

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**UNIVERSITÉ MOHAMED KHIDER, BISKRA**

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

**DÉPARTEMENT DE MATHÉMATIQUES**



Thèse pour obtenir le titre de

**DOCTEUR en Sciences**

Option : **Statistique**

Présentée par

**BENELMIR Imane**

Titre

**Modélisation de la Dépendance par les Copules**

*Membres du Comité d'Examen*

Pr. <b>Necir</b>	<b>Abdelhakim</b>	Université de Biskra	Président
Pr. <b>Meraghni</b>	<b>Djamel</b>	Université de Biskra	Rapporteur
Pr. <b>Fellag</b>	<b>Hocine</b>	Université de Tizi Ouzou	Examineur
Pr. <b>Mohdeb</b>	<b>Zaher</b>	Ecole Nationale Polytechnique de Constantine	Examineur
Pr. <b>Yousfate</b>	<b>Abderrahmane</b>	Université de Sidi-Bel-Abbes	Examineur
Dr. <b>Benatia</b>	<b>Fatah</b>	Université de Biskra	Examineur

*A mes très chers parents,*

*A mon mari et ma fille,*

*A mon frère et ma soeur,*

*A mes proches, amis et collègues.*

## REMERCIEMENTS

*Je tiens à remercier "Allah" le tout puissant pour toute la volonté et le courage qu'il m'a donné pour l'achèvement de ce travail.*

*La réalisation de cette thèse a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma reconnaissance.*

*Je tiens d'abord à exprimer toute ma reconnaissance à mon Directeur de thèse le professeur **MERAGHNI Djamel**. Je le remercie de m'avoir encadré, orienté, aidé et conseillé.*

*J'adresse mes sincères remerciements au professeur **NECIR Abdelhakim** qui fut le premier à me faire découvrir l'histoire des copules, pour sa gentillesse et ses judicieux conseils.*

*Je souhaite renouveler mes remerciements au Dr. **BRAHIMI Brahim** qui par ses paroles, ses écrits, ses conseils et ses critiques a guidé mes réflexions et a accepté de me rencontrer et de répondre à mes questions durant mes recherches.*

*Je tiens aussi à remercier Monsieur le Président le Pr. **NECIR Abdelhakim** et les membres du jury : le Pr. **FELLAG Hocine**, le Pr. **MOHDEB Zaher**, le Pr. **YOUSFATE Abderrahmane** et le Dr. **BENATIA Fatah**, qui ont bien daigné siéger la soutenance de cette thèse.*

*Je remercie mes très chers parents qui ont toujours été là pour moi, "Vous avez tout sacrifié pour vos enfants n'épargnant ni santé ni efforts. Vous m'avez donné un magnifique modèle de labeur et de persévérance. Je vous suis redevable d'une éducation dont je suis fier".*

*Un grand remerciement à mon mari, mon frère et ma soeur pour leurs soutiens inconditionnels et leurs encouragements. Je vous dois toute ma reconnaissance et mon attachement.*

*Enfin j'adresse mes plus sincères remerciements à tous mes proches, **amies**  
**et collègues**, qui m'ont toujours soutenu et encouragé au cours de la réalisation de cette thèse.*

*À tous ces intervenants, je présente mes remerciements, mon respect et ma gratitude.*

# Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iv
Liste des figures	vii
Liste des tableaux	viii
Introduction	1
<b>1 Théorie des copules</b>	<b>5</b>
1.1 Théorème de Sklar . . . . .	6
1.2 Densité de la copule . . . . .	8
1.3 Propriétés d'une copule . . . . .	8
1.4 Types de copules . . . . .	13
1.4.1 Copules usuelles . . . . .	13
1.4.2 Copules elliptiques . . . . .	13
1.4.3 Copule de Plackett . . . . .	17
1.4.4 Copules Archimédiennes . . . . .	18
1.4.5 Copules de valeurs extrêmes . . . . .	21
1.4.6 Copules Archimax . . . . .	22
1.4.7 Copules empiriques . . . . .	24

1.5	Copules associées à une copule . . . . .	25
1.5.1	Copule de survie . . . . .	26
1.5.2	Copule duale . . . . .	27
1.5.3	Co-copule . . . . .	27
1.5.4	Copule mixte . . . . .	28
1.6	Construction de la copule . . . . .	28
1.7	Applications des Copules . . . . .	31
1.8	Mise en oeuvre des copules . . . . .	31
1.8.1	Inférence statistique . . . . .	32
1.8.2	Simulation de la copule . . . . .	55
<b>2</b>	<b>Mesures de dépendance et de concordance</b>	<b>62</b>
2.1	Fonction de concordance . . . . .	62
2.1.1	Corrélation linéaire de Pearson . . . . .	63
2.1.2	Concept de concordance . . . . .	63
2.1.3	Propriétés de la fonction de concordance . . . . .	64
2.2	Mesures de concordance . . . . .	69
2.2.1	Tau de Kendall . . . . .	70
2.2.2	Rho de Spearman . . . . .	71
2.3	Autres mesures de corrélation . . . . .	74
2.3.1	Beta de Blomqvist . . . . .	74
2.3.2	Gamma de Gini . . . . .	76
2.4	Copule de Marshall-Olkin . . . . .	77
2.4.1	Propriétés de la copule . . . . .	79
2.5	Copule de Farlie-Gumbel-Morgenstern . . . . .	83
2.5.1	Itération des distributions de FGM . . . . .	84
2.5.2	Propriétés de la copule . . . . .	85
<b>3</b>	<b>Association measures and estimation of copula parameters</b>	<b>88</b>

3.1	Introduction . . . . .	89
3.1.1	Marshall-Olkin copula . . . . .	91
3.1.2	Farlie-Gumbel-Morgenstern copula . . . . .	92
3.2	Association measures . . . . .	93
3.2.1	Kendall's tau . . . . .	94
3.2.2	Spearman's rho . . . . .	94
3.2.3	Gini's gamma . . . . .	94
3.2.4	Blomqvist's beta . . . . .	95
3.3	Simulation comparative study . . . . .	95
3.4	Appendix . . . . .	97
	<b>Conclusion générale</b>	<b>100</b>
	<b>Annexe : Abréviations et Notations</b>	<b>103</b>
	<b>Bibliographie</b>	<b>106</b>

# Table des figures

1.1	Densité $c_\rho$ de trois copules Gaussiennes selon $\rho$ . . . . .	15
1.2	Nuages de points de trois copules Gaussiennes selon $\rho$ . . . . .	15
1.3	Densité $c_{\rho,\nu}$ de trois copules de Student selon $\rho$ avec $\nu = 5$ . . . . .	16
1.4	Nuages de points de trois copules de Student selon $\rho$ avec $\nu = 5$ . . . . .	17
1.5	Nuages de points de trois copules Archimédiennes de paramètres $\theta$ . . . . .	20
1.6	Densité $c_\theta(u, v)$ de trois copules Archimédiennes de paramètres $\theta$ . . . . .	20
1.7	Fonction $K(t)$ pour différentes copules et différentes valeurs de $\theta$ [5]. . . . .	34
1.8	Fonction $J(t)$ pour différentes copules et différentes valeurs de $\theta$ [5]. . . . .	36
1.9	Fonction $M(t)$ pour différentes copules et différentes valeurs de $\theta$ [5]. . . . .	37
1.10	Fonctions $L(t)$ et $R(t)$ pour différentes copules et différentes valeurs de $\theta$ [5]. . . . .	39
1.11	Kendall plots comparant un échantillon de données simulées à partir d'une copule Student avec $\rho = 0.5$ et $\nu = 3$ à plusieurs copules estimées sur ce même échantillon[5]. . . . .	52
1.12	Nuages de points de la copule MO selon les trois cas de dépendance. . . . .	57
1.13	Nuages de points de la copule FGM selon les trois cas de dépendance. . . . .	60
2.1	Les quatre quadrants. . . . .	75
2.2	Copule de MO. Gauche : Densité de la copule. Droite : Simulation de 2000 vas avec $a = 0.25$ et $b = 0.75$ . . . . .	79
2.3	Relation entre le paramètre de la copule et les quatre mesures de concordance. . . . .	82
2.4	Copule de FGM. Gauche : Densité de la copule. Droite : Simulation de 2000 vas avec $a = 0.5$ et $b = 1.2$ . . . . .	85

# Liste des tableaux

1.1	Dépendance de queue de quelques familles classiques de copules. . . . .	12
1.2	Générateurs des principales copules Archimédiennes. . . . .	19
1.3	Copules de valeurs extrêmes les plus utilisées. . . . .	23
1.4	Fonction de Kendall de trois copules Archimédiennes. . . . .	34
1.5	Tableau de contingence donnant les effectifs empiriques issus de la copule empirique (respectivement les effectifs théoriques issus de la copule théorique). . . . .	54
2.1	Le tau de Kendall et le rho de Spearman de quelques familles classiques. . . . .	73
3.1	Association measures of FGM and MO copulas with two parameters. . . . .	94
3.2	Estimation biases and rmse's for MO copula parameters under strong dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel). . . . .	97
3.3	Estimation biases and rmse's for MO copula parameters under moderate dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel). . . . .	97
3.4	Estimation biases and rmse's for MO copula parameters under weak dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel). . . . .	98
3.5	Estimation biases and rmse's for FGM copula parameters under strong dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel). . . . .	98
3.6	Estimation biases and rmse's for FGM copula parameters under moderate dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel). . . . .	98
3.7	Estimation biases and rmse's for FGM copula parameters under weak dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel). . . . .	99

# Introduction générale

Mesurer la dépendance entre deux ou plusieurs variables aléatoires (vas) est une pratique largement répandue par les statisticiens. Un riche ensemble de mesures de dépendance entre les vas a été proposé comme le coefficient de corrélation de Pearson, le tau de Kendall, le rho de Spearman,...etc. Bien que ces mesures soient simples à calculer et peuvent être facilement interprétées, elles ne sont pas en mesure de détecter toutes les formes de dépendances. Donc il était indéniable de trouver un autre moyen pour résoudre ce problème. En effet, la fonction "copule" a l'avantage de modéliser complètement la dépendance entre les vas.

L'introduction des copules à des fins statistiques est un phénomène relativement récent qui trouve sa source à la fin des années 50, dans des recherches portant initialement sur les tables de contingence. Au début, il était difficile de trouver des traces de la notion copule dans la littérature statistique. Mais à partir des années 70 et avec le développement de la théorie des processus empiriques, d'autres auteurs ont redécouvert le concept des fonctions copules sous d'autres appellations en établissant, de manière originale certaines de leurs propriétés. C'est ainsi qu'elles sont nommées "représentation uniforme" par Kimeldorf et Sampson [62], "fonction de dépendance" par Galambos [40] et Deheuvels [24], ou encore "la forme standard" par Cook et Johnson [19].

Les copules sont devenues en quelques années un outil important, pour modéliser les risques multivariés (entre autres). Elles permettent de coupler les lois marginales afin d'obtenir une loi multivariée, d'où le nom latin **copula** qui signifie au sens figuré, lien, alliance ou union. Cette notion apparaît sous d'autres appellations dans certains travaux de Frechet [39], Feron [37] et Dall'Aglio [21] portant sur l'étude des tables de contingence et lois multidimensionnelles

à structures marginales fixées. Mais c'était en 1959 que le terme "copule" a été utilisé pour la première fois grâce à Sklar [75] dans la théorie des lois multidimensionnelles.

La théorie des copules permet de construire des modèles paramétriques pour la loi jointe de  $d$  variables réelles dépendantes  $(X_1, \dots, X_d)$ . En pratique, d'après Fisher [38], elle est de plus en plus utilisée en assurance, finance, environnement, fiabilité, la biologie... etc, pour modéliser de manière plus réaliste la loi jointe de plusieurs "risques". Dans ce travail on se restreindra au cas bivarié ( $d = 2$ ) mais les méthodes peuvent aisément se généraliser en dimension supérieure (au moins en théorie).

Ainsi, à chaque fois qu'il est nécessaire de modéliser une structure de dépendance entre plusieurs variables, nous pouvons faire appel à la fonction copule qui possède les avantages suivants :

- Elle est une fonction qui modélise les propriétés des structures de dépendance telles que la forme et l'intensité.
- Elle est capable de mesurer la dépendance pour les distributions à queues lourdes.
- Elle permet de construire des modèles de distributions multidimensionnelles.
- Elle autorise l'élaboration des modèles non gaussiens, i.e. qu'avec une copule, on peut construire par exemple une distribution avec une marginale gaussienne et une autre uniforme.

**Ce travail est constitué de cinq parties :**

**La première partie** est consacrée à l'introduction générale où on présente les grandes lignes de ce travail, et on fournit un exposé de synthèse sur ce sujet. On commence par donner un historique des copules et expliquer le choix de la copule pour modéliser la structure de dépendance. Ensuite on rappelle le contenu des chapitres suivants et la contribution de ce travail.

**Le premier chapitre** est une introduction mathématique aux copules. On y retrouve les principales définitions et théorèmes qui serviront de base aux chapitres suivants. On expose d'abord les propriétés fondamentales et on présente les principales familles de copules, à commencer par les grandes familles qui sont les copules usuelles et elliptiques, Archimédiennes, ...etc, accompagnées de présentations graphiques de quelques-unes d'entre elles. Puis, dans la partie copules associées à une copule, on va exprimer les probabilités au terme de celles-ci, on va aussi

apprendre à les construire. **Dans la septième section** une discussion sera amorcée sur ses différents domaines d'applications des copules.

**La huitième section** est réparti en deux sous-sections : inférence statistique et simulation de la copule. La première sous-section est divisée en trois parties, dans la première on présente des fonctions qui permettent de faire un choix entre les copules et de n'en retenir que celles qui conviennent à une certaine série de données.

L'objectif principal de la deuxième partie est de fournir différentes méthodes d'estimation des copules à savoir l'approche paramétrique, semi-paramétrique et non-paramétrique.

La première approche consiste à estimer le paramètre de la copule par l'une des méthodes suivantes :

- Maximum de vraisemblance exacte (EML).
- Fonctions d'inférence des marginales (IFM).
- Moments.

