



# Generalized FGM copulas : Properties and applications

ACT-7102 : Modèles avancés de la théorie du risque

Réalisé par :

**David YIGBE** et **Stanislas MBOUNGOU**

Sous la supervision de :

**Hélène Cossette**, professeure à l'Université Laval et **Ilie Radu Mitric**,  
professeur à l'Université Laval

Date : 2 avril 2024

# Table des matières

<b>1</b>	<b>INTRODUCTION</b>	<b>3</b>
<b>2</b>	<b>REVUE DE LITTERATURE</b>	<b>4</b>
<b>3</b>	<b>DEFINITIONS</b>	<b>5</b>
3.1	Généralités des copules . . . . .	5
3.2	Les copules Farlie-Gumbel-Morgenstern . . . . .	6
3.3	Sous-familles de la classe générale . . . . .	9
3.4	La corrélation rho de Spearman . . . . .	14
<b>4</b>	<b>ESTIMATION DES PARAMÈTRES DES COPULES</b>	<b>21</b>
4.1	La méthode du maximum de vraisemblance . . . . .	21
<b>5</b>	<b>Extension de certaines copules FGM connues</b>	<b>23</b>
5.1	Extension de la copule FGM . . . . .	23
5.2	Extension de la copule Hung-Kotz . . . . .	24
5.3	Extension de la copule Bairamov-Kotz . . . . .	25
<b>6</b>	<b>Application sur les données réelles</b>	<b>25</b>
6.1	Présentation de l'ensemble de données . . . . .	25
6.2	Régression de BMI et PED . . . . .	26
6.3	Analyse de la structure de dépendance entre BMI et PED . . . . .	26
6.3.1	Test de Kolmogorov-Smirnov . . . . .	32
6.3.2	Comparaison des fonctions de répartition conjointes . . . . .	37
<b>7</b>	<b>CONCLUSION</b>	<b>38</b>
<b>A</b>	<b>Représentation 3D de la densité de la copule FGM et de ses extensions</b>	<b>39</b>
	<b>Bibliographie</b>	<b>44</b>

# 1 INTRODUCTION

Cet article traite des copules Farlie-Gumbel-Morgensten (FGM) et s'intéresse particulièrement à une limite de celles-ci. En effet, les copules FGM sont connues pour modéliser les variables présentant une faible dépendance. Cela représente une limite en modélisation puisqu'on n'est pas toujours assuré d'avoir une dépendance faible entre les variables. Rappelons qu'une copule est une fonction qui décrit complètement la structure de dépendance entre deux ou plusieurs variables aléatoires. Elle contient toutes les informations nécessaires pour relier les distributions marginales à leur distribution conjointe, grâce à un ou plusieurs coefficients de dépendances. Chacun de coefficient de dépendance est inférieur à un en valeur absolue. C'est d'ailleurs ce qui empêche le coefficient de corrélation de dépasser certaines valeurs, vu la relation qui le lie au paramètre de dépendance. Cette relation sera définie en fonction de la sous-famille de copules considérée. La plus connue des mesures d'association est :

$$\rho_s = \frac{\theta}{3}$$

pour une copule FGM bivariée, où  $\theta$  est le coefficient de dépendance. Le coefficient de corrélation ne peut donc pas excéder la valeur 0,33 en valeur absolue. *Hakim Bekrizadeh*, dans son article qui fait l'objet de notre étude, part du théorème de *Sklar* et des copules *FGM* pour construire des sous-familles de copules ayant une plage de valeurs plus grande pour les coefficients de corrélation de Spearman. La nouvelle classe :

$$C_{\theta}^{\psi,p}(s, t) = st[1 + \theta\psi(s)\psi(t)]^p, p \in [0, \infty), \forall (s, t) \in [0, 1]^2, \theta \in \Theta \in [-1, 1]$$

définie par notre auteur, permet de modéliser une dépendance élevée, ce qui donne la possibilité d'élargir l'éventail des applications potentielles de la classe de copule FGM dans diverses branches de la technologie et de la science, dont la médecine en particulier. C'est d'ailleurs à ce domaine que s'intéressera *Bekrizadeh*, qui illustre l'efficacité de son modèle en appliquant quelques sous-familles de sa nouvelle classe pour modéliser un ensemble de données sur le diabète.

## 2 REVUE DE LITTERATURE

Dans la littérature, on trouve plusieurs travaux sur les copules qui ont essayé d'élargir la plage de valeurs admissibles du coefficient de corrélation de *Spearman* dans le cadre de la copule FGM. *Bekrizadeh* s'est beaucoup inspiré des travaux menés par ses pairs, dont nous présentons les plus saillants dans les lignes qui suivent. En 1999, *Huang et Kotz* ont mis au point des extensions à paramètre unique de type polynomial de la copule FGM. Ils ont montré que la borne supérieure du coefficient de corrélation de *Spearman* peut être augmentée jusqu'à environ 0.375 tandis que la borne inférieure reste à 0.33. Un an plus tard, les chercheurs *Lai et Xie* ont étudié une extension de la distribution uniforme bivariée qui a des effets positifs sur la distribution uniforme en généralisant la représentation uniforme de la distribution des FGM. Puis, par une simple transformation, ils ont également obtenu des familles de distribution bivariées avec des marges prédéfinies. Cependant, leur coefficient de corrélation ne variait qu'entre  $-0.103$  et  $+0.128$ . En 2002, *Kotz* qui avait travaillé avec *Huang* en 1999, a décidé de tenter une nouvelle approche. Ainsi, en collaboration avec *Baimarov*, il a pu élargir la famille qu'il a définie avec *Kotz* par le passé, et a pu obtenir des résultats plus satisfaisants. Le coefficient de corrélation pouvait aller jusqu'à 0.516 et sa borne inférieure pouvait descendre à  $-0.48$ . Tout cela démontre les efforts fournis par les scientifiques, mais aussi l'importance de l'utilisation des copules dans la modélisation de nos jours. D'ailleurs, ces travaux ont mené à plusieurs applications récentes dont la simulation des tempêtes marines multivariées tentée par *Corbella et Stretch* en 2013, la structure de dépendance entre les marchés boursiers et des changes (*Wang, Wu et Lai, 2013*), la gestion des risques opérationnels (*Arbenz 2013*), et l'optimisation de portefeuille en présence de rendements financiers dépendants à long terme effectuée par *Boubaker et Sghaier* en 2013 également.

Nous ne pouvons finir sans mentionner que le travail de *Bekrizadeh* a, à son tour, inspiré d'autres auteurs tels que *Christopher Blier-Wong, Hélène Cossette, Sébastien Legros et Etienne Marceau*, qui dans un article commun publié en 2023, ont présenté par une approche stochastique, une nouvelle méthode de construction de copules de grande dimension basée sur des mélanges de fonctions de puissance, qui peut être adaptée à des contextes plus généraux pour construire des familles plus larges de copules. Plus récemment, en janvier 2024, sur la base des travaux de *Bekrizadeh, Susanne Saminger-Platz, Anna Kolesárová, Adam Šeliga, Radko Mesiar et Erich Peter Klement* ont pu définir ensemble de nouvelles copules bivariées (et même ce qu'ils appellent des quasi-copules) avec des propriétés intéressantes, en appliquant sur plusieurs fonctions (dont certaines définies par notre auteur) des transformations paramétrées et d'autres constructions telles que la transposition, le retournement et la troncature.

### 3 DEFINITIONS

#### 3.1 Généralités des copules

Une copule est une fonction qui permet de décrire complètement la structure de dépendance entre différents risques. Elle contient toutes les informations permettant de relier des distributions marginales à leur distribution conjointe.

Une copule multivariée est donc essentiellement une fonction de répartition conjointe sur  $[0, 1]^n$  d'un vecteur de variables aléatoires  $U = (U_1, \dots, U_n)$  dont chaque composante  $U_i$ , ( $i = 1, \dots, n$ ) est de loi uniforme sur  $[0, 1]$ . Elle satisfait aux propriétés suivantes :

- $C(u_1, u_2)$  est non décroissante sur  $[0, 1]^2$ .
- $C(u_1, u_2)$  est continue sur  $[0, 1]^2$ .
- $\lim_{u_i \rightarrow 0} C(u_1, u_2) = 0$ ,  $i = 1, 2$
- $\lim_{u_1 \rightarrow 1} C(u_1, u_2) = u_2$  et  $\lim_{u_2 \rightarrow 1} C(u_1, u_2) = u_1$
- Pour tout rectangle  $(a_1, b_1] \times (a_2, b_2] \in [0, 1]^2$ , où on a :  $a_1 \leq b_1$  et  $a_2 \leq b_2$ , on a :  
 $P_r(U_1 \in (a_1, b_1], U_2 \in (a_2, b_2]) = C(b_1, b_2) - C(a_1, b_2) - C(a_2, b_1) + C(a_1, a_2)$ .

Dans notre contexte, nous considérerons  $(x, s, t, s_1, s_2, t_1, t_2) \in [0, 1]^7$ . On aura simplement :

1.  $C(x, 0) = C(0, x) = 0$  et  $C(x, 1) = C(1, x) = x, \forall x \in [0, 1]$
2.  $\forall (s_1, s_2, t_1, t_2) \in [0, 1]^4$ , tel que  $s_1 \leq s_2$  et  $t_1 \leq t_2$ ,  
 $C(s_2, t_2) - C(s_2, t_1) - C(s_1, t_2) + C(s_1, t_1) \geq 0$  (P 2)

Éventuellement, pour des fonctions deux fois différentiables, cette dernière propriété peut être remplacée par la condition

$$c(s, t) = \frac{\partial^2 C(s, t)}{\partial s \partial t} \quad (1)$$

Ceci se justifie par le fait qu'on peut recourir à l'expression ci-contre :

$$f_{X_1 X_2}(x_1, x_2) = \frac{\partial^2 F(X_1, X_2(x_1, x_2))}{\partial x_1 \partial x_2} \quad (2)$$

où  $c(s, t)$  est aussi appelé la densité de la copule. Une copule  $C$  est symétrique si  $C(s, t) = C(t, s)$ , pour tout  $(s, t) \in [0, 1]^2$ , sinon  $C$  est antisymétrique. Autrement dit, la relation 1 devient triviale dès qu'on reconnaît les termes de la relation (P-2) comme définissant une probabilité.

**Définition 1.** *Vecteur aléatoire*

Soit  $X = (X_1, \dots, X_n)$  un vecteur de  $n$  variables aléatoires continues avec une fonction de distribution cumulative jointe notée  $F_X$  et des marginales univariées  $F_{X_i}$ ,  $i = 1, \dots, n$ ,  $F_X$  peut être écrit comme une fonction des marginales univariées  $F_{X_i}$ ,  $i = 1, \dots, n$ , et la copule  $C$  décrivant la structure de dépendance est défini comme suit :

$$F_X(x_1, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)).$$

et la fonction de densité conjointe

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) c(F_{X_1}(x_1), \dots, F_{X_n}(x_n)),$$

**Théorème 1.** *Théorème de Sklar*

Soit un vecteur de variables aléatoires  $X = (X_1, \dots, X_n)$  avec fonction de répartition conjointe  $F_X$  et marginales univariées  $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ . Alors, il existe une copule  $C$  qui permet de représenter  $F_X$  en fonction de  $C$  et des marginales

$$F_X(x_1, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n))$$

Cette copule est unique si les variables aléatoires  $X_1, \dots, X_n$  sont continues. Soit une copule  $C$  et des marginales  $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ . Alors,

$$C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)) = F_X(x_1, \dots, x_n)$$

où la fonction de répartition multivariée  $F_X$  appartient à la classe de Fréchet  $\mathcal{CF}(F_{X_1}, \dots, F_{X_n})$ . Cette construction est valide peu importe que les variables aléatoires  $X_1, \dots, X_n$  soient continues ou discrètes.

En guise de commentaire, le théorème de Sklar a deux sections, d'abord, il existe toujours une copule qui permet de représenter une fonction de distribution conjointe en fonction de ses marginales ; et ensuite, en ayant une copule et des marginales, il est toujours possible de créer une loi multivariée. Dans le cadre de ce travail, nous nous focaliserons essentiellement sur les copules de *Farlie-Gumbel-Morgenstern* (FGM).

### 3.2 Les copules Farlie-Gumbel-Morgenstern

**Définition 2.** Soit un vecteur de deux variables aléatoires  $X = (X_1, X_2)$  avec fonction de répartition conjointe

$$F_{X_1, X_2}(x_1, x_2) = (1 - e^{-\beta_1 x_1})(1 - e^{-\beta_2 x_2}) + \theta(1 - e^{-\beta_1 x_1})(1 - e^{-\beta_2 x_2})e^{-\beta_1 x_1}e^{-\beta_2 x_2}, \quad (3)$$

où  $-1 \leq \theta \leq 1$ ,  $\beta_i > 0$ ,  $i = 1, 2$ , et  $x_1 > 0$ ,  $x_2 > 0$ .

Le vecteur  $X = (X_1, X_2)$  est dit obéir à une loi exponentielle bivariée (Eyrard) Farlie-Dumbel-Morgenstern (FGM). Dans le cadre du cours ACT-7017, nous avons vu que si un vecteur  $X = (X_1, X_2)$  obéit à une loi exponentielle bivariée, alors les lois marginales de  $X_i$ ,  $i = 1, 2$  sont exponentielles de paramètres  $\beta_i > 0$ ,  $i = 1, 2$ .

**Définition 3.** *Mesures d'associations*

La copule FGM modélise les dépendances faibles; cela se constate par les mesures d'associations. Parmi ces mesures, on trouve le coefficient de corrélation de Pearson entre deux variables aléatoires, le tau de Kendall et le rho de Spearman. Le coefficient de corrélation de Pearson peut être déduit à partir de la fonction de copule et de ses distributions marginales Nelsen(2006).

