \in \Omega^1 \times \Omega^2$ . Il faut remarquer qu'un MMM à  $K \times L$  classes peut toujours s'écrire sous forme d'un MM à  $KL$  classes

$$f(\mathbf{y}; \boldsymbol{\theta}) = \sum_{m=1}^{KL} \pi_m f_m(\mathbf{y}; \boldsymbol{\theta}_m).$$

Mais l'intérêt d'un MMM provient des résultats de classification aux échelles supérieures et donc de sa capacité à regrouper des classes à l'échelle deux pour constituer des « super » classes à l'échelle un.

6. encore décuplée par le potentiel de modélisation des densités à copule

#### 4.1. Restauration bayésienne

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

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

Contrairement au MM, nous pouvons ici considérer deux fonctions de coût :

$$\begin{aligned} L_1(\mathbf{x}_n, \tilde{\mathbf{x}}_n) &= \mathbb{1}_{x_n^1 \neq \tilde{x}_n^1} + \mathbb{1}_{x_n^2 \neq \tilde{x}_n^2} \\ L_2(\mathbf{x}_n, \tilde{\mathbf{x}}_n) &= \mathbb{1}_{\mathbf{x}_n \neq \tilde{\mathbf{x}}_n} = \mathbb{1}_{x_n^1 \neq \tilde{x}_n^1 \text{ ou } x_n^2 \neq \tilde{x}_n^2} \end{aligned}$$

$L_1$  est appelée « Maximum a Posteriori des Marges » (MPM) alors que  $L_2$  est appelée « Maximum A Posteriori » (MAP). Après calculs, nous aboutissons ici à

$$MPM : \begin{cases} \hat{x}_n^1 = \arg \max_{k \in \Omega^1} p(x_n^1 = k | \mathbf{y}_n) \\ \hat{x}_n^2 = \arg \max_{l \in \Omega^2} p(x_n^2 = l | \mathbf{y}_n) \end{cases} \quad [20]$$

$$MAP : \quad \hat{\mathbf{x}}_n = \arg \max_{(k,l) \in \Omega^1 \times \Omega^2} p(\mathbf{x}_n = (k,l) | \mathbf{y}_n) \quad [21]$$

où

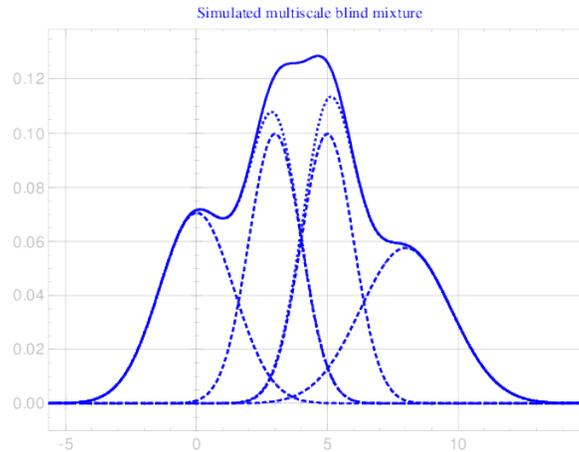
$$\begin{aligned} p(x_n^1 = k | \mathbf{y}_n) &= \sum_{l=1}^L p(\mathbf{x}_n = (k,l) | \mathbf{y}_n) \\ p(x_n^2 = l | \mathbf{y}_n) &= \sum_{k=1}^K p(\mathbf{x}_n = (k,l) | \mathbf{y}_n). \end{aligned}$$

Étant donné l'ensemble des paramètres d'un modèle de mélange de mélanges, la simulation des deux couples de données  $(y^1, y^2)$  et  $(x^1, x^2)$  peut se réaliser de la manière suivante

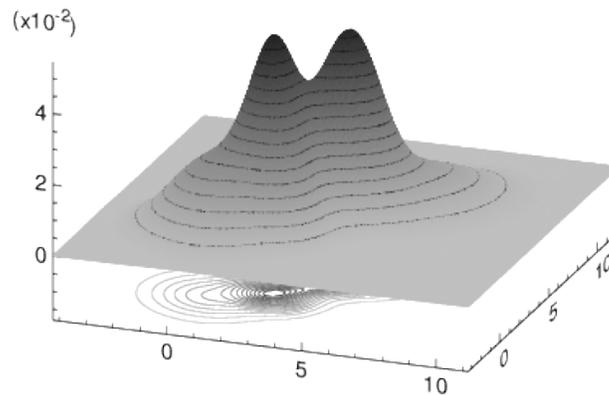
- tirage de  $x^1$ , selon la loi  $p(x^1) = \sum_{l=1}^L \pi_{x^1, l}$  ;
- tirage de  $x^2$ , selon la loi  $p(x^2 | x^1) = \frac{\pi_{x^1, x^2}}{p(x^1)}$  ;
- tirage de  $y^1$ , selon la loi marginale  $f_{x^1, x^2}^{(1)}(y^1; \boldsymbol{\theta}_{x^1, x^2})$  ;
- tirage de  $y^2$ , selon la loi conditionnelle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

Si le MMM est gaussien, alors la loi conditionnelle l'est également, et le tirage de  $y^2$  se fait selon une loi gaussienne. Si le MMM est constitué de densités à copules, le tirage de  $y^2$  s'effectue selon le principe décrit dans la section 2.2.

20 1<sup>re</sup> soumission à *Cahier Scientifique*



(a) 1D



(b) 2D

Figure 5 – Deux mélanges de mélange à deux échelles. Dans le mélange 1D, on peut distinguer les quatre lois gaussiennes  $f_{k,l}$  de la seconde échelle, les deux mélanges  $f_k$  à la première échelle et le mélange global. Dans le cas 2D, les marginales sont toutes gaussiennes mais les copules ne le sont pas, avec des taux de Kendall différents :  $c_1(\tau = -0.1)$ ,  $c_1(\tau = -0.2)$ ,  $c_6(\tau = 0.3)$  et  $c_7(\tau = 0.3)$ .

