

# Modèles de Markov triplets à observations non gaussiennes

Dans le chapitre précédent, nous nous sommes concentrés sur la loi du processus caché et nous avons étendu le modèle de chaînes de Markov cachées à celui des chaînes semi-markoviennes cachées  $M$ -stationnaires. Dans le présent chapitre, nous étudions la loi d'observation. Pour cela, nous rappelons diverses lois non gaussiennes telles que les lois de type exponentiel, elliptiques ou les “Vecteurs aléatoires sphériquement invariants” (abréviation anglaise SIRV) fréquemment rencontrés en traitement du signal radar. Nous nous intéressons ensuite aux copules, qui sont tout d'abord présentées brièvement dans leur contexte historique. Ensuite nous décrivons leur introduction récente dans le contexte des chaînes de Markov cachées et couples [21, 23, 24, 98], qui permettent la conception de très nombreux modèles particuliers du bruit. Pour finir, nous présentons dans la dernière sous-section un modèle triplet de chaînes de Markov non stationnaires cachées avec du bruit corrélé non gaussien original. Une méthode originale d'estimation des paramètres de type ICE est proposée et la méthode non supervisée correspondante de segmentation est validée par des expériences informatiques. En particulier, nous montrons l'importance du choix de la bonne copule ; toute chose égale par ailleurs (en particulier, les mêmes marginales des observations conditionnellement aux classes), l'utilisation d'une copule différente de celle correspondant aux données peut dégrader de manière significative les résultats des segmentations.

## 4.1 Lois elliptiques, modèles exponentiels et lois de Von Mises-Fisher

Dans cette section, nous donnons deux familles de lois généralisant la loi normale. La première est celle des lois de type exponentiel et la seconde est celle des lois elliptiques. Nous commençons par donner la définition générale d'un modèle exponentiel. Dans un second temps, nous donnons l'exemple de la loi de Von Mises-Fisher qui sera utilisée dans les applications au radar au chapitre 6.

### 4.1.1 Modèles exponentiels

Soit  $(\mathcal{Y}, \mathcal{B}, \nu)$  un ensemble mesuré, où  $\mathcal{B}$  est une tribu sur  $\mathcal{Y}$  et  $\nu$  une mesure définie sur  $\mathcal{B}$ .

**Définition 4.1.1** (Modèles exponentiels). *Soit  $a$  une application de  $\mathcal{Y}$  dans  $\mathbb{R}^k$ ,  $\Theta$  un sous-ensemble de  $\mathbb{R}^d$ ,  $\theta \in \Theta$  et  $\alpha$  une application de  $\Theta$  dans  $\mathbb{R}^k$ . Une variable aléatoire à valeurs dans  $(\mathcal{Y}, \mathcal{B})$  suit une loi de type exponentiel de paramètre numérique  $\theta$  et de paramètres fonctionnels  $a$  et  $\alpha$  si sa densité par rapport à  $\nu$  s'écrit :*

$$\forall y \in \mathcal{Y}, p(y) \propto \exp(\langle \alpha(\theta), a(y) \rangle),$$

où  $\langle \cdot, \cdot \rangle$  est le produit scalaire euclidien de  $\mathbb{R}^k$ .

Lorsque le modèle exponentiel est paramétré par  $\lambda = \alpha(\theta)$ , on dit que le paramétrage est canonique.

### 4.1.2 Loi de Von Mises-Fisher

Rappelons d'abord comment est définie la mesure de Lebesgue sur la sphère  $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ , que l'on notera  $\sigma_{\mathcal{S}^{d-1}}$ . La sphère sera munie de sa tribu borélienne  $\mathcal{B}_{\mathcal{S}^{d-1}}$ . Notons également  $\mathcal{B}_{[0, +\infty[}$  la tribu borélienne de  $[0, +\infty[$ . La tribu borélienne de  $\mathbb{R}^d$  est la tribu produit  $\mathcal{B}_{[0, +\infty[} \otimes \mathcal{B}_{\mathcal{S}^{d-1}}$ ; soit celle engendrée par les boréliens, dit élémentaires,  $B_1 \times B_2$  où  $B_1 \in \mathcal{B}_{[0, +\infty[}$  et  $B_2 \in \mathcal{B}_{\mathcal{S}^{d-1}}$ . La mesure de Lebesgue sur la sphère est définie par :

$$\sigma_{\mathcal{S}^{d-1}}(B_2) = \frac{\lambda_{\mathbb{R}^d}(B_1 \times B_2)}{\int_{B_1} r^{d-1} dr},$$

qui est aussi l'intégrale de la fonction indicatrice de  $B_2$  par rapport à la mesure  $\sigma_{\mathcal{S}^{d-1}}$ . On définit ensuite l'intégrale des fonctions étagées (combinaison de fonctions indicatrices), l'intégrale d'une fonction mesurable de  $\mathcal{B}_{\mathcal{S}^{d-1}}$  dans  $\mathbb{R}$  est finalement définie en utilisant le théorème classique de Beppo-Levy. On a la formule de changement de variable suivante :

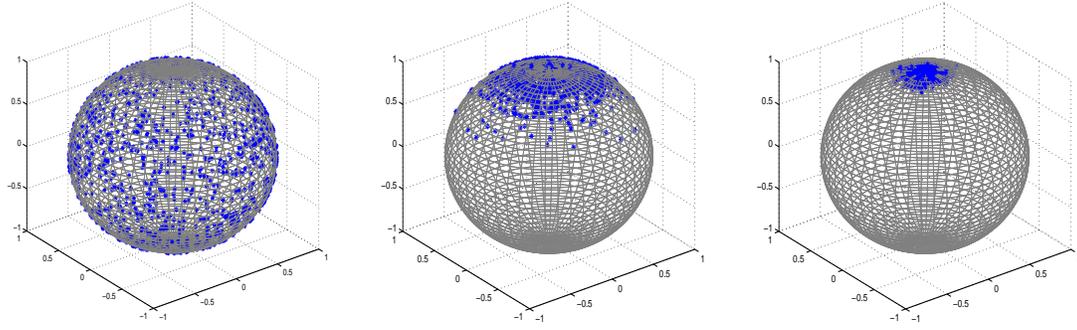
$$\int_B f(y) d\lambda_{\mathbb{R}^d}(y) = \int_{\psi^{-1}(B)} f(ru) r^{d-1} dr \otimes d\sigma_{\mathcal{S}^{d-1}}(u),$$

où  $\psi$  est l'application de  $[0, +\infty[ \times \mathcal{S}^{d-1}$  dans  $\mathbb{R}^d$  qui à  $(r, u)$  associe  $y = ru$  et  $B$  est un borélien de  $\mathbb{R}^d$ .

Soit  $\mu \in \mathcal{S}^{d-1}$  et  $\kappa \in \mathbb{R}^+$ . La loi de Von Mises-Fisher de paramètre de direction  $\mu$  et de paramètre de concentration  $\kappa$  est une loi sur la sphère de  $\mathcal{S}^{d-1}$  de densité par rapport à la mesure de Lebesgue  $\sigma_{\mathcal{S}^{d-1}}$  donnée par :

$$u \in \mathcal{S}^{d-1} \rightarrow \frac{\kappa^{\frac{d}{2}-1}}{(2\pi)^{\frac{d}{2}} I_{\frac{d}{2}-1}(\kappa)} \exp(\kappa \langle \mu, u \rangle), \quad (4.1)$$

où  $I_\nu$  est la fonction de Bessel modifiée de première espèce de paramètre  $\nu$  (voir Annexe A) et  $\langle \cdot, \cdot \rangle$  est le produit scalaire de  $\mathbb{R}^d$ .



(a) Loi uniforme sur la sphère  $\kappa = 0$ .      (b) Concentration  $\kappa = 10$ .      (c) Concentration  $\kappa = 100$ .

FIG. 4.1 – Exemples de trois lois de Von Mises-Fisher.

### Estimation des paramètres

Nous rappelons ci-après la méthode d'estimation par maximum de vraisemblance proposée dans [5], des paramètres d'une loi de Von Mises-Fisher.

Soit  $u_{1:N} = (u^{(1)}, \dots, u^{(N)})$  un échantillon de la loi de Von Mises-Fisher. La vraisemblance s'écrit :

$$p(u_{1:N}) = C \frac{\kappa^{N(\frac{d}{2}-1)}}{\left(I_{\frac{d}{2}-1}(\kappa)\right)^N} \exp\left(\kappa \left\langle \mu, \sum_{n=1}^N u^{(n)} \right\rangle\right),$$

où  $C$  est une constante indépendante de  $\kappa$  et de  $\mu$ . Ainsi la log-vraisemblance est :

$$L(u_{1:N}) = \text{Cste} + N \left(\frac{d}{2} - 1\right) \log(\kappa) - N \log\left(I_{\frac{d}{2}-1}(\kappa)\right) + \kappa \left\langle \mu, \sum_{n=1}^N u^{(n)} \right\rangle.$$

La maximisation en  $\mu$  ne dépend pas du paramètre  $\kappa$  et est atteinte en :

$$\hat{\mu}_{MV} = \frac{\sum_{n=1}^N u^{(n)}}{\left\| \sum_{n=1}^N u^{(n)} \right\|}.$$

Pour l'estimation du paramètre  $\kappa$ , on doit résoudre :

$$\frac{\partial}{\partial \kappa} \frac{I_{\frac{d}{2}-1}(\kappa)}{I_{\frac{d}{2}-1}(\kappa)} - \left(\frac{d}{2} - 1\right) \times \frac{1}{\kappa} = \frac{1}{N} \left\langle \hat{\mu}_{MV}, \sum_{n=1}^N u^{(n)} \right\rangle,$$

soit

$$\frac{\partial}{\partial \kappa} \frac{I_{\frac{d}{2}-1}(\kappa)}{I_{\frac{d}{2}-1}(\kappa)} - \left(\frac{d}{2} - 1\right) \times \frac{1}{\kappa} = \frac{1}{N} \left\| \sum_{n=1}^N u^{(n)} \right\|.$$

En utilisant le développement de Laurent de la fonction de Bessel figurant en annexe A, on montre que la dérivée de  $I_\nu$  satisfait :

$$I'_\nu(x) = \frac{\nu I_\nu(x)}{x} + I_{\nu+1}(x),$$

ainsi le maximum de vraisemblance  $\hat{\kappa}_{MV}$  satisfait :

$$\frac{I_{\frac{d}{2}}(\hat{\kappa}_{MV})}{I_{\frac{d}{2}-1}(\hat{\kappa}_{MV})} = \frac{1}{N} \left\| \sum_{n=1}^N u^{(n)} \right\|,$$

il est unique car la fonction  $\kappa \rightarrow \frac{I_{\frac{d}{2}}(\kappa)}{I_{\frac{d}{2}-1}(\kappa)}$  est croissante.