■ Le  $\tau$  de Kendall nous permet de mieux le voir

$$\tau = \frac{2\theta}{9}$$

■ Le  $\rho$  de Spearman est définie comme suit :

$$\rho_S(X_1, X_2) = \frac{\theta}{3}, \quad -1 < \theta < 1$$

■ Le  $\rho$  du coefficient de corrélation de Pearson est définie comme suit :

$$\rho_P(X_1, X_2) = \frac{(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \frac{\theta}{4}, \quad -1 < \theta < 1$$

Une dépendance modérée positive ou négative peut être modélisée avec la copule FGM.

Dans le cadre de cet article, nous définirons plutôt la copule FGM de la façon générale :

$$C^{FGM}(s, t) = st[1 + \theta(1 - s)(1 - t)], \theta \in [-1, 1], \forall (s, t) \in [0, 1]^2$$

En posant  $s = (1 - e^{-\beta_1 x_1})$  et  $t = (1 - e^{-\beta_2 x_2})$  pour  $x_1, x_2 \in [0, +\infty[$ , on retrouve notre relation 3. Le cas exponentiel bivarié n'est donc qu'un cas particulier de ce qui a été défini ci-haut.

Voulant étendre la plage de corrélation dans la copule FGM, Bekrizadeh, Parham et Zadkarami ont proposé une nouvelle famille symétrique de FGM généralisées dont la dépendance est la suivante :

$$C^{BPZ}(s, t) = st(1 + \theta(1 - s^\alpha)(1 - t^\alpha))^n \quad (4)$$

$\alpha > 0$ ,  $n = 0, 1, 2, \dots, \forall (s, t) \in [0, 1]^2$ , où  $-\min\{1, \frac{1}{p^2}\} \leq \theta \leq \frac{1}{p^\alpha}$ .

L'auteur de cet article procède à une nouvelle extension, que nous présenterons après les définitions ci-dessous.

**Définition 4.** *Considérons une fonction strictement continues  $\psi : [0, 1] \rightarrow [0, 1]$  différentiable sur  $(0, 1)$ . La fonction  $C_\theta^{\psi, p}(s, t) : [0, 1]^2 \rightarrow [0, 1]$  est définie comme*

$$C_\theta^{\psi, p}(s, t) = st[1 + \theta\psi(s)\psi(t)]^p, p \in [0, \infty), \forall (s, t) \in [0, 1]^2, \quad (5)$$

où  $\theta \in \Theta \in [-1, 1]$ .

Le nombre de valeurs que peut prendre  $\theta$  dans l'espace des paramètres  $\Theta$  dépend des propriétés de la fonction  $\psi$  considérée.

À quelle(s) condition(s) la fonction définie ci-dessus est-elle une fonction de répartition ? Supposons que leur signe sur  $[0, 1]$  ne change pas afin d'obtenir une dépendance déterminée unique structure ; notons que la copule est limitée à la plage de  $[0, 1]$  et donc,  $[1 + \theta\psi(s)\psi(t)]^p$  doit être borné sur  $[0, 1]$  . Le théorème suivant donne des conditions nécessaires et suffisantes sur  $\psi$  pour s'assurer que  $C_{\theta}^{\psi,p}$  est une copule bivariée.

**Théorème 2.** Soit  $\psi : [0, 1] \rightarrow [0, 1]$  différentiable et monotone sur  $(0, 1)$  et  $|p\theta| \leq 1$  remplissant les conditions suivantes :

1.

$$\psi(1) = 0$$

2.

$$|x\psi(x)| \leq 1, \forall x \in [0, 1], \text{ où } \psi'(x) = \frac{\partial\psi(x)}{\partial x}.$$

Alors, la copule définie en 5 est une copule bivariée si  $\theta \in [-1, 1]$  ;  $\Theta \in [-1, 1]$

*Démonstration.* La preuve s'établit en deux étapes.

Premièrement, il est clair que :

$$C_{\theta}^{\psi,p}(x, 1) = C_{\theta}^{\psi,p}(x, 1) = x, \forall x \in [0, 1] \Leftrightarrow (P1).$$

Deuxièmement, comme  $\phi$  est différentiable et monotone sur  $[0, 1]$ , éventuellement, pour  $C_{\delta}^{\psi,p}$  deux fois différentiable et deux fois croissante, la propriété (P2) peut être remplacée par

$$\begin{aligned} C_{\delta}^{\psi}(s, t) &= \frac{\partial^2 C(s, t)}{\partial s \partial t} \\ &= [1 + \theta\phi(s)\phi(t)]^{p-2} \\ &\times \{ (+\theta\phi(t) [\phi(s) + ps\phi'(s)]) (1 + \theta\phi(s) [\phi(t) + \phi'(t)]) + pst\theta\phi'(s)\phi'(t) \}. \end{aligned} \quad (6)$$

Cette expression est non négatif, si  $|x\phi'(x)| \leq 1, x \in [0, 1]$  et  $p\theta \geq 0$ , alors,

$$1 + \theta\psi(t) [\psi(s) + ps\psi'(s)] \geq 1, \forall s, t \in [0, 1]$$

et

$$1 + \theta\psi(s) [\psi(t) + ps\psi'(t)] \geq 1, \forall s, t \in [0, 1]$$

avec  $|x\psi'(x)| \leq 1, \forall x \in [0, 1]$ , on obtient :

$$|pst\psi'(t)\psi'(s)| \leq 1, \forall s, t \in [0, 1].$$

Il en découle ce qui suit,

$$\{(1 + \theta\psi(t) [\psi(s) + ps\psi'(s)])(1 + \theta\psi(s) [\psi(t) + ps\psi'(t)]) + pst\psi'(t)\psi'(s)\} \geq 0 \forall s, t \in [0, 1].$$



Si  $p\theta \leq 0$ , alors nous avons :

$$\{(1 + \theta\psi(t) [\psi(s) + p\psi'(s)])(1 + \theta\psi(s) [\psi(t) + p\psi'(t)]) + p\theta st\psi'(t)\psi'(s)\} \leq 1 \quad \forall s, t \in [0, 1].$$

et

$$|p\theta st\psi'(t)\psi'(s)| \leq 1, \quad \forall s, t \in [0, 1]. \quad (7)$$

En conséquence, la fonction  $C_{\theta}^{\psi,p}$  est deux fois différentiable.

D'abord, notons que  $\theta$  est le paramètre décrivant la dépendance de la famille  $C_{\theta}^{\psi,p}$  et  $\theta = 0$ , conduit à l'indépendance de  $T$  et  $S$ . Ensuite, la famille  $C_{\theta}^{\psi,p}$  est une classe symétrique de copule bivariée, i.e.  $\forall s, t \in [0, 1], C_{\theta}^{\psi,p}(s, t) = C_{\theta}^{\psi,p}(t, s)$ . Enfin, on détermine la plage de valeurs admissibles du paramètre  $\theta$ , telle que  $C_{\theta}^{\psi,p}(s, t) \geq 0$  par rapport à  $\psi$  et  $p$ .

Par conséquent, la fonction  $C_{\theta}^{\psi,p}(s, t)$  satisfait deux fois la propriété de croissance.

Tout d'abord, on note que  $\theta$  est un paramètre qui décrit la dépendance de la famille  $C_{\theta}^{\psi,p}$  et  $\theta = 0$ , conduit à l'indépendance de  $S$  et  $T$ . Ensuite, la famille  $C_{\theta}^{\psi,p}(s, t)$  est une classe symétrique des copules bivariées, i.e.,  $\forall s, t \in [0, 1], C_{\theta}^{\psi,p}(s, t) = C_{\theta}^{\psi,p}(t, s)$ . Par suite, il est déterminé la plage admissible du paramètre  $\theta$ ,  $C_{\theta}^{\psi,p}(s, t) \geq 0$ .

### 3.3 Sous-familles de la classe générale

L'un des avantages de la nouvelle classe est de générer certaines sous-familles symétriques de classe  $C_{\theta}^{\psi,p}$ . Certaines de ces sous-familles sont mentionnées dans les exemples qui suivent.

**Exemple 1.** Soit  $\phi(x) = 1 - \sqrt{x}, \forall x \in [0, 1]$ . Comme,  $\phi(1) = 0, 0 \leq \phi(x) \leq 1$ , et  $-1 \leq x\phi'(x)' \leq 0, \forall x \in [0, 1]$ , par le théorème 2, on obtient :

$$C_{\theta}^{\psi,p} = st[1 + \theta(1 - \sqrt{s})(1 - \sqrt{t})]^p, \quad \forall s, t \in [0, 1]$$

est une copule pour  $\theta \in [-4p^{-1}, 2p^{-1}]$  et  $p \geq 4$ .

*Démonstration.* La relation 20 du théorème 2 nous fournit :

$$|p\theta st\psi'(t)\psi'(s)| \leq 1, \quad \forall s, t \in [0, 1].$$

Comme ,

$$\psi'(x) = -\frac{1}{2\sqrt{x}},$$

alors pour  $s = t = x$  ,

$$\left| p\theta x^2 \times \frac{1}{4x} \right| \leq 1.$$

Comme  $st < 1$ , alors,

$$\left| p\theta st \times \frac{1}{4x} \right| \leq \left| \frac{p\theta}{4x} \right| \leq 1,$$

i.e.

$$|p\theta| \leq 4x \leq 4; \text{ car } 0 \leq x \leq 1.$$

i.e.

$$-4 \leq p\theta \leq 4 \implies -4p^{-1} \leq \theta \leq 4p^{-1}.$$

□

**Exemple 2.** Soit  $\phi(x) = x(1-x), \forall x \in [0, 1]$ . Comme  $\phi(1) = 0, 0 \leq \phi(x) \leq 1$ , et  $|x\phi(x)'| = |x(1-2x)| \leq 1, \forall x \in [0, 1]$ , par le théorème 2, on obtient :

$$C_{\theta}^{\psi,p} = st[1 + \theta(1-s)(1-t)]^p, \forall s, t \in [0, 1]$$

est une copule pour  $\theta \in [-p^{-1}, 1]$  et  $p \geq 1$ .

*Démonstration.* On a

$$\psi'(x) = -2x + 1 \text{ et } x\psi'(x) = x(1-2x).$$

$$\text{Or, } 0 \leq x \leq 1 \implies -2 \leq -2x \leq 0 \implies -1 \leq 1-2x \leq 1.$$

Alors,

$$|1-2x| \leq 1 \text{ et } |x(1-2x)| \leq 1.$$

Montrons que  $\theta \in [-p^{-1}, 1]$  Rappelons que  $\forall s, t \in [0, 1], |p\theta st\psi'(t)\psi'(s)| \leq 1$ .

Pour  $s = t = x$ , on a :

$$|p\theta[x(1-2x)]^2| \leq 1.$$

Comme,

$$\forall x \in [0, 1], |1-2x| \leq 1, \text{ alors } (x(1-2x))^2 \leq 1.$$

Ainsi,

$$|p\theta[x(1-2x)]^2| \leq 1 \implies |p\theta| \leq 1 \Leftrightarrow -p^{-1} \leq \theta \leq p^{-1}.$$

On vient de trouver la borne inférieure de  $\theta$  qui est  $-p^{-1}$ . Quant à la borne supérieure, on ne trouve pas de justification rigoureuse pour laquelle la borne supérieure de  $\theta$  1.

Si :

- $\theta \in [-p^{-1}, p^{-1}] \implies |p\theta| \leq 1 \text{ et } p \in [1, +\infty[.$
- $\theta \in [-1, 1] \implies |p\theta| \leq 1 \text{ et } p \in [0, 1].$

□

**Exemple 3.** Soit  $\phi(x) = e^{-x} - e^{-1}, \forall x \in [0, 1]$ . Comme  $\phi(1) = 0, 0 \leq \phi(x) \leq 1$ , et  $|x\phi(x)'| = |xe^{-x}| \leq 1, \forall x \in [0, 1]$ , par le théorème 2, on obtient :

$$C_{\theta}^{\psi,p} = st[1 + \theta(e^{-s} - e^{-1})(e^{-t} - e^{-1})]^p, \forall s, t \in [0, 1]$$

est une copule pour  $\theta \in [-e^2p^{-1}, -\frac{e^2p^{-1}}{1-e}]$  et  $p \geq 8$ .

*Démonstration.* On a  $\psi(x) = -e^{-x} \implies x\psi'(x) = -xe^{-x}, \forall x \in [0, 1], e^{-x} \in [e^{-1}, 1] \implies |xe^{-x}| \in [0, 1]$ .

Pour  $p = s = x$ , la relation  $|p\theta st\psi'(t)\psi'(s)| \leq 1$  nous donne :

$$|p\theta x^2 e^{-2x}| \leq 1 \implies |p\theta e^{-2x}| \leq 1, \forall x \in [0, 1].$$

Pour  $x = 1$ , on a

$$-1 \leq p\theta e^{-2} \leq 1 \implies -p^{-1}e^2 \leq \theta \leq p^{-1}e^2.$$

□

**Exemple 4.** Soit  $\phi(x) = \frac{1-x}{1+x}, \forall x \in [0, 1]$ . Comme  $\phi(1) = 0, 0 \leq \phi(x) \leq 1$ , et  $|x\phi(x)'| = \frac{-2x^2}{(1+x)^2} \leq 1, \forall x \in [0, 1]$ , par le théorème 2, on obtient :

$$C_{\theta}^{\psi,p} = st[1 + \theta(\frac{1-s}{1+s})(\frac{1-t}{1+t})]^p, \forall s, t \in [0, 1]$$

est une copule pour  $\theta \in [-4p^{-1}, 2p^{-1}]$  et  $p \geq 4$ .