À titre d'illustration, nous avons représenté, dans la figure 5, un mélange 1D et un mélange 2D constitués de deux densités, elles-mêmes composées d'un mélange de deux densités (le poids des quatre densités est fixé à 0.25). Les taux d'erreur de restauration supervisée de  $N = 2000$  échantillons dans le cas 1D et de  $N = 20000$

Tableau 3 – Taux d’erreur MAP et MPM de la restauration supervisée d’un échantillon d’observations générées selon les mélanges de la figure 5.

	Échelle 1	Échelle 2		Échelle 1	Échelle 2
MPM	8.27	19.09	MPM	5.69	18.27
MAP	8.23	19.07	MAP	5.67	18.30
(a) 1D			(b) 2D		

échantillons dans le cas 2D sont reportés dans le tableau 3. La première remarque est de constater que les taux d’erreur MPM et MAP sont presque identiques aux deux échelles. Le taux d’erreur à l’échelle 1 s’interprète comme le taux d’erreur entre les deux sous-mélanges, alors que le taux d’erreur à l’échelle 2 s’interprète comme la somme des taux d’erreur entre les composantes des sous-mélanges.

#### 4.2. Estimation des paramètres

Tout comme nous l’avons fait dans le cas des MM, distinguons l’estimation EM de l’estimation SEM et ECI.

##### 4.2.1. Algorithme EM

L’algorithme EM pour l’estimation d’un MMM gaussien ressemble fortement à celui exposé précédemment dans le cas du MM. La fonction auxiliaire de l’eq. [12] s’écrit ici

$$\mathcal{Q}(\Theta|\Theta^{(\ell)}) = \sum_{n=1}^N \sum_{k=1}^K \sum_{l=1}^L \xi_n^{(\ell)}(k, l) \ln(\pi_{k,l} f_{k,l}(\mathbf{y}_n; \theta_{k,l})) \quad [22]$$

avec les probabilités conjointes *a posteriori* définies par

$$\xi_n^{(\ell)}(k, l) \propto \pi_{k,l}^{(\ell)} f_{k,l}(\mathbf{y}_n; \theta_{k,l}^{(\ell)}) \quad [23]$$

Selon le principe EM, la maximisation de  $\mathcal{Q}(\Theta|\Theta^{(\ell)})$  conduit aux formules de ré-estimation suivantes :

22 1<sup>re</sup> soumission à *Cahier Scientifique*

⇨ **Vecteurs moyens et matrices de covariance**

$$\boldsymbol{\mu}_k^{(\ell+1)} = \frac{\sum_{n=1}^N \xi_n^{(\ell)}(k, l) \mathbf{y}_n}{\sum_{n=1}^N \xi_n^{(\ell)}(k, l)}, \quad [24]$$

$$\Gamma_k^{(\ell+1)} = \frac{\sum_{n=1}^N \xi_n^{(\ell)}(k, l) \left( \mathbf{y}_n - \boldsymbol{\mu}_{k,l}^{(\ell+1)} \right) \left( \mathbf{y}_n - \boldsymbol{\mu}_{k,l}^{(\ell+1)} \right)^T}{\sum_{n=1}^N \xi_n^{(\ell)}(k, l)}. \quad [25]$$

⇨ **Poids du mélange**

$$\pi_k^{(\ell+1)} = \frac{1}{N} \sum_{n=1}^N \xi_n^{(\ell)}(k, l). \quad [26]$$

4.2.2. *Algorithmes SEM et ECI*

Selon le principe SEM, l'estimation de tous les paramètres est réalisée à partir d'estimateurs empiriques sur les données complétées  $(\hat{\mathbf{x}}^{(\ell)}, \mathbf{y})$ , où  $\hat{\mathbf{x}}^{(\ell)} = \{\mathbf{x}_1^{(\ell)}, \dots, \mathbf{x}_N^{(\ell)}\}$  est obtenu par simulation, en effectuant des tirages selon

$$p \left( x_n^1 = k \mid \mathbf{y}_n; \boldsymbol{\Theta}^{(\ell)} \right) = \sum_{l=1}^L \xi_n^{(\ell-1)}(k, l), \quad [27]$$

$$p \left( x_n^2 = l \mid x_n^1 = k, \mathbf{y}_n; \boldsymbol{\Theta}^{(\ell)} \right) = \frac{\xi_n^{(\ell-1)}(k, l)}{p \left( x_n^1 = k \mid \mathbf{y}_n; \boldsymbol{\Theta}^{(\ell)} \right)}. \quad [28]$$

Selon le principe ECI, la ré-estimation d'un paramètre  $\delta$  à l'itération  $\ell + 1$  s'écrit

$$\delta^{(\ell+1)} = \mathbb{E} \left[ \hat{\delta}(\mathbf{x}, \mathbf{y}) \mid \mathbf{y}; \delta^{(\ell)} \right]. \quad [29]$$

Là encore, le principe ECI conduit à deux situations différentes selon que l'on cherche à estimer les poids du mélange ou les paramètres des densités. Les premiers sont obtenus grâce à des formules de ré-estimation similaires à [26], tandis que les seconds sont obtenus par simulation, selon la même démarche que celle présentée pour SEM.