La loi de Von Mises-Fisher étant de type exponentiel, l'étape de maximisation de l'algorithme EM aboutit à des équations similaires. Détaillons l'algorithme EM dans le cas des chaînes de Markov cachées à bruit indépendant.

Soit  $u_{1:N} = (u^{(1)}, \dots, u^{(N)})$  la réalisation observée et  $x_{1:N} = (x_1, \dots, x_N)$  la réalisation cachée,  $\mu_j, \kappa_j$  les paramètres de la loi de Von Mises-Fisher  $p(u^{(n)}|x_n = \omega_j)$  et  $\theta_q$  le vecteur paramètre obtenu à l'étape  $q$  de EM. Alors les paramètres  $\mu_{q+1,j}$  et  $\kappa_{q+1,j}$  obtenus à l'étape  $q+1$  sont :

$$\mu_{q+1,j} = \frac{\sum_{n=1}^N u^{(n)} p(x_n = \omega_j | u_{1:N}; \theta_q)}{N \left\| \sum_{n=1}^N u^{(n)} p(x_n = \omega_j | u_{1:N}; \theta_q) \right\|} ;$$

$$\frac{I_{\frac{d}{2}}(\kappa_{q+1,j})}{I_{\frac{d}{2}-1}(\kappa_{q+1,j})} = \frac{\left\| \sum_{n=1}^N u^{(n)} p(x_n = \omega_j | u_{1:N}; \theta_q) \right\|}{\sum_{n=1}^N p(x_n = \omega_j | u_{1:N}; \theta_q)}.$$

Divers résultats concernant la segmentation et l'estimation des paramètres dans les modèles où les observations suivent des lois de Von Mises-Fisher sont présentés au chapitre 6 sur des données réelles issues de radar bande HF.

### 4.1.3 Lois elliptiques

Une variable aléatoire prenant ses valeurs dans  $\mathbb{R}^d$  suit une loi elliptique si les isodensités sont des ellipsoïdes de  $\mathbb{R}^d$ . En d'autres termes, la densité d'une loi elliptique est définie par :

**Définition 4.1.2.** Soit  $m \in \mathbb{R}^d$ ,  $\Sigma$  une matrice réelle symétrique définie positive de dimension  $d \times d$  et  $h$  une fonction de  $\mathbb{R}^+$  dans  $\mathbb{R}^+$ . Une variable aléatoire  $Y$  à valeurs dans  $\mathbb{R}^d$  suit une loi elliptique de paramètres euclidiens  $m \in \mathbb{R}^d$  et  $\Sigma$  et de paramètre fonctionnel  $h$  si sa densité s'écrit :

$$\forall y \in \mathbb{R}^d, p(y) = \frac{1}{\sqrt{\det \Sigma}} \times h \left( \left\| \Sigma^{-\frac{1}{2}}(y - m) \right\|^2 \right).$$

La proposition suivante donne le lien entre lois elliptiques et lois uniformes sur la sphère :

**Proposition 4.1.1.** *Une variable aléatoire  $Y$  suit une loi elliptique de paramètres euclidiens  $m$  et  $\Sigma$  si et seulement si  $Y$  s'écrit :*

$$Y = R\Sigma^{\frac{1}{2}}U + m,$$

où  $U$  et  $R$  sont indépendantes,  $U$  suit une loi uniforme sur la sphère  $\mathcal{S}^{d-1}$  et  $R$  est une variable aléatoire à valeurs dans  $\mathbb{R}^+$ .

Dans ce cas, le paramètre fonctionnel de la loi de  $Y$  est donné par :

$$\forall r \in \mathbb{R}^+, h(r) = \frac{\Gamma\left(\frac{d}{2} + 1\right)}{d\pi^{\frac{d}{2}}r^{\frac{d-1}{2}}}f(\sqrt{r}),$$

où  $f$  est la densité de la variable aléatoire  $R$  et  $\Gamma$  est la fonction eulérienne rappelée en annexe A.

Par ailleurs, il est possible de calculer les moments d'une loi elliptique. En effet, nous avons :

**Proposition 4.1.2.** *Soit  $Y$  une variable aléatoire elliptique de paramètres euclidiens  $m$  et  $\Sigma$  et de paramètre fonctionnel  $h$ . Sous la condition :*

$$\int_0^{+\infty} h(r^2)r^{d+1}dr < +\infty,$$

$Y$  est de carré intégrable, sa moyenne et sa matrice de covariance sont alors données par :

$$\mathbb{E}(Y) = m,$$

$$\text{Cov}(Y) = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2} + 1\right)} \left( \int_0^{+\infty} h(r^2)r^{d+1}dr \right) \Sigma.$$

Dans la section suivante, nous présentons des lois elliptiques particulières : les vecteurs aléatoires sphériquement invariants (SIRV). Ces vecteurs aléatoires sont le produit d'une variable aléatoire positive et d'un vecteur gaussien. Un vecteur gaussien est elliptique ; il est le produit de la loi uniforme sur la sphère et de la racine carrée d'une loi du  $\chi^2$  ; il s'ensuit que la loi d'un SIRV est elliptique.

## 4.2 Vecteurs aléatoires sphériquement invariants

Dans cette section, nous étudions un cas particulier de distributions elliptiques, celui des vecteurs aléatoires sphériquement invariants (SIRV). Un SIRV de dimension  $d$  est le produit de deux variables aléatoires : la “texture” qui est une variable aléatoire à valeurs positives, et le “speckle” qui est un vecteur aléatoire gaussien de  $\mathbb{R}^d$  ou  $\mathbb{C}^d$ , selon que le SIRV soit réel ou complexe. La terminologie est empruntée à celle des spécialistes du signal radar [55, 105, 118, 119]. En traitement du signal radar, à une distance et un angle de visée donnés, le signal réfléchi est un vecteur complexe appelé “données In Phase-Quadrature (IQ)”. Ce vecteur complexe est souvent considéré comme un SIRV, le “speckle” pouvant s'interpréter physiquement comme le chatoiement optique dû à l'excitation des électrons et la texture comme les fluctuations spatiales macroscopiques du signal réfléchi. Nous détaillerons l'acquisition des données IQ au chapitre 6.

### 4.2.1 Lois SIRV à valeurs dans $\mathbb{R}^d$ : définition et exemples

**Définition 4.2.1** (Vecteurs aléatoires sphériquement invariants réels). *Une variable aléatoire  $Y$  à valeurs dans  $\mathbb{R}^d$  est un vecteur sphériquement invariant s'il s'écrit :*

$$Y = R \times Z + m,$$

où  $R$  est une variable aléatoire réelle positive appelée *texture*,  $Z$  est un vecteur aléatoire gaussien centré à valeurs dans  $\mathbb{R}^d$  appelé “*speckle*”, et  $m$  est un vecteur réel appelé *paramètre de position*.

A partir de maintenant, on notera  $\Sigma$  la matrice de covariance du “*speckle*”. Lorsque le paramètre de position  $m = 0$  et  $\Sigma$  est la matrice identité, on dira que le SIRV est centré réduit. Les deux principaux exemples de SIRV réels sont :

#### Lois de Student sur $\mathbb{R}^d$

La texture est l'inverse de la racine carré d'une variable aléatoire de loi  $\Gamma$  de paramètre de forme  $\nu$ , sa densité est définie sur  $\mathbb{R}^+$  par :

$$p(r) = \frac{2}{\Gamma(\nu)} \frac{1}{r^{2\nu+1}} \exp\left(-\frac{1}{r^2}\right).$$

La loi de Student de dimension  $d$  de paramètres  $\nu$ ,  $m$  et  $\Sigma$  est notée  $\mathcal{S}_{\mathbb{R}^d}(\nu, m, \Sigma)$  et sa densité est :

$$y \in \mathbb{R}^d \rightarrow \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} \times \frac{\Gamma(\nu + \frac{d}{2})}{\Gamma(\nu)} \times \frac{1}{[1 + \frac{1}{2}(y - m)^T \Sigma^{-1}(y - m)]^{\nu + \frac{d}{2}}}$$

#### Lois $K$ sur $\mathbb{R}^d$

La texture est la racine carrée d'une loi  $\Gamma$  de paramètre de forme  $\nu$ , sa densité est définie sur  $\mathbb{R}^+$  par :

$$p(r) = \frac{2}{\Gamma(\nu)} r^{2\nu-1} \exp(-r^2).$$

La densité d'une loi  $K$  de paramètres  $\nu$ ,  $m$  et  $\Sigma$ , notée  $K_{\mathbb{R}^d}(\nu, m, \Sigma)$ , est :

$$y \in \mathbb{R}^d \rightarrow \frac{2}{(2\pi)^{\frac{d}{2}} \sqrt{\det(\Sigma)} \Gamma(\nu)} \left( \frac{(y - m)^T \Sigma^{-1}(y - m)}{2} \right)^{\frac{\nu-d}{2}} K_{\nu-\frac{d}{2}} \left( 2\sqrt{\frac{(y - m)^T \Sigma^{-1}(y - m)}{2}} \right),$$

où  $K_\nu$  désigne la fonction de Bessel de deuxième espèce modifiée à  $\nu$  degrés de liberté (voir Annexe A).