*Démonstration.* La fonction  $\psi$  est dérivable sur  $[0, 1]$  et  $\forall x \in [0, 1]$ . On a :

$$\psi'(x) = \frac{-(1+x) - (1-x)}{(1+x)^2} \implies \psi'(x) = \frac{-2}{(1+x)^2} \implies x\psi'(x) = -\frac{2x}{(1+x)^2}.$$

On a :

$$1 \leq 1+x \leq 2 \implies 1 \leq (1+x)^2 \leq 4 \implies \frac{1}{4} \leq \frac{1}{(1+x)^2} \leq 1, \forall x \in [0, 1], \text{ et } -2 \leq -2x \leq 0.$$

On a donc :

$$-\frac{1}{2} \leq -\frac{2x}{(1+x)^2} \leq 0 \implies |x\psi'(x)| \leq \frac{1}{2} < 1.$$

Pour  $p = s = x$ , la relation  $|p\theta st\psi'(t)\psi'(s)| \leq 1$  nous donne :

$$\left| p\theta x^2 \times \frac{4}{(1+x)^4} \right| \leq 1 \implies \frac{|p\theta|}{(1+x)^4} \leq \frac{1}{4} \implies |p\theta| \leq \frac{(1+x)^4}{4}.$$

Alors,  $\forall x \in [0, 1]$ , on a :

$$|p\theta| \leq \frac{1}{4} \times 2^4 = 4 \implies -4 \leq p\theta \leq 4.$$

D'où :

$$-4p^{-1} \leq \theta \leq 4p^{-1}.$$

□

**Remarque 1.** La famille  $C_{\theta}^{\psi,p}$  comprend certaines familles connues de copules de FGM introduites par les chercheurs au cours des dernières années, qui sont les suivantes :

— Farlie(1960), Gumbel(1960), et Morgenstern(1956) :

$$C_{\theta}^p(s, t) = st[1 + \theta st(1 - s)(1 - t)], p = 1, -1 \leq \theta \leq 1.$$

— Huang et Kotz (1999) :

$$C_{\theta}^p(s, t) = st[1 + \theta st(1 - s^{\alpha})(1 - t^{\alpha})], p = 1, -(\max\{1, p\alpha^2\})^{-1} \leq \theta \leq (p\alpha)^{-1}.$$

— Lai et Xie (2000) :

$$C_{\theta}^p(s, t) = st[1 + \theta s^a t^b (1 - s^a)(1 - t^b)], p = 1, a, b \geq 1.$$

$$-(\max\{p(m^+)^2, p(m^-)^2\})^{-1} \leq \theta \leq -(pm^+m^-)^{-1},$$

où,

$$m^+ = \left[ \frac{a(a+b-1) + \sqrt{ab(a+b-1)}}{(a+b)(a+b-1)} \right]^{a-1} \left[ 1 - \frac{a(a+b-1) + \sqrt{ab(a+b-1)}}{(a+b)(a+b-1)} \right]^{b-1} \\ \times \left[ \frac{a(a+b-1) + \sqrt{ab(a+b-1)}}{(a+b-1)} - a \right]$$

$$m^- = \left[ \frac{a(a+b-1) - \sqrt{ab(a+b-1)}}{(a+b)(a+b-1)} \right]^{a-1} \left[ 1 - \frac{a(a+b-1) - \sqrt{ab(a+b-1)}}{(a+b)(a+b-1)} \right]^{b-1} \\ \times \left[ \frac{a(a+b-1) - \sqrt{ab(a+b-1)}}{(a+b-1)} - a \right]$$

Bairamov et Kotz (2002) :

$$C_{\theta}^p(s, t) = st[1 + \theta(1 - s^{\gamma})^{\lambda}(1 - t^{\gamma})^{\lambda}], p = 1, \gamma \geq 0, \lambda \geq 1, -(\max(1, pm^2))^{-1} \leq \theta \leq (pm)^{-1},$$

où,

$$m = \frac{\gamma^{\lambda}(\lambda - 1)^{\lambda-1}}{\lambda + 1}.$$

Bekrizadeh, Parham, and Zadkarami (2012) :

$$C_{\theta}^p(s, t) = st(1 + \theta st(1 - s^{\alpha})(1 - t^{\alpha}))^p, p = 0, 1, 2, \dots, -\min(1, \frac{1}{p\alpha^2}) \leq \theta \leq \frac{1}{p\alpha}.$$

Pathak et Vellaisamy (2015) :

$$C_{\theta}^p(s, t) = st(1 + \theta(1 - s^{\alpha})(1 - t^{\beta}))^p, \alpha, \beta \geq 0, p \geq 1,$$

$$-\min\{1, \frac{1}{p^2\alpha\beta}\} \leq \theta \leq \min\{\frac{1}{p\alpha}, \frac{1}{p\beta}\}.$$

**Proposition 1.** Deux propriétés asymptotiques de la famille  $C_{\theta}^{\psi,p}$  sont données ci-dessous :

$$\begin{aligned} \lim_{p \rightarrow 0} C_{\theta}^{\psi,p}(s, t) &= \lim_{p \rightarrow 0} st[1 + \theta\psi(s)\psi(t)]^p \\ &= st \\ &= \Pi(s, t), \end{aligned}$$

où  $\Pi(s, t)$  est la copule indépendante.

Soit  $\theta = \frac{\delta}{p}$ , où  $\delta \leq p$ , alors

$$\begin{aligned} \lim_{p \rightarrow 0} C_{\theta}^{\psi,p}(s, t) &= \lim_{p \rightarrow 0} st[1 + \frac{\delta}{p}\psi(s)\psi(t)]^p \\ &= st \lim_{p \rightarrow 0} st[1 + \frac{\delta}{p}\psi(s)\psi(t)]^p \\ &= st \exp \delta\psi(s)\psi(t) \\ &= C_{\delta}^{\psi}(s, t) \end{aligned} \tag{8}$$

La nouvelle famille de copule  $C_{\delta}^{\psi}(s, t)$  peut être la nouvelle extension de la copule de Gumbel-Barnet(GB) discuté par *Hutchinson* et *Lai* (1990), quand  $\psi(x) = \log(x), \forall x \in [0, 1]$  et la copule de Celebioglu-Cuadras introduite par *Cuadras* (2009, 2015), quand  $\psi(x) = 1 - x, \forall x \in [0, 1]$ .

### 3.4 La corrélation rho de Spearman

Les mesures de dépendance sont des instruments courants pour résumer une structure de dépendance complexe dans le cas bivarié. Pour un examen historique des mesures de dépendance, voir Joe (1997) et Nelsen (2006). L'une des mesures non paramétriques les plus importantes de la dépendance est le rho de Spearman  $\rho_s$  entre les variables aléatoires. Soit  $X$  et  $Y$  deux variables aléatoires continues liées par la copule  $C$ . Alors, la version populaire du rho de Spearman pour  $X$  et  $Y$  est donnée par

$$\rho_S = 12 \int_0^1 \int_0^1 C(s, t) ds dt - 3 \quad (9)$$

Nous avons donc obtenu une formule générale pour le rho de Spearman associé pour la famille  $C_\theta^{\psi, p}(s, t)$ . Puisque nous ne pouvons pas donner des formules pour les propriétés de dépendance en termes de fonction élémentaire, la famille  $C_\theta^{\psi, p}$  est remplacée par la série d'expansion

$$\Omega = (\theta, \psi) : |\theta\psi(s)\psi(t)| \leq 1, \forall s, t \in [0, 1].$$

Basée sur le  $\Omega$ , la famille  $C_\theta^{\psi, p}$  en 5 peut également être réécrite en utilisant le théorème du binôme de Newton qui s'écrit comme suit : pour deux quantités non nulles  $X$  et  $Y$  dans  $\mathbb{R}$  et  $n \in \mathbb{N}$  on a,

$$(X + Y)^n = \sum_{k=0}^n \binom{n}{k} X^k Y^{n-k}.$$

Notre fonction devient alors :

$$\begin{aligned} C_\theta^{\psi, p}(s, t) &= st \sum_{k=0}^m \binom{p}{k} \theta^k \psi^k(s) \psi^k(t) \\ &= st + \sum_{k=1}^m \binom{p}{k} \theta^k \psi^k(s) \psi^k(t). \end{aligned} \quad (10)$$

Notons que, dans 10,  $m = p$ , quand  $p$  est entier, sinon,  $m \rightarrow +\infty$ . Les copules avec des sections cubiques sont ainsi introduites, avec la conclusion que les copules avec des sections polynomiales d'ordre supérieur augmenteraient les degrés de dépendance.

**Proposition 2.** *Soit  $(X, Y)$  une paire de variables aléatoires de la famille  $C_\theta^{\psi, p}$ . Le rho de Spearman de la famille  $C_\theta^{\psi, p}$  est donné par*

$$\rho_S = 12 \sum_{k=1}^m \binom{p}{k} \theta^k [\Psi(k)]^2, \quad (11)$$

où

$$\Psi(k) = \int_0^1 x \psi^k(x) dx$$

*Démonstration.*

$$\begin{aligned} \rho_S &= \int_0^1 \int_0^1 C_\theta^{\psi, p}(s, t) ds dt - 3 \\ &= 12 \left\{ \int_0^1 \int_0^1 s t ds dt + \int_0^1 \int_0^1 \sum_{k=1}^m \binom{p}{k} \theta^k s \psi^k(s) t \psi^k(t) ds dt \right\} - 3 \\ &= 12 \left\{ \sum_{k=1}^m \binom{p}{k} \theta^k \left( \int_0^1 s \psi^k(s) ds \right) \left( \int_0^1 t \psi^k(t) dt \right) \right\} \\ &= 12 \sum_{k=1}^m \binom{p}{k} \theta^k \left[ \int_0^1 x \psi^k(x) dx \right]^2 \\ &= 12 \sum_{k=1}^m \binom{p}{k} \theta^k [\Psi(k)]^2. \end{aligned}$$

□

**Remarque 2.** Vu que  $\rho_S$  peut être exprimée en termes d'intégrale au sens de Lebesgue, on peut intégrer terme par terme (en faisant attention au cas  $|\theta| = 1$ ). Cette remarque nous sera utile au niveau de l'exemple présenté un peu plus tard. Pour l'instant, nous allons nous intéresser à la relation suivante, où on essaie de borner la mesure  $\rho_S$  :

$$\left| \rho_S - 12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 \right| \leq 12 \sum_{k=p+1}^{\infty} [\Psi(k)]^2 \quad (12)$$

*Démonstration.* Rappelons que :  $\rho_S = 12 \sum_{k=1}^m \binom{p}{k} \theta^k [\Psi(k)]^2$ , où

$$\Psi(k) = \int_0^1 x \psi^k(x) dx.$$

On pose  $p = m$  pour  $p$  entier. Considérons le cas général où  $p$  est non entier. Comme  $p \leq m$ , on a :

$$\begin{aligned} \sum_{k=1}^p \binom{p}{k} \theta^k [\Psi(k)]^2 &\leq \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 \\ \Rightarrow -12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 &\leq -12 \sum_{k=1}^p \binom{p}{k} \theta^k [\Psi(k)]^2 \end{aligned}$$

On a alors :

$$\rho_S - 12 \sum_{k=1}^m \binom{p}{k} \theta^k [\Psi(k)]^2 \leq \rho_S - 12 \sum_{k=1}^p \binom{p}{k} \theta^k [\Psi(k)]^2 \quad (13)$$

Vu que  $m \leq \infty$ , on a :

$$\sum_{k=1}^m \binom{p}{k} \theta^k [\Psi(k)]^2 \leq \sum_{k=1}^{\infty} \binom{p}{k} \theta^k [\Psi(k)]^2,$$

Donc,

$$\rho_S - 12 \sum_{k=0}^{\infty} \binom{p}{k} \theta^k [\Psi(k)]^2 \leq \rho_S - 12 \sum_{k=1}^m \binom{p}{k} \theta^k [\Psi(k)]^2.$$



Cette dernière inégalité, jointe à la relation 13 devient :

$$\begin{aligned}
\rho_S - 12 \sum_{k=1}^{\infty} \binom{p}{k} \theta^k [\Psi(k)]^2 &\leq \rho_S - 12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 \\
&\leq \rho_S - 12 \sum_{k=1}^p \binom{p}{k} \theta^k \Psi(k)^2 \\
&= 12 \sum_{k=p+1}^{\infty} \binom{p}{k} \theta^k [\Psi(k)]^2.
\end{aligned} \tag{14}$$

En remplaçant  $\rho_S$  par son expression dans le premier terme de l'inégalité, on trouve :

$$\begin{aligned}
12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 - 12 \sum_{k=1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2 &\leq \rho_S - 12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 \\
&\leq 12 \sum_{k=p+1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2
\end{aligned} \tag{15}$$

Comme  $p \leq m$ , on a :

$$12 \sum_{k=1}^p \binom{p}{k} \theta^k \Psi(k)^2 \leq 12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2$$

Donc,

$$12 \sum_{k=1}^p \binom{p}{k} \theta^k \Psi(k)^2 - 2 \sum_{k=1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2 \leq \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 - 12 \sum_{k=1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2$$

Le dernier terme à droite de cette dernière inégalité est le même que celui à gauche de l'expression (14). Il en découle que :

$$\begin{aligned}
12 \sum_{k=1}^p \binom{p}{k} \theta^k \Psi(k)^2 - \frac{1}{2} \sum_{k=1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2 &\leq \rho_S - \frac{1}{2} \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 \\
&\leq \frac{1}{2} \sum_{k=p+1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2.
\end{aligned}$$

On déduit ainsi :

$$\begin{aligned}
-12 \sum_{k=p+1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2 &\leq \rho_S - 12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 \\
&\leq 12 \sum_{k=p+1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2.
\end{aligned}$$

Alors,

$$\left| \rho_S - 12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 \right| \leq 12 \sum_{k=p+1}^{\infty} \binom{p}{k} \theta^k \Psi(k)^2. \quad (16)$$

Or  $\forall k > p$ ,  $\binom{p}{k} = 0$ . Donc,  $\forall k > p$ ,  $\binom{p}{k} \leq 1$ . Il advient que  $\binom{p}{k} \theta^k \leq \theta^k \leq 1$ ; car  $|\theta| \leq 1$ .  
D'où,

$$\left| \rho_S - 12 \sum_{k=1}^m \binom{p}{k} \theta^k \Psi(k)^2 \right| \leq 12 \sum_{k=p+1}^{\infty} \Psi(k)^2. \quad (17)$$

□

Nous venons ainsi de trouver une borne du rho de Spearman. Le terme à droite de l'inégalité est le reste d'une série. Nous allons voir ce qu'il vaut, avec un exemple concret.