À titre d'illustration, nous avons représenté les courbes de convergence de quelques paramètres par SEM pour le mélange 1D dans la figure 6 et pour le mélange 2D dans la figure 7. Malgré une convergence relativement lente pour certains paramètres, SEM est capable de retrouver la valeur des paramètres à l'origine du mélange multi-échelles, malgré une initialisation grossière par l'algorithme des k-moyennes.

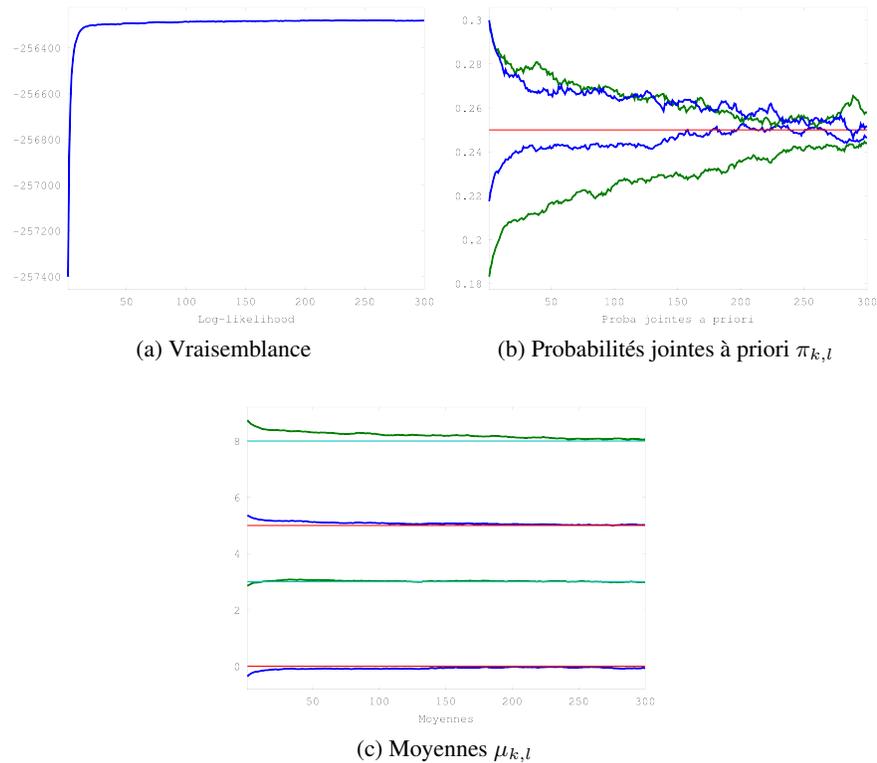


Figure 6 – Tracés de l'évolution des paramètres estimés d'un MMM Gaussien par SEM pour le mélange 1D. Les traits horizontaux représentent les valeurs attendues.

Avec le modèle 1D, le taux d'erreur de restauration MPM obtenu est de 8.2% à l'échelle 1 et de 19.5% à l'échelle 2. Avec le modèle 2D, le taux d'erreur de restauration MPM obtenu est de 5.5% à l'échelle 1 et de 18.6% à l'échelle 2. Ces valeurs sont presque similaires aux taux obtenus par restauration supervisée.

## 5. Segmentation non supervisée d'images vectorielles

Ce chapitre rapporte les résultats de segmentations obtenues avec les modèles de mélange pour différentes expériences portant sur des images multi-bandes (image couleur et images IRM de modalités différentes), et son extraits de (Derrode et Pieczynski, 2012). Pour ces images, nous n'avons pas de « vérités-terrain » à notre disposition ou d'expert pour nous aider à interpréter les résultats. Aussi, il faut considérer les résultats comme une illustration des algorithmes présentés. Nous n'avons pas été en mesure d'évaluer la sélection de copules dans ces applications de segmentation

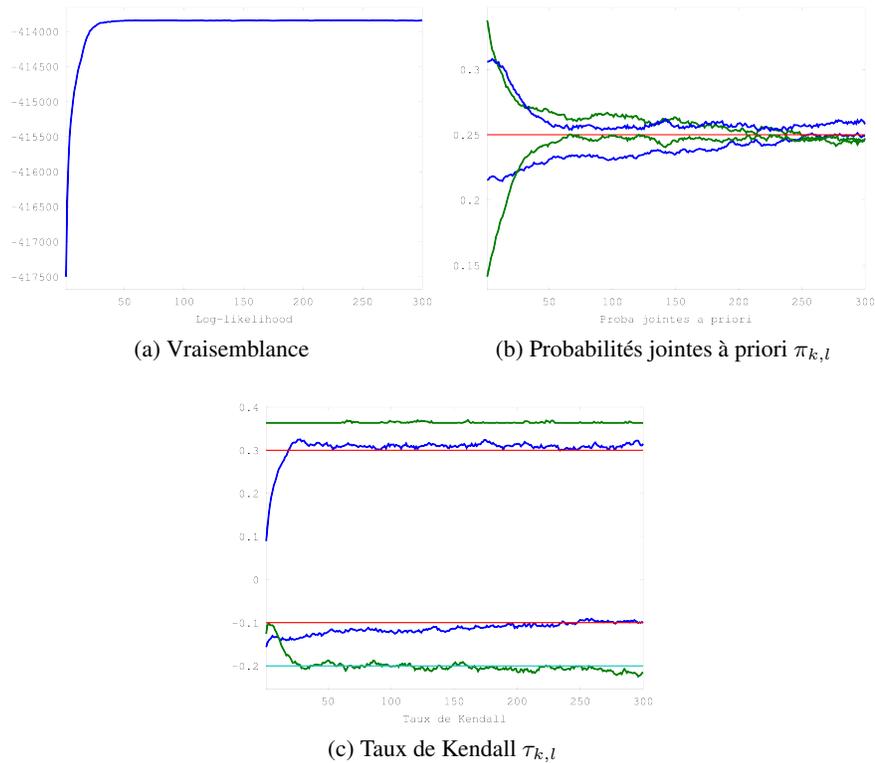


Figure 7 – Tracés de l'évolution des paramètres estimés d'un MMM Gaussien par SEM pour le mélange 2D. Les traits horizontaux représentent les valeurs attendues.

d'images car l'algorithme proposé par D. Huard ne permet que de traiter un nombre restreint de données du a des problèmes de dépassements des capacités numériques.