### 4.2.2 Lois gaussiennes complexes circulaires et lois SIRV complexes

Nous introduisons ci-après, les lois gaussiennes complexes circulaires, importantes pour la suite, notamment dans les algorithmes de détection sur les données radar. A partir de ces lois gaussiennes complexes, nous définirons les lois SIRV à valeurs dans un  $\mathbb{C}$ -espace vectoriel.

### Lois gaussiennes complexes circulaires

**Définition 4.2.2** (Variable aléatoire gaussienne complexe circulaire). *Une variable aléatoire  $Z$  à valeurs dans  $\mathbb{C}$  est gaussienne si le vecteur  $(\operatorname{Re}(Z), \operatorname{Im}(Z))$  est un vecteur gaussien de  $\mathbb{R}^2$ . Elle est dite :*

- centrée si  $\mathbb{E}(Z) = 0$  ;
- circulaire si  $\operatorname{Re}(Z)$  et  $\operatorname{Im}(Z)$  sont indépendantes et de même variance.

On note  $\sigma_{\mathbb{C}}^2$  la variance  $\mathbb{E}(|Z - \mathbb{E}(Z)|^2)$  et  $m = \mathbb{E}(Z) = \mathbb{E}(\operatorname{Re}(Z)) + i\mathbb{E}(\operatorname{Im}(Z))$  la moyenne de  $Z$ .

La densité d'une variable aléatoire gaussienne complexe circulaire  $Z$  de moyenne  $m$  et de variance  $\sigma_{\mathbb{C}}^2$  est :

$$z \in \mathbb{C} \rightarrow \frac{1}{\pi\sigma_{\mathbb{C}}^2} \exp\left(-\frac{|z - m|^2}{\sigma_{\mathbb{C}}^2}\right).$$

On définit également :

**Définition 4.2.3** (Vecteurs aléatoires gaussiens complexes circulaires). *Un vecteur aléatoire complexe  $Z = (Z_1, \dots, Z_d)$  est gaussien si pour tout  $(\lambda_1, \dots, \lambda_d) \in \mathbb{C}^d$ , la variable aléatoire  $\lambda_1 Z_1 + \dots + \lambda_d Z_d$  est une variable aléatoire gaussienne complexe.*

*Le vecteur aléatoire  $Z$  est circulaire si  $\mathbb{E}((Z - \mathbb{E}(Z))(Z - \mathbb{E}(Z))^T) = 0$ . On définit sa matrice de covariance complexe par :*

$$\Sigma_{\mathbb{C}} = \mathbb{E}((Z - \mathbb{E}(Z))(\bar{Z} - \mathbb{E}(\bar{Z}))^T).$$

La densité d'un vecteur aléatoire gaussien complexe circulaire de moyenne  $m$  et de covariance complexe  $\Sigma_{\mathbb{C}}$  est donnée par :

$$z \in \mathbb{C}^d \rightarrow \frac{1}{\pi^d \det \Sigma_{\mathbb{C}}} \exp\left(-\overline{(z - m)}^T \Sigma_{\mathbb{C}}^{-1} (z - m)\right).$$

Les SIRV réels sont étendus aux SIRV complexes pour lesquels le “speckle” est une variable aléatoire gaussienne complexe circulaire. Nous donnons ci-dessous les densités des lois de Student et  $K$  complexes.

#### Lois de Student sur $\mathbb{C}^d$

La densité d'une loi de Student complexe est donnée par :

$$y \in \mathbb{C}^d \rightarrow \frac{\Gamma(\nu + d)}{\pi^d \Gamma(\nu) \det(\Sigma_{\mathbb{C}})} \times \frac{1}{\left[1 + \overline{(y - m)}^T \Sigma_{\mathbb{C}}^{-1} (y - m)\right]^{\nu + d}}.$$

#### Lois $K$ sur $\mathbb{C}^d$

La densité d'une loi  $K$  complexe est donnée par :

$$y \in \mathbb{C}^d \rightarrow \frac{2}{\pi^d \Gamma(\nu) \det(\Sigma_{\mathbb{C}})} \left(\overline{(y - m)}^T \Sigma_{\mathbb{C}}^{-1} (y - m)\right)^{\frac{\nu - d}{2}} K_{\nu - d} \left(2\sqrt{\overline{(y - m)}^T \Sigma_{\mathbb{C}}^{-1} (y - m)}\right).$$

### 4.3 Copules et lois multivariées non gaussiennes

Nous allons présenter dans cette section des lois multivariées qui peuvent se définir explicitement à partir des lois marginales et d'un terme d'agrégation appelé "copule".

L'introduction des copules est historiquement issue des travaux de M. Fréchet et de ceux de A. Sklar. Les travaux de M. Fréchet [45] concernaient les familles de vecteurs aléatoires ayant mêmes lois marginales. Il a ainsi donné le nom de "classes de Fréchet" aux classes de la relation d'équivalence  $[Y = (Y_1, \dots, Y_d) \sim \tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_d)] \Leftrightarrow [Y \text{ et } \tilde{Y} \text{ ont mêmes lois marginales}]$ . Cependant, M. Fréchet n'a pas établi la forme générale des fonctions de répartition jointes. Il a fallu attendre le théorème établi par A. Sklar [110, 112], qui porte actuellement son nom, pour pouvoir exprimer la fonction de répartition jointe en fonction des fonctions de répartition marginales. Nous présenterons dans la sous-section 4.3.1 ci-après, la problématique qui a conduit A. Sklar à son théorème.

#### 4.3.1 Copules et théorème de Sklar

Les travaux de A. Sklar ne concernaient pas directement l'écriture de la loi jointe fonction des lois marginales. Il travaillait avec B. Schweizer sur la généralisation de la conjonction "et" en logique floue [43, 50, 63, 88] et sur les espaces métriques probabilisés [80, 110]. En logique floue, la valeur de vérité d'une formule est étendue en une fonction à valeurs dans  $[0, 1]$  mesurant la croyance que l'on a d'une proposition. Cette fonction de croyance peut être une probabilité, auquel cas on parle de logique floue probabilisée. La valeur de vérité de la conjonction a été étendue par A. Sklar et B. Schweizer à l'aide des normes triangulaires :

**Définition 4.3.1** (Normes triangulaires). *Une application  $T : [0, 1]^2 \rightarrow [0, 1]$  est une norme triangulaire bivariée si elle satisfait :*

- $T(u, v) = T(v, u)$  (commutativité) ;
- $T(T(u, v), w) = T(u, T(v, w))$  (associativité) ;
- $T(u, 1) = u$  ;
- Si  $u_1 \leq u_2$  et  $v_1 \leq v_2$  alors  $T(u_1, v_1) \leq T(u_2, v_2)$ .

*Une norme triangulaire de dimension  $d$  est une application  $T_d$  de  $[0, 1]^d$  dans  $[0, 1]$  définie récursivement à partir d'une norme triangulaire bivariée  $T$  par :*

1.  $T_3(u, v, w) = T(T(u, v), w)$  ;
2.  $T_{n+1}(u_1, \dots, u_{n+1}) = T(T_n(u_1, \dots, u_n), u_{n+1})$ .

La croyance de l'expression "A et B", notée  $A \wedge B$ , est alors donnée par  $\text{Bel}(A \wedge B) = T(\text{Bel}(A), \text{Bel}(B))$ , où  $\text{Bel}(A)$  (resp.  $\text{Bel}(B)$ ) désigne le degré de croyance de A (resp. B). Les espaces métriques probabilisés sur lesquels travaillèrent A. Sklar et B. Schweizer sont des ensembles de variables aléatoires  $(W_{x,y})_{(x,y) \in \mathcal{X}^2}$  à valeurs dans  $\mathbb{R}^+$  et indexées sur un espace métrique  $\mathcal{X}$ , tels que  $W_{x,y}$  a les propriétés d'une distance :

- $\forall (x, y, t), \mathbb{P}(W_{x,y} \leq t) = \mathbb{P}(W_{y,x} \leq t)$  (symétrie) ;
- $\mathbb{P}(W_{x,y} = 0) = 1$  si et seulement si  $x = y$  ;
- $\forall (x, y, z, t, s), \mathbb{P}(W_{x,z} \leq t; W_{z,y} \leq s) \leq \mathbb{P}(W_{x,y} \leq t + s)$  (inégalité triangulaire).

La distance entre deux points  $x$  et  $y$  de  $\mathcal{X}$  est ainsi généralisée en utilisant la fonction de répartition  $t \rightarrow F_{x,y}(t) = \mathbb{P}(W_{x,y} \leq t)$ . La problématique qui a conduit A. Sklar à son

théorème est la suivante. Dans un espace métrique classique, si on connaît les distances  $d(x, z)$  et  $d(z, y)$ , on sait majorer la distance  $d(x, y)$ . Dans un espace métrique probabilisé, on voudrait également minorer la fonction de répartition  $F_{x,y}$  lorsque l'on connaît les fonctions de répartitions  $F_{x,z}$  et  $F_{z,y}$ . L'idée est alors d'exprimer  $\mathbb{P}(W_{x,z} \leq t; W_{z,y} \leq s)$  en fonction de  $F_{x,z}$  et  $F_{z,y}$ . Le théorème formulé par A. Sklar affirme qu'il est possible d'écrire  $\mathbb{P}(W_{x,z} \leq t; W_{z,y} \leq s)$  en fonction de  $F_{x,z}$  et  $F_{z,y}$  grâce à un terme d'agrégation appelée copule.