Considérons  $\forall x \in [0, 1]$ ,  $\phi(x) = 1 - x$ .

On a :

$$\left| \rho_S - 12 \sum_{k=1}^{10} \binom{p}{k} \theta^k \Psi(k)^2 \right| \leq 12 \sum_{k=11}^{\infty} \Psi(k)^2 = 4\pi^2 - \frac{252778139}{6403320} \approx 0.002.$$

*Démonstration.* On a :

$$\Psi(k) = \int_0^1 x \psi(x)^k dx = \int_0^1 x(1-x)^k dx = \int_0^1 x^{2-1} (1-x)^{(k+1)-1} dx = I(1; 2; k+1) \quad (18)$$

Où I est la fonction bêta incomplète définie par :

$$I(1; \alpha; \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Dans notre situation, on a donc :

$$\Psi(k) = \frac{\Gamma(1)\Gamma(k+1)}{\Gamma(1+k+2)} = \frac{1 \cdot k!}{(k+2)!} \Rightarrow \Psi(k) = \frac{1}{(k+2)(k+1)} \Rightarrow \Psi(k)^2 = \frac{1}{(k+2)^2(k+1)^2}$$

À présent, il nous faut calculer  $\sum_{k=11}^{\infty} \Psi(k)^2 = \sum_{k=11}^{\infty} \frac{1}{(k+2)^2(k+1)^2}$ .

On a :

$$\sum_{k=11}^{\infty} \Psi(k)^2 = \sum_{k=0}^{\infty} \Psi(k)^2 - \sum_{k=0}^{10} \Psi(k)^2 \quad (19)$$

Le deuxième terme est calculable à la machine et même à la main. On s'intéresse donc un peu plus au premier terme (qu'on notera A) pour le moment.

$$A = \sum_{k=0}^{\infty} \Psi(k)^2 = \sum_{k=0}^{\infty} \frac{1}{(k+2)^2(k+1)^2}$$

Vu que  $\frac{1}{x+2}(x+1) = -\frac{1}{x+2} + \frac{1}{x+1}$ ,  
On peut écrire,

$$\begin{aligned} A &= \sum_{k=0}^{\infty} \left( -\frac{1}{k+2} + \frac{1}{k+1} \right)^2 \\ &= \sum_{k=0}^{\infty} \left( \frac{1}{(k+2)^2} + \frac{1}{(k+1)^2} - 2 \frac{1}{(k+2)(k+1)} \right) \\ &= \sum_{k=0}^{\infty} \left( \frac{1}{(k+2)^2} + \frac{1}{(k+1)^2} - 2 \left( -\frac{1}{k+2} + \frac{1}{k+1} \right) \right). \end{aligned}$$

En décomposant les termes, on obtient :

$$\begin{aligned} A &= \sum_{k=0}^{\infty} \frac{1}{(k+2)^2} + \sum_{k=0}^{\infty} \frac{1}{(k+1)^2} \\ &\quad + 2 \left( \sum_{k=0}^{\infty} \left( \frac{1}{k+2} - \frac{1}{k+1} \right) \right). \end{aligned}$$

Rappelons l'expression d'une somme télescopique :  $\sum_{k=0}^{n-1} (a_{k+1} - a_k) = a_n - a_0$ . Alors,

$$\sum_{k=0}^{\infty} \left( \frac{1}{k+2} - \frac{1}{k+1} \right) = \frac{1}{\infty} - \frac{1}{0+1} = -1$$

On en déduit que :

$$A = -2 + \sum_{k=0}^{\infty} \left( \frac{1}{(k+2)^2} + \frac{1}{(k+1)^2} \right)$$

Or,

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{1}{(k+2)^2} &= \sum_{k=2}^{\infty} \frac{1}{k^2} \\ &= \sum_{k=1}^{\infty} \frac{1}{k^2} - \frac{1}{1^2} \\ &= \sum_{k=1}^{\infty} \frac{1}{k^2} - 1 = \zeta(2) - 1 \\ \sum_{k=0}^{\infty} \frac{1}{(k+1)^2} + \sum_{k=1}^{\infty} \frac{1}{k^2} &= \zeta(2). \end{aligned}$$

Il vient alors que :

$$A = -2 + \left( \frac{\pi^2}{6} - 1 \right) + \zeta(2); \text{ avec } \zeta(2) = \frac{\pi^2}{6}$$

Ainsi :  $A = \frac{\pi^2}{3}$

On en déduit (voir la relation 19) que :

$$\sum_{k=11}^{\infty} [\Psi(k)]^2 = \frac{\pi^2}{3} - 3 - \sum_{k=0}^{10} \frac{1}{(k+2)^2(k+1)^2}$$

On obtient donc notre borne supérieure définie comme suit :

$$\begin{aligned} 12 \sum_{k=11}^{\infty} [\Psi(k)]^2 &= 12 \left( \frac{\pi^2}{3} - 3 - \sum_{k=0}^{10} \frac{1}{(k+2)^2(k+1)^2} \right) \\ &= \pi^2 - 36 - 12 \sum_{k=0}^{10} \left( \frac{1}{(k+1)(k+2)} \right)^2 \\ &= 0.002311616 \end{aligned}$$

□

**Remarque 3.** Pour  $\phi(x) = 1 - x$  et  $p = 1$ , on a :  $\rho_S = \frac{1}{2}\theta^1 [\Psi(1)]^2$  avec  $\Psi(1) = \int_0^1 x\psi(x)dx$ .  
Donc,  $\Phi(1) = \int_0^1 x - x^2 dx = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$ . On en déduit que :  $\rho_S = 12\theta^1 \left(\frac{1}{1}\right) \left(\frac{1}{6}\right)^2 = \left[\frac{12}{6} \times \frac{1}{6}\right] \theta = \frac{\theta}{3}$

## 4 ESTIMATION DES PARAMÈTRES DES COPULES

Dans cette section, on explique la méthode d'estimation des paramètres de la copule utilisée dans cet article, en l'occurrence la méthode de maximum de vraisemblance. Notons qu'il existe plusieurs autres méthodes d'estimation des paramètres d'une copule. Dont :

- Méthode du maximum de vraisemblance exacte ;
- Méthode des fonctions d'inférence des marginales (IFM) ;
- Méthode de pseudo-maximum de vraisemblance (maximum de vraisemblance canonique) ;
- Méthode d'inversion du tau du Kendall et du rho de Spearman ;
- Méthode de la distance minimale ;
- La méthode basée sur l'inférence bayésienne. Smith (2011) et Michael et al. (2006) ont suggéré l'approche bayésienne pour estimer les paramètres de la copule gaussienne en se basant sur l'algorithme de simulation des méthodes de Monte-Carlo par chaîne de Markov (MCMC).

L'objectif de l'estimation des paramètres des copules est de construire une inférence sur les mesures de dépendances, afin de mesurer l'intensité de dépendance entre différentes variables aléatoires. On distingue deux types d'estimation en statistique : l'estimation ponctuelle et l'estimation par intervalle de confiance. Dans cet article, nous effectuons des estimations ponctuelles pour différents paramètres de la copule FGM et FGM généralisée.

### **Définition 5.** *Estimation ponctuelle*

*L'estimation ponctuelle d'un paramètre de population se produit lorsqu'une valeur unique est utilisée pour estimer ce paramètre, c'est-à-dire qu'un point spécifique de l'échantillon est utilisé pour estimer la valeur souhaitée. En d'autres termes, c'est une estimation précise à un instant donné. Plus précisément, en statistique, l'estimation ponctuelle consiste à déterminer un estimateur (une fonction de l'échantillon).*

*Un estimateur est une fonction des données de l'échantillon dont la valeur constitue l'estimation d'un paramètre. Par exemple, si nous voulons estimer la moyenne ou la proportion dans une population, nous utilisons des échantillons pour calculer un estimateur qui nous donne une valeur approximative pour ce paramètre. De façon mathématique :*

*Soit  $T$  une statistique de l'échantillon  $(X_1, X_2, \dots, X_n)$ , alors il existe une fonction  $\phi$  telle que :  $T = \phi(X_1, X_2, \dots, X_n)$ .*

**Définition 6.** *Estimation par intervalles de confiance* Soit un paramètre de la population dénoté  $\theta$ . Pour estimer son intervalle de confiance, on doit déterminer un intervalle  $[a, b]$  qui contient sa vraie valeur à un seuil de confiance  $\alpha$  connu d'avance. Ce type d'estimation est plus robuste que l'estimation ponctuelle.

### 4.1 La méthode du maximum de vraisemblance

Soit  $C_\theta$  une copule paramétrique multivariée de paramètre  $\theta$ . On estime sous les deux hypothèses suivantes :

$H_0 : C \in C_0$ , telle que  $C_0 = \{C_\theta : \theta \in \Omega\}$ , où  $\Omega \subseteq \mathbb{R}^p$  pour tout entier  $p \geq 1$ ,  
 $H_0 : F_1 \in F_1, \dots, F_d \in F_d$ , telle que  $F_j$  pour  $j = 1, \dots, d$  sont les lois marginales de  $X_1, \dots, X_d$  et  
 $\mathbf{F}_j = \{\mathbf{F}_{\gamma_j} : \gamma_j \in \Gamma_j\}$ , où  $\Gamma_j$  est un sous-ensemble de  $\mathbf{R}$ .

Le problème du maximum de vraisemblance revient à maximiser la fonction log-vraisemblance  $\mathbf{L}(\theta)$  définie par :

$$L(\theta) = \sum_{j=1}^n \ln f(x_j), \quad (20)$$

telle que  $f(x_1, \dots, x_d)$  est la fonction de densité conjointe de  $X_1, \dots, X_d$ .

En remplaçant la fonction de densité conjointe dans la relation 20, on obtient :

$$L(\theta) = \sum_{i=1}^n \ln (c_\theta(F_1(x_i1), \dots, F_d(x_id))) + \sum_{i=1}^n \sum_{j=1}^d \ln f_i(x_j),$$

où  $c_\theta$  est la fonction de densité de la copule  $C_\theta$ . L'estimateur du maximum de vraisemblance de  $\theta$ , noté  $\hat{\theta}^{EMV}$  est donné par :

$$\hat{\theta}^{EMV} = \arg \max L(\theta).$$

D'un point de vue statistique, cet estimateur est robuste et en converge asymptotiquement vers la loi normale.

$$\sqrt{n} \left( \hat{\theta}^{EMV} - \theta \right) \rightarrow N \left( 0, I^{-1}(\theta) \right),$$

telle que  $I(\theta)$  est la matrice d'information de Fischer. Cette matrice est estimée par l'inverse de la matrice Hessian de la fonction de vraisemblance.

## 5 Extension de certaines copules FGM connues

Un des avantages de la famille  $C_{\theta}^{\psi,p}$  est que nous pouvons utiliser ce modèle et ses nouvelles extensions par introduction de paramètres supplémentaires  $\mathbf{p}$  pour certaines familles de FGM généralisées qui ont été présentées à la section précédente. Dans la suite, nous donnons quelques sous-classes de la famille  $\{C_{\theta}^{\psi,p}\}$  de type 5 qui est généré par une famille multi-paramètres  $\{\psi\}$ . Certaines extensions sont mentionnées dans les sous-sections suivantes.