### 5.1. Segmentation d'une image couleur

La première expérience concerne la segmentation d'une image couleur, cf figure 8. Nous avons segmenté les 3 bandes colorimétriques avec, d'une part, des mélanges classiques gaussiens sur chaque bande (de manière indépendant les uns par rapport aux autres) et, d'autre part, avec le modèle de mélange de mélange à 3 échelles gaussien. Les résultats sont reportés dans les figures 9 et 10 respectivement, en considérant à chaque fois 2 classes par échelle. 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

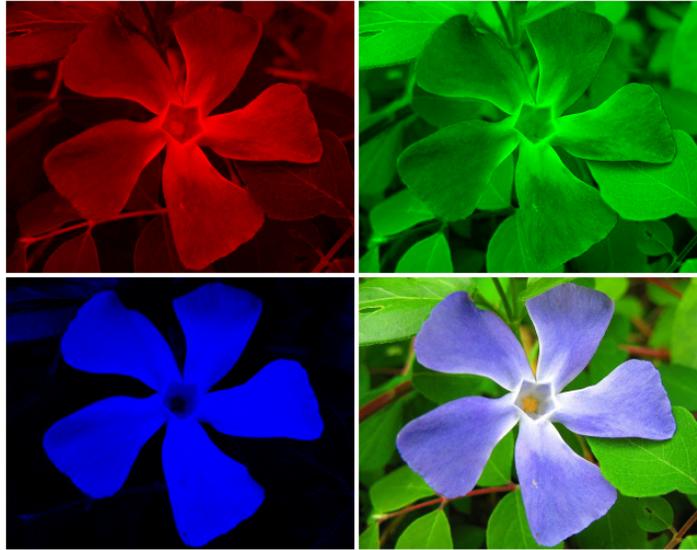


Figure 8 – Une image couleur de fleur (bas droit), et sa décomposition R (haut gauche), V (haut droit) et B (bas gauche). Dimensions :  $320 \times 250$ .

segmentations avec l'image originale. Comme les 3 échelles sont segmentées en 2 classes chacune, 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 clairement que le MM gaussien 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 MMM gaussien est plus satisfaisant et plus riche en détails que son équivalente pour le MM gaussien. On peut pointer en exemple les fines branches marron et le cœur de la fleur.

## 5.2. 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 11, ont été générées par le simulateur « BrainWeb » ([www.bic.mni.mcgill.ca/brainweb/](http://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→verte et DP→bleue.

26 1<sup>re</sup> soumission à *Cahier Scientifique*

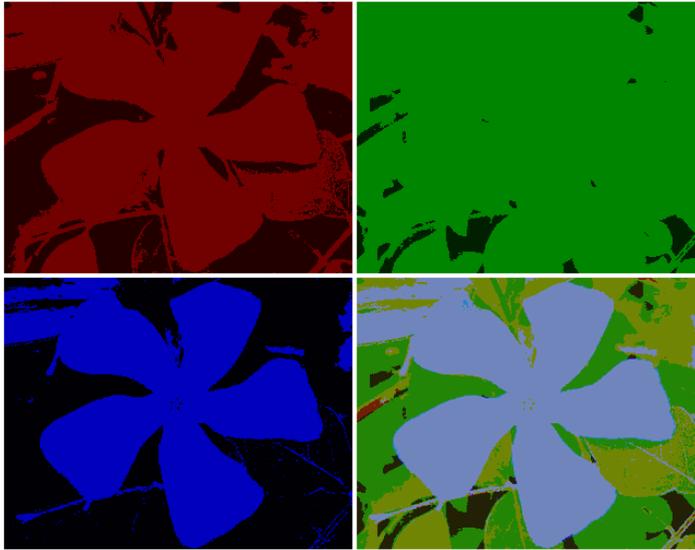


Figure 9 – Segmentations individuelles des bandes en 2 classes chacune par un MM, et recomposition couleur (bas droit).

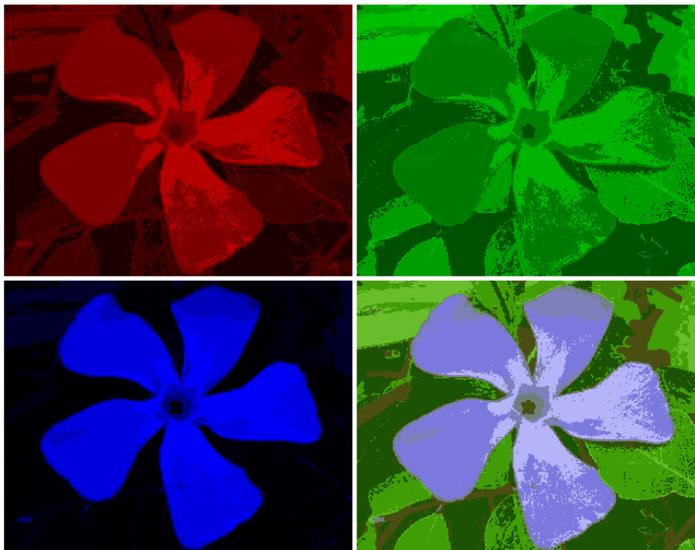


Figure 10 – Segmentations des bandes en 3 échelles et 2 classes par échelle par un MMM gaussien, et recomposition couleur (bas droit).