Dans la seconde approche on utilise :

- Maximum de vraisemblance canonique (CML).
- L-moments bivariés (BLM).

Quand à la troisième approche on utilise les méthodes suivantes :

- Copule empirique.
- Procédure de GENEST et RIVEST.

Dans la troisième partie, on propose une méthode de sélection de la meilleure copule basée sur la méthode graphique et les tests d'adéquation à savoir : la statistique du chi-deux et le test de Cramér-von Mises.

On termine (deuxième sous-section) par quelques points sur la mise en oeuvre des copules en présentant les diverses techniques de simulation.

**Dans le second chapitre**, on va voir comment quantifier la dépendance, en présentant des mesures de corrélation, ou plus précisément des mesures de concordance à savoir la corrélation de Pearson, le tau de Kendall et le rho de Spearman. D'autres mesures d'association vont aussi

être présentées comme le gamma de Gini et le beta de Blomqvist. On va prendre comme exemple pour le calcul de ses différentes mesures, deux familles de copules à deux paramètres : la copule de Marshall-Olkin et la copule de Farlie-Gumbel-Morgenstern dont on va étudier leurs propriétés fondamentales.

Notre contribution se trouve dans le **troisième chapitre** qui est la partie application. Dans cette partie, on étudie la méthode d'inversion (semi-paramétrique) des différentes mesures d'association (quatre mesures), pour estimer les paramètres de deux copules choisies. Ceci dans le but de comparer la robustesse de ces estimateurs en terme du biais et de la racine de l'erreur quadratique moyenne. Pour un meilleur résultat, on va confronter cette méthode à une autre méthode semi-paramétrique qui est la méthode des BLM. Sachant que tout cela va se faire grâce à une étude de simulation.

On achève notre travail par une conclusion générale.

# Chapitre 1

## Théorie des copules

La construction des copules repose sur les propriétés des fonctions de répartition (fds). On rappelle ci-dessous quelques propriétés importantes des fds bivariées. Soit  $(X, Y)$  un couple de vas, la loi du couple est caractérisée par la fd bivariée  $H$  définie comme ceci

$$\forall (x, y) \in \mathbb{R}^2 : H(x, y) = \mathbb{P}[X \leq x, Y \leq y].$$

On appelle lois marginales, les lois de  $X$  et de  $Y$  prises séparément. On peut exprimer les fds de ces lois marginales en fonction de  $H$ . Par exemple, pour  $X$  on obtient

$$F(x) := \mathbb{P}[X \leq x] = \lim_{y \rightarrow +\infty} H(x, y).$$

On rappelle que les vas  $X$  et  $Y$  sont indépendantes si et seulement si

$$\forall (x, y) \in \mathbb{R}^2 : H(x, y) = F(x)G(y),$$

où  $G$  est la fd de  $Y$ .

On note par la suite

$$F^{-1}(t) := \inf\{x \in \mathbb{R} \mid F(x) \geq t\}; \text{ pour } t \in [0, 1],$$

la fonction quantile qui est l'inverse de la fd  $F$ . On rappelle le lemme d'inversion qui est couramment utilisé pour générer des échantillons d'une loi dont la fonction quantile est connue : si  $U$  suit une loi uniforme sur l'intervalle  $\mathbf{I} = [0, 1]$ <sup>1</sup>, alors  $F^{-1}(U)$  suit la même loi que  $X$ . Les fds des lois marginales sont alors des fonctions continues et les vas  $F(x)$  et  $G(y)$  suivent des lois uniformes sur l'intervalle  $\mathbf{I}$  notée  $\mathcal{U}_{\mathbf{I}}$ .

On vérifie alors que les lois marginales possèdent également une densité par rapport à la mesure de Lebesgue. Par exemple, la densité de la loi de  $X$  est donnée par

$$\forall x \in \mathbb{R} : f(x) = \int_{\mathbb{R}} h(x, y) dy,$$

où  $h$  est la densité bivariée du couple  $(X, Y)$ .

On est alors ramené à modéliser un couple de vas dont les lois marginales sont des lois  $\mathcal{U}_{\mathbf{I}}$ . La fd d'un tel couple est appelée "copule". Ceci est formalisé plus précisément ci-dessous.

## 1.1 Théorème de Sklar

Ce théorème est fondamentale dans la théorie des copules. Il a été établi par Sklar en 1959 où il précise le lien défini par la copule  $C$ , déterminée à partir de la distribution jointe  $H$ , entre les fds marginales univariées  $F$  et  $G$  et la distribution complète bivariée  $H$ .

**Théorème 1.1.1** *Si  $H$  est la fd de  $(X, Y)$  alors il existe une copule  $C$  de dimension 2 telle que*

$$\boxed{\forall (x, y) \in \mathbb{R}^2 : H(x, y) = C(F(x), G(y))}. \quad (1.1)$$

---

<sup>1</sup>Rappelons qu'une va  $U$  qui suit la loi uniforme sur  $\mathbf{I}$  a pour fd :

$$\mathbb{P}(U \leq u) = \begin{cases} 0 & \text{si } u < 0 \\ u & \text{si } 0 \leq u < 1 \\ 1 & \text{si } u \geq 1 \end{cases}$$

**Preuve.** Soit  $C$  la fd du vecteur  $(U, V)$  où  $U = F(X)$  et  $V = G(Y)$ .

$$\begin{aligned}
 H(x, y) &= \mathbb{P}[X \leq x, Y \leq y] \\
 &= \mathbb{P}[F^{-1}(U) \leq x, G^{-1}(V) \leq y] \\
 &= \mathbb{P}[U \leq F(x), V \leq G(y)] \\
 &= C(F(x), G(y)).
 \end{aligned}$$

Ce qu'il fallait démontrer (CQFD). ■

**Théorème 1.1.2** *Soit  $H$  la fd bivariée de marginales  $F$  et  $G$ . La copule  $C$  associée à  $H$  est donnée par*

$$\boxed{\forall (u, v) \in \mathbf{I}^2 : C(u, v) = H(F^{-1}(u), G^{-1}(v)).} \quad (1.2)$$

**Preuve.** En appliquant le théorème 1.1 en posant  $u = F(x)$  et  $v = G(y)$ , on trouve que

$$\begin{aligned}
 C(u, v) &= \mathbb{P}(U \leq u, V \leq v) \\
 &= \mathbb{P}(F(X) \leq u, G(Y) \leq v) \\
 &= \mathbb{P}(X \leq F^{-1}(u), Y \leq G^{-1}(v)) \\
 &= H(F^{-1}(u), G^{-1}(v)).
 \end{aligned}$$

CQFD. ■

Si  $F$  et  $G$  sont continues, alors  $C$  est unique. Réciproquement, si  $C$  est la copule des vas  $X$  et  $Y$ , alors  $H$  est une fd jointe avec des marginales  $F$  et  $G$  respectivement.

Le théorème de Sklar suggère donc de modéliser la loi jointe d'un couple de vas en deux étapes :

- Modélisation des lois marginales.
- Modélisation de la structure de dépendance après transformation des lois marginales en lois uniformes.

## 1.2 Densité de la copule

Les copules admettent des densités de probabilités. Si la densité  $c$  associée à la copule  $C$  existe, alors elle est définie par

$$c(u, v) := \frac{\partial^2 C(u, v)}{\partial u \partial v},$$

où  $c : \mathbf{I}^2 \rightarrow \mathbb{R}^+$ . Si la fd jointe  $H$  est absolument continue, en utilisant le théorème de Sklar, on peut exprimer la densité d'un couple aléatoire  $(X, Y)$  en fonction de la densité de sa copule et de ses marginales  $f$  et  $g$  par

$$h(x, y) := c(F(x), G(y))f(x)g(y). \quad (1.3)$$

## 1.3 Propriétés d'une copule

On note tout d'abord que les copules vérifient certaines propriétés.

**a.** Les marges suivent des lois  $\mathcal{U}_1$  :

- $C(u, v) = 0$  si  $u \leq 0$  ou  $v \leq 0$ .
- $C(u, v) = 1$  si  $u \geq 1$  et  $v \geq 1$ .
- $C(u, v) = u$  si  $v \geq 1$  et  $C(u, v) = v$  si  $u \geq 1$ .

**b.** La continuité :

On note que les copules sont des fonctions continues. Plus précisément, elles vérifient une condition de Lipschitz.  $\forall u_1, u_2, v_1, v_2 \in \mathbf{I}$  on a

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

**c.** La symétrie :

On dit que  $C$  est symétrique si

$$\forall (u, v) \in \mathbf{I}^2 : C(u, v) = C(v, u).$$

d. L'ordre :

On dit que  $C_1$  est plus petite que  $C_2$  et inversement si

$$\forall (u, v) \in \mathbf{I}^2 : C_1(u, v) \leq C_2(u, v).$$

e. La croissance :

$C$  est 2-croissante :  $\forall (u_1, u_2), (v_1, v_2) \in \mathbf{I}^2$  vérifiant  $u_1 \leq u_2$  et  $v_1 \leq v_2$  on a

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

f. Le théorème d'invariance :

L'un des théorèmes essentiels à la théorie des copules est celui de l'invariance par transformations strictement croissantes. Soient  $X$  et  $Y$  deux v.a.s continues de marginales  $F$  et  $G$  et de copule  $C_{XY}$ . Si  $\alpha$  et  $\beta$  sont deux fonctions strictement croissantes, alors

$$C_{\alpha(X)\beta(Y)} := C_{XY}.$$

**Preuve.** On note respectivement  $F$  et  $G$  les f.d.s jointes des vecteurs  $(X, Y)$  et  $(\alpha(X), \beta(Y))$ . Par la suite on note que  $F_1, F_2$  sont les f.d.s marginales de  $F$  et par  $G_1, G_2$  les f.d.s marginales de  $G$ . On remarque que les marges de  $G$  sont

$$\begin{aligned} G_1(x) &= \mathbb{P}[\alpha(X) \leq x] \\ &= \mathbb{P}[X \leq \alpha^{-1}(x)], \text{ car } \alpha \text{ est une fonction croissante} \\ &= F_1(\alpha^{-1}(x)). \end{aligned}$$

On fait la même chose pour la deuxième marge de  $G$ , on obtient alors  $G_2(y) = F_2(\beta^{-1}(y))$ .

On pose  $G_1(x) = u$  et  $G_2(y) = v$ , on obtient donc  $G_1^{-1}(u) = \alpha F_1^{-1}(u)$  et  $G_2^{-1}(v) = \beta F_2^{-1}(v)$ .

On en déduit alors que

$$\begin{aligned}
C_{\alpha(X)\beta(Y)}(u, v) &= G(x, y) \\
&= G(G_1^{-1}(u), G_2^{-1}(v)) \\
&= \mathbb{P}[\alpha(X) \leq G_1^{-1}(u), \beta(Y) \leq G_2^{-1}(v)] \\
&= \mathbb{P}[X \leq \alpha^{-1}(G_1^{-1}(u)), Y \leq \beta^{-1}(G_2^{-1}(v))] \\
&= \mathbb{P}[X \leq F_1^{-1}(u), Y \leq F_2^{-1}(v)] \\
&= F(x, y) \\
&= C_{XY}(u, v).
\end{aligned}$$

CQFD. ■

Par exemple, la copule de la distribution lognormale est la même que celle de la loi normale (en effet la première est une transformation strictement croissante " $y = \log x$ " de la seconde). Dans le cas de transformations strictement décroissantes des distributions marginales d'une copule  $C$  on a le résultat suivant :

### **Théorème 1.3.1**

*Si  $\alpha$  est strictement croissante ( $\nearrow$ ) et  $\beta$  strictement décroissante ( $\searrow$ ) :*

$$C_{\alpha(X)\beta(Y)}(u, v) = u - C_{XY}(u, 1 - v).$$

*- Si  $\alpha$  est strictement décroissante ( $\searrow$ ) et  $\beta$  strictement croissante ( $\nearrow$ ) :*

$$C_{\alpha(X)\beta(Y)}(u, v) = v - C_{XY}(1 - u, v).$$

*- Si  $\alpha$  et  $\beta$  sont strictement décroissantes ( $\searrow$ ) :*

$$C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 - C_{XY}(1 - u, 1 - v).$$

**g.** La dépendance de queue :

Cette notion est très importante dans l'étude de la dépendance asymptotique entre deux vas. Cela va nous permettre de voir le niveau de dépendance dans les valeurs extrêmes (upper tail

dependence) et dans les petites valeurs (lower tail dependence).

L'indicateur de dépendance de queue d'une copule se déduit des probabilités conditionnelles suivantes :

$$\begin{aligned}
- \mathbb{P}[U \leq u | V \leq v] &= \frac{\mathbb{P}[U \leq u, V \leq v]}{\mathbb{P}[V \leq v]} = \frac{C(u, v)}{v}. \\
- \mathbb{P}[U > u | V > v] &= \frac{\mathbb{P}[U > u, V > v]}{\mathbb{P}[V > v]} \\
&= \frac{\mathbb{P}[U \leq 1, V > v] - \mathbb{P}[U \leq u, V > v]}{\mathbb{P}[V \leq 1] - \mathbb{P}[V \leq v]} \\
&= \frac{\mathbb{P}[U \leq 1, V \leq 1] - \mathbb{P}[U \leq 1, V \leq v] - \mathbb{P}[U \leq u, V > v]}{1 - v} \\
&= \frac{1 - v - \mathbb{P}[U \leq u, V \leq 1] + \mathbb{P}[U \leq u, V \leq v]}{1 - v} \\
&= \frac{1 - v - u + C(u, v)}{1 - v}.
\end{aligned}$$

Cette mesure de la dépendance de queue est la probabilité qu'une variable est extrême, étant donné que l'autre est extrême. On peut alors définir les dépendances de queue à gauche et à droite. Pour plus de détail voir Joe [57].

**Définition 1.3.1** (*Lower tail dependence*)

Une copule  $C$  a une dépendance de queue à gauche si :

$$\lambda_L := \lim_{u \rightarrow 0^+} \mathbb{P}[X \leq F^{-1}(u) | Y \leq G^{-1}(u)] = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u} \text{ existe et } \lambda_L \in ]0, 1].$$

Si  $\lambda_L = 0$  alors elle n'a pas de dépendance de queue à gauche.