### 5.1 Extension de la copule FGM

Soit  $\psi(x) = 1 - x$ ,  $\forall x \in [0, 1]$ , alors la famille  $C_{\theta}^{\psi,p}$  réduit à une nouvelle copule FGM symétrique généralisée comme suit :

$$C_{\theta}^{FGM,p}(s, t) = st [1 + \theta(1 - s)(1 - t)]^p, p \in (0, \infty), \forall s, t \in [0, 1]. \quad (21)$$

En utilisant la relation 5, on obtient l'expression de la fonction de densité de  $C_{\theta}^{FGM,p}$  définie comme :

$$\begin{aligned} c_{\theta}^{FGM,p}(s, t) &= [1 + \theta(1 - s)(1 - t)]^{p-2} \\ &\times \{(1 + \theta(1 - t)[1 - s(p + 1)])(1 + \theta(1 - s)[1 - t(p + 1)]) + p\theta st\} \end{aligned} \quad (22)$$

Comme la fonction densité  $c_{\theta}^{FGM,p}$ , pour tout  $s, t \in [0, 1]$  est convexe, alors le maximum et le minimum de  $c_{\theta}^{FGM,p}$  doivent être pris sur les limites du carré  $\Delta = (s, t) : 0 \leq s \leq 1, 0 \leq t \leq 1$ . De plus, de chaque côté du carré  $\Delta$ , soit  $s$  ou  $t$  est constant et la fonction  $c_{\theta}^{FGM,p}$  est décroissante. Il en résulte que le maximum et le minimum de la fonction  $c_{\theta}^{FGM,p}$ , doivent être pris dans les points d'angle du carré  $\Delta$ . Par conséquent, nous déterminons la plage admissible du paramètre associé  $\theta$  dans l'expression 22 sur le carré  $\Delta$ , de sorte que  $c_{\theta}^{FGM,p} \geq 0, \forall s, t \in [0, 1]$  comme suit :

$$-(\max\{1, p\})^{-1} \leq \theta \leq p^{-1}.$$

Comme  $\Psi(k) = \int_0^1 x\psi^k(x)dx = \frac{1}{(k+1)(k+2)}$ , en utilisant la relation 11, nous obtenons le rho de Spearman comme suit :

$$\rho_S = 12 \sum_{k=1}^m \binom{p}{k} \theta^k \left[ \frac{1}{(k+1)(k+2)} \right]^2 \quad (23)$$

où la limite supérieure de  $\rho_S$  dans 23 peut être augmentée jusqu'à environ, une valeur de 0.3805 lorsque  $p \rightarrow \infty$ . Tandis que la limite inférieure 0,3333 reste inchangée. Par conséquent, la plage admissible de  $\rho_S$  dans la nouvelle famille des FGM généralisées et symétriques est  $[-0.3333, 0.3805]$ . Ainsi, la copule  $c_{\theta}^{FGM,p}$  a amélioré la plage de valeurs de la corrélation, car la section polynomiale a un ordre plus élevé par rapport à la copule FGM.

## 5.2 Extension de la copule Hung-Kotz

*Huang et Kotz (1999)* ont développé des extensions à un seul paramètre de type polynomial de la famille des FGM. Il est montré que  $\rho_S$  peut être augmenté jusqu'à environ +0.375 alors que la limite inférieure reste -0,33. Maintenant, dans la famille  $C_\theta^{\psi,p}$ , supposons que  $\psi(x) = 1 - x^\alpha$ ,  $\alpha \geq 0, \forall x \in [0, 1]$ . Alors la famille  $C_\theta^{\psi,p}$  se restreint à une nouvelle famille de copules Hung-Kotz généralisée symétrique comme suit définie comme suit :

$$C_\theta^{HK,p}(s, t) = st[1 + \theta(1 - s^\alpha)(1 - t^\alpha)]^p, \alpha \geq 0, -(max\{1, p\alpha^2\})^{-1} \leq \theta \leq (p\alpha)^{-1}, \forall s, t \in [0, 1]. \quad (24)$$

En utilisant la relation 5, la fonction de densité résultant de 24 est donnée par :

$$c_\theta^{HK,p}(s, t) = [1 + \theta(1 - s^\alpha)(1 - t^\alpha)]^{p-2} \times \{(1 + \theta(1 - t^\alpha)[1 - s^\alpha(1 + \alpha p)]) (1 + \theta(1 - s^\alpha)[1 - t^\alpha(1 + \alpha p)]) + p\theta s^\alpha t^\alpha\} \quad (25)$$

Comme  $\Psi(k) = \int_0^1 x\psi^k(x)dx = \frac{\Gamma(k+1)\Gamma(\frac{2}{\alpha})}{\alpha\Gamma(k+1+\frac{2}{\alpha})}$ , alors,

$$\rho_S = 12 \sum_{k=1}^m \binom{p}{k} \theta^k \left[ \frac{\Gamma(k+1)\Gamma(\frac{2}{\alpha})}{\alpha\Gamma(k+1+\frac{2}{\alpha})} \right]. \quad (26)$$

En prenant  $\alpha \approx 1.85$  et  $p \geq 500$  dans 26, on obtient  $\rho_{S,max} = 0.43$ . De façon similaire, en prenant  $\alpha \approx 0.1$  et  $p \geq 500$ , on arrive à  $\rho_{S,max} = -0.50$ . Par conséquent, la plage de valeurs admissibles de  $\rho_S$  dans la généralisation de la copule *Hung-Kotz* est  $[-0.50, 0.43]$ . Il en découle que la copule généralisée de *Hung-Kotz* améliore l'amplitude  $\rho_S$  de la famille *Hung-Kotz*. Ainsi, la famille de *Hung-Kotz* généralisée en 25 montre que les bornes supérieure et inférieure  $\rho_S$  ont été améliorées par rapport à l'amplitude de  $\rho_S$  dans la famille de *Hung-Kotz*.

**Remarque 4.** Pour  $p = 1$ , la relation 26 donne  $\rho_S = 3\theta[\frac{\alpha}{\alpha+2}]^2$ , avec la plage  $-3(max\{1, \alpha^2\})^{-1} (\frac{\alpha}{\alpha+2})^2 \leq \rho_S \leq 3\frac{\alpha}{(\alpha+2)^2}$ .

C'est le même résultat que celui discuté par Huang et Kotz (1999).



### 5.3 Extension de la copule Bairamov-Kotz

*Bairamov et Kotz (2002)* ont fourni plusieurs théorèmes caractérisant les propriétés de symétrie et de dépendance des copules FGM. Ils ont également proposé une modification en introduisant deux paramètres additionnels. Pour la classe  $C_\theta^{\psi,p}$ , une nouvelle famille de *Bairamov-Kotz* généralisée symétrique lorsque  $\psi(x) = (1 - x^\gamma)^\lambda$ ,  $\gamma > 0$ ,  $\lambda \geq 1$ ,  $\forall x \in [0, 1]$  est donnée comme suit :

$$C_\theta^{HK,p}(s, t) = st [1 + \theta(1 - s^\gamma)^\delta(1 - t^\gamma)^\delta]^p, \quad \gamma > 0, \lambda \geq 1, p \in [0, \infty), \quad (27)$$

$$p \in [0, \infty), -(\max(1, pm^2))^{-1} \leq \theta \leq (pm)^{-1} \text{ où } m = \gamma^\delta \left( \frac{\lambda - 1}{\gamma\delta + 1} \right).$$

La plage de paramètres admissible en 27 peut être déterminée par une procédure similaire suivie par *Bairamov et Kotz (2002)* incluant la valeur de  $p$ .

Comme  $\Psi(k) = \int_0^1 x\psi^k(x)dx = \frac{\Gamma(\lambda k + 1)\Gamma(\frac{2}{\gamma})}{\gamma\Gamma(\lambda k + 1 + \frac{2}{\gamma})}$ , nous avons

$$\rho_S = 12 \sum_{k=1}^{\infty} \binom{p}{k} \theta^k \left( \frac{\Gamma(\lambda k + 1)\Gamma(\frac{2}{\gamma})}{\gamma\Gamma(\lambda k + 1 + \frac{2}{\gamma})} \right). \quad (28)$$

En prenant  $\gamma \approx 2.3915$ ,  $\delta = 1.415$ , et  $p \rightarrow \infty$  dans la relation 28, on obtient  $\rho_{S,max} \approx 0.60$ . En prenant  $\gamma \approx 0.31$ ,  $\delta = 1.15$ , et  $p = 29$  on arrive à  $\rho_{S,min} \approx -0.6198$ . Par conséquent, la plage admissible de  $\rho_S$  dans la famille généralisée de *Bairamov-Kotz* est  $[-0.6198, 0.60]$ . Donc, la famille généralisée de *Bairamov-Kotz* donnée par la 27 améliore l'amplitude  $\rho_S$  *Bairamov-Kotz (2002)*.

**Remarque 5.** Pour  $p = 1$ , 28 devient  $\rho_S = 12\theta \left( \frac{\Gamma(\lambda+1)\Gamma(\frac{2}{\gamma})}{\gamma\Gamma(\lambda+1+\frac{2}{\gamma})} \right)^2$ . C'est le même résultat discuté par *Bairamov-Kotz (2002)*.

## 6 Application sur les données réelles

### 6.1 Présentation de l'ensemble de données

Dans cette section, nous appliquons les sous-familles mentionnées de notre copule bivariable symétrique présentée à un ensemble de données réel en science médicale. Selon le manuel du paquet MASS de R, l'Institut national américain du diabète et des maladies digestives et rénales a recueilli un ensemble de données auprès d'une population de femmes (au moins 21 ans, d'origine indienne Pima et vivant près de Phoenix, en Arizona) qui ont été testés pour le diabète selon les critères de l'Organisation Mondiale de la Santé (OMS). Cet ensemble de données comprend 332 enregistrements complets après avoir supprimé les données (principalement manquantes) sur l'insuline sérique. Cet ensemble de données a ensuite été réanalysé par *Li et Fang (2012)*.

Il existe huit(08) variables considérées comme des facteurs de risque du diabète chez les Indiens Pima, où l'indice de masse corporelle (IMC) et la fonction de pedigree du diabète (DEP) ont été diagnostiqués comme les plus importantes d'entre elles. Parmi les huit (08) variables, seule la variable *type* est catégorielle et les autres sont quantitatives.

FIGURE 1 – Présentation de l'ensemble de données

	npreg	glu	bp	skin	bmi	ped	age	type
1	6	148	72	35	33.6	0.627	50	Yes
2	1	85	66	29	26.6	0.351	31	No
3	1	89	66	23	28.1	0.167	21	No
4	3	78	50	32	31.0	0.248	26	Yes
5	2	197	70	45	30.5	0.158	53	Yes
6	5	166	72	19	25.8	0.587	51	Yes
7	0	118	84	47	45.8	0.551	31	Yes
8	1	103	30	38	43.3	0.183	33	No
9	3	126	88	41	39.3	0.704	27	No
10	9	119	80	35	29.0	0.263	29	Yes
11	1	97	66	15	23.2	0.487	22	No
12	5	109	75	26	36.0	0.546	60	No
13	3	88	58	11	24.8	0.267	22	No
14	10	122	78	31	27.6	0.512	45	No
15	4	103	60	33	24.0	0.966	33	No
16	9	102	76	37	32.9	0.665	46	Yes

## 6.2 Régression de BMI et PED

L'analyse de régression de BMI et PED révèle que le coefficient de dépendance linéaire entre les deux variables est significativement non nul. Cela est confirmé par le test de Student, au seuil 5%, et la  $p - value = 0.02092$ . En outre, la qualité de l'ajustement ne semble pas être importante. En effet, le modèle n'explique que 1.61% de la variabilité de BMI; car le coefficient de détermination  $\mathbf{R}^2 = 0.01606$ .

Au regard de la corrélation linéaire entre ces deux facteurs,  $Rho_{pearson} = 0.13$ , nous nous devons de faire appel à certains outils pour pouvoir décrire la relation entre ces deux variables, et analyser leur effet mutuel. Il est donc nécessaire de déterminer la distribution conjointe des deux facteurs. Cette détermination se fait par l'ajustement des sous-familles de copule FGM (avec extensions déjà présentées).

## 6.3 Analyse de la structure de dépendance entre BMI et PED

Lorsqu'on parle de copule, par le fait même, on fait explicitement appelle aux distributions marginales (première partie du théorème de Sklar). Il nous faut donc estimer les distributions marginales de nos deux facteurs, afin de pouvoir les joindre et ensuite trouver la copule correspondante. Il nous faut avoir les lois auxquelles s'adaptent bien nos données.

FIGURE 2 – Régression de BMI et PED

```
> verif_relation <- lm(bmi~ped,data=DM)
> summary(verif_relation)
```

Call:  
lm(formula = bmi ~ ped, data = DM)

Residuals:

Min	1Q	Median	3Q	Max
-13.776	-4.956	-0.756	4.024	34.392

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	31.8975	0.7016	45.464	<2e-16 ***
ped	2.5404	1.0947	2.321	0.0209 *

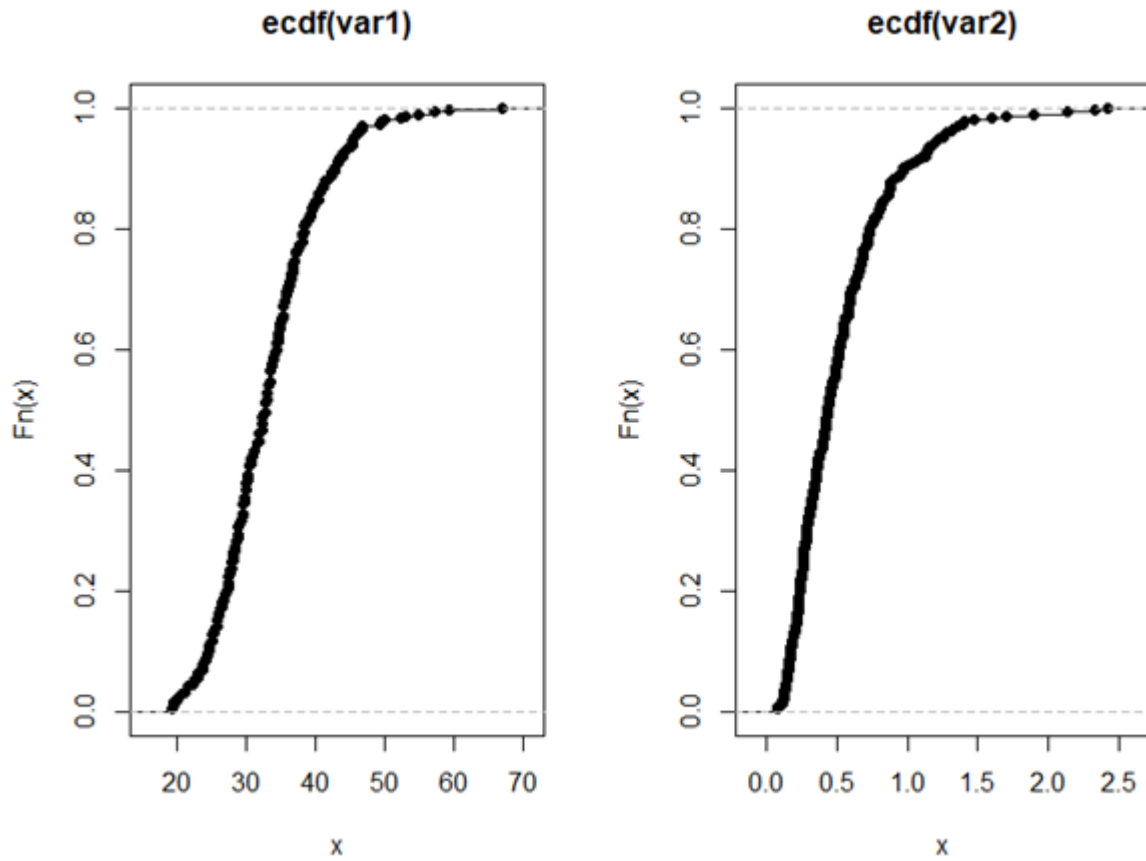
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.235 on 330 degrees of freedom  
Multiple R-squared: 0.01606, Adjusted R-squared: 0.01308  
F-statistic: 5.385 on 1 and 330 DF, p-value: 0.02092

Nous considérerons les lois gamma, Weibull, Pareto, exponentielle, normale et lognormale pour chacun des deux facteurs. Comment identifier la plus adéquate à notre jeu de données ? Il nous faut comparer les fonctions de répartition empirique des deux facteurs avec celles des lois susmentionnées, dont on trouvera les paramètres assez aisément sous R.