Les résultats de segmentation par des MM gaussiens indépendants à  $K = 2$  classes sont reportés dans la figure 12 ; 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 MMM sont reportés dans la figure 13a lorsque l'on considère des copules gaussiennes et des marges gaussiennes. La différence entre les deux résultats est particulièrement notable. Le MMM 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 13b. 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 et Pieczynski, 1997)–, ou bien encore en considérant les lois candidates dans le système de Pearson (Derrode, 2009).

## 6. Conclusion

Ce papier décrit des modèles de mélange vectoriel spécifiés par copules. Nous avons également présenté des méthodes de restauration supervisée de données et d'estimation des paramètres pour la restauration non supervisée (méthodes EM, SEM et ICE). Les expériences illustrent le bon comportement des algorithmes sur des données simulées, même lorsque les formes des copules ne sont pas connues. Il faut cependant noter que l'algorithme de sélection de copules proposé par D. Huard (Huard *et al.*, 2006) est restrictif puisqu'il ne permet de traiter que des jeux de données de taille relativement faible ( $N < 2500$ ) suite à des problèmes de dépassement de capacité numérique. Ainsi il ne nous a pas été possible d'évaluer la sélection de copules en classification d'images. Une première perspective consistera à la remplacer par des méthodes concurrentes, parmi lesquels l'algorithme proposé dans (dos Santos Silva et Lopes, 2008) semble un bon candidat.

Nous avons ensuite développé une démarche équivalente pour les modèles de mélange multi-échelles où chaque classe est représentée par un mélange (potentiellement constitué de densités à copules). Ce modèle est susceptible de prendre en compte des classes multi-modales et des expériences systématiques plus poussées sont requises pour comprendre les limites des algorithmes d'estimation non supervisée à retrouver les paramètres du modèle. En utilisant des densités à copules, nous obtenons ainsi des mélanges d'une grande complexité qui restent pourtant numériquement identifiables si le mélange n'est pas trop difficile.

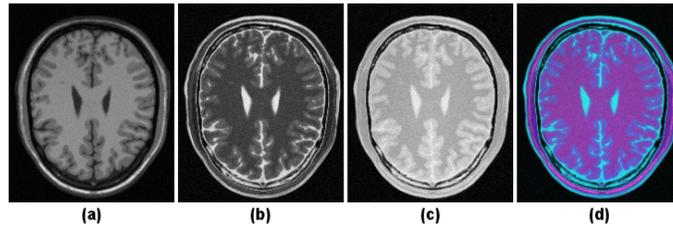


Figure 11 – Une coupe IRM du cerveau en modalités (a) T1, (b) T2, (c) DP (Densité de Protons) et (d) rendu coloré des 3 images. Dimensions des images :  $180 \times 216$ .

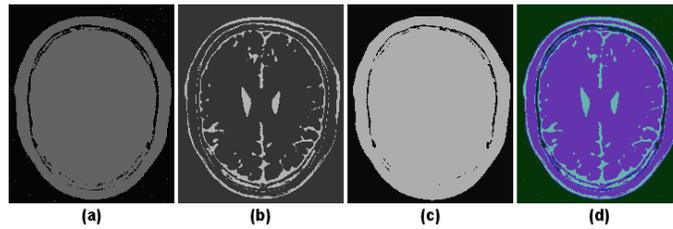
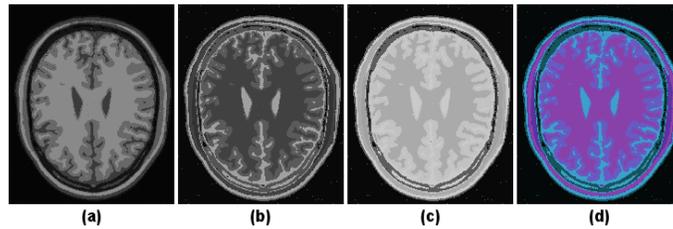
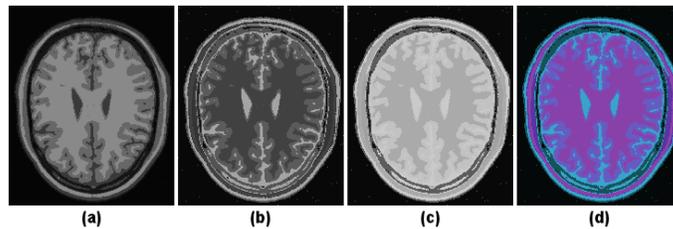


Figure 12 – Résultats de segmentation (MPM) des 3 modalités par des MM gaussiens indépendants, avec 2 classes pour chaque coupe.



(a) Marges gaussiennes



(b) Marges gamma

Figure 13 – Résultats de segmentation (MPM) des 3 modalités IRM par un MMM avec des copules gaussiennes et (a) des marges gaussiennes, (b) des marges gammas, avec 2 classes pour chaque échelle.

Enfin, toujours dans la perspective de prise en compte du voisinage, il serait intéressant d'étudier le développement d'un modèle de « mélange de chaînes de Markov cachées », en s'inspirant du modèle de « chaînes de Markov triplet » (Pieczynski, 2002) qui introduit un processus latent sans nécessairement de signification physique, en supplément des processus observé et des états cachés.