**Définition 1.3.2** (*Upper tail dependence*)

Une copule  $C$  a une dépendance de queue à droite si :

$$\lambda_U := \lim_{u \rightarrow 1^-} \mathbb{P}[X > F^{-1}(u) | Y > G^{-1}(u)] = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u} \text{ existe et } \lambda_U \in ]0, 1].$$

Si  $\lambda_U = 0$  alors elle n'a pas de dépendance de queue à droite.

Les paramètres de la dépendance de queue  $\lambda_L$  et  $\lambda_U$  sont facilement évalués pour certaines familles de copules dont nous citons quelques unes dans le tableau (1.1) ci-dessous :

Famille	$\lambda_L$	$\lambda_U$
Clayton	$2^{-1/\theta}$	0
Gumbel	0	$2 - 2^{1/\theta}$
Frank	0	0
Joe	0	$2 - 2^{1/\theta}$
AMH	0	0
Gaussienne	0	0
FGM	0	0

TAB. 1.1 – Dépendance de queue de quelques familles classiques de copules.

Les cinq premières copules sont de type Archimédiennes voir la section 1.4.4.

La copule de Farlie-Gumbel-Morgenstern (FGM) est définie comme suit

$$C_\theta(u, v) = uv [1 + \theta (1 - u) (1 - v)], \theta \in [-1, 1].$$

**Remarque 1.3.1** *Pour une copule Archimédienne, les paramètres de la dépendance de queue peuvent être exprimés en termes de limites concernant le générateur et son inverse voir Nelsen [66] :*

$$\lambda_U = 2 - \lim_{u \rightarrow 1^-} \frac{1 - \phi^{-1}(2\phi(u))}{1 - u} \text{ et } \lambda_L = \lim_{u \rightarrow 0^+} \frac{\phi^{-1}(2\phi(u))}{u}.$$

Ces mesures de la dépendance de queue sont non paramétriques et ne dépendent que de la copule  $C$ .

## 1.4 Types de copules

De nombreuses familles de copules ont été proposées dans la littérature. Les plus utilisées, sont introduites dans cette section. Elles sont toutes disponibles dans le package `copula` sur le logiciel R. Une liste plus complète peut être trouvée dans Nelsen [66].

### 1.4.1 Copules usuelles

La famille des copules est bornée par ce qu'on appelle par les **bornes de Fréchet-Hoeffding**.

On a

$$\forall (u, v) \in \mathbf{I}^2 : W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) = M(u, v).$$

- $M$  est la fd du couple  $(U, U)$ , elle est appelée copule minimum ou copule comonotone.
- $W$  est la fd du couple  $(U, 1 - U)$ , elle est appelée copule maximum ou copule anticomonotone.

On note que  $M$  est une copule, en revanche  $W$  est une copule uniquement dans le cas où  $d = 2$ .

Un autre cas particulier est celui des variables indépendantes : la copule associée est la copule  $\Pi$  définie par

$$\forall (u, v) \in \mathbf{I}^2 : \Pi(u, v) = uv,$$

appelée copule indépendante ou copule produit.

### 1.4.2 Copules elliptiques

En travaillant dans le cadre des copules multivariées, on a été confronté dès le début à un problème. Toute la littérature abondante de ces dernières années sur les copules était entièrement dévolue au cas bivarié et lorsque le nombre de dimensions est supérieur à deux, les auteurs se ramènent toujours à une copule elliptique (normale ou Student) pour la simple et bonne raison que les densités de ces copules sont faciles à calculer et donc plus aisées à simuler. Les copules elliptiques sont des copules associées aux distributions elliptiques.

**Définition 1.4.1** *On appelle copule elliptique toute copule qui s'écrit de la forme suivante*

$$C_\rho(u, v) := \frac{1}{\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi_{g,1}^{-1}(u)\Phi_{g,2}^{-1}(v)} \int_{-\infty}^{\Phi_{g,1}^{-1}(u)\Phi_{g,2}^{-1}(v)} g\left(\frac{x^2 - 2\rho xy + y^2}{\sqrt{1-\rho^2}}\right) dx dy = H_\rho(\Phi_{g,1}^{-1}(u), \Phi_{g,2}^{-1}(v)),$$

où  $H_\rho$  est la distribution jointe des vas  $X$  et  $Y$ ,  $\Phi_{g,1}^{-1}(u)$  et  $\Phi_{g,2}^{-1}(v)$  sont leurs fonctions quantiles respectives et  $\rho$  leurs coefficients de corrélation. Dans cette famille se trouve entre autres la copule Gaussienne et la copule de Student.

### Copule normale

Cette copule ne permet pas de mesurer la dépendance entre les queues des distributions marginales. Il s'agit d'une propriété contraignante lorsqu'on veut évaluer la dépendance entre les événements rares. Un des types de copules les plus utilisées dans la modélisation est la copule normale bivariée.

Il s'agit de la copule la plus couramment utilisée, elle est donnée par

$$C_\rho(u, v) := \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)\Phi^{-1}(v)} \int_{-\infty}^{\Phi^{-1}(u)\Phi^{-1}(v)} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) dx dy,$$

où  $\Phi^{-1}$  est la fonction quantile de la loi normale standard  $\mathcal{N}(0, 1)$ . On retrouve les cas particuliers suivants comme cas limites :  $C_{-1} = W$ ,  $C_0 = \Pi$  et  $C_1 = M$ .

La figure (1.1) montre le graphique de la densité de trois copules Gaussienne qui diffèrent selon la valeur du  $\rho$ .

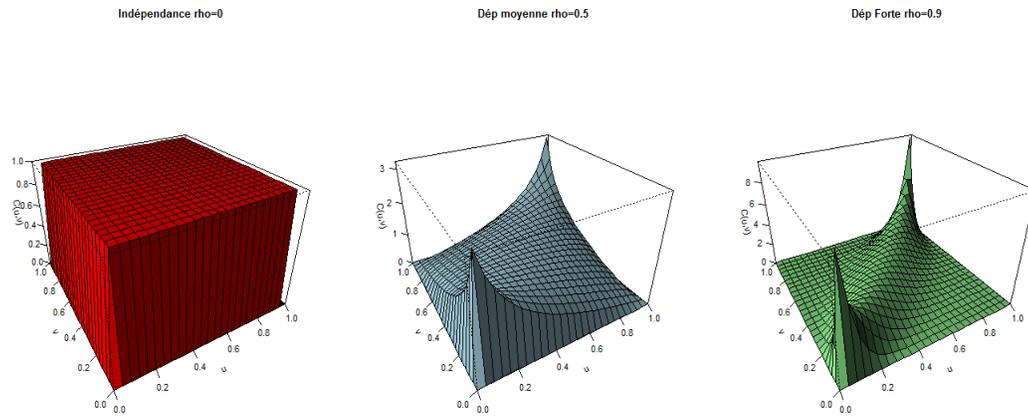


FIG. 1.1 – Densité  $c_\rho$  de trois copules Gaussiennes selon  $\rho$ .

On remarque que pour  $\rho = 0$ , on a  $C = \Pi$ . Pour les mêmes valeurs de  $\rho$ , la figure (1.2), quant à elle, illustre ce à quoi pourrait ressembler un nuage de points d'une copule Gaussienne.

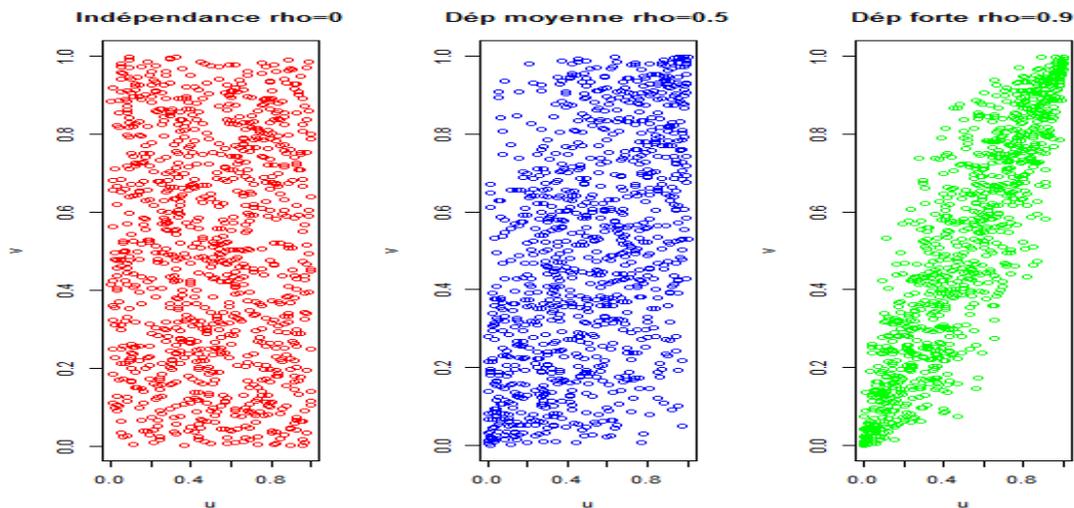


FIG. 1.2 – Nuages de points de trois copules Gaussiennes selon  $\rho$ .

Ne présentant pas une dépendance de queue, la copule Gaussienne n'est pas adaptée pour modéliser l'interaction de valeurs extrêmes. Dans ce cas, mieux vaudra utiliser d'autres copules, tout comme pour caractériser une dépendance non linéaire.

De la même manière, on peut définir la copule de Student basée sur la loi de Student bivariée.

## Copule de Student

Comparativement à la copule gaussienne, la copule de Student permet grâce à son degré de liberté, de mieux tenir compte des queues de distribution épaisses. En outre, lorsque le degré de liberté tend vers l'infini, la copule de Student est égale à la copule Gaussienne. La copule de Student est définie de la façon suivante

$$C_{\rho, \nu}(u, v) := \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{T_{\nu}^{-1}(u)} \int_{-\infty}^{T_{\nu}^{-1}(v)} \left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{-\left(\frac{\nu}{2}+1\right)} dx dy,$$

avec une matrice diagonale définie positive,  $\text{diag } \rho = 1$  et  $T_{\nu}^{-1}$  est la fonction inverse de la distribution standard de Student à  $\nu$  degrés de liberté où  $\nu > 0$ .

**Remarque 1.4.1** *Les copules elliptiques sont des copules symétriques et relativement simples d'utilisation du fait que l'on connaît bien les distributions auxquelles elles sont associées. Elles sont souvent appelées copules implicites car n'ayant pas de forme analytique explicite et s'expriment par conséquent en fonction de leurs distributions bivariées.*

En comparaison avec la copule Gaussienne, la figure (1.3) montre le graphique de la densité de trois copules de Student de degrés de liberté  $\nu = 5$  qui differt par leur valeur de  $\rho$ .

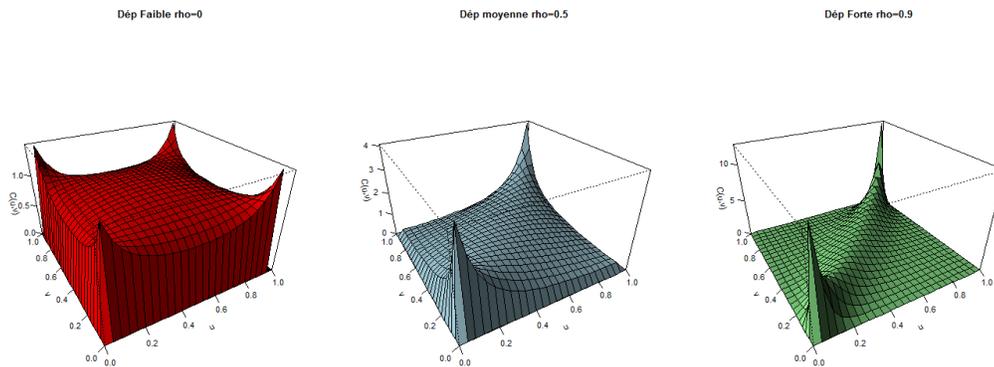


FIG. 1.3 – Densité  $c_{\rho, \nu}$  de trois copules de Student selon  $\rho$  avec  $\nu = 5$ .

Contrairement à la copule Gaussienne, la copule de Student montre des dépendances de queue à droite et à gauche pour  $\rho \neq \pm 1$ . De plus on remarque que l'indépendance n'est pas caractérisée par une valeur nulle de  $\rho$ , car la densité  $c_{\rho,\nu}$  ne vaut pas uniquement 1 sur  $\mathbf{I}^2$ . Cela sera le cas notamment pour de grandes valeurs de degré de liberté  $\nu$ . Ceci est également visible en observant les nuages de points de la figure (1.4) ci-dessous, où pour  $\rho = 0$  on remarque une plus grande concentration de points dans le coin du graphique.

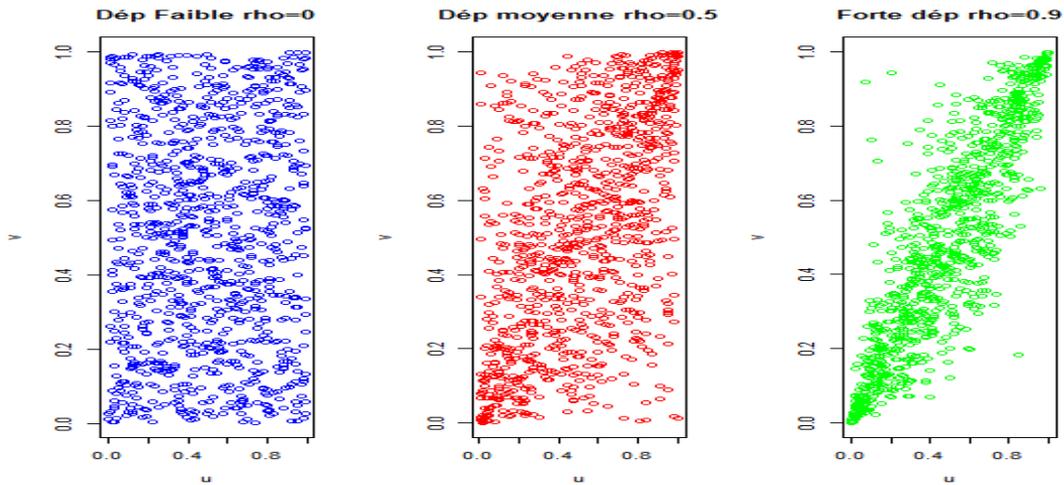


FIG. 1.4 – Nuages de points de trois copules de Student selon  $\rho$  avec  $\nu = 5$ .