Nous donnons ci-dessous la définition d'une copule telle qu'elle est énoncée par R. B. Nelsen [85]. On définit pour cela l'opérateur  $\Delta_{u_k}^{v_k}$  qui à une fonction  $C'$  de  $[0, 1]^{d'}$  dans  $[0, 1]$ , pour  $d' \leq d$ , associe la fonction de  $[0, 1]^{d'-1}$  dans  $[0, 1]$  définie par :

$$\forall x \in [0, 1]^{d'-1}, \Delta_{u_k}^{v_k} C'(x) = C'(x_1, \dots, v_k, \dots, x_{d'}) - C'(x_1, \dots, u_k, \dots, x_{d'}).$$

**Définition 4.3.2** (Copule de dimension  $d$ ). *Une fonction  $C : [0, 1]^d \rightarrow [0, 1]$  est une copule si elle vérifie :*

1. pour tout  $i$  et pour tout  $(u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_d) \in [0, 1]^{d-1}$ ,  
 $C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_d) = 0$ ;
2. pour tout  $i$  et tout  $u_i \in [0, 1]$ ,  $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ ;
3. pour tout  $u, v \in [0, 1]^d$  tel que pour tout  $i$ ,  $u_i \leq v_i$ , on a  $C = \Delta_{u_d}^{v_d} \Delta_{u_{d-1}}^{v_{d-1}} \dots \Delta_{u_1}^{v_1} C \geq 0$ .

**Remarque :** Une copule de dimension  $d$  peut s'interpréter comme la fonction de répartition d'une loi à valeurs dans  $[0, 1]^d$  dont les lois marginales sont uniformes sur  $[0, 1]$ .

Enonçons maintenant le théorème de Sklar :

**Théorème 4.3.1** (Théorème de Sklar). *Soit  $F$  une fonction de répartition sur  $\mathbb{R}^d$  de fonctions de répartitions marginales  $F_1, \dots, F_d$ . Alors il existe une copule  $C$  sur  $[0, 1]^d$  telle que :*

$$\forall y \in \mathbb{R}^d, F(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d)). \quad (4.2)$$

*De plus, si les marges  $F_1, \dots, F_d$  sont continues, alors la copule est unique.*

La réciproque de ce théorème est donnée par le théorème de Deheuvels :

**Théorème 4.3.2** (Théorème de Deheuvels). *Si  $F_1, \dots, F_d$  sont  $d$  fonctions de répartition à support réel et si  $C$  est une copule de dimension  $d$ , alors il existe un vecteur aléatoire  $Y$  de dimension  $d$  de marginales  $F_j$  et de copule  $C$  tels que la fonction de répartition  $F$  de  $Y$  soit donnée par :*

$$\forall y \in \mathbb{R}^d, F(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d)).$$

*Preuve.* Voir [37]. □

**Remarque :** En plus de répondre aux attentes de A. Sklar, les copules répondent aux attentes de M. Fréchet. Elles permettent d'exprimer la loi jointe de vecteurs aléatoires à partir des lois marginales. Sous l'hypothèse de continuité des marginales, l'unique copule est

la fonction de répartition de  $U = (F_1(Y_1), \dots, F_d(Y_d))$ . Si de plus, la fonction de répartition  $F$  est dérivable, nous avons :

$$\forall y \in \mathbb{R}^d, f(y_1, \dots, y_d) = f_1(y_1) \dots f_d(y_d) c(F_1(y_1), \dots, F_d(y_d)) \quad (4.3)$$

où  $f$  est la densité jointe,  $f_1, \dots, f_d$  sont les densités marginales et  $c$  est la densité de la copule, soit la densité du vecteur  $U = (F_1(Y_1), \dots, F_d(Y_d))$ .

Après avoir formulé son théorème, A. Sklar voulait trouver des conditions pour que  $\mathbb{P}(W_{x,z} \leq t, W_{z,y} \leq s)$  représente la croyance sur  $(W_{x,z} \leq t, W_{z,y} \leq s)$ . Il voulait ainsi trouver des conditions pour que la copule  $C$  soit une norme triangulaire. Dans ce cas, l'espace métrique probabilisé est appelé espace métrique flou ou espace de Menger [80]. Les conditions pour qu'une copule soit une norme triangulaire sont données par la proposition 4.3.1. Rappelons d'abord la définition d'application lipschitzienne.

**Définition 4.3.3** (Application lipschitzienne). *Une application  $T : I \subset \mathbb{R}^d \rightarrow \mathbb{R}$  est lipschitzienne si :*

$$\text{Pour tout } u \in I \text{ et tout } v \in I, |T(u) - T(v)| \leq \sum_{k=1}^d |u_k - v_k|.$$

Nous avons :

**Proposition 4.3.1.**

- Une copule bivariée est une norme triangulaire si et seulement si elle satisfait  $C(C(u, v), w) = C(u, C(v, w))$  ;
- Une norme triangulaire est une copule si et seulement si elle est lipschitzienne.

Nous donnons dans la sous-section suivante les différents exemples de copules que nous utiliserons par la suite.

## 4.3.2 Exemples de copules

### Copule produit

La copule produit est la copule d'un vecteur aléatoire à composantes indépendantes. Cette copule est donnée par :

$$C(u_1, \dots, u_d) = \prod_{j=1}^d u_j.$$

### Bornes de Fréchet

L'expression de la borne supérieure de Fréchet est :

$$C(u_1, u_2, \dots, u_d) = \min(u_1, u_2, \dots, u_d).$$

La borne inférieure de Fréchet est une copule uniquement dans le cas bivarié et est donnée par :

$$C(u_1, u_2, \dots, u_d) = \max(u_1 + \dots + u_d - d + 1, 0).$$

La terminologie “supérieure” et “inférieure” provient de l’inégalité suivante :

$$\max(u_1 + \dots + u_d - d + 1, 0) \leq C(u_1, \dots, u_d) \leq \min(u_1, \dots, u_d),$$

où  $C$  est une copule de dimension  $d$ .

### Copule gaussienne

La copule gaussienne est l’unique copule telle que si chaque marginale  $Y_k$  d’un vecteur aléatoire  $(Y_1, \dots, Y_d)$  suit une loi normale  $\mathcal{N}_{\mathbb{R}}(m_k, \sigma_k^2)$ , alors le vecteur joint suit une loi normale  $\mathcal{N}_{\mathbb{R}^d}((m_1, \dots, m_d), \Sigma)$ .

Elle est donnée par :

$$c(u_1, \dots, u_d) = \frac{1}{\sqrt{\det R}} \exp \left( -\frac{1}{2} (\phi^{-1}(u_1), \dots, \phi^{-1}(u_d))^T (R^{-1} - \text{Id}) (\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)) \right),$$

où  $\phi$  est la fonction de répartition d’une loi normale réelle centrée et réduite et  $R$  est la matrice de corrélation de Pearson donnée par :

$$R = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \dots & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & \frac{1}{\sigma_d} \end{pmatrix} \Sigma \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \dots & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & 0 & \frac{1}{\sigma_d} \end{pmatrix}.$$

### Copule de Student

La copule de Student est l’unique copule telle que si pour tout  $k \in \{1, \dots, d\}$ ,  $Y_k$  suit une loi de Student de paramètre de forme  $\nu$  alors le vecteur aléatoire  $(Y_1, \dots, Y_d)$  suit une loi de Student de même paramètre de forme.

La densité de la copule de Student de paramètre de forme  $\nu$  et de matrice de corrélation  $R$  est donnée par :

$$c(u_1, \dots, u_d) = \frac{1}{\sqrt{\det R}} \times \frac{\Gamma(\nu + \frac{d}{2}) \Gamma(\nu)^{d-1}}{\Gamma(\nu + \frac{1}{2})^d} \times \frac{\left[ \left( 1 + \frac{(\phi^{-1}(u_1))^2}{2} \right) \dots \left( 1 + \frac{(\phi^{-1}(u_d))^2}{2} \right) \right]^{\nu + \frac{1}{2}}}{\left[ 1 + \frac{1}{2} (\phi^{-1}(u_1), \dots, \phi^{-1}(u_d))^T R^{-1} (\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)) \right]^{\nu + \frac{d}{2}}}. \quad (4.4)$$

où  $\phi$  est la fonction de répartition de la loi de Student monovariée de paramètre de forme  $\nu$ , centrée et réduite.

## Copule $K$

La copule  $K$  est l'unique copule telle que si les lois marginales sont des lois  $K$  de paramètre de forme  $\nu$ , alors le vecteur joint suit encore une loi  $K$  de même paramètre de forme.

La densité de la copule  $K$  s'écrit :

$$c(u_1, \dots, u_d) = \frac{\Gamma(\nu)^{d-1}}{2^{d-1}} \times \frac{2^{\frac{\nu}{2}(d-1)}}{\sqrt{\det R}} \times \frac{[\xi^T R^{-1} \xi]^{\frac{\nu}{2} - \frac{d}{4}}}{[\xi_1 \dots \xi_d]^{\nu - \frac{1}{2}}} \times \frac{K_{\nu - \frac{d}{2}}(\sqrt{2\xi^T R^{-1} \xi})}{\prod_{j=1}^d K_{\nu - \frac{1}{2}}(\sqrt{2}\xi_j)}, \quad (4.5)$$

où  $\xi = (\phi_\nu^{-1}(u_1), \dots, \phi_\nu^{-1}(u_d))$  avec  $\phi_\nu$  fonction de répartition d'une loi  $K$  monovariée de paramètre de forme  $\nu$ , centrée et réduite et  $K_\nu$  est la fonction de Bessel modifiée de seconde espèce de paramètre de forme  $\nu$ .