## 7. Bibliographie

- Balakrishnan N., Lai C. D., *Continuous Bivariate Distributions*, Wiley Series in Probability and Statistics : Applied Probability and Statistics, Wiley-Interscience, 2009.
- Banerjee A., Dhillon I. S., Ghosh J., Sra S., « Clustering on the unit hypersphere using von Mises-Fisher distributions », *J. of Machine Learning Research*, vol. 6, p. 1345-1382, 2005.
- Bohning D., Dietz E., Schlattmann P., « Recent developments in computer-assisted analysis of mixtures », *Annals of the Institute of Statistical Mathematics*, vol. 54, n° 2, p. 525-536, 1998.
- Bougarradh A., Mhiri S., Ghorbel F., « Segmentation non supervisée d'images angiographiques de la rétine par le système de Pearson et le ré-échantillonnage bootstrap », *22<sup>e</sup>me colloque Traitement du Signal et des Images (GRETSI'09)*, Dijon, France, 2009.
- Brunel N., Pieczynski W., « Unsupervised signal restoration using hidden Markov chains with copulas », *Signal Processing*, vol. 85, p. 2304-2315, 2005.
- Celeux G., Diebolt J., « The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem », *Computational Statistics Quarterly*, vol. 2, p. 73-82, 1985.
- Delignon Y., Marzouki A., Pieczynski W., « Estimation of Generalised Mixture and Its Application in Image Segmentation », *IEEE Transactions on Image Processing*, vol. 6, n° 10, p. 1364-1375, 1997.
- Dempster A., Laird N., Rubin D., « Maximum likelihood from incomplete data via the EM algorithm », *J. of the Royal Statistical Society, series B*, vol. 39, p. 1-38, 1977.
- Derrode S., 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](http://www.fresnel.fr/perso/derrode/HMC_Ext/Doc/Pearson.pdf), École Centrale Marseille & Institut Fresnel, 2009.
- Derrode S., Pieczynski W., « Segmentation d'images par modèle de mélange conjoint non gaussien », *Traitement du Signal, accepté en janvier 2012*, 2012.
- Devroye L., *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, New York, U.S.A., 1986.
- Di Zio M., Guarnera U., Rocci R., « A mixture of mixture models for a classification problem : The unity measure error », *Computational Statistics and Data Analysis*, vol. 51, n° 5, p. 2573-2585, 2007.
- dos Santos Silva R., Lopes H. F., « Copula, marginal distributions and model selection : a Bayesian note », *Statistics and Computing*, vol. 18, p. 313-320, 2008.
- Frühwirth-Schnatter S., *Finite mixture and Markov switching models*, Springer Series in Statistics, Springer, New York, U.S.A., 1985.

30 1<sup>re</sup> soumission à *Cahier Scientifique*

- Genest C., Masiello E., Tribouley K., « Estimating copula densities through wavelets », *Insurance : Mathematics and Economics*, vol. 44, p. 170-181, 2009a.
- Genest C., Rémillard B., Beaudoin D., « Goodness-of-fit tests for copulas : A review and a power study », *Insurance : Mathematics and Economics*, vol. 44, p. 199-213, 2009b.
- Giordana N., Pieczynski W., « 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, 1997.
- Huard D., Évin G., Favre A.-C., « Bayesian copula selection », *Computational Statistics & Data Analysis*, vol. 51, p. 809-822, 2006.
- Iyengar S., Varshney P., Damarla T., « A Parametric Copula-Based framework for Hypothesis Testing Using Heterogeneous Data », *IEEE Trans. on Signal Processing*, vol. 59, n° 5, p. 2308-2319, May, 2011.
- Jörnsten R., Keleş S., « Mixture models with multiple levels, with application to the analysis of multi-factor gene expression data », *Biostatistics*, vol. 9, n° 3, p. 540-554, 2008.
- Ketchantang W., Derrode S., Martin L., Bourennane S., « Pearson-based mixture model for color object tracking », *Machine Vision and Applications, special issue on video surveillance research in industry and academia*, 2008.
- Kotz S., Balakrishnan N., Johnson N. L., *Continuous Multivariate Distributions, Vol. 1 : Models and Applications*, Springer Series in Statistics, Springer, New York, U.S.A., 2008.
- Kwan R.-S., Evans A., Pike G., « MRI-simulation based evaluation of image processing and classification methods », *IEEE trans. on Medical Imaging*, vol. 18, n° 11, p. 1085-1097, 1999.
- Li J., « Clustering based on a multi-layer mixture model », *J. of Computational and Graphical Statistics*, vol. 14, n° 3, p. 547-568, 2005.
- Li J., Ray S., Lindsay B., « A nonparametric statistical approach to clustering via mode identification », *The Journal of Machine Learning Research*, vol. 8, p. 1687-1723, 2007.
- Li J., Zha H., « Two-way Poisson mixture models for simultaneous document classification and word clustering », *Computational Statistics and Data Analysis*, vol. 50, n° 1, p. 163-180, 2006.
- Mallapragada P. K., Jin R., Jain A., « Non-parametric mixture models for clustering », *Structural and Syntactic Pattern Recognition (SSPR'10)*, August 18-20, Izmir, Turkey, 2010.
- McKenna S., Raja Y., Gong S., « Tracking colour objects using adaptive mixture models », *Image and Vision Computing*, vol. 17, n° 3/4, p. 225-231, 1999.
- McLachlan G., Krishnan T., *The EM algorithm and extensions*, Wiley Series in Probability and Statistics : Applied Probability and Statistics, Wiley-Interscience, New York, U.S.A., 2008.
- McLachlan G., Peel D., *Finite mixture models*, Wiley Series in Probability and Statistics : Applied Probability and Statistics, Wiley-Interscience, New York, U.S.A., 2000.
- Melnykov V., Maitra R., « Finite mixture models and model-based clustering », *Statistics Surveys*, vol. 4, p. 80-116, 2010.
- Mercier G., Mesures de Dépendance entre Images RSO, Rapport de recherche rr-2005003-iti, <http://public.enst-bretagne.fr/~mercierg/articles/2005/mesureDep.pdf>, GET/ENST Bretagne, 2005.