### 1.4.3 Copule de Plackett

La copule de Plackett s'écrit de la forme suivante

$$C_{\theta}(u, v) = \frac{[1 + (\theta - 1)(u + v)] - \sqrt{[1 + (\theta - 1)(u + v)]^2 - 4uv\theta(\theta - 1)}}{2(\theta - 1)},$$

avec  $\theta \geq 0$  et  $\theta \neq 1$ , pour  $\theta = 1$  on a  $C_1(u, v) = \Pi$ .

Il s'agit d'une famille de copule assez souple qui donne de bons résultats pour de nombreuses applications. On retrouve les cas particuliers suivants comme cas limites :  $C_0 = W$ ,  $C_1 = \Pi$  et  $C_{\infty} = M$ .

### 1.4.4 Copules Archimédiennes

La classe des copules Archimédiennes, définies par Genest et MacKay [41], joue un rôle très important. D'une part, elles permettent de construire une grande variété de familles de copules, donc de représenter une grande variété de structures de dépendance. D'autres part, les copules ainsi générées ont des formes analytiques fermées et sont faciles à simuler. En effet, contrairement aux copules Gaussiennes et aux copules de Student, les copules Archimédiennes ont le grand avantage de décrire des structures de dépendance très diverses dont notamment les dépendances dites asymétriques, où les coefficients de queue inférieure et de queue supérieure sont différentes. Pour plus d'éléments sur cette famille de copules le lecteur peut se référer à l'excellent ouvrage de Nelsen [66].

Plusieurs raisons justifient l'utilisation de ce type de copules entre autres :

- La grande variété de familles paramétriques.
- Les propriétés particulières et intéressantes que cette classe possède.
- La grande variété des différentes structures de dépendance.
- La facilité avec laquelle elles peuvent être construites et simulées.

Les copules Archimédiennes ont été introduite et développée par le professeur Christian Genest [42] de l'université de Laval au Québec. Un nombre important de familles qui lui appartiennent et qui possèdent des propriétés intéressantes seront mentionnées dans la Table 1.2.

**Définition 1.4.2** *On appelle copule Archimédienne de générateur  $\phi$  la copule définie par*

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), \quad (1.4)$$

avec  $\phi : \mathbf{I} \rightarrow [0, +\infty[$  une fonction continue, strictement décroissante vérifiant  $\phi(1) = 0$ . On définit l'inverse de  $\phi$  par  $\phi^{[-1]}$  tel que

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & \text{si } 0 \leq t \leq \phi(0), \\ 0 & \text{si } \phi(0) \leq t < \infty. \end{cases}$$

Si  $\phi(0) = \infty$ , alors  $\phi^{[-1]} = \phi^{-1}$ .

$\phi$  est au moins deux fois continument dérivable telle que  $\phi'(u) < 0$  et  $\phi''(u) > 0$  pour tout  $u \in \mathbf{I}$ .

**Remarque 1.4.2** Une autre caractérisation des copules Archimédiennes peut se faire à l'aide de la fonction de Kendall,

$$K(t) = \mathbb{P}(C(U, V) \leq t) = t - \frac{\phi(t)}{\phi'(t)}; t \in \mathbf{I}. \quad (1.5)$$

Le tableau (1.2) présente quelques familles Archimédiennes classiques.

Copule	$\phi_\theta(t)$	$\mathbf{C}_\theta(\mathbf{u}, \mathbf{v})$
Clayton	$(t^{-\theta} - 1) / \theta$	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}; \theta \in [-1, 0[ \cup ]0, \infty[$
Gumbel	$(-\ln t)^\theta$	$\exp \left\{ - [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} \right\}; \theta \geq 1$
Frank	$-\ln \left( \frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right)$	$-\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right); \theta \in \mathbb{R}$
Joe	$-\ln (1 - (1 - t)^\theta)$	$1 - [(1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta(1 - v)^\theta]^{1/\theta}; \theta \geq 1$
AMH	$\ln \left( \frac{1 - \theta(1 - t)}{t} \right)$	$uv / (1 - \theta(1 - u)(1 - v)); \theta \in [-1, 1[$

TAB. 1.2 – Générateurs des principales copules Archimédiennes.

- Pour la **Copule de Clayton** on retrouve les cas particuliers suivants comme cas limites :

$$C_{-1} = W, C_0 = \Pi \text{ et } C_\infty = M.$$

- Pour la **Copule de Gumbel Hougard** on retrouve les cas particuliers suivants comme cas limites :  $C_1 = \Pi$  et  $C_\infty = M$ .

- Pour la **Copule de Franck** on retrouve les cas particuliers suivants comme cas limites :

$$C_{-\infty} = W, C_0 = \Pi \text{ et } C_\infty = M.$$

- Pour la **Copule de Joe** on retrouve les cas particuliers suivants comme cas limites :  $C_1 = \Pi$  et  $C_\infty = M$ .

- Pour la **Copule de Ali-Mikhail-Haq (AMH)** on retrouve les cas particuliers suivants comme cas limites :  $C_{-1} = \Pi / (2 - \Sigma + \Pi)$ ,  $C_0 = \Pi$  et  $C_1 = \Pi / (\Sigma - \Pi)$  avec  $\Sigma = u + v$ .

Les figures (1.5) et (1.6) ci-dessous représentent les nuages de points et les densités de trois copules Archimédiennes de paramètres  $\theta$ .

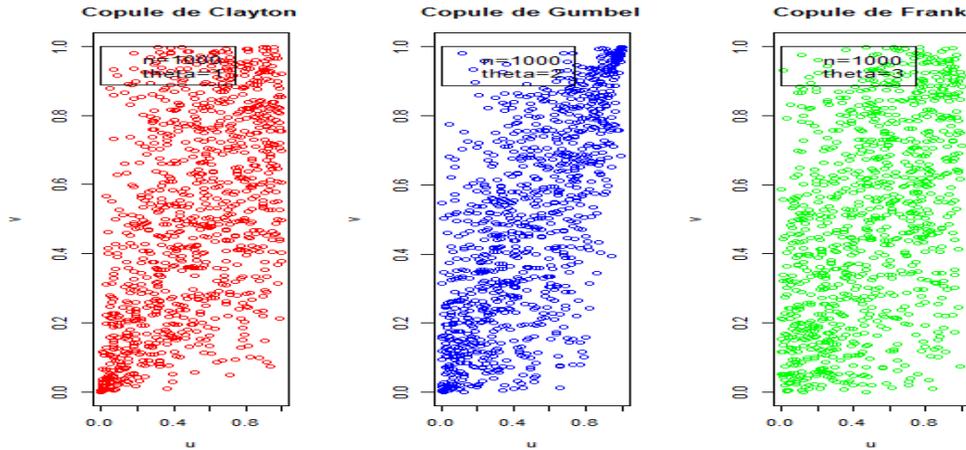


FIG. 1.5 – Nuages de points de trois copules Archimédiennes de paramètres  $\theta$ .

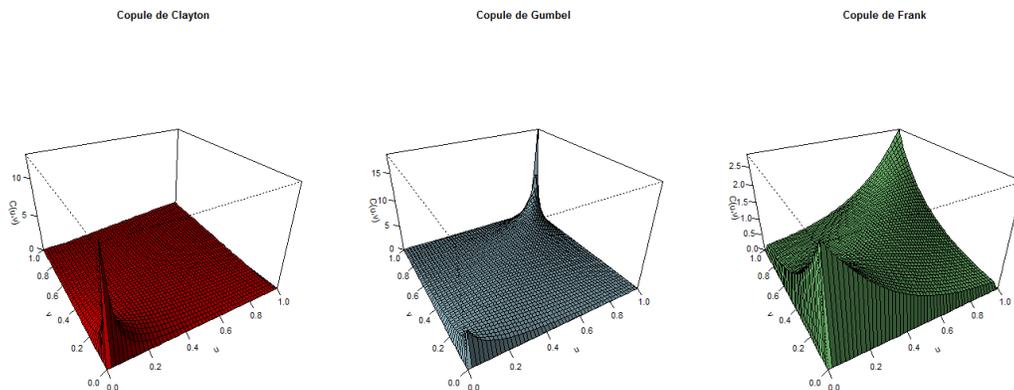


FIG. 1.6 – Densité  $c_\theta(u, v)$  de trois copules Archimédiennes de paramètres  $\theta$ .

D'après les graphiques (1.5) et (1.6) on remarque d'abord que la **copule de Frank** (copule symétrique) représente une plus forte concentration de points aux coins  $(0, 0)$ ,  $(1, 1)$ , comparativement au reste du nuages où l'étalement des points est plus prononcé, ce que corrobore le graphique de la densité ci-dessous.

On remarque que la **copule de Gumbel** a une dépendance de queue à droite (supérieure), tandis que la **copule de Clayton** a une forte dépendance de queue à gauche (inférieure). Ces deux dernières copules auront donc des valeurs de  $\lambda_U$  et  $\lambda_L$  non nulles, respectivement.

### 1.4.5 Copules de valeurs extrêmes

Une autre classe particulière de copules est celle de valeurs extrêmes. Comme leur nom l'indique, les copules de valeurs extrêmes sont issues de la théorie de valeurs extrêmes car il est naturel d'utiliser une loi (les marginales) de valeurs extrêmes généralisée (GEV) lorsqu'on cherche à modéliser la loi de  $M_{X,n} = \max_{1 \leq i \leq n} (X_i)$  et de  $M_{Y,n} = \max_{1 \leq i \leq n} (Y_i)$  avec  $n$  grand.

La distribution de probabilité de la variable  $M_n$  tel que  $M_n = (M_{X,n}, M_{Y,n})$  est la copule de valeurs extrêmes qui vérifie

$$\forall n \in \mathbb{R}^+ : C(u^n, v^n) = C^n(u, v).$$

La copule de Gumbel est un exemple de copule extrême. Contrairement au cas unidimensionnel où la loi GEV est l'unique loi limite possible. Il n'existe pas d'écriture unifiée des lois limites possibles dans le cas bidimensionnel. En pratique, cela signifie que l'utilisateur doit choisir une famille de copules extrêmes à partir des informations dont il dispose.

**Exemple 1.4.1** *On montre que la copule de Gumbel est une copule de valeurs extrêmes*

$$\begin{aligned} C(u^n, v^n) &= \exp \left\{ -[(-\ln u^n)^\theta + (-\ln v^n)^\theta]^{1/\theta} \right\} \\ &= \exp \left\{ -[n^\theta ((-\ln u)^\theta + (-\ln v)^\theta)]^{1/\theta} \right\} \\ &= \exp \left\{ -n [(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} \right\} \\ &= \left( \exp \left\{ -[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta} \right\} \right)^n \\ &= C^n(u, v). \end{aligned}$$

Pour plus de détail sur les copules de valeurs extrêmes se référer à Geoffroy [48], Tiago de Olivera [78] et Sibuya [74].

### 1.4.6 Copules Archimax

On considère une nouvelle famille de copules introduites par Caperaa et al. [13] qui englobent la plupart des familles connues des copules, notamment les copules Archimédiennes et les copules de valeurs extrêmes. Cette nouvelle famille offre plus de flexibilité pour la modélisation.

Une fonction bivariée est une copule Archimax si et seulement si elle s'écrit de la forme suivante

$$C_{\phi,A}(u, v) := \phi^{-1} \left( (\phi(u) + \phi(v)) A \left( \frac{\phi(u)}{\phi(u) + \phi(v)} \right) \right),$$

avec  $A : \mathbf{I} \rightarrow [1/2, 1]$  une fonction de dépendance Archimax tel que

$$\forall t \in \mathbf{I} : \max(t, 1 - t) \leq A(t) \leq 1.$$

**Remarque 1.4.3** En choisissant  $A(t) = 1$  on retrouve la forme générale des copules Archimédiennes :

$$C_{\phi,1}(u, v) = C_{\phi}(u, v) = \phi^{-1}(\phi(u) + \phi(v)).$$

**Remarque 1.4.4** Pour  $\phi(t) = -\ln t$  on obtient la forme générale des copules de valeurs extrêmes. Elles s'expriment en terme de la fonction de dépendance Archimax  $A$  comme ci-dessous

$$C_{\phi,A}(u, v) = C_A(u, v) = \exp \left\{ (\ln u + \ln v) A \left( \frac{\ln u}{\ln u + \ln v} \right) \right\}.$$

**Exemple 1.4.2** Voici un exemple de familles de lois de valeurs extrêmes. On cherche à calculer la fonction de dépendance Archimax  $A$ , associée à la copule de Galambos  $C_{\theta}$ . On a pour  $\theta \geq 1$

$$\begin{aligned} C_{\theta}(u, v) &= uv \exp \left\{ - [(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{-1/\theta} \right\} \\ &= \exp \left\{ [(\ln u + \ln v) - [(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{-1/\theta}] \right\} \\ &= C_A(u, v) \\ &= \exp \left\{ (\ln u + \ln v) A \left( \frac{\ln u}{\ln u + \ln v} \right) \right\}. \end{aligned}$$

On pose :  $(\ln u + \ln v) A \left( \frac{\ln u}{\ln u + \ln v} \right) = (1)$ , alors

$$\begin{aligned}
(1) &= (\ln u + \ln v) - [(-\ln u)^{-\theta} + (-\ln v)^{-\theta}]^{-1/\theta} \\
&= (\ln u + \ln v) \left[ 1 - \left[ \left( \frac{\ln u}{\ln u + \ln v} \right)^{-\theta} + \left( \frac{\ln v}{\ln u + \ln v} \right)^{-\theta} \right]^{-1/\theta} \right].
\end{aligned}$$

On obtient  $A\left(\frac{\ln u}{\ln u + \ln v}\right) = 1 - \left[ \left( \frac{\ln u}{\ln u + \ln v} \right)^{-\theta} + \left( \frac{\ln v}{\ln u + \ln v} \right)^{-\theta} \right]^{-1/\theta}$ .