Parmi les autres exemples de copules, on peut citer également les copules archimédiennes [46, 76, 117], qui présentent surtout un intérêt dans le cas bivarié.

### 4.3.3 Mesures de dépendance

Dans cette sous-section, nous abordons les différentes mesures de dépendance, appelées aussi mesures d'association entre variables aléatoires. Une mesure de dépendance est un moyen de quantifier la dépendance entre variables aléatoires, ou encore la manière dont les deux variables aléatoires sont liées. La mesure de dépendance la plus classique entre deux variables aléatoires réelles  $Y_1$  et  $Y_2$  est le coefficient de corrélation de Pearson donné par :

$$\rho = \frac{\sigma_{Y_1, Y_2}}{\sqrt{\sigma_{Y_1}^2 \sigma_{Y_2}^2}},$$

où  $\sigma_{Y_i}^2$  est la variance de la variable aléatoire  $Y_i$  et  $\sigma_{Y_1, Y_2}$  est la covariance entre  $Y_1$  et  $Y_2$ .

Cependant, le coefficient de corrélation de Pearson a ses limites. Premièrement, il n'existe que si les variables aléatoires sont de carré intégrable. Deuxièmement, il s'avère inefficace pour modéliser certaines situations de dépendance. Par exemple, on aimerait que si  $Y_2 = f(Y_1)$ , alors la corrélation vaut 1 si  $f$  est croissante ou  $-1$  si  $f$  est décroissante, mais ce n'est pas le cas. En effet, à titre d'exemple, si  $Y_1$  suit une loi uniforme sur  $[1, 2]$  et si  $Y_2 = \frac{1}{Y_1}$ ,

alors  $\rho = \sqrt{12} \times \frac{2 - 3 \log(2)}{\sqrt{2 - 4 (\log(2))^2}} \simeq -0.98$ . La corrélation de Pearson s'avère être une bonne

mesure de corrélation si  $f$  est linéaire.

### Corrélation $\rho_S$ de Spearman

**Définition 4.3.4** (Corrélation de Spearman). *Soit  $(Y_1, Y_2)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^2$ , de copule  $C$  et de fonctions de répartition marginales continues  $F_1$  et  $F_2$ . La corrélation de Spearman entre  $Y_1$  et  $Y_2$  est la corrélation de Pearson entre les variables aléatoires  $U_1 = F_1(Y_1)$  et  $U_2 = F_2(Y_2)$  définie par :*

$$\rho_S = 12\sigma_{U_1, U_2},$$

où  $\sigma_{U_1, U_2}$  est la covariance entre  $U_1$  et  $U_2$ .

**Remarque :** Lorsque la copule  $C$  d'un vecteur aléatoire  $(Y_1, Y_2)$  admet une densité  $c$ , la corrélation de Spearman est également donnée par :

$$\rho_S = \rho(U, V) = 12 \int_{[0,1]^2} uvc(u, v) dudv - 3.$$

La proposition suivante nous montre qu'une dépendance du type  $Y_2 = f(Y_1)$ , où  $f$  est un  $\mathcal{C}^1$ -difféomorphisme de  $\mathbb{R}$  dans  $\mathbb{R}$ , est équivalente à  $\rho_S = 1$  ou  $-1$ .

**Proposition 4.3.2.** *Soit  $\rho_S$  la corrélation de Spearman entre deux variables aléatoires réelles  $Y_1$  et  $Y_2$  de fonctions de répartition respectives  $F_1$  et  $F_2$ , qui sont des  $\mathcal{C}^1$ -difféomorphismes de  $\mathbb{R}$  dans  $[0, 1]$ . Nous avons :*

- $\rho_S = 1$  si et seulement si  $Y_2 = f(Y_1)$  où  $f$  est un  $\mathcal{C}^1$ -difféomorphisme de  $\mathbb{R}$  dans  $\mathbb{R}$  strictement croissant ;
- $\rho_S = -1$  si et seulement si  $Y_2 = g(Y_1)$  où  $g$  est un  $\mathcal{C}^1$ -difféomorphisme de  $\mathbb{R}$  dans  $\mathbb{R}$  strictement décroissant.

*Preuve.* Notons  $U_1 = F_1(Y_1)$  et  $U_2 = F_2(Y_2)$ . Si  $\rho_S = 1$ ,  $\rho_S$  étant la corrélation de Pearson du couple  $(U_1, U_2)$ , alors  $U_2 = \lambda U_1 + b$ , avec  $\lambda > 0$ . Comme  $U_1$  et  $U_2$  suivent toutes les deux la même loi, alors  $\lambda = 1$  et  $b = 0$ . On en déduit que  $Y_2 = (F_2^{-1} \circ F_1)(Y_1)$ . Si  $\rho_S = -1$ , alors par le même raisonnement,  $U_2 = 1 - U_1$ , d'où  $Y_2 = F_2^{-1}(1 - F_1(Y_1))$ .

Pour la réciproque, on se sert du lemme :

**Lemme 4.3.1.** *Soient  $U$  et  $V$  deux variables aléatoires suivant une loi uniforme sur  $[0, 1]$ . Si  $V = g(U)$  où  $g$  est un  $\mathcal{C}^1$ -difféomorphisme de  $[0, 1]$  dans  $[0, 1]$ , alors nécessairement  $g(u) = u$  pour tout  $u \in [0, 1]$  ou  $g(u) = 1 - u$  pour tout  $u \in [0, 1]$ .*

*Preuve du lemme.* Il suffit d'écrire la densité de  $U$  fonction de celle de  $V$ . Si  $V$  suit la loi uniforme, pour que  $U$  suive la loi uniforme, il est nécessaire que  $|g'(u)| = 1$  pour tout  $u$  et que  $g([0, 1]) = [0, 1]$ . On en déduit le résultat.  $\square$

Ainsi si  $Y_2 = f(Y_1)$ , où  $f$  est un  $\mathcal{C}^1$ -difféomorphisme, alors  $U_2 = (F_2 \circ f \circ F_1^{-1})(U_1)$  d'où, en utilisant le lemme,  $U_2 = U_1$  si  $f$  est croissante et  $U_2 = 1 - U_1$  si  $f$  est décroissante. Ainsi  $\rho_S = 1$  si  $f$  est croissante et  $-1$  si  $f$  est décroissante.  $\square$

**Remarque :** Par construction, la corrélation de Spearman ne dépend que de la copule. Ainsi, la corrélation de Spearman de  $(f(Y_1), g(Y_2))$  où  $f$  et  $g$  sont deux  $\mathcal{C}^1$ -difféomorphismes croissants est égale à la corrélation de Spearman de  $(Y_1, Y_2)$ , les deux vecteurs aléatoires ayant même copule.

### Mesure de concordance-discordance de Kendall et corrélation $\tau$ de Kendall

La corrélation de Kendall dérive de la notion de concordance-discordance définie ci-après et est très utilisée en finance et économétrie.

**Définition 4.3.5** (Corrélation de Kendall). *Soient deux vecteurs aléatoires  $(Y_1, Y_2)$  et  $(\tilde{Y}_1, \tilde{Y}_2)$  de  $\mathbb{R}^2$  indépendants suivant la même loi bivariée. On appelle “coefficient de corrélation de Kendall” de  $(Y_1, Y_2)$ , la quantité :*

$$\tau = \mathbb{P}((Y_1 - \tilde{Y}_1)(Y_2 - \tilde{Y}_2) > 0) - \mathbb{P}((Y_1 - \tilde{Y}_1)(Y_2 - \tilde{Y}_2) < 0).$$

**Remarque :** Si deux variables aléatoires  $Y_1$  et  $Y_2$  sont liées par une relation du type  $Y_2 = f(Y_1)$  où  $f$  est une fonction monotone de  $\mathbb{R}$  dans  $\mathbb{R}$ , alors  $\tau = 1$  si  $f$  est croissante et  $\tau = -1$  si  $f$  est décroissante.

Si la copule d’un vecteur aléatoire bivarié  $(Y_1, Y_2)$  admet une densité, alors sa corrélation de Kendall est donnée par :

$$\tau = 4 \int_{[0,1]^2} C(u, v)c(u, v)dudv - 1.$$

La corrélation de Spearman et celle de Kendall sont des cas particuliers d’une mesure de dépendance appelée mesure de concordance-discordance et définie par :

**Définition 4.3.6.** *Soient  $(Y_1, Y_2)$  et  $(\tilde{Y}_1, \tilde{Y}_2)$  deux vecteurs aléatoires bivariés indépendants de mêmes lois marginales et de copules respectives  $C$  et  $\tilde{C}$ . La mesure de concordance-discordance  $Q(C, \tilde{C})$  est définie par*

$$Q(C, \tilde{C}) = \mathbb{P}((Y_1 - \tilde{Y}_1)(Y_2 - \tilde{Y}_2) > 0) - \mathbb{P}((Y_1 - \tilde{Y}_1)(Y_2 - \tilde{Y}_2) < 0).$$

**Remarque :** Si les copules respectives  $C$  et  $\tilde{C}$  de deux vecteurs aléatoires bivariés  $(Y_1, Y_2)$  et  $(\tilde{Y}_1, \tilde{Y}_2)$  de mêmes lois marginales admettent des densités, alors la mesure de concordance-discordance est donnée par :

$$Q(C, \tilde{C}) = 4 \int_{[0,1]^2} C(u, v)\tilde{c}(u, v)dudv - 1.$$

On en déduit les propriétés suivantes :

- $\rho_S = 3Q(C, \Pi)$ , où  $\Pi$  désigne la copule produit ;
- $\tau = Q(C, C)$ .