Voici, tout d'abord, comment se présentent les fonctions de répartition empiriques de nos facteurs (BMI à gauche, PED à droite) :

FIGURE 3 – Fonctions de répartition empiriques de BMI et de PED



Comparons ces distributions empiriques avec celles des différentes lois retenues. Ci-dessous les valeurs des paramètres obtenues à partir de R.

FIGURE 4 – Calibrage des paramètres de BMI et de PED

```

> pargamma_var1=fitdistr(var1,"gamma")$estimate;pargamma_var1
      shape      rate
21.8241415  0.6565676
> pargamma_var2=fitdistr(var1,"gamma")$estimate;pargamma_var2
      shape      rate
21.8241415  0.6565676
> pargamma_var2=fitdistr(var2,"gamma")$estimate;pargamma_var2
      shape      rate
2.584001 4.890340
> par(mfrow=c(1,2))
> plot(ecdf(var1))
> plot(ecdf(var2))
> pargamma_var1=fitdistr(var1,"gamma")$estimate;pargamma_var1
      shape      rate
21.8241415  0.6565676
> parexpo_var1=fitdistr(var1,"exponential")$estimate;parexpo_var1
      rate
0.03008445
> parlognorm_var1=fitdistr(var1,"lognormal")$estimate;parlognorm_var1
      meanlog      sdlog
3.4806611 0.2143401
> parnorm_var1=fitdistr(var1,"normal")$estimate;parnorm_var1
      mean      sd
33.239759  7.271925
> parweib_var1=fitdistr(var1,"weibull")$estimate;parweib_var1
Messages d'avis :
1: Dans densfun(x, parm[1], parm[2], ...) : Production de NaN
2: Dans densfun(x, parm[1], parm[2], ...) : Production de NaN
      shape      scale
4.527982 36.173817

```

Nous venons ainsi de calibrer nos différents modèles probables, en estimant leurs paramètres. Grâce à ces paramètres, nous arrivons à visualiser, déjà, l'évolution comparative des fonctions de répartition empiriques avec celles des lois considérées.

FIGURE 5 – Adéquation entre la fonction de répartition empirique de BMI et les distributions considérées

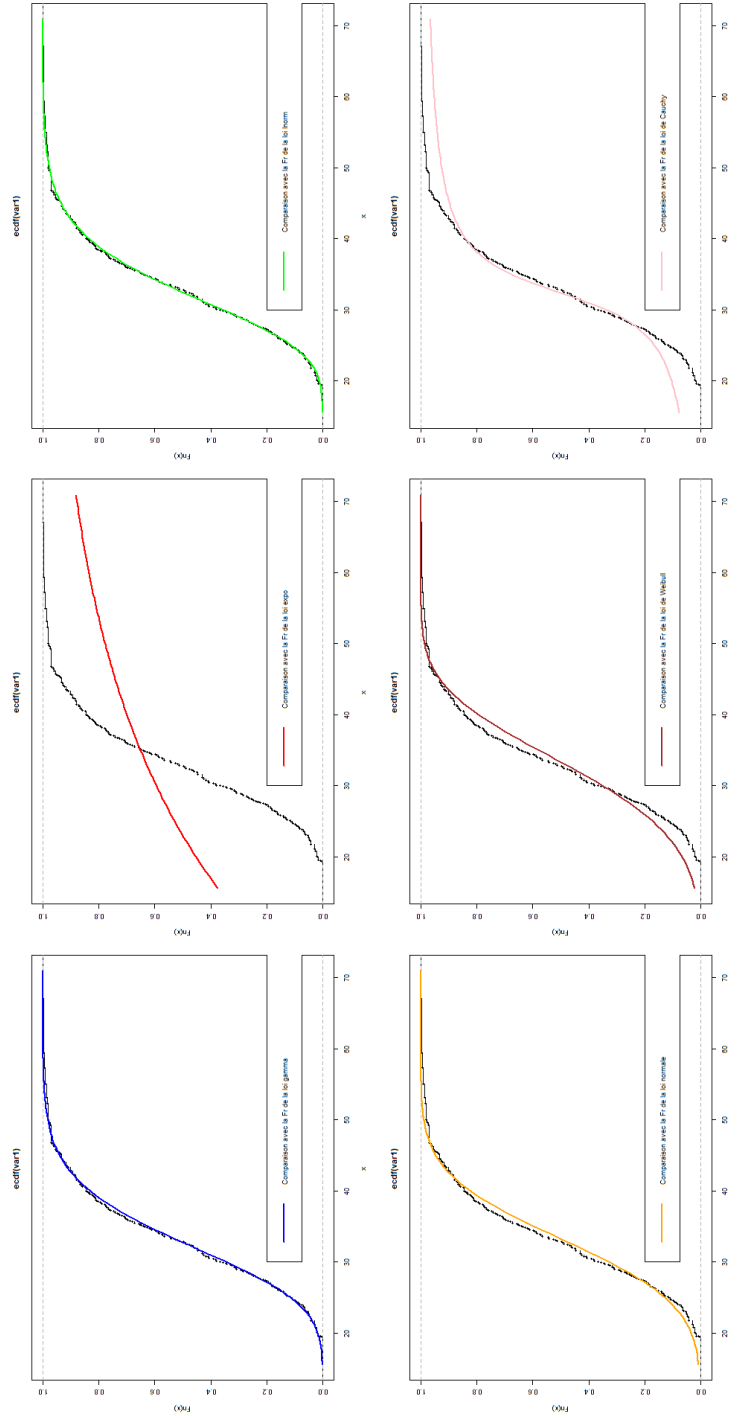
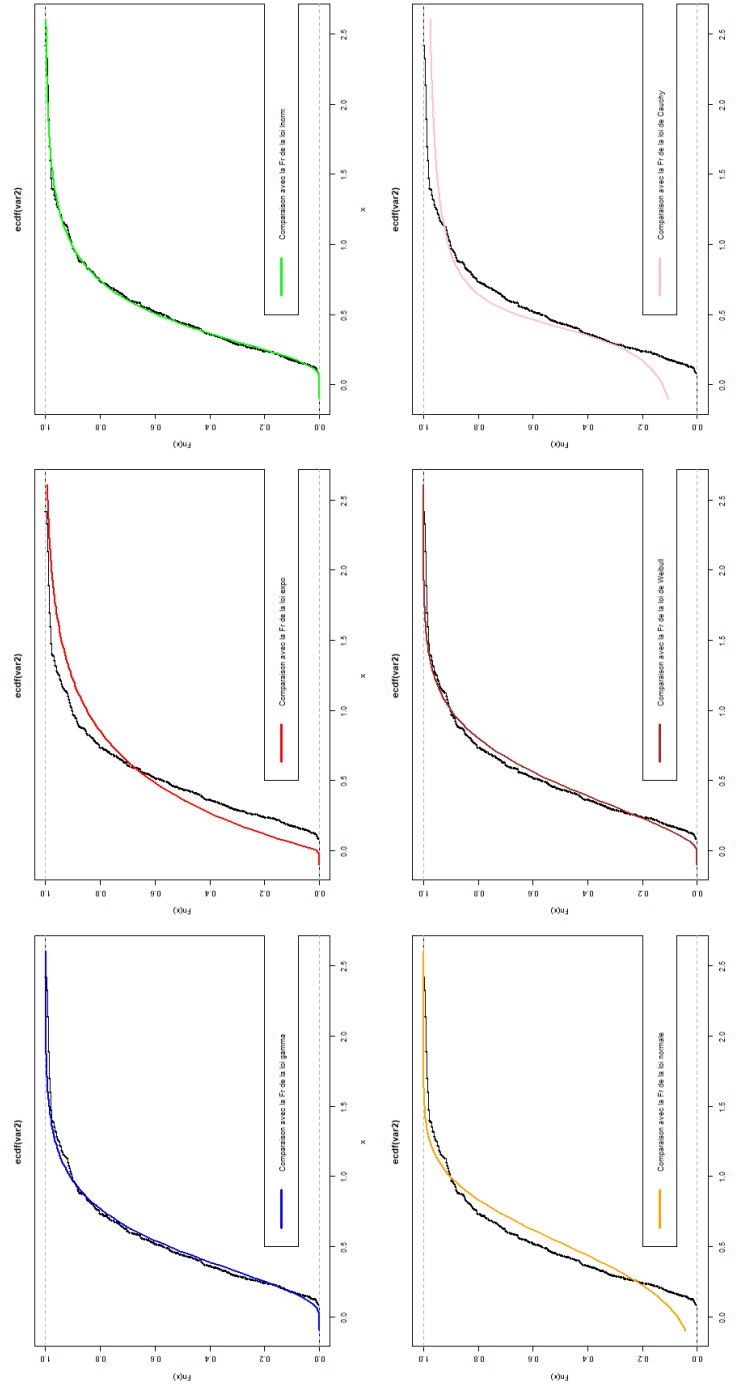


FIGURE 6 – Adéquation entre la fonction de répartition empirique de PED et les distributions considérées



Les figures 5 et 6 suggèrent que les lois gamma et log-normale semblent mieux adaptées pour décrire nos données. Mais, il nous faut valider ces hypothèses à l'aide d'un test statistique, celui de Kolmogorov-Smirnov le cas échéant. Après validation, nous choisirons le modèle le mieux adapté selon un critère d'information.

### 6.3.1 Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov (ou test K-S) est une méthode statistique utilisée pour comparer deux distributions de probabilité.

Le test de Kolmogorov-Smirnov permet de déterminer si deux échantillons proviennent de la même distribution (test à deux échantillons) ou si un échantillon donné provient d'une distribution de probabilité de référence (test à un échantillon).

La statistique du test de Kolmogorov-Smirnov,  $D_n$ , est directement construite à partir de la notion de distance Kolmogorov-Smirnov entre deux fonctions de répartition. Cette distance est définie comme la plus grande différence entre les deux fonctions. La statistique est donc :

$$D_n = \max |F_n(x) - F_0(x)|.$$

Lors de nos tests, nous nous intéresserons simplement aux p-values, qui nous permettront de rejeter ou non le statu quo  $H_0 : D_n=0$ . Naturellement, nous ne retiendrons que les modèles donnant une p-value  $> 5$

Il est vrai que lorsque nous obtenons des ex-aequo (valeurs identiques) dans nos données, le test KS pourrait rencontrer des problèmes pour calculer les p-values correctes, mais il n'empêche que même avec des ex-aequo, nos valeurs de p-values sont valables et peuvent être utilisées pour évaluer la significativité. On aurait pu éviter les ex-aequo en multipliant nos données par un facteur pour éviter les valeurs nulles, mais cela peut être risqué. Les résultats obtenus sont illustrés par la figure 7 et la figure 8.



FIGURE 7 – Résultat des tests de KS entre BMI et les distributions considérées.

```
> ks.test(x,pgamma,pargamma_var1[1],pargamma_var1[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.033749, p-value = 0.8439
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, pgamma, pargamma_var1[1], pargamma_var1[2]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,pexp,parexpo_var1[1])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.45087, p-value < 2.2e-16
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, pexp, parexpo_var1[1]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,plnorm,parlognorm_var1[1],parlognorm_var1[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.028475, p-value = 0.9506
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, plnorm, parlognorm_var1[1], parlognorm_var1[2]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,pnorm,parnorm_var1[1],parnorm_var1[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.05345, p-value = 0.2991
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, pnorm, parnorm_var1[1], parnorm_var1[2]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,pweibull,parweib_var1[1],parweib_var1[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.080448, p-value = 0.02721
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, pweibull, parweib_var1[1], parweib_var1[2]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,pcauchy,parcauchy_var1[1],parcauchy_var1[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.099622, p-value = 0.002748
alternative hypothesis: two-sided
```

FIGURE 8 – Résultat des tests de KS entre PED et les distributions considérées.

```
> ks.test(x,pgamma,pargamma_var2[1],pargamma_var2[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.051293, p-value = 0.3468
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, pgamma, pargamma_var2[1], pargamma_var2[2]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,pexp,parexpo_var2[1])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.19963, p-value = 6.442e-12
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, pexp, parexpo_var2[1]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,plnorm,parlognorm_var2[1],parlognorm_var2[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.033282, p-value = 0.8557
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, plnorm, parlognorm_var2[1], parlognorm_var2[2]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,pnorm,parnorm_var2[1],parnorm_var2[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.11923, p-value = 0.000159
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, pnorm, parnorm_var2[1], parnorm_var2[2]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,pweibull,parweib_var2[1],parweib_var2[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.065199, p-value = 0.1189
alternative hypothesis: two-sided

Message d'avis :
Dans ks.test.default(x, pweibull, parweib_var2[1], parweib_var2[2]) :
aucun ex-aequo ne devrait être présent pour le test de Kolmogorov-Smirnov
> ks.test(x,pcauchy,parcauchy_var2[1],parcauchy_var2[2])

Asymptotic one-sample Kolmogorov-Smirnov test

data: x
D = 0.16096, p-value = 6.758e-08
alternative hypothesis: two-sided
```

Ils indiquent que les seuls modèles adaptés aux distributions BMI et PED sont la gamma, lognormale et normale.