En posant :  $t = \frac{\ln u}{\ln u + \ln v}$  et  $1 - t = \frac{\ln v}{\ln u + \ln v}$ , on retrouve la fonction de dépendance Archimax

$$A(t) = 1 - \{t^{-\theta} + (1-t)^{-\theta}\}^{-1/\theta}.$$

Le tableau (1.3) présente quelques familles de copules de valeurs extrêmes.

Copule	$\theta$	$\mathbf{C}(\mathbf{u}, \mathbf{v})$	$\mathbf{A}(\mathbf{t})$
$\Pi$	/	$uv$	1
Gumbel A	$[1, \infty[$	$\exp\left\{-\left(\tilde{u}^\theta + \tilde{v}^\theta\right)^{1/\theta}\right\}$	$\left(t^\theta + (1-t)^\theta\right)^{1/\theta}$
Gumbel B	$[0, 1]$	$uv \exp\left\{\theta\left(\tilde{u}\tilde{v}/\tilde{u} + \tilde{v}\right)\right\}$	$\theta t^2 - \theta t + 1$
Galambos	$[0, \infty[$	$uv \exp\left\{\left(\tilde{u}^{-\theta} + \tilde{v}^{-\theta}\right)^{-1/\theta}\right\}$	$1 - \left(t^{-\theta} + (1-t)^{-\theta}\right)^{-1/\theta}$
Logistique de Joe	$[1, \infty]^2$	$\exp\left\{-\left(\tilde{u}^{\theta_1} + \tilde{v}^{\theta_1} - \mathcal{V}(\tilde{u}, \tilde{v}; \theta)\right)^{1/\theta_1}\right\}$	$\left(t^{\theta_1} + (1-t)^{\theta_1} - \zeta(t; \theta)\right)^{1/\theta_1}$
Marshall-Olikn	$[0, 1]^2$	$u^{1-\theta_1} v^{1-\theta_2} \min(u^{\theta_1}, v^{\theta_2})$	$\max(1 - \theta_1 t, \theta_2(1-t))$

TAB. 1.3 – Copules de valeurs extrêmes les plus utilisées.

avec :  $\tilde{u} = -\ln u$ ,  $\tilde{v} = -\ln v$ ,  $\mathcal{V}(\tilde{u}, \tilde{v}; \theta) = \left(\tilde{u}^{-\theta_1\theta_2} + \tilde{v}^{-\theta_1\theta_2}\right)^{-1/\theta_2}$

et  $\zeta(t; \theta) = \left(t^{\theta_1\theta_2} + (1-t)^{-\theta_1\theta_2}\right)^{-1/\theta_2}$ ;  $\theta = (\theta_1, \theta_2)$ .

### 1.4.7 Copules empiriques

La notion de la copule empirique a été introduite par Deheuvels [24] à partir des versions empiriques des fds. On peut construire la copule empirique  $\hat{C}_n$  en fonction des marginales empiriques  $F_n$ ,  $G_n$  et  $H_n$  comme ceci

$$\hat{C}_n(u, v) := H_n(F_n^{-1}(u), G_n^{-1}(v)),$$

où  $F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}$  et  $G_n(y) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq y\}}$ .

Dans le cas d'une distribution bivariee par un echantillon note  $\{(x_i, y_i)\}_{1 \leq i \leq n}$ , celle ci s'ecrit de la fasson suivante

$$H_n(x, y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x, Y_i \leq y\}}. \quad (1.6)$$

La copule empirique bivariee notee  $\hat{C}_n$  est definie sur l'ensemble

$\mathcal{L} = \left\{ \left( \frac{j}{n}, \frac{k}{n} \right); j, k = 1, \dots, n \right\}$ . La copule empirique  $\hat{C}_n$  est

$$\hat{C}_n \left( \frac{j}{n}, \frac{k}{n} \right) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(X_i \leq x_{(j)}, Y_i \leq y_{(k)})}, \quad (1.7)$$

tel que  $x_{(j)}$  et  $y_{(k)}$  representent les statistiques de rang de l'echantillon

$(X_1, X_2, \dots, X_n)$  et  $(Y_1, Y_2, \dots, Y_n)$  respectivement.

La copule empirique peut aussi s'ecrire comme suit

$$\hat{C}_n(u, v) = \begin{cases} \hat{C}_n \left( \frac{j-1}{n}, \frac{k-1}{n} \right) & \text{si } \frac{j-1}{n} \leq u < \frac{j}{n} \text{ et } \frac{k-1}{n} \leq v < \frac{k}{n}, \\ 1 & \text{si } u = 1 \text{ et } v = 1. \end{cases}$$

La fonction de densite empirique de la copule  $\hat{C}_n$  parfois dite copule de frequence empirique notee  $\hat{c}_n$  est donnee par

$$\hat{c}_n \left( \frac{j}{n}, \frac{k}{n} \right) = \begin{cases} \frac{1}{n} & \text{si } (x_{(j)}, y_{(k)}) \in \{(x_i, y_i)\}_{1 \leq i \leq n}, \\ 0 & \text{si non.} \end{cases}$$

Il existe une relation entre  $\hat{C}_n$  et  $\hat{c}_n$  donnée par :  $\hat{C}_n \left( \frac{j}{n}, \frac{k}{n} \right) = \sum_{p=1}^j \sum_{q=1}^k \hat{c}_n \left( \frac{p}{n}, \frac{q}{n} \right)$  et

$$\hat{c}_n \left( \frac{j}{n}, \frac{k}{n} \right) = \hat{C}_n \left( \frac{j}{n}, \frac{k}{n} \right) - \hat{C}_n \left( \frac{j-1}{n}, \frac{k}{n} \right) - \hat{C}_n \left( \frac{j}{n}, \frac{k-1}{n} \right) + \hat{C}_n \left( \frac{j-1}{n}, \frac{k-1}{n} \right).$$

## 1.5 Copules associées à une copule

Dans cette partie on va exprimer les probabilités  $\mathbb{P}(X \leq x \text{ ou } Y \leq y)$ ,  $\mathbb{P}(X > x \text{ ou } Y > y)$ ,  $\mathbb{P}(X > x \text{ ou } Y \leq y)$  et  $\mathbb{P}(X \leq x \text{ ou } Y > y)$  en terme de copules.

**Proposition 1.5.1** *On note  $\bar{F}$  la fonction de survie d'une va  $X$  associée à  $F$  tel que*

$$\forall x \in \mathbb{R} : \bar{F}(x) = \mathbb{P}(X > x) = 1 - F(x).$$

*La fonction de survie  $\bar{H}$  d'un couple  $(X, Y)$  associé à la fd  $H$  est donnée par*

$$\boxed{\bar{H}(x, y) = \mathbb{P}(X > x, Y > y) = 1 - F(x) - G(y) + H(x, y).} \quad (1.8)$$

*Cette dernière peut également s'écrire comme ci dessous :*

$$\begin{aligned} \bar{H}(x, y) &= \mathbb{P}(X > x, Y > y) \\ &= 1 - \mathbb{P}(X \leq x \text{ ou } Y \leq y) \\ &= 1 - [F(x) + G(y) - \mathbb{P}(X \leq x, Y \leq y)] \\ &= 1 - F(x) - G(y) + H(x, y). \end{aligned}$$

Et donc

$$\bar{H}(x, y) = \mathbb{P}(X > x, Y > y) = 1 - F(x) - G(y) + H(x, y).$$

**Remarque 1.5.1** *Existe-t-il une relation analogue reliant les fonctions de survies marginales à leurs fonction de survie jointe ? Pour répondre à cette question on a*

$$\begin{aligned}\bar{H}(x, y) &= 1 - F(x) - G(y) + C(F(x), G(y)) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(F(x), G(y)) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(1 - \bar{F}(x), 1 - \bar{G}(y)).\end{aligned}$$

### 1.5.1 Copule de survie

Soit  $\hat{C}$  une copule de survie associée à la copule  $C$ , elle est définie par

$$\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

On pose  $\bar{C}(u, v) = 1 - u - v + C(u, v)$ .

On peut aussi écrire la distribution de survie en fonction de la copule de survie comme ci dessous

$$\bar{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y)),$$

ou bien

$$\bar{C}(u, v) = \hat{C}(1 - u, 1 - v).$$

### 1.5.2 Copule duale

La fonction  $\tilde{C}$  est dite copule duale associée à la copule  $C$  si elle s'écrit de la façon suivante

$$\tilde{C}(u, v) = u + v - C(u, v),$$

telle que

$$\begin{aligned} \tilde{C}(u, v) &= \mathbb{P}(X \leq x \text{ ou } Y \leq y) \\ &= \mathbb{P}(X \leq x) + \mathbb{P}(Y \leq y) - \mathbb{P}(X \leq x, Y \leq y) \\ &= F(x) + G(y) - C(F(x), G(y)) \\ &= u + v - C(u, v). \end{aligned}$$

**Remarque 1.5.2**  $\tilde{C}$  n'est pas une copule car  $\tilde{C}(1, 0) = 1 \neq 0$ .

### 1.5.3 Co-copule

La co-copule notée par  $\overset{*}{C}$  est exprimée par

$$\overset{*}{C}(u, v) = 1 - C(u, v),$$

telle que

$$\begin{aligned} \overset{*}{C}(u, v) &= \mathbb{P}(X > x \text{ ou } Y > y) \\ &= \mathbb{P}(X > x) + \mathbb{P}(Y > y) - \mathbb{P}(X > x, Y > y) \\ &= 1 - F(x) + 1 - G(y) - [1 - H(x, y)] \\ &= \bar{F}(x) + \bar{G}(y) - \bar{H}(x, y) \\ &= 1 - C(1 - \bar{F}(x), 1 - \bar{G}(y)) \\ &= 1 - C(u, v). \end{aligned}$$

**Remarque 1.5.3**  $\overset{*}{C}$  n'est pas une copule car  $\overset{*}{C}(1, 0) = 1 \neq 0$ .

### 1.5.4 Copule mixte

La copule mixte  $\acute{C}$  associée avec la copule  $C$  est définie par

$$\boxed{\acute{C}(u, v) = 1 - u + C(u, v).}$$

tel que

$$\begin{aligned} \acute{C}(u, v) &= \mathbb{P}(X > x \text{ ou } Y \leq y) \\ &= \bar{F}(x) + G(y) - \mathbb{P}(X > x, Y \leq y) \\ &= \bar{F}(x) + C(F(x), G(y)) \\ &= 1 - u + C(u, v). \end{aligned}$$

On a  $\mathbb{P}(X > x, Y \leq y) = G(y) - C(F(x), G(y))$  car  $\mathbb{P}(\bar{A} \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .

**Remarque 1.5.4** Dans le cas ou  $\mathbb{P}(X \leq x \text{ ou } Y > y)$ , on obtient

$$\acute{C}(u, v) = 1 - v + C(u, v),$$

car  $\mathbb{P}(X \leq x, Y > y) = F(x) - C(F(x), G(y))$ .

## 1.6 Construction de la copule

Les copules permettent de résoudre un autre problème qui est l'élaboration des modèles non gaussiens. En effet il est très difficile de construire des modèles non gaussiens. La famille des distributions gaussiennes est non seulement réduite mais présente l'inconvénient que les marges sont identiques. Or avec les copules on peut construire par exemple une distribution avec une marge gaussienne et une marge uniforme.

On va présenter trois exemples, dans les deux premiers exemples on va construire une copule à partir de marges identiques. Tandis que dans le dernier exemple on va faire le contraire i.e. construire une distribution jointe à partir d'une copule avec des marges différentes.

**Exemple 1.6.1** (*Copule normale*)

Soit  $H_\rho$  la distribution normale jointe, telle que

$$H_\rho(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^x \int_{-\infty}^y \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( \left( \frac{s-\mu_1}{\sigma_1} \right)^2 - \frac{2\rho(s-\mu_1)(t-\mu_2)}{\sigma_1\sigma_2} + \left( \frac{t-\mu_2}{\sigma_2} \right)^2 \right) \right\} ds dt,$$

avec  $-1 \leq \rho \leq +1$ ,  $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$  et  $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ . La copule associée à  $H_\rho$  est alors donnée

par

$$C_\rho(u, v) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{F^{-1}(u)} \int_{-\infty}^{G^{-1}(v)} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( \left( \frac{x-\mu_1}{\sigma_1} \right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left( \frac{y-\mu_2}{\sigma_2} \right)^2 \right) \right\} dx dy.$$

En particulier si  $X = Y \sim \mathcal{N}(0, 1)$  de corrélation  $\rho$  on a la copule de Gauss

$$C_\rho(u, v) := \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{F^{-1}(u)} \int_{-\infty}^{G^{-1}(v)} \exp \left( -\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} \right) dx dy,$$

Comme sa distribution, la copule de Gauss est souvent caractérisée par sa densité  $c_\rho$ . D'après

1.3 on en déduit que

$$\begin{aligned} h_\rho(x, y) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} \right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} \right\} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{y^2}{2} \right\} c_\rho(F_X(x), G_Y(y)). \end{aligned}$$

Par la suite, on a

$$c_\rho(F_X(x), G_Y(y)) = \frac{1}{\sqrt{1-\rho^2}} \exp \left\{ -\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)} + \frac{x^2 + y^2}{2} \right\}.$$

**Exemple 1.6.2** (*Copule de Galambos*)

Soit  $H$  la distribution bivariable suivante

$$\forall x, y \in [0, \infty[ : H(x, y) = \exp \left\{ - \left[ (x + y) - (x^{-\theta} + y^{-\theta})^{-1/\theta} \right] \right\}; \quad \theta \geq 1,$$

avec  $F(x) = \exp \{-x\}$  et  $G(y) = \exp \{-y\}$ .

Leurs fonctions quantiles respectives sont  $F^{-1}(u) = -\ln u$  et  $G^{-1}(v) = -\ln v$ .