### Coefficients de mélangeance

Les coefficients de mélangeance sont des mesures de dépendance asymptotique. L’étude de la mélangeance de processus stationnaires permet de déduire leurs propriétés asymptotiques [58, 116]. Soient  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé,  $\mathcal{F}$  et  $\mathcal{G}$  deux sous-tribus de  $\mathcal{A}$ . Notons  $\mathbb{P}(A|\mathcal{G})$  l’espérance sachant  $\mathcal{G}$  de la variable aléatoire qui à  $\omega \in \Omega$  associe 1 si  $\omega \in A$  et 0 sinon. Les différents coefficients de mélangeance entre  $\mathcal{F}$  et  $\mathcal{G}$  sont définis par :

- $\alpha$ -mélangeance, ou mélangeance forte :  $\alpha(\mathcal{F}, \mathcal{G}) = \sup_{A \in \mathcal{F}, B \in \mathcal{G}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$  ;
- $\beta$ -mélangeance :  $\beta(\mathcal{F}, \mathcal{G}) = \mathbb{E} \left( \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{P}(A|\mathcal{G})| \right)$  ;
- $\rho$ -mélangeance :  $\rho(\mathcal{F}, \mathcal{G}) = \sup \{ |\rho(X, Y)| : X \text{ } \mathcal{F}\text{-mesurable, } Y \text{ } \mathcal{G}\text{-mesurable} \}$ , où  $\rho$  est la corrélation de Pearson ;

$$- \phi\text{-mélangeance} : \phi(\mathcal{F}, \mathcal{G}) = \sup_{A \in \mathcal{F}, B \in \mathcal{G}, \mathbb{P}(B) > 0} |\mathbb{P}(A) - \mathbb{P}(A|B)|.$$

Nous avons les inégalités  $2\alpha(\mathcal{F}, \mathcal{G}) \leq \beta(\mathcal{F}, \mathcal{G}) \leq \phi(\mathcal{F}, \mathcal{G})$  et  $4\alpha(\mathcal{F}, \mathcal{G}) \leq \rho(\mathcal{F}, \mathcal{G}) \leq 2\sqrt{\phi(\mathcal{F}, \mathcal{G})}$ . Soit  $c$  un des coefficients de mélangeance. On dit qu'un processus  $(Y_n)_{n \geq 0}$  est  $c$ -mélangeant, si  $\lim_{n \rightarrow +\infty} c(\sigma(Y_0), \sigma((Y_k)_{k \geq n})) = 0$ , où  $\sigma(Z)$  est la tribu engendrée par la variable aléatoire  $Z$ .

Soient  $f$  et  $g$  deux  $\mathcal{C}^1$ -difféomorphismes croissants. Comme  $Y$  et  $f(Y)$  engendrent la même tribu et  $(Y_1, Y_2)$  et  $(f(Y_1), g(Y_2))$  ont la même copule, ainsi les coefficients de mélangeance entre deux tribus  $\sigma(Y_1)$  et  $\sigma(Y_2)$  ne dépendent que de la copule.

Par conséquent, la corrélation d'un processus  $Y = (Y_n)_{n \geq 1}$  aura le même comportement asymptotique que celle de tout processus  $Z = (f_n(Y_n))_{n \geq 1}$ , où chaque  $f_n$  est une fonction croissante. Nous reviendrons sur cette remarque au chapitre 5 lorsque l'on définira la dépendance longue dans les processus non gaussiens.

## 4.4 Chaînes de Markov cachées $M$ -stationnaires à bruit corrélé non gaussien

Nous proposons dans cette section un modèle original et montrons son intérêt en segmentation non supervisée des données non stationnaires, corrélées et non gaussiennes.

### 4.4.1 Copules dans les chaînes de Markov cachées $M$ -stationnaires

Considérons un processus triplet  $(X, U, Y) = (X_n, U_n, Y_n)_{1 \leq n \leq N}$  tel que chaque  $X_n$  prend ses valeurs dans l'ensemble fini  $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$ , chaque  $U_n$  prend ses valeurs dans l'ensemble fini  $\Lambda = \{\lambda_1, \dots, \lambda_M\}$ , et chaque  $Y_n$  prend ses valeurs dans  $\mathbb{R}$ . Nous proposons de généraliser dans cette sous-section le modèle de chaînes de Markov cachées  $M$ -stationnaires vu au chapitre 3 au cas d'observations corrélées et non gaussiennes. La loi du triplet  $(X, U, Y)$  sera de la forme :

$$\begin{aligned} p(x_{1:N}, u_{1:N}, y_{1:N}) &= p(u_1)p(x_1|u_1)p(y_1|x_1) \\ &\times \prod_{n=1}^{N-1} p(u_{n+1}|u_n)p(x_{n+1}|x_n, u_{n+1})p(y_{n+1}|x_n, x_{n+1}, y_n). \end{aligned} \quad (4.6)$$

On remarque que les processus  $U$  et  $(X, U)$  sont des chaînes de Markov. On supposera que la chaîne de Markov  $(X, U)$  est stationnaire réversible, soit  $p(x_n = \omega_i, u_n = \lambda_k, x_{n+1} = \omega_j, u_{n+1} = \lambda_l) = p(x_n = \omega_j, u_n = \lambda_l, x_{n+1} = \omega_i, u_{n+1} = \lambda_k)$  et ne dépend pas de  $n$ . Les lois initiales  $p(u_1)$  et  $p(x_1|u_1)$  et les transitions  $p(u_{n+1}|u_n)$  et  $p(x_{n+1}|x_n, u_{n+1})$  sont définies à partir de  $p(x_n, u_n, x_{n+1}, u_{n+1})$ . On supposera également que  $p(y_n, y_{n+1}|x_n = \omega_i, x_{n+1} = \omega_j) = p(y_{n+1}, y_n|x_n = \omega_j, x_{n+1} = \omega_i)$  et ne dépend pas de  $n$ , ainsi la chaîne de Markov  $(X, U, Y)$  est également stationnaire et réversible. Plus exactement, la loi jointe  $p(y_n, y_{n+1}|x_n, x_{n+1})$  sera donnée par :

$$p(y_n, y_{n+1}|x_n, x_{n+1}) = p(y_n|x_n)p(y_{n+1}|x_{n+1})c_{x_n, x_{n+1}}(F_{x_n}(y_n), F_{x_{n+1}}(y_{n+1})), \quad (4.7)$$

où  $c_{x_n, x_{n+1}}$  est la copule de la loi jointe  $p(y_n, y_{n+1}|x_n, x_{n+1})$  et  $F_{x_n}$  (resp.  $F_{x_{n+1}}$ ) est la fonction de répartition de la loi marginale  $p(y_n|x_n)$  (resp.  $p(y_{n+1}|x_{n+1})$ ).

Afin d'avoir l'égalité  $p(y_n, y_{n+1}|x_n = \omega_i, x_{n+1} = \omega_j) = p(y_{n+1}, y_n|x_n = \omega_j, x_{n+1} = \omega_i)$ ,

on supposera  $c_{\omega_i, \omega_j}(F_{\omega_i}(y_n), F_{\omega_j}(y_{n+1})) = c_{\omega_j, \omega_i}(F_{\omega_j}(y_{n+1}), F_{\omega_i}(y_n))$ . La loi conditionnelle  $p(y_{n+1}|x_n, x_{n+1}, y_n)$  est exprimée par :

$$p(y_{n+1}|x_n, x_{n+1}, y_n) = p(y_{n+1}|x_{n+1})c_{x_n, x_{n+1}}(F_{x_n}(y_n), F_{x_{n+1}}(y_{n+1})). \quad (4.8)$$

#### 4.4.2 Estimation des paramètres

Nous proposons dans cette sous-section de détailler l'algorithme ICE dans le cas du modèle de chaînes de Markov cachées  $M$ -stationnaires à bruit corrélé non gaussien. Les paramètres à estimer sont les lois  $p(x_n, u_n, x_{n+1}, u_{n+1})$ , les lois marginales  $p(y_n|x_n)$  et la copule de chaque loi jointe  $p(y_n, y_{n+1}|x_n, x_{n+1})$ . L'estimation de  $p(x_n, u_n, x_{n+1}, u_{n+1})$  a déjà été étudiée dans la sous-section 3.2.2 du chapitre 3, ainsi nous nous consacrons uniquement à l'estimation de la loi jointe  $p(y_n, y_{n+1}|x_n, x_{n+1})$ . Celle-ci se déroule en deux temps. Notons  $\theta_q$  le paramètre estimé à l'étape  $q$  de l'algorithme ICE. Nous commençons par ré-estimer les lois marginales, puis avec les nouvelles valeurs courantes des lois marginales, nous ré-estimons la copule. Considérons le cas où les marginales  $p(y_n|x_n = \omega_i)$  sont des lois normales  $\mathcal{N}_{\mathbb{R}}(m_i, \sigma_i^2)$ . Soit  $y_{1:N} = (y_1, \dots, y_N)$  la réalisation observée. Les estimateurs de  $m_i$  et  $\sigma_i^2$  à partir des données complètes  $(x_{1:N}, u_{1:N}, y_{1:N})$  sont :

$$\begin{aligned} \hat{m}_i(x_{1:N}, y_{1:N}) &= \frac{\sum_{n=1}^N y_n I(x_n = \omega_i)}{\sum_{n=1}^N I(x_n = \omega_i)}, \\ \hat{\sigma}_i^2(x_{1:N}, y_{1:N}) &= \frac{\sum_{n=1}^N (y_n - \hat{m}_i(x_{1:N}, y_{1:N}))^2 I(x_n = \omega_i)}{\sum_{n=1}^N I(x_n = \omega_i)}. \end{aligned}$$