- Nelsen R. B., *An introduction to copulas*, Springer, Lecture Notes in Statistics, 2nd edition, New York, U.S.A., 2006.
- Nguyen T. M., Wu Q. M., « Robust Student's-t mixture model with spatial constraints and its application in medical image segmentation », *IEEE Transactions on Medical Imaging*, vol. 31, n° 1, p. 103-116, 2012.
- Nguyen T. M., Wu Q. M., Ahuja S., « An extension of the standard mixture model for image segmentation », *IEEE Transactions on Neural Networks*, vol. 21, n° 8, p. 1326-1338, 2012.
- Nikoloulopoulos A., Karlis D., « Copula model evaluation based on parametric bootstrap », *Computational Statistics & Data Analysis*, vol. 52, p. 3342-3353, 2008.
- Noh Y., Choi K., Lee I., « Identification of marginal and joint CDFs using Bayesian method for RBDO », *Structural and Multidisciplinary Optimization*, vol. 40, p. 35-51, 2010.
- Oliveira-Brochado A., Martins F. V., Assessing the Number of Components in Mixture Models : a Review, Fep working papers, Universidade do Porto, Faculdade de Economia do Porto, November, 2005.
- Pearson K., « Contribution to the mathematical theory of evolution », *C. Philosophical Transactions of the Royal Society*, vol. 185, p. 71-110, 1894.
- Permuter H., Francos J., Jermyn I., « A study of Gaussian mixture models of color and texture features for image classification and segmentation », *Pattern Recognition*, vol. 39, p. 695-706, 2006.
- Pieczynski W., « Statistical image segmentation », *Machine Graphics and Vision*, vol. 1, n° 1-2, p. 261-268, 1992.
- Pieczynski W., « Chaînes de Markov Triplet, Triplet Markov Chains », *Comptes Rendus de l'Académie des Sciences - Mathématique*, vol. 335, n° 3, p. 275-278, 2002.
- Roncalli T., Gestion des risques multiples ou copules et aspects multidimensionnels du risque, Rapport de recherche, [www.thierry-roncalli.com/download/Lecture-Notes-Copula-Ensayo.pdf](http://www.thierry-roncalli.com/download/Lecture-Notes-Copula-Ensayo.pdf), Groupe de Recherche Opérationnelle - Crédit Lyonnais, 2002.
- Sakji-Nsibi S., Benazza-Benyahia A., « Copula-based statistical models for multicomponent image retrieval using a Bayesian copula selection », *Proc. of the IEEE Int. Conf. on Image Processing (ICIP'09)*, 7-10 November, Cairo, Egypt, p. 253-256, 2009.
- Sanjay-Gopal S., Hebert T. J., « Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm », *IEEE Transactions on Image Processing*, vol. 7, n° 7, p. 1014-1028, 1998.
- Sfikas G., Nikou C., Galatsanos N., Heinrich C., « MR brain tissue classification using an edge-preserving spatially variant Bayesian mixture model », *11th Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI'08)*, New York, U.S.A., 2008.
- Stauffer C., « Adaptive background mixture models for real-time tracking », *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'99)*, Ft. Collins, Colorado, U.S.A., p. 246-252, 1999.
- Sundaresan A., Varshney P., « Location Estimation of a Random Signal Source Based on Correlated Sensor Observations », *IEEE Trans. on Signal Processing*, vol. 59, n° 2, p. 787-799, February, 2011.

32 1<sup>re</sup> soumission à *Cahier Scientifique*

Titterington D. M., Smith A. F. M., Makov U. E., *Statistical analysis of finite mixture distributions*, Wiley Series in Probability and Statistics : Applied Probability and Statistics, Wiley-Interscience, Chichester, U.K., 1985.

Vardi Y., Shepp L. A., Kaufman L., « A Statistical Model for Positron Emission Tomography », *J. of the American Statistical Association*, vol. 80, n° 389, p. 89-20, 1985.

Vermunt J. K., Magidson J., *Classification : The Ubiquitous Challenge*, Heidelberg : Springer, C. Weihs und W. Gaul (eds.), chapter Hierarchical mixture models for nested data structures, p. 176-183, 2005.

Wu C. F. J., « On convergence properties of the EM algorithm », *The Annals of Statistics*, vol. 11, p. 95-103, 1983.

Xiaomei Q., Jie Z., Xiaojing S., « Archimedean copula estimation and model selection via 11-norm symmetric distribution », *Insurance : Mathematics and Economics*, vol. 46, p. 406-414, 2010.

**A. Liste de copules utilisées dans la partie expérimentale.**

**Tableau 4 – Copules  $C_p(u_1, u_2; \theta)$  à un paramètre  $\theta$  utilisées dans la partie expérimentale (FGM signifie Farlie-Gumbel-Morgenstern).**