On remplace les vas  $x$  et  $y$  par leurs fonctions quantiles  $F^{-1}$  et  $G^{-1}$  respectivement, afin d'obtenir la copule  $C$  associée à  $H$  qui s'appelle la copule de Galambos. Cette dernière est donnée par

$$\boxed{C_{\theta}(u, v) = uv \exp \left\{ - \left[ (-\ln u)^{-\theta} + (-\ln v)^{-\theta} \right]^{-1/\theta} \right\} .}$$

**Exemple 1.6.3** (*Construction d'une distribution jointe*)

Partant maintenant de la copule de Galambos  $C_{\theta}$ , on veut construire la distribution jointe  $H$  ayant une marge normale  $\mathcal{N}(0, 1)$  et une marge uniforme  $\mathcal{U}_{\mathbf{I}}$  respectivement, i.e.

$$\forall x \geq 0 : F(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} \right\} \text{ et } \forall y \in \mathbf{I} : G(y) = y, \text{ alors}$$

$$\begin{aligned} H(x, y) &= C_{\theta}(F(x), G(y)) \\ &= C_{\theta}(F(x), y) \\ &= F(x)y \exp \left\{ - \left[ (-\ln F(x))^{-\theta} + (-\ln y)^{-\theta} \right]^{-1/\theta} \right\} \\ &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} \right\} y \exp \left\{ - \left[ (-\ln F(x))^{-\theta} + (-\ln y)^{-\theta} \right]^{-1/\theta} \right\}. \end{aligned}$$

Donc

$$\boxed{\forall x, y \in [0, \infty[ : H(x, y) = \frac{y}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} - \left[ (-\ln F(x))^{-\theta} + (-\ln y)^{-\theta} \right]^{-1/\theta} \right\} .}$$

## 1.7 Applications des Copules

Il est une industrie en croissance rapide pour les copules. Le commercial logiciel R fournit plusieurs de ces modèles. On peut également obtenir ces derniers dans d'autres logiciels majeurs tels que Mathematica, Matlab, ...etc.

Il existe de nombreuses applications de celles-ci, en particulier la copule Gaussienne, les copules de valeurs extrêmes et les copules Archimédiennes. Elles sont utiles en assurance (Darlap et Mayr [23]), finance (Dobric et Schmid [28]), économie (Purcaru [69]), gestion des risques (Van Der Hoek et Sherris [79]), hydrologie (Wist et al. [82]), environnement (Salvadori et De Michele [70]), sciences de Gestion (Abbas [1]), recherche opérationnelle (Acciolya et Chiyoshi [2]), fiabilité (Chen et Fan [15]), analyse de la survie (Viswanathan et Manatunga [81]), ingénierie (Srinivas et al. [77]), sciences médicales (Escarela et al.) [33], ...etc.

## 1.8 Mise en oeuvre des copules

On a défini les copules, énoncé leurs propriétés fondamentales et décrit les copules les plus utilisées dans la pratique. On va maintenant présenter deux méthodes essentielles aux statisticiens qui utilisent les copules : les méthodes d'inférence statistique et les méthodes de simulation.

Bien que la fonction copule a l'avantage de modéliser complètement la dépendance entre les variables, elle reste inconnue en pratique, d'où le besoin de l'estimer. Dans cette section, on va présenter quelques méthodes d'estimation de copules. Il existe plusieurs approches bien connues pour l'estimation d'une copule. Elles sont divisées en approches paramétriques, semi-paramétriques et non-paramétriques.

Premièrement l'approche paramétrique impose un modèle spécifique pour la copule et les distributions marginales à la fois. Habituellement on utilise par exemple la méthode du maximum de vraisemblance exacte (EML) pour obtenir des estimateurs des paramètres voir Shih et Louis [73], Joe [57] et [58]. On peut estimer conjointement les paramètres des distributions marginales et les paramètres de la copule. L'inconvénient de cette méthode c'est qu'elle a besoin de calculs intensifs et parfois le problème d'optimisation est difficile à résoudre. Joe et Xu [56] ont proposé

une procédure en deux étapes appelée fonction d'inférence pour les marges (IFM). D'abord on estime les paramètres des marginales, ensuite les paramètres de la copule en utilisant la méthode du maximum de vraisemblance. Tout comme la EML, l'inconvénient de cette méthode est qu'elle dépend des hypothèses de la distribution marginale. Cette approche est largement utilisée dans la pratique en raison de sa simplicité, pour plus de détails voir Oakes [67] et Joe [58]. On note que cette approche a été proposée en premier par Genest et al. [43].

Deuxièmement, l'approche semi-paramétrique, suppose un modèle paramétrique pour la copule et non-paramétrique pour les distributions marginales. Par exemple la méthode CML (canonical maximum likelihood) avec les distributions marginales estimées par la distribution empirique, et la méthode du maximum de vraisemblance utilisée pour estimer le vecteur des paramètres de la copule paramétrique. Cette approche a été étudiée par Oakes [68], Genest et Rivest [42] et Genest et al. [43]. Récemment, Kim et al. [61] ont fait une étude comparative entre les méthodes semi-paramétrique et paramétrique pour l'estimation des copules.

Finalement, l'approche non-paramétrique tient compte à la fois des modèles non-paramétriques pour la copule et pour les distributions marginales. Deheuvels [24] a suggéré la distribution multidimensionnelle empirique pour estimer la fonction copule. Gijbels et Mielniczuk [49] ont estimé une copule bidimensionnelle en utilisant une méthode basée sur le lissage à noyau. En outre ils ont suggéré la méthode de réflexion afin de résoudre le problème du biais aux frontières. Plus récemment Chen et Huang [16] ont proposé un estimateur bidimensionnel basé sur l'estimateur local linéaire. Pour les données i.i.d., Sancetta et Satchell [71] ont développé un estimateur basé sur les polynômes de Bernstein de la fonction copule appelé par la suite estimateur de Bernstein.

### 1.8.1 Inférence statistique

Aujourd'hui, les statisticiens utilisent les copules pour modéliser la dépendance entre des données. C'est pourquoi les études sur l'estimation des paramètres des copules se développent sans cesse et font l'objet de nombreux articles. Lorsqu'on observe plusieurs phénomènes aléatoires dont on veut étudier les liens de dépendance à partir d'une copule, on a le choix parmi plusieurs copules possibles. Il s'agit alors de voir quel type de copule est le mieux adapté, puis, une fois le

type de copule choisi, il s'agit d'estimer le (ou les paramètres) qui caractérise la copule retenue. Pour estimer le(s) paramètre(s) il existe différentes méthodes. Les méthodes d'inférence statistique utilisées sont des déclinaisons de méthodes standards paramétriques, semi paramétriques et non paramétriques selon le cas.

Dans le cas où on cherche à modéliser la loi jointe de  $(X, Y)$ . Les étapes de la modélisation consistent à :

1. Choisir des formes paramétriques pour les lois marginales et la copule.
2. Estimer les paramètres.
3. Valider le modèle ajusté.

### 1. Choix de la bonne copule

Au vu de toute la théorie exposée précédemment, nous sommes en droit de nous demander quelle copule pourrait le mieux correspondre à une certaine série de données? Deux copules ayant le même coefficient de corrélation (e.g. tau de Kendall) peuvent en effet avoir des comportements tout à fait différents.

On va donc dans cette partie donner quelques méthodes nous permettant de faire le tri parmi les copules afin de ne choisir que celles qui pourraient présenter des caractéristiques semblables à celles de notre série de données.

En fait, on va présenter plusieurs fonctions, dont certaines ont été exposées par Venter [80] et qui auront des caractéristiques tout à fait différentes selon la copule choisie.

Ces fonctions peuvent également être établies de façon empirique, uniquement à partir du rang de chaque observation. Une simple comparaison graphique nous permet de ne retenir qu'une ou deux familles de copules pour la poursuite de notre étude.

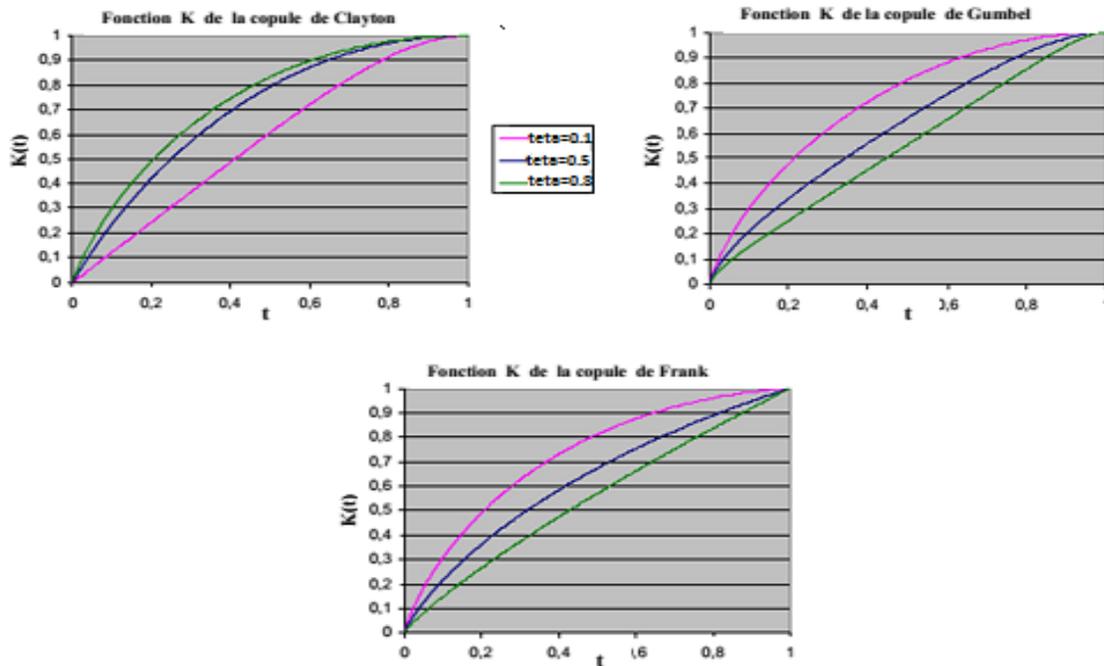
- **Fonction  $K(t)$**  : Cette fonction qui a déjà été évoquée dans notre présentation de la copule Archimédienne, est la fonction de Kendall. Il a été démontré que pour une copule de type Archimédienne, cette fonction se définit comme dans 1.5.

Dans le cadre des copules Archimédiennes présentées auparavant, la fonction  $K$  est la suivante

Copule	$K(t)$
Clayton	$t(1 + \theta(1 - t^{1/\theta}))$
Gumbel	$t(1 - \frac{\ln t}{\theta})$
Frank	$t + \frac{1}{\theta} \ln \left( \frac{1 - \exp(-\theta t)}{1 - \exp(-\theta)} \right)$

TAB. 1.4 – Fonction de Kendall de trois copules Archimédiennes.

Les graphiques de la fonction  $K(t)$  pour différentes copules et différentes valeurs de  $\theta$  sont présentées dans la figure (1.7).

FIG. 1.7 – Fonction  $K(t)$  pour différentes copules et différentes valeurs de  $\theta$  [5].

On suppose maintenant qu'on dispose d'un échantillon de  $n$  observations  $\{(x_i, y_i)\}_{1 \leq i \leq n}$  issu d'un vecteur aléatoire  $(X, Y)$ . Pour établir un estimateur non-paramétrique de la fonction  $K$  (qui est ni plus ni moins que la fd de la va  $C(u, v)$ ) à partir de cet échantillon, la procédure à suivre consiste à :

a. Définir la pseudo-observation  $t_i$  pour chaque  $i = 1 \dots n$  tel que

$$t_i = \frac{\text{nombre de paires } (x_j, y_j) \text{ telles que } x_j \leq x_i \text{ et } y_j \leq y_i}{n-1} = \frac{1}{n-1} \sum_{j=1}^n \mathbf{1}_{\{x_j \leq x_i, y_j \leq y_i\}},$$

où  $x_j$  et  $y_j$  représentent les statistiques d'ordre associées à l'échantillon.

b. Définir l'estimateur non-paramétrique de  $K$  comme suit

$$K_n(t) = \frac{\text{nombre de } t_i \leq t}{n} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{t_i \leq t\}}.$$

Cette estimation non-paramétrique de  $K$  pourra ensuite être comparée graphiquement aux versions paramétriques de  $K$  pour les différentes copules Archimédiennes.

Le paramètre empirique  $\hat{\theta}$  de la copule pourra être établi, par exemple à partir du coefficient de Kendall empirique de l'échantillon. En effet, on va voir dans le second chapitre qu'il existe une relation directe entre les mesures de concordance et le paramètre de la copule. Il suffit donc simplement de résoudre une équation pour déterminer le paramètre empirique de la copule, même si cette opération peut se révéler un peu plus délicate comme le cas de la copule de Frank par exemple.

- **Fonction  $J(t)$  ou tau cumulatif :** Cette fonction est basée sur le coefficient du tau de Kendall  $\tau$  2.3. L'idée consiste à définir une fonction de tau cumulatif en prenant :

$$J(t) = \frac{4}{C^2(t, t)} \int_0^t \int_0^t C(u, v) c(u, v) du dv - 1, \quad J(1) = \tau.$$

Les graphiques de la fonction  $J(t)$  pour différentes copules et différentes valeurs de  $\theta$  sont représentés dans la figure (1.8).