L'espérance sachant  $y_{1:N}$  sous le paramètre  $\theta_q$  n'est pas calculable explicitement. Ainsi, on simule  $L$  échantillons indépendants  $(x_{1:N}^{(m)}, u_{1:N}^{(m)})$  pour  $1 \leq m \leq L$  selon la loi a posteriori  $p(x_{1:N}, u_{1:N}|y_{1:N}; \theta_q)$ . Les ré-estimation de  $m_i$  et  $\sigma_i^2$  sont alors données par :

$$\begin{aligned} m_i^{q+1} &= \frac{1}{L} \sum_{m=1}^L \hat{m}_i(x_{1:N}^{(m)}, u_{1:N}^{(m)}, y_{1:N}^{(m)}), \\ (\sigma_i^{q+1})^2 &= \frac{1}{L} \sum_{m=1}^L \hat{\sigma}_i^2(x_{1:N}^{(m)}, u_{1:N}^{(m)}, y_{1:N}^{(m)}). \end{aligned}$$

Dans les expérimentations présentées dans la sous-section suivante, nous prendrons  $L = 1$ . Notons maintenant  $F_{\omega_i}^{q+1}$  la fonction de répartition de la loi normale  $\mathcal{N}_{\mathbb{R}}(m_i^{q+1}, (\sigma_i^{q+1})^2)$ . La deuxième étape de ICE consiste à estimer la copule. Dans le cas des copules gaussiennes (resp. de Student), notons  $\phi$  la fonction de répartition de la loi normale centrée réduite (resp. loi de Student centrée réduite). Le paramètre de corrélation  $\rho_{x_n, x_{n+1}}$  de la copule  $c_{x_n, x_{n+1}}$  de  $p(y_n, y_{n+1}|x_n, x_{n+1})$  est estimé à partir de l'échantillon  $z_{1:N} = (z_1, \dots, z_N)$  où  $z_n = \phi^{-1} \circ F_{\omega_i}^{q+1}(y_n)$ . En effet,  $p(z_n, z_{n+1}|x_n, x_{n+1})$  suit une loi normale (resp. de Student) de corrélation

$\rho_{x_n, x_{n+1}}$ . Un estimateur de  $\rho_{\omega_i, \omega_j}$  à partir des données complètes  $(x_{1:N}, u_{1:N}, z_{1:N})$  est donné par :

$$\hat{\rho}_{\omega_i, \omega_j}(x_{1:N}, z_{1:N}) = \frac{\hat{\sigma}_{\omega_i, \omega_j}^{(1,2)}}{\hat{\sigma}_{\omega_i, \omega_j}^{(1)} \hat{\sigma}_{\omega_i, \omega_j}^{(2)}},$$

où

$$\begin{aligned} (\hat{\sigma}_{\omega_i, \omega_j}^{(1)})^2 &= \frac{\sum_{n=1}^{N-1} (z_n - \hat{m}_{\omega_i, \omega_j}^{(1)})^2 I(x_n = \omega_i, x_{n+1} = \omega_j)}{\sum_{n=1}^{N-1} I(x_n = \omega_i, x_{n+1} = \omega_j)}, \\ (\hat{\sigma}_{\omega_i, \omega_j}^{(2)})^2 &= \frac{\sum_{n=1}^{N-1} (z_{n+1} - \hat{m}_{\omega_i, \omega_j}^{(2)})^2 I(x_n = \omega_i, x_{n+1} = \omega_j)}{\sum_{n=1}^{N-1} I(x_n = \omega_i, x_{n+1} = \omega_j)}, \\ \hat{\sigma}_{\omega_i, \omega_j}^{(1,2)} &= \frac{\sum_{n=1}^{N-1} (z_n - \hat{m}_{\omega_i, \omega_j}^{(1)})(z_{n+1} - \hat{m}_{\omega_i, \omega_j}^{(2)}) I(x_n = \omega_i, x_{n+1} = \omega_j)}{\sum_{n=1}^{N-1} I(x_n = \omega_i, x_{n+1} = \omega_j)}, \end{aligned}$$

avec

$$\begin{aligned} \hat{m}_{\omega_i, \omega_j}^{(1)} &= \frac{\sum_{n=1}^{N-1} z_n I(x_n = \omega_i, x_{n+1} = \omega_j)}{\sum_{n=1}^{N-1} I(x_n = \omega_i, x_{n+1} = \omega_j)}, \\ \hat{m}_{\omega_i, \omega_j}^{(2)} &= \frac{\sum_{n=1}^{N-1} z_{n+1} I(x_n = \omega_i, x_{n+1} = \omega_j)}{\sum_{n=1}^{N-1} I(x_n = \omega_i, x_{n+1} = \omega_j)}. \end{aligned}$$

Comme pour l'estimation des lois marginales, l'espérance sachant  $y_{1:N}$  (resp.  $z_{1:N}$ ) de  $\hat{\rho}_{\omega_i, \omega_j}$  n'est pas calculable explicitement. Ainsi on simule  $L$  échantillons indépendants  $(x_{1:N}^{(m)}, u_{1:N}^{(m)})$  pour  $1 \leq m \leq L$  selon la loi a posteriori  $p(x_{1:N}, u_{1:N} | y_{1:N}; \theta_q)$  et la ré-estimation de  $\rho_{\omega_i, \omega_j}$  est :

$$\rho_{\omega_i, \omega_j}^{q+1} = \frac{1}{L} \sum_{m=1}^L \hat{\rho}_{\omega_i, \omega_j}(x_{1:N}^{(m)}, u_{1:N}^{(m)}, z_{1:N}^{(m)}).$$

### 4.4.3 Expérimentations

Nous présentons dans cette sous-section deux expériences, dans lesquelles nous considérons une chaîne de Markov M-stationnaire  $(X, U, Y)$  dont la distribution est donnée par (4.6) et (4.7). La réalisation de processus mono-dimensionnels sera transformée en image bi-dimensionnelle grâce au parcours d'Hilbert-Peano figurant en annexe B. En reprenant les notations de la sous-section 4.4.1, on prendra pour ces deux expériences,  $K = 2$  et  $M = 2$ . Les paramètres de  $(X, U)$  seront donnés par :

$$\begin{aligned} p(u_n = \lambda_1, u_{n+1} = \lambda_1) &= p(u_n = \lambda_2, u_{n+1} = \lambda_2) = 0.49995, \\ p(u_n = \lambda_1, u_{n+1} = \lambda_2) &= p(u_n = \lambda_2, u_{n+1} = \lambda_1) = 0.00005, \end{aligned}$$

$$\begin{aligned} p(x_n = \omega_1, x_{n+1} = \omega_1 | u_{n+1} = \lambda_1) &= p(x_n = \omega_2, x_{n+1} = \omega_2 | u_{n+1} = \lambda_1) = 0.495, \\ p(x_n = \omega_2, x_{n+1} = \omega_1 | u_{n+1} = \lambda_1) &= p(x_n = \omega_2, x_{n+1} = \omega_1 | u_{n+1} = \lambda_1) = 0.005, \\ p(x_n = \omega_1, x_{n+1} = \omega_1 | u_{n+1} = \lambda_2) &= p(x_n = \omega_2, x_{n+1} = \omega_2 | u_{n+1} = \lambda_2) = 0.45, \\ p(x_n = \omega_2, x_{n+1} = \omega_1 | u_{n+1} = \lambda_2) &= p(x_n = \omega_2, x_{n+1} = \omega_1 | u_{n+1} = \lambda_2) = 0.05. \end{aligned}$$

Dans les deux expériences, les lois marginales  $p(y_n | x_n)$  seront respectivement la loi normale  $\mathcal{N}_{\mathbb{R}}(0, 1)$  lorsque  $x_n = \omega_1$  et la loi normale  $\mathcal{N}_{\mathbb{R}}(2, 1)$  lorsque  $x_n = \omega_2$ . Dans la première expérience, nous simulons les données avec une copule gaussienne et dans la seconde, nous simulons avec une copule de Student. Les données seront ensuite segmentées, dans chacune de ces expériences, en utilisant le modèle avec copule gaussienne et le modèle avec copule de Student. Le but de ces expériences est de savoir si le choix de la copule a de l'importance lorsque l'on ne connaît pas la manière dont les données sont corrélées, et si le choix d'une copule différente de celle du vrai modèle dégrade considérablement les résultats.

#### Segmentation de données simulées avec une copule gaussienne

Dans cette première expérience, les données sont simulées à l'aide d'une copule gaussienne de paramètres  $\rho_{\omega_1, \omega_1} = \rho_{\omega_2, \omega_2} = 0.9$  et  $\rho_{\omega_1, \omega_2} = \rho_{\omega_2, \omega_1} = 0$ . Ainsi,  $p(y_n, y_{n+1} | x_n, x_{n+1})$  est une distribution normale de corrélation  $\rho_{x_n, x_{n+1}}$ . Les données sont ensuite segmentées par trois méthodes. La première méthode utilise les vrais paramètres du modèle et les états cachés sont estimés par MPM. La seconde méthode utilise le vrai modèle, celui dont la copule est gaussienne, les paramètres sont estimés par ICE et les états cachés estimés par MPM. Quant à la dernière méthode, elle utilise le modèle dont la copule est de Student de paramètre de forme  $\nu = 10$ , les autres paramètres étant estimés par ICE, et les états cachés par MPM.