#p	Nom	FdR $C_p$	Dfp $C_p$	$\tau$ de Kendall
0	Produit	$C_0 = u_1 u_2$	$c_0 = 1$	0
1	Gaussienne <sup>d</sup>	$C_1 = \int_0^{u_1} \int_0^{u_2} \left( \frac{\phi^{-1}(u_2) - \rho\phi^{-1}(u)}{\sqrt{1-\rho^2}} \right) \phi \left( \frac{\phi^{-1}(u_1) - \rho\phi^{-1}(u)}{\sqrt{1-\rho^2}} \right) du$	$c_1 = \frac{1}{\sqrt{1-\theta^2}} \exp\left(-\frac{1}{2}\xi^T(\rho - I)\xi\right)$	$\frac{2}{\pi} \arcsin \theta$ [-1, 1]
		où $\xi_i = \phi^{-1}(u_i)$ avec $\phi$ la distribution gaussienne standard, $\rho = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ et $I$ sont les matrices identité et de corrélation.		
2	Student <sup>d</sup>	$C_2 = \int_0^{u_1} \int_0^{u_2} \left( \frac{\sqrt{\nu+1} \left( \frac{t_\nu^{-1}(u_2) - \rho t_\nu^{-1}(u)}{\sqrt{1-\rho^2}} \right) dt_\nu}{\nu + (t_\nu^{-1}(u_1))^2} \right) dt_\nu$	$c_2 = \frac{1}{\sqrt{1-\theta^2}} \frac{\Gamma(\frac{\nu}{2}+1) \Gamma(\frac{\nu}{2})}{\Gamma^2(\frac{\nu+1}{2})} \left( 1 + \frac{1}{\nu} \xi^T \rho^{-1} \xi \right) \prod_{i=1}^2 \left( 1 + \frac{\xi_i^2}{\nu} \right)^{-\frac{\nu+1}{2}}$	$\frac{2}{\pi} \arcsin \theta$ [-1, 1]
		où $\xi_i = t_\nu^{-1}(u_i)$ avec $t_\nu$ la distribution $t$ avec $\nu$ degrés de liberté, $\rho = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix}$ est la matrice de corrélation et $\Gamma(\cdot)$ la fonction Gamma.		
3	Gumbel <sup>b</sup>	$C_3 = \exp\left(-\frac{(U_1 + U_2)^\theta}{U_1 \ln(u_1)} - \frac{U_2}{u_2 \ln(u_2)}\right)$	$c_3 = \frac{U_1}{u_1 \ln(u_1)} - \frac{U_2}{u_2 \ln(u_2)} (\theta - 1 + U_1 + U_2)^{\frac{1}{\theta}} \exp\left(-\frac{(U_1 + U_2)^\theta}{\theta}\right)$	$1 - \frac{1}{\theta}$ [0, 1]
		où $U_1 = (-\ln(u_1))^\theta$ et $U_2 = (-\ln(u_2))^\theta$ .		
4	FGM	$C_4 = u_1 u_2 (1 + \theta(1 - u_1)(1 - u_2))$	$c_4 = 1 + \theta(1 - 2u_1)(1 - 2u_2)$	$\frac{2\theta}{9}$ $\left[ -\frac{2}{9}, \frac{2}{9} \right]$
5	A section cubique	$C_5 = u_1 u_2 (1 + 2\theta(1 - u_1)(1 - u_2)(1 + u_1 + u_2 - 2u_1 u_2))$	$c_5 = 1 + 2\theta \left( (1 - u_1)(1 - u_2)(-8u_2 u_1 + 2u_1 + 2u_2 + 1) + u_1(1 - u_2)(4u_2 u_1 - u_1 - 2u_2 - 1) + (1 - u_1)u_2(4u_2 u_1 - 2u_1 - u_2 - 1) + u_1 u_2(-2u_2 u_1 + u_1 + u_2 + 1) \right)$	$\frac{2}{3}\theta - \frac{6}{225}\theta^2$ $\left[ 0, \frac{33}{200} \right]$
6	Clayton <sup>b</sup>	$C_6 = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$	$c_6 = (1 + \theta) u_1^{-1-\theta} u_2^{-1-\theta} (-1 + u_1^{-\theta} + u_2^{-\theta})^{-\frac{1}{\theta}-2}$	$\frac{\theta}{\theta+2}$ ]0, 1[
7	A12 <sup>b,c</sup>	$C_7 = \left( 1 + (U_1 + U_2)^{\frac{1}{\theta}} \right)^{-1}$	$c_7 = \frac{U_1}{u_1(u_1 - 1)} - \frac{U_2}{u_2(u_2 - 1)} (\theta - 1 + (\theta + 1)(U_1 + U_2)^{\frac{1}{\theta}}) \frac{(U_1 + U_2)^{\frac{1}{\theta}-2}}{(1 + (U_1 + U_2)^{\frac{1}{\theta}})^3}$	$1 - \frac{2}{3\theta}$ $\left[ \frac{1}{3}, 1 \right]$
		où $U_1 = \left( \frac{1}{u_1} - 1 \right)^\theta$ et $U_2 = \left( \frac{1}{u_2} - 1 \right)^\theta$		
8	A14 <sup>b,c</sup>	$C_8 = \left( 1 + (U_1 + U_2)^{\frac{1}{\theta}} \right)^{-\theta}$	$c_8 = U_1 U_2 (U_1 + U_2)^{\frac{1}{\theta}-2} \left( 1 + (U_1 + U_2)^{\frac{1}{\theta}} \right)^{-2-\theta} \frac{(\theta - 1 + 2\theta(U_1 + U_2)^{\frac{1}{\theta}})}{\theta u_1 u_2 \left( u_1^{\frac{1}{\theta}} - 1 \right) \left( u_2^{\frac{1}{\theta}} - 1 \right)}$	$1 - \frac{2}{3\theta}$ $\left[ \frac{1}{3}, 1 \right]$
		où $U_1 = \left( u_1^{-\frac{1}{\theta}} - 1 \right)^\theta$ et $U_2 = \left( u_2^{-\frac{1}{\theta}} - 1 \right)^\theta$		

6. Famille des copules elliptiques.

6. Famille des copules archimédiennes.

6. dénommé ainsi selon l'ordre d'apparence dans (Nelsen, 2006).