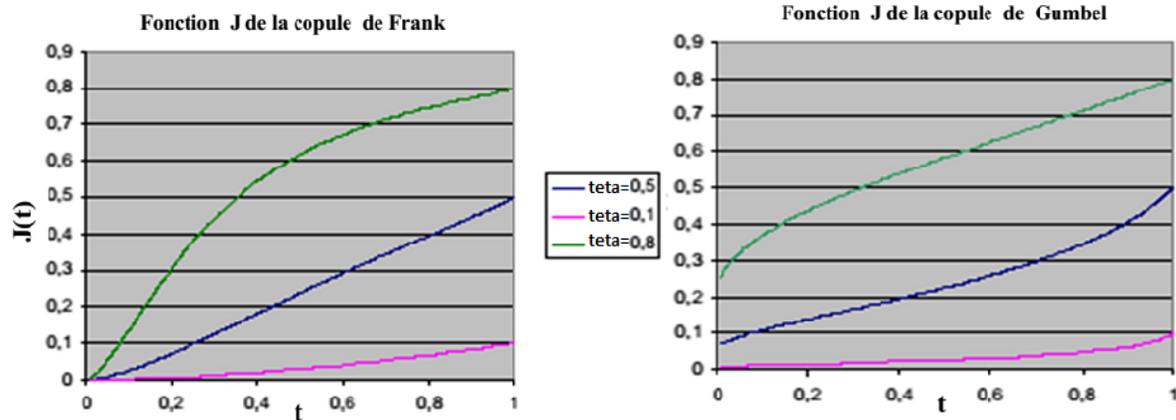


FIG. 1.8 – Fonction  $J(t)$  pour différentes copules et différentes valeurs de  $\theta$  [5].

On constate des différences notables dans le comportement de la fonction  $J$  pour les différentes copules. La fonction  $J$  peut démarrer avec une faible corrélation (e.g. copule de Frank) ou une corrélation plus importante (e.g. copule de Gumbel). On a pas représenté la copule de Clayton dans la mesure où  $J$  est représentée très rapidement par une ligne horizontale dont la valeur est bien entendu égale à  $\tau$ . Ceci s'explique aisément dans la mesure où la copule de Clayton possède un coefficient de "lower tail dependence" et donc la dépendance est surtout concentrée près du point  $(0, 0)$ .

L'intérêt de cette fonction réside là encore dans le fait qu'une version empirique de celle-ci puisse être établie. On va, comme pour la fonction  $K$ , donner la procédure pour établir la fonction  $J$  de manière empirique i.e.  $J_n(t)$ . Considérons le même échantillon d'observations  $\{(x_i, y_i)\}_{1 \leq i \leq n}$ . On note par  $x_{(i)}$  (respectivement  $y_{(i)}$ ) le rang de la  $i^{\text{ème}}$  observation dans l'ensemble des valeurs  $x_1, \dots, x_n$  (respectivement  $y_1, \dots, y_n$ ). La procédure à suivre consiste à :

- Transformer les données  $(x_i, y_i)$  en  $(u_i, v_i)$  en prenant :  $(u_i, v_i) = \left( \frac{x_{(i)}}{n+1}, \frac{y_{(i)}}{n+1} \right)$ .
- Définir  $C(t, t)$  en prenant pour chaque  $j = 1 \dots n$  :

$$C(t, t) = \frac{\text{nombre de paires } (u_j, v_j) \text{ telles que } u_j \leq t \text{ et } v_j \leq t}{n}.$$

Ce qui revient à prendre la fd empirique des données transformées en étape "a".

- Définir la pseudo-observation  $t_i$  pour chaque  $i = 1 \dots n$  comme précédemment

$$t_i = \frac{\text{nombre de paires } (x_j, y_j) \text{ telles que } x_j \leq x_i \text{ et } y_j \leq y_i}{n-1},$$

où  $x_j$  et  $y_j$  représentent les statistiques d'ordre associées à l'échantillon.

d. Définir ensuite la quantité :  $I(t) = \frac{1}{n} \sum_{i=1}^n t_i \times \mathbf{1}_{\{u_i \leq t, v_i \leq t\}}$ .

e. Enfin, définir  $J_n(t)$  par

$$J_n(t) = -1 + \frac{4I(t)}{C^2(t, t)}.$$

Il ne reste plus qu'à comparer la version empirique ainsi obtenue à la version théorique de  $J(t)$  pour différentes copules.

- **Fonction  $M(t)$**  : On pose

$$M(t) = \mathbb{E}(V/U < t) = \frac{1}{t} \int_0^t \int_0^1 vc(u, v) du dv.$$

Comme  $\mathbb{E}[V] = 1/2$ , quelle que soit la copule choisie, la fonction  $M$  vérifiera  $M(1) = 1/2$ .

Les graphiques de la fonction  $M(t)$  pour différentes copules et différentes valeurs de  $\theta$  sont présentées dans la figure (1.9).

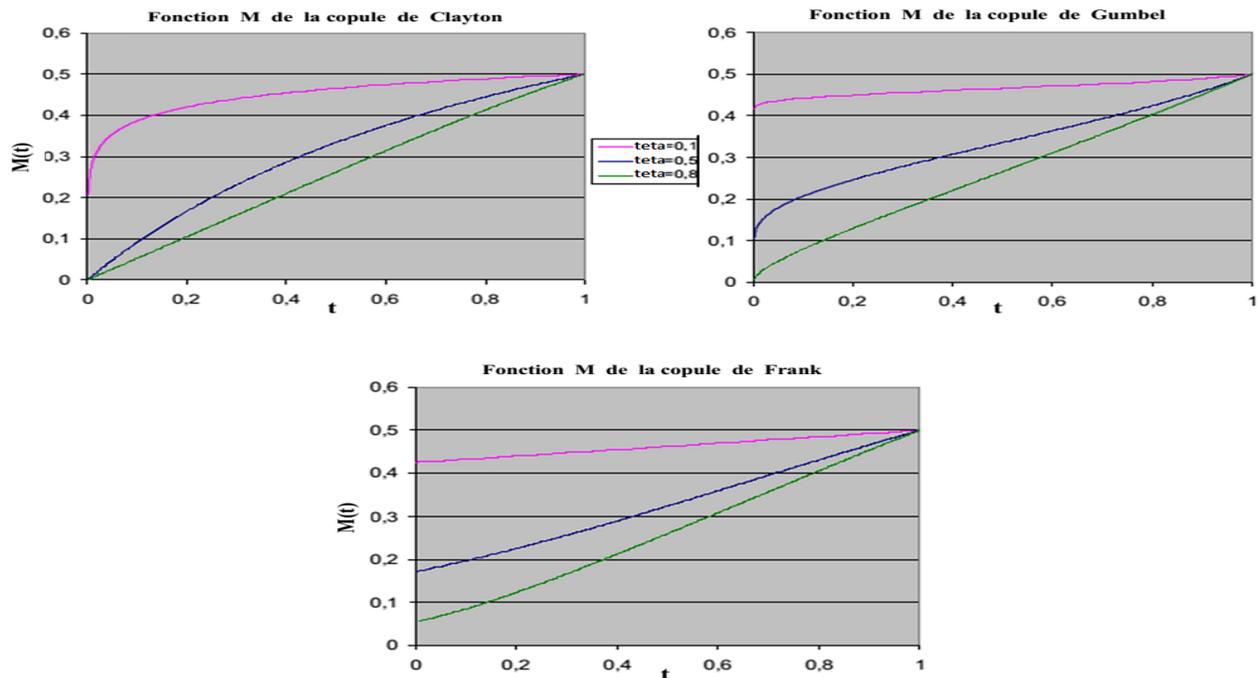


FIG. 1.9 – Fonction  $M(t)$  pour différentes copules et différentes valeurs de  $\theta$  [5].

La différence entre les copules se fera donc au niveau des petites valeurs de  $t$  et à l'allure de la courbe au fur et à mesure que l'on approche du point  $t = 1$ .

La procédure à suivre pour calculer cette fonction à partir des données  $\{(x_i, y_i)\}_{1 \leq i \leq n}$  consiste à :

a. Définir  $D_n(t)$  en prenant : 
$$D_n(t) = \sum_{i=1}^n \mathbf{1}_{\{u_i \leq t\}}.$$

b. Définir  $N_n(t)$  suivant la formule : 
$$N_n(t) = \sum_{i=1}^n v_i \times \mathbf{1}_{\{u_i \leq t\}}.$$

c. Soit  $M_n(t) = N_n(t)/D_n(t)$ .

Il est évident que  $M_n(1) = 1/2$  puisque :

$$\begin{cases} N_n(1) = \sum_{i=1}^n v_i = \frac{1}{n+1} \sum_{i=1}^n i = \frac{1}{n+1} \frac{n(n+1)}{2} = \frac{n}{2}, \\ D_n(1) = n. \end{cases}$$

- **Fonctions  $L(t)$  et  $R(t)$  ou "tail concentration functions"** : On s'est déjà penché précédemment sur les concepts de "upper tail" et "lower tail" dépendance. Les fonctions qu'on va présenter maintenant sont totalement basées sur ces concepts.

Pour tout  $t$  appartenant à  $\mathbf{I}$ , on définit

$$L(t) = \mathbb{P}(U < t, V < t)/t^2 \text{ et } R(t) = \mathbb{P}(U > t, V > t)/(1-t)^2.$$

Ces fonctions mettent donc en valeur la "concentration de probabilité" aux abords des points  $(0,0)$  et  $(1,1)$ . Celles-ci peuvent donc se traduire très facilement en termes de copules, puisque

$$L(t) = C(t,t)/t^2 \text{ et } R(t) = (1 - 2t + C(t,t))/(1-t)^2.$$

On a expliqué précédemment (lors de la présentation de la fonction  $J$ ) comment calculer  $C(t,t)$  à partir des données. Etablir la version non-paramétrique de ces fonctions ne pose donc a priori aucun problème.

On sait que  $R(0) = L(1) = 1$ , donc ces fonctions peuvent très facilement être différenciées l'une de l'autre.

Les graphiques des fonction  $L(t)$  et  $R(t)$  pour différentes copules et différentes valeurs de  $\theta$  sont présentées dans la figure (1.10).

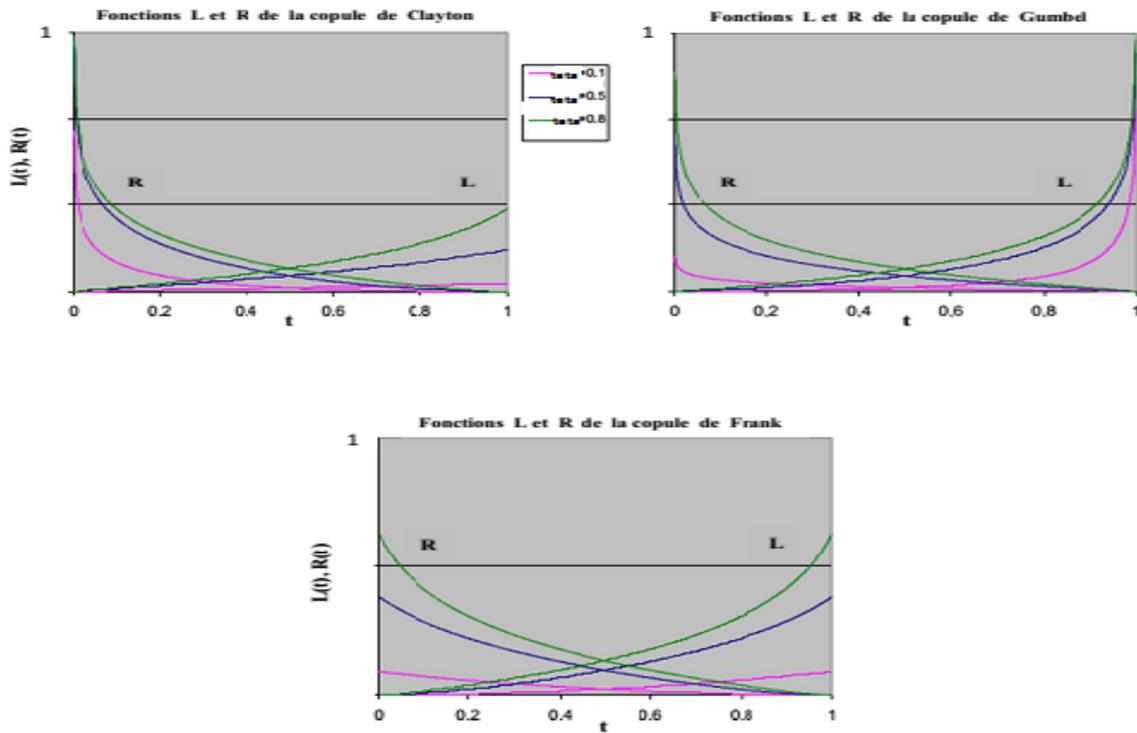


FIG. 1.10 – Fonctions  $L(t)$  et  $R(t)$  pour différentes copules et différentes valeurs de  $\theta$  [5].

Une constatation s'impose d'emblée au vu de ces graphiques : plus  $\theta$  est important, plus les valeurs de ces fonctions sont élevées aux abords des points dits intéressants (à savoir 0 pour  $R(t)$  et 1 pour  $L(t)$ )

- La copule de Clayton présente une concentration à gauche, mais à droite est quasiment nulle.
- La copule de Gumbel présente une forte concentration dans les deux queues, même si la droite est de loin la plus épaisse. Au fur et à mesure que la corrélation diminue, l'importance relative de la queue à gauche décroît par rapport à celle de droite.
- La copule de Frank est, quant à elle, totalement symétrique.

Nous venons donc de passer en revue les différentes fonctions.

## 2. Estimation des paramètres de la copule

Lorsqu'il s'agit d'une estimation paramétrique d'une copule, on impose un modèle paramétrique pour la copule et pour les distributions marginales. Dans ce qui suit, on présente les principales méthodes d'estimation des

### Estimation paramétrique

- **Maximum de vraisemblance exacte (EML)** : Si la copule appartient à une famille paramétrique, il est fréquent d'utiliser la méthode EML pour obtenir l'estimation des paramètres.