Pour départager entre ces trois modèles, nous utilisons le critère d'information de Akaike (AIC). En utilisant le logiciel R, on obtient les résultats suivants pour l'AIC :

FIGURE 9 – AIC des modèles avec BMI

```
> vrgamma_var1=fitdistr(x,"gamma")$loglik;vrgamma_var1
[1] -1117.424
> vnrm_var1=fitdistr(x,"normal")$loglik;vnrm_var1
[1] -1129.783
> vlnrm_var1=fitdistr(x,"lognormal")$loglik;vlnrm_var1
[1] -1115.324
```

On devra donc logiquement garder la log-normale, mais l'auteur de l'article a plutôt porté son choix sur la loi gamma. Cela ne change pas grand-chose, puisque les deux modèles ont été validés et qu'en plus, leurs AIC sont très proches. On note aussi que la lognormale a une queue plus lourde, ce qui pourrait ne pas être très avantageux si on veut approfondir les analyses.

Rappelons que les paramètres de BMI pour la loi gamma sont :  $\alpha_1 = 21.8241415$  (contre 21.82 dans l'article) et  $\beta_1 = 0.6565676$  (contre 0.66 dans l'article). Pour la variable PED, ce sont les lois gamma, Weibull et lognormale qui sont adéquates. Tout comme avec le BMI, on retiendra la loi gamma. Les paramètres de PED pour la gamma sont :  $\alpha_2 = 2.584001$  (contre 2.58 dans l'article) et  $\beta_2 = 4.890340$  (contre 4.89 dans l'article).

Après ces estimations, nous nous sommes attaqués aux paramètres des copules FGM et FGM généralisée. Nous tenions à voir à quel point la nouvelle classe de copules se démarque de celles déjà connues. On s'est alors aperçus qu'il y a une différence significative au niveau des valeurs de  $\theta$  et  $p$  estimées. Déjà, nous avons rencontré un problème avec le  $\theta$  de la FGM dont les valeurs variaient vers la fin de l'article (tantôt 0.243 tantôt 0.326). Nos estimations nous ont mené à une valeur de  $\theta = 0.2908512$  pour la copule FGM. Quant à la copule FGM généralisée, nous avons obtenu  $\theta = 0.2310301$  (contre 0.174 dans l'article) et  $p = 1.259995$  (contre 2.6854 dans l'article).

C'est avec ces valeurs estimées que nous avons par la suite calculé les valeurs correspondantes du coefficient de corrélation de Spearman.

Une première remarque est que notre valeur du  $\rho_s$  des données empiriques diffèrent de celle fourni dans l'article. Nous avons trouvé une valeur de 0.09697264 contre 0.1385 dans l'article. Le fait est que ce sont des données réelles. On ne peut logiquement pas trouver différentes valeurs, ce qui nous a amenés à nous demander s'il n'y a pas eu d'omission dans notre article. D'ailleurs la valeur fournie dans l'article s'approche beaucoup plus du coefficient de corrélation de Pearson (0.1267162).

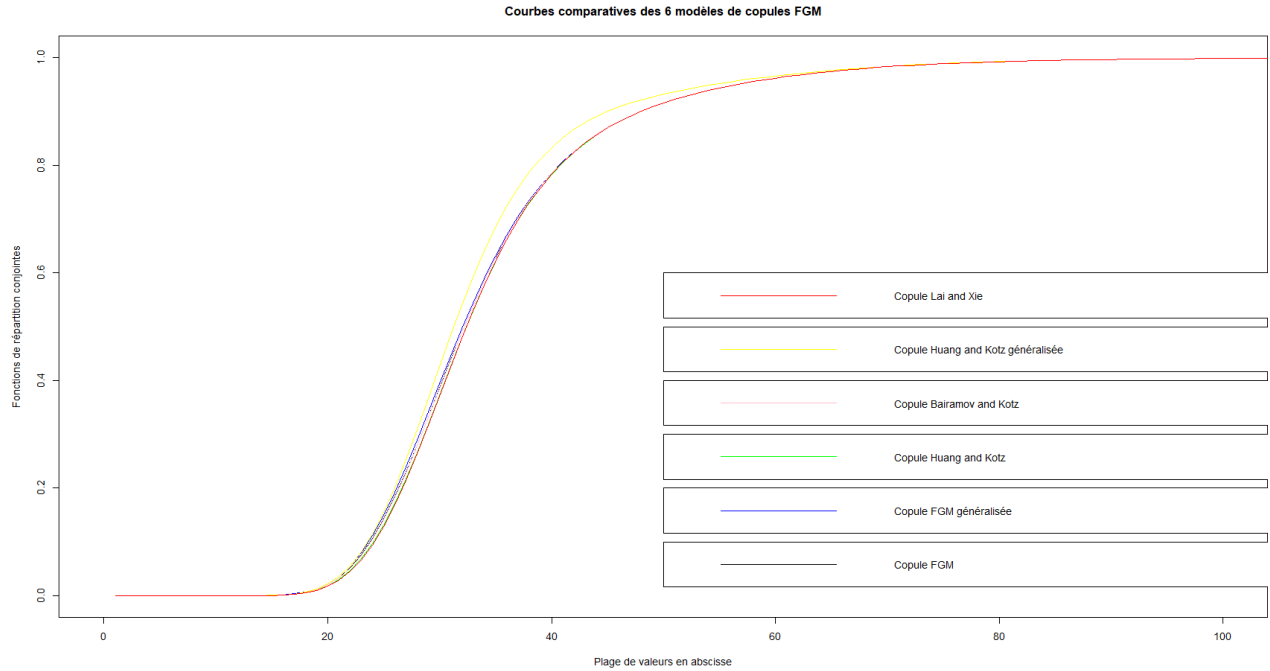
Vu que nous avons obtenu, entre-temps, une nouvelle formule pour le rho de Spearman, nous allons à présent l'utiliser pour effectuer une comparaison. Dès le début de notre travail, on a souligné que l'auteur cherchait à étendre la plage de valeurs possibles du coefficient de corrélation de Spearman. Nous avons donc utilisé la relation 23 pour calculer la

valeur de  $\rho_s$  dans le cas de la copule FGM généralisée, puis nous comparons cette valeur à celle du  $\rho_s$  de la FGM standard (donc  $p=1$ ). Nous obtenons  $\rho_s(FGM) = 0.0969504 < \rho_s(FGM \text{ généralisée}) = 0.09703226$ .

### 6.3.2 Comparaison des fonctions de répartition conjointes

Voici un aperçu des courbes comparatives des fonctions de répartition conjointes des six (06) différents modèles de copules qui ont été utilisés.

FIGURE 10 – Application des modèles sur les données BMI



Ici, nous avons utilisé les différentes sous-familles de copules obtenues par application de la nouvelle classe définie par notre auteur. On arrive à visualiser l'évolution des fonctions de répartition conjointes, dont certaines évoluent naturellement plus vite que d'autres.

Nous avons, une fois de plus, été confrontés à un problème : les courbes présentées par l'auteur. En effet, l'auteur affirme avoir construit les fonctions de répartition empiriques, mais elles décollent trop tôt, ce qui diffère assez de nos observations. On aurait pu se dire qu'il a zoomé sur une portion des graphes, mais cela n'est pas plausible vu que l'origine des axes est en zéro dans l'article. Aussi, aucune précision n'a été donnée sur la méthodologie de construction des courbes : nous avons  $y = f(s,t)$  donc une courbe en trois dimensions, qui finalement se retrouve sur un plan, dans l'article. Ainsi les données en abscisse n'ont pas d'interprétation concrète.

Il est vrai que nos graphes se présentent aussi en deux dimensions, mais c'est parce que nous avons fait coïncider systématiquement les valeurs de  $s$  et  $t$  sur l'intervalle considéré. Nous proposons des visualisations en 3D en annexe, ce qui nous permet principalement de cerner la notion de convexité d'une fonction à deux paramètres (nous avons utilisé le logiciel GeoGebra). Si notre auteur a aussi utilisé la même hypothèse que nous (c'est-à-dire  $s=t$ ), on s'attend à ce que les écarts entre les représentations graphiques ne soient pas aussi significatifs.

## 7 CONCLUSION

En somme, cet article, fort d'innovations intéressantes, a proposé, de manière rigoureuse, une nouvelle classe de copules FGM permettant d'élargir la plage de valeurs d'une mesure d'association, le rho de Spearman. L'auteur s'est d'abord assuré de pouvoir définir une copule bivariée avec la nouvelle fonction qu'il a introduite, et a ainsi pu trouver des conditions auxquelles les différents paramètres doivent satisfaire. Cela est d'autant plus utile que nous arrivons aisément à repérer si des estimations sont correctes ou non. Ensuite, *Bekrizadeh* a trouvé une expression fermée permettant de calculer le coefficient de corrélation de Spearman, suivant la sous-famille de copules considérée. Cela permet d'obtenir des approximations fiables comme l'illustrent les applications de cet article. Et, comme annoncé dans l'article, la plage de valeurs du coefficient de corrélation de Spearman a bel et bien pu être étendue.

Nous soulignons toutefois le manque criant de précision à propos de certains outils mobilisés par l'auteur pour faire des estimations, et à propos des considérations ayant permis d'obtenir certains graphiques représentés dans cet article. Aussi, la sélection des meilleures lois pour l'ajustement aux données, sur la base du critère d'information d'Akaike, n'a pas été respectée scrupuleusement puisque les lois lognormale et Weibull permettaient un meilleur fitting aux données sur le BMI et le PED respectivement. La loi gamma venait juste en deuxième position et s'ajustait bien aux données, mais il aurait été utile de justifier en quelques mots ce qui a motivé ce choix.

## A Représentation 3D de la densité de la copule FGM et de ses extensions



L'équation de la courbe ci-dessous (figure 11) est :  $f(x,y)=x y (1+0.2 (1-x) (1-y)) ((1+0.2 (1-y) (1-4 x)) (1+0.2 (1-x) (1-4 y))+3*0.2 x y)$

FIGURE 11 – Représentation 3D de la densité de la copule FGM

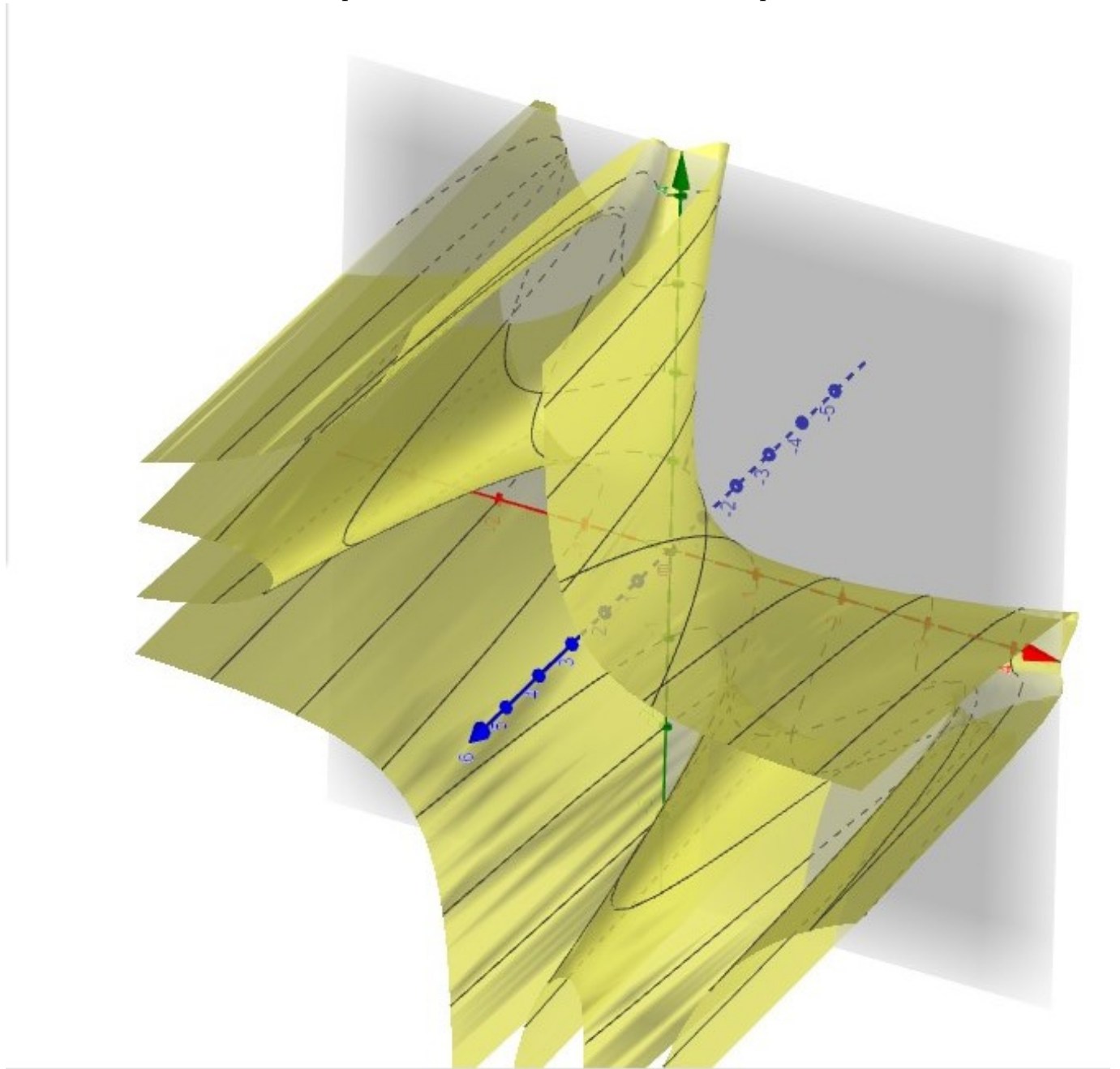




FIGURE 12 – Représentation de la densité de la copule FGM généralisée  $\theta = 0.8$  et  $p = 1$