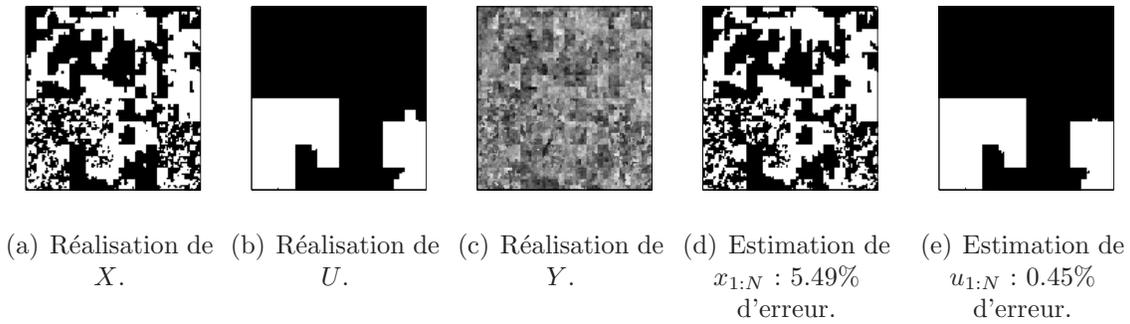


FIG. 4.2 – Segmentation avec les vrais paramètres de données simulées avec la copule gaussienne.

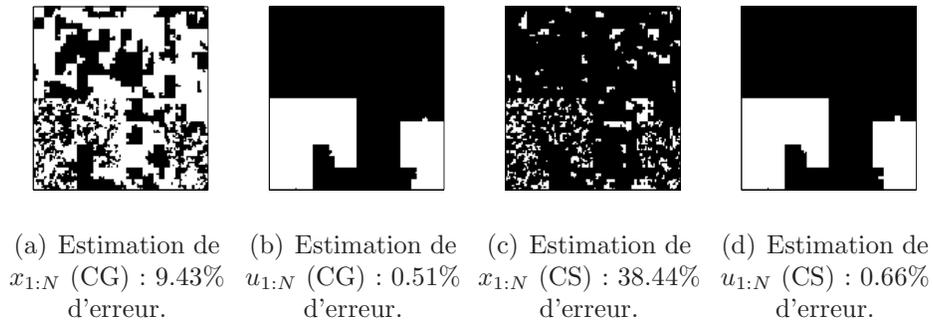


FIG. 4.3 – Segmentation non supervisée de données simulées avec la copule gaussienne en utilisant le vrai modèle avec copule gaussienne (CG) et le modèle avec copule de Student (CS).

		CG		CS		Vraies valeurs	
		$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$
$m_{\omega_i}$		-0.05	2.04	0.41	1.75	0	2
$\sigma_{\omega_i}^2$		0.91	0.96	1.63	1.04	1	1
$\rho$	$\omega_1$	0.89	-0.05	0.93	0.18	0.9	0
	$\omega_2$	-0.05	0.92	0.18	0.78	0	0.9
Taux d'erreur pour $X$		9.43%		38.44%		5.49%	
Taux d'erreur pour $U$		0.51%		0.66%		0.45%	

TAB. 4.1 – Estimation de la loi d'observation en utilisant le vrai modèle avec copule gaussienne (CG) et le modèle avec copule de Student (CS).

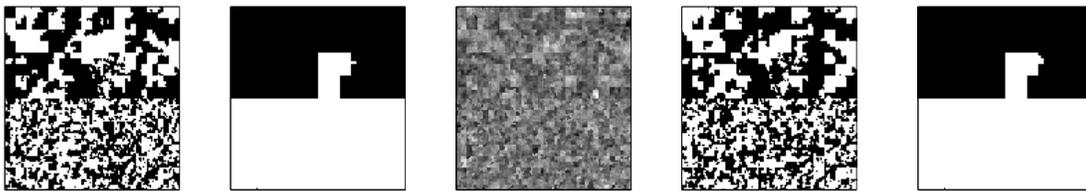
		CG				CS				
		$\lambda_1$		$\lambda_2$		$\lambda_1$		$\lambda_2$		
$p(u_n = \lambda_k, u_{n+1} = \lambda_l)$		$\lambda_1$	0.4999		0.0001		0.4999		0.0001	
		$\lambda_2$	0.0001		0.4999		0.0001		0.4999	
$p(x_n = \omega_i, x_{n+1} = \omega_j   u_{n+1} = \lambda_l)$		$\omega_1$	0.49	0.01	0.45	0.05	0.49	0.01	0.48	0.02
		$\omega_2$	0.01	0.49	0.05	0.45	0.01	0.49	0.02	0.48

TAB. 4.2 – Estimation de la loi de  $(X, U)$  en utilisant le vrai modèle avec copule gaussienne (CG) et le modèle avec copule de Student (CS).

De ces résultats, nous constatons que le choix d'une copule différente de celle qui a permis de simuler les données peut dégrader fortement les résultats. La moyenne et la variance sont mal estimées par le modèle utilisant une copule de Student, ce qui a pour conséquence une mauvaise estimation des états cachés. Cependant, on peut remarquer que la loi de  $U$  ainsi que sa réalisation demeurent bien estimées.

### Segmentation de données simulées avec une copule de Student

Dans cette expérience, les données sont simulées à l'aide d'une copule de Student de paramètre de forme  $\nu = 10$  et de paramètres de corrélation  $\rho_{\omega_1, \omega_1} = \rho_{\omega_2, \omega_2} = 0.9$  et  $\rho_{\omega_1, \omega_2} = \rho_{\omega_2, \omega_1} = 0$ . Les données sont ensuite segmentées par trois méthodes : la première utilise les vrais paramètres, la seconde utilise le modèle avec copule gaussienne et les paramètres sont estimés avec ICE et la dernière méthode utilise le vrai modèle avec copule de Student et les paramètres également estimés par ICE. Dans tous les cas, la segmentation est faite par l'estimateur du MPM.



(a) Réalisation de  $X$ . (b) Réalisation de  $U$ . (c) Réalisation de  $Y$ . (d) Estimation de  $x_{1:N}$  : 15.39% d'erreur. (e) Estimation de  $u_{1:N}$  : 0.73% d'erreur.

FIG. 4.4 – Segmentation avec les vrais paramètres de données simulées avec la copule de Student.

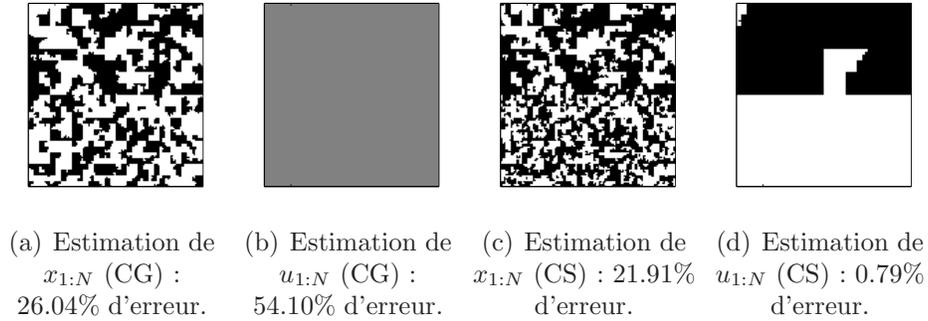


FIG. 4.5 – Segmentation de données simulées avec la copule de Student en utilisant le modèle avec copule gaussienne (CG) et le vrai modèle avec copule de Student (CS).

		CG		CS		Vraies valeurs	
		$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$
$m_{\omega_i}$		0.36	1.61	-0.13	1.94	0	2
$\sigma_{\omega_i}^2$		2.66	2.88	0.91	1.18	1	1
$\rho$	$\omega_1$	0.92	-0.01	0.89	0.01	0.9	0
	$\omega_2$	-0.01	0.94	0.01	0.91	0	0.9
Taux d'erreur pour $X$		26.4%		21.91%		15.39%	
Taux d'erreur pour $U$		54.10%		0.79%		0.73%	

TAB. 4.3 – Estimation de la loi d'observation en utilisant le modèle avec copule gaussienne (CG) et le vrai modèle avec copule de Student (CS).

		CG				CS			
		$\lambda_1$		$\lambda_2$		$\lambda_1$		$\lambda_2$	
$p(u_n = \lambda_k, u_{n+1} = \lambda_l)$	$\lambda_1$	$\sim 0.5$		$\sim 0$		0.4999		0.0001	
	$\lambda_2$	$\sim 0$		$\sim 0.5$		0.0001		0.4999	
		$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$
$p(x_n = \omega_i, x_{n+1} = \omega_j   u_{n+1} = \lambda_l)$	$\omega_1$	0.46	0.04	-	-	0.49	0.01	0.45	0.05
	$\omega_2$	0.04	0.46	-	-	0.01	0.49	0.05	0.45

TAB. 4.4 – Estimation de la loi de  $(X, U)$  en utilisant le modèle avec copule gaussienne (CG) et le vrai modèle avec copule de Student (CS).

On constate que remplacer la copule de Student par la copule gaussienne peut, en mode non supervisé, dégrader notablement les résultats. La nature dont les marginales sont corrélées doit être considérée avec précaution. Dans les deux expériences que nous avons présentées, les marginales étaient des lois normales, seule la copule était différente.

## Conclusion

Dans ce chapitre, nous nous sommes intéressés, afin de mieux modéliser les différentes variétés des bruits, aux diverses extensions des lois normales. Nous avons tout d'abord rappelé différentes généralisations de la loi normale telles que les modèles exponentiels, les lois elliptiques ou sphériquement invariantes (SIRV). En guise d'exemple de loi de type exponentiel, nous avons cité la loi de Von Mises-Fisher qui sera utilisée dans les applications radar au chapitre 6. Un autre moyen de considérer des lois multivariées non gaussiennes est d'utiliser des copules, qui permettent de modéliser un grand nombre de lois multivariées à lois marginales données. Nous avons ensuite utilisé les copules pour modéliser des observations dépendantes dans le cadre du modèle de Markov caché non stationnaire. Nous avons présenté des simulations, au sein desquelles nous avons utilisé une méthode d'estimation des paramètres de type ICE originale, montrant qu'il est important, dans le cadre de ce modèle et en non supervisé, d'utiliser la vraie copule. Nous verrons au chapitre suivant, comment utiliser les copules pour générer des observations à dépendance longue.