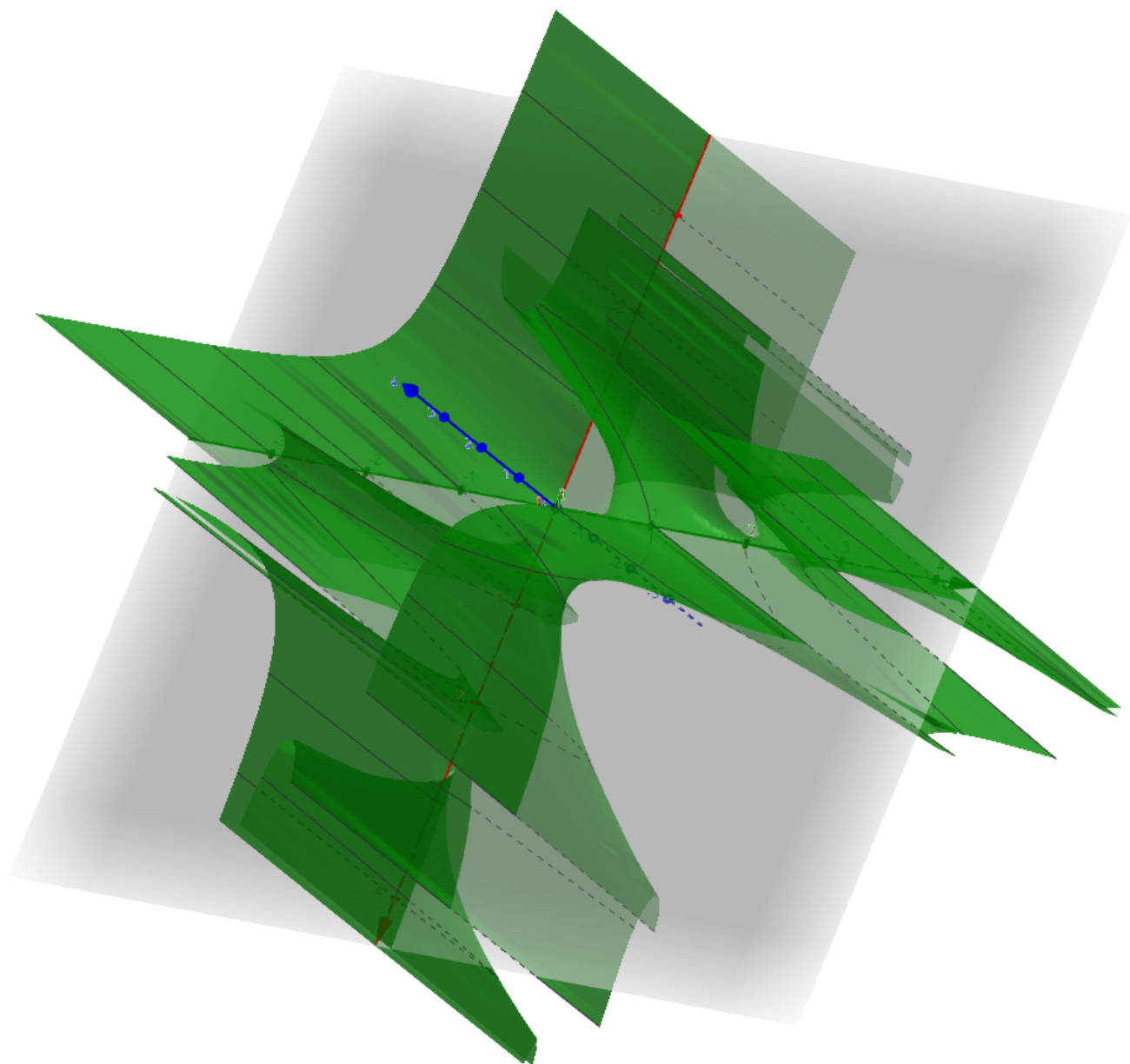


FIGURE 13 – Représentation de la densité de la copule FGM généralisée  $\theta = 0.5$  et  $p = 1$

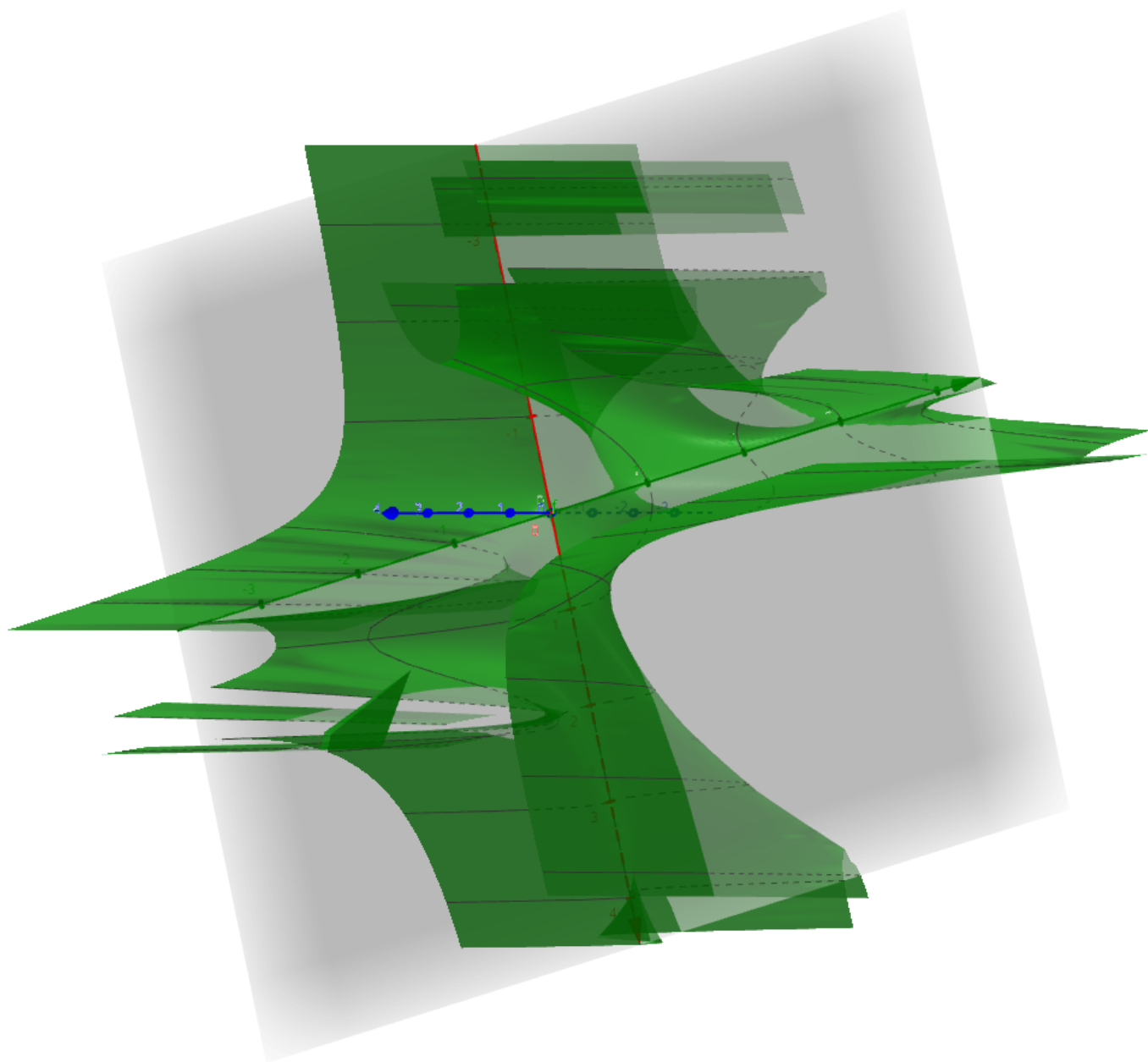
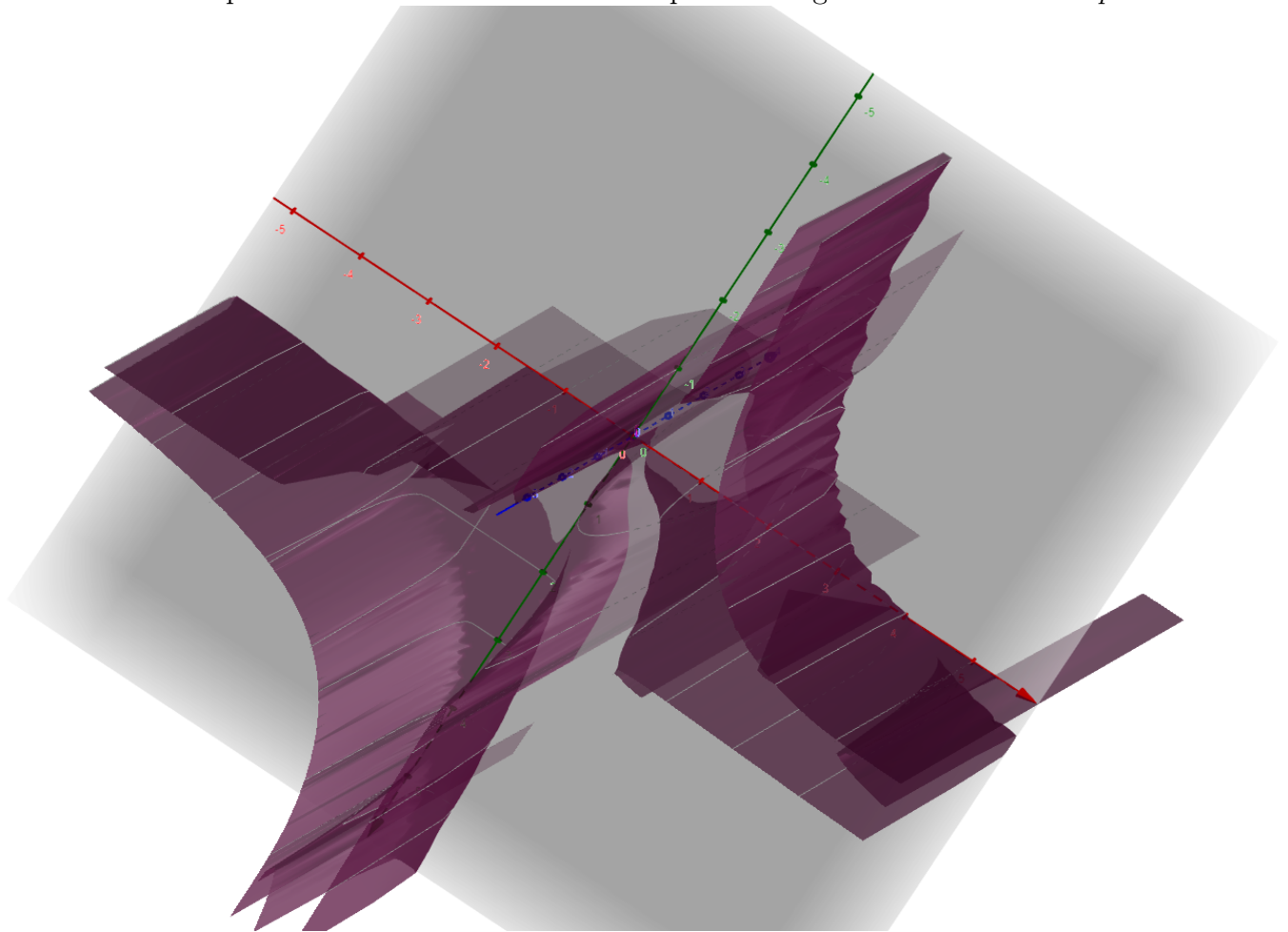


FIGURE 14 – Représentation de la densité de la copule FGM généralisée  $\theta = 0.5$  et  $p = 20$



Sur ces figures présentées ci-haut, nous arrivons à visualiser facilement la convexité d'une fonction à deux paramètres, sur des intervalles bien définis.

# Bibliographie

- Hakim Bekrizadeh (2024) Generalized FGM copulas : Properties and applications, Communications in Statistics - Simulation and Computation, 53 :2, 744-755.
- Michael Stanley Smith : Bayesian approaches to copula modelling. Methodology (stat.ME), 2011.
- Michael Pitt, David Chan et Robert Kohn (2006) : Efficient bayesian inference for gaussian copula regression models. Biometrika, 93 :537–554.
- Nelsen,R,B. (2006)- An introduction to copulas. second ed. Springer, New York.