On suppose qu'on a un échantillon  $\{(x_i, y_i)\}_{1 \leq i \leq n}$  généré à partir d'une distribution  $H$ . Les fds marginales sont  $F(x; \beta_1)$  et  $G(y; \beta_2)$ , où  $\beta_1$  (respectivement  $\beta_2$ ) est un vecteur des paramètres. On suppose aussi que les fds marginales existent et seront notées  $f(x; \beta_1)$  et  $g(y; \beta_2)$ , et que la fonction copule  $C$  appartient à une famille paramétrique sera notée  $C(u, v; \theta)$  tel que  $\theta$  est un vecteur des paramètres. Le vecteur des paramètres à estimer est  $\vartheta = (\beta_1, \beta_2, \theta)^t$ . A partir du théorème de Sklar 1.1 la distribution jointe peut être écrite comme suit :

$$H(x, y; \vartheta) := C(F(x; \beta_1), G(y; \beta_2); \theta).$$

À partir de 1.2, sous des conditions de continuité, la densité jointe  $h$  du vecteur aléatoire  $(X, Y)$  qui s'écrit en fonction de la densité de sa copule  $c$  et de ses marginales  $f$  et  $g$  est donnée par

$$h(x, y; \vartheta) := c(F(x; \beta_1), G(y; \beta_2); \theta) f(x; \beta_1) g(y; \beta_2).$$

La fonction log-vraisemblance s'exprime alors

$$L(\vartheta) := \sum_{i=1}^n \log h(x_i, y_i; \vartheta) = \sum_{i=1}^n \log c(F(x_i; \beta_1), G(y_i; \beta_2); \theta) + \sum_{i=1}^n [\log f(x_i; \beta_1) + \log g(y_i; \beta_2)],$$

et peut se décomposer en deux termes : celui de gauche est associé à la structure de dépendance, et le second est associé aux lois marginales. Notons que ce second terme est le seul qui apparaît

si l'on suppose que les composantes du vecteur  $(X, Y)$  sont indépendantes.

On peut estimer les paramètres  $\vartheta$  par maximisation de la fonction log-vraisemblance par rapport à  $\vartheta$ . On obtient l'estimateur EML de  $\vartheta$ , noté  $\hat{\vartheta}_{EMV}$  donné par

$$\hat{\vartheta}_{EML} = \arg \max_{\vartheta \in \Theta} L(\vartheta), \forall \vartheta \in \Theta,$$

où  $\Theta$  est l'espace des paramètres. Sous des conditions de régularité, on peut montrer que l'estimateur  $\hat{\vartheta}_{EML}$  existe, qu'il est convergent, asymptotiquement efficace et vérifie la propriété de la normalité asymptotique :

$$\sqrt{n} \left( \hat{\vartheta}_{EML} - \vartheta \right) \xrightarrow{L} \mathcal{N}(0, I^{-1}(\vartheta)), \text{ avec } I(\vartheta) \text{ l'information de Fisher }^2.$$

Le problème avec cette méthode est qu'elle peut engendrer des temps de calcul très longs dans le cas d'une grande dimension, car elle nécessite d'estimer conjointement les paramètres des lois marginales et les paramètres de la structure de dépendance. De plus, l'estimation de la copule est sensible à une éventuelle erreur d'estimation des marginales car elles interviennent directement dans le calcul de la log-vraisemblance.

**- Fonctions d'inférence des marginales (IFM) :** Cette méthode estime séparément les paramètres des distributions marginales à ceux de la fonction copule, donc elle exige des calculs intensifs parce que l'estimation des paramètres des lois marginales et des paramètres de la copule se fait d'une façon simultanée. Joe et Xu [56] ont proposé la méthode IFM et ont estimé les paramètres séparément. Cette méthode repose sur le fait que la représentation en copule permet de séparer les paramètres spécifiques des distributions marginales  $(\beta_1, \beta_2)$  de ceux de la structure de dépendance  $\theta$ , cette méthode se fait en deux étapes :

a. Estimation des paramètres des marginales

- On estime les paramètres  $\beta_1$  et  $\beta_2$  des distributions marginales, en maximisant la fonction

---

<sup>2</sup>Information de Fisher :  $I(\vartheta) = \mathbb{E} \left[ \left( \frac{d}{d\vartheta} L(\vartheta) \right)^2 \right]$ .

log-vraisemblance

$$L(\beta_1) = \sum_{i=1}^n \log f(x_i; \beta_1) \text{ et } L(\beta_2) = \sum_{i=1}^n \log g(y_i; \beta_2).$$

- Les estimateurs  $\hat{\beta}_1$  et  $\hat{\beta}_2$  des paramètres  $\beta_1$  et  $\beta_2$  sont donnés par

$$\hat{\beta}_1 = \arg \max_{\beta_1 \in \Theta} L(\beta_1) \text{ et } \hat{\beta}_2 = \arg \max_{\beta_2 \in \Theta} L(\beta_2).$$

b. Estimation du vecteur des paramètres de la copule

Compte tenu des estimateurs de l'étape précédente  $\hat{\beta}_1$  et  $\hat{\beta}_2$ . La fonction log-vraisemblance est définie par

$$L(\theta) = \sum_{i=1}^n \log c \left( F(x_i; \hat{\beta}_1), G(y_i; \hat{\beta}_2); \theta \right).$$

L'estimateur de  $\theta$  noté par  $\hat{\theta}$  est donné par

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta).$$

Cette méthode d'estimation est appelée par Shih et Louis [73] "two-stage parametric maximum likelihood method", alors que Joe et Xu [56] utilisent le terme "Inference Functions for Margins".

L'estimateur IFM est défini par

$$\hat{\vartheta}_{IFM} = \left( \hat{\beta}_1, \hat{\beta}_2, \hat{\theta} \right)^t.$$

Joe [57] a montré que sous des conditions de régularité, l'estimateur IFM vérifie la propriété de la normalité asymptotique :

$\sqrt{n} \left( \hat{\vartheta}_{IFM} - \vartheta \right) \xrightarrow{L} \mathcal{N}(0, \zeta^{-1}(\vartheta))$ , où  $\zeta(\vartheta)$  est la matrice de l'information de Godambe <sup>3</sup>.

---

<sup>3</sup>La matrice de Godambe :  $\zeta(\vartheta) = D^{-1}M(D^{-1})^t$ , où  $D = \mathbb{E} \left[ \frac{d}{d\vartheta} s(\vartheta)^t \right]$ ,  $M = \mathbb{E} [s(\vartheta)^t s(\vartheta)]$  et

$s(\vartheta) = \left( \frac{dL(\beta_1)}{d\beta_1}, \frac{dL(\beta_2)}{d\beta_2} \right)$ .

Cette méthode pourrait présenter l'avantage de reposer sur des calculs plus légers que ceux générés par la méthode EML, mais la détermination de la matrice de Godambe peut s'avérer très complexe en raison des multiples calculs des dérivées. Là encore, la méthode est sensible à une éventuelle erreur de spécification des marginales pour la même raison que celle évoquée dans la méthode précédente.

- **Méthode des moments :** Cette méthode revient à se donner une mesure de concordance  $\kappa_C$  et à considérer que la valeur du paramètre  $\theta$  de la copule  $C$  est celle qui égalise la valeur théorique  $\kappa_C$  à la valeur estimée  $\hat{\kappa}_C$ . Elle n'assure aucune robustesse de l'estimateur. Dans la pratique, le tau de Kendall est le plus souvent utilisé comme mesure de concordance étant donné la simplicité de l'estimation de ce tau. Cette méthode consiste à estimer séparément les paramètres  $\beta_1$  et  $\beta_2$  des lois marginales à celui de la copule i.e.  $\theta$ , où ce dernier ce calcul en inversant la mesure de concordance. Pour cela on doit suivre les étapes suivantes :

a. Estimer les paramètres des marginales

Soit  $X$  une va qui suit une loi de probabilité  $F$  a un paramètre (qui peut être un ensemble de plusieurs nombres réels,  $\beta_1 = (b_1, \dots, b_p)$ ). Il est possible de calculer en fonction de  $\beta_1$  les  $p$  premiers moments de  $X$ , notés  $\mu_{1,n}, \mu_{2,n}, \dots, \mu_{p,n}$ . Ces  $p$  équations permettent de calculer l'estimateur de  $\beta_1$  :

$$\left\{ \begin{array}{l} \mu_{1,n} := \mathbb{E}(X) = f_1(b_1, \dots, b_p) \\ \mu_{2,n} := \mathbb{E}(X^2) = f_2(b_1, \dots, b_p) \\ : \\ \mu_{p,n} := \mathbb{E}(X^p) = f_p(b_1, \dots, b_p) \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} b_1 = \psi_1(\mu_{1,n}, \mu_{2,n}, \dots, \mu_{p,n}) \\ b_2 = \psi_2(\mu_{1,n}, \mu_{2,n}, \dots, \mu_{p,n}) \\ : \\ b_p = \psi_p(\mu_{1,n}, \mu_{2,n}, \dots, \mu_{p,n}) \end{array} \right.$$

L'estimateur de  $\beta_1$  est obtenu en remplaçant dans ces expressions les moments théoriques par les moments empiriques de  $X$  calculés à partir de l'échantillon de taille  $n$  :

$$\left\{ \begin{array}{l} \hat{\mu}_{1,n} = \frac{1}{n} \sum_{i=1}^n X_i \\ \hat{\mu}_{2,n} = \frac{1}{n} \sum_{i=1}^n X_i^2 \\ \vdots \\ \hat{\mu}_{p,n} = \frac{1}{n} \sum_{i=1}^n X_i^p \end{array} \right. \implies \left\{ \begin{array}{l} \hat{b}_1 = \psi_1(\hat{\mu}_{1,n}, \hat{\mu}_{2,n}, \dots, \hat{\mu}_{p,n}) \\ \hat{b}_2 = \psi_2(\hat{\mu}_{1,n}, \hat{\mu}_{2,n}, \dots, \hat{\mu}_{p,n}) \\ \vdots \\ \hat{b}_p = \psi_p(\hat{\mu}_{1,n}, \hat{\mu}_{2,n}, \dots, \hat{\mu}_{p,n}) \end{array} \right.$$

Si les fonctions  $\psi_i$  sont continues en  $(\mu_{1,n}, \mu_{2,n}, \dots, \mu_{p,n})$ , alors les estimateurs obtenus sont des estimateurs convergents.

Répéter cette étape pour la deuxième marginale.

b. Inverser la mesure d'association pour obtenir le paramètre  $\theta$  de la copule

Supposant qu'il existe une relation entre la mesure d'association et le paramètre  $\theta$ . Cette relation est définie par

$$\kappa_C = J(\theta),$$

où  $J$  est une fonction continue et dérivable, alors l'estimateur par la méthode des moments  $\hat{\theta}_M$  de  $\theta$  est défini par

$$\hat{\theta}_M = J^{-1}(\hat{\kappa}_C),$$

telle que  $\hat{\kappa}_C$  est l'estimateur empirique (non paramétrique) de  $\kappa_C$ .

**Exemple 1.8.1** (*Copule de Clayton*)

D'après (2.1) le tau de Kendall de la copule de Clayton est  $\theta / (\theta + 2)$  avec  $\theta \in [-1, 0[ \cup ]0, \infty[$ .

L'estimateur par la méthode des moments est

$$\hat{\theta}_M = \frac{2\hat{\tau}}{1 - \hat{\tau}}.$$

**Exemple 1.8.2** (*Copule de Gumbel*)

De la même manière on peut trouver l'estimateur de  $\theta$  de la copule de Gumbel tel que son tau de Kendall est égale à  $(\theta - 1) / \theta$  avec  $\theta \geq 1$ , ayant des marginales exponentielles de paramètre  $\beta$ , alors

$$\hat{\beta} = \frac{1}{\bar{X}_n} \text{ et } \hat{\theta}_M = \frac{1}{1 - \hat{\tau}},$$

où  $\hat{\beta}$  est l'estimateur de  $\beta$ .

Pour plus de détail sur l'estimation des paramètres de la copule par la méthode des moments voir Brahim et Necir [11].

**Estimation semi-paramétrique**

L'approche semi-paramétrique suppose un modèle paramétrique pour la copule  $C$  et non-paramétrique pour les distributions marginales. La méthode d'estimation semi-paramétrique qu'on va étudier ici est aussi connue sous le nom de pseudo maximum de vraisemblance (PML) ou encore de maximum de vraisemblance canonique (CML).

- **Maximum de vraisemblance canonique (CML)** : Dans la méthode IFM, l'estimation du paramètre de la copule dépend de l'estimation des paramètres des distributions marginales. Genest et al. [43] ont proposé la méthode CML, qui ne fait aucune hypothèse sur les distributions marginales. En utilisant les fds empiriques, on construit un nouveau échantillon :

$$F_n(x_i) := \frac{1}{n+1} \sum_{j=1}^n \mathbf{1}(X_j \leq x_i) \text{ et } G_n(y_i) := \frac{1}{n+1} \sum_{j=1}^n \mathbf{1}(Y_j \leq y_i).$$

Dans les équations ci-dessus, on utilise  $n+1$ , au lieu de  $n$ , pour éviter le problème d'absence de limites de la fonction de densité de la copule. Après la transformation, l'estimation du paramètre de la copule paramétrique est donnée par

$$\hat{\theta}_{CML} = \arg \max_{\theta \in \Theta} L(\theta),$$

tel que

$$L(\theta) = \sum_{i=1}^n \log c(F_n(x_i), G_n(y_i); \theta).$$

Les temps de calculs sont raisonnables et le paramètre de la copule est estimé indépendamment de la spécification des lois marginales.

Kim et al. [61] ont comparé les méthodes paramétriques et semi-paramétriques pour l'estimation des copules, leur étude a conclu que :

- Les méthodes IFM et EML ne sont pas robustes contre les erreurs de spécification des distributions marginales.
- La méthode CML est conceptuellement presque la même que l'IFM, mais elle surmonte sa non robustesse contre une mauvaise spécification des distributions marginales.
- En termes de calculs statistiques et d'analyse des données, la CML est aussi facile à implémenter que la méthode IFM.
- Un avantage de plus pour la CML sur l'IFM, est que la CML n'impose pas de modèles paramétriques pour les distributions marginales.
- Leurs résultats de simulation montrent que l'estimateur de CML est meilleur que ceux de IFM et EML dans la plupart des situations pratiques.

- **L-moments bivariés (BLM) :** La méthode BLM est un outil simple d'utilisation pour l'estimation des paramètres de la copule. On peut donner des formes explicites des L-moments en fonction de la copule.

Le  $k^{\text{ième}}$  BLM d'une copule  $C$  s'écrit comme suit

$$\forall k \geq 1 : \delta_k = \int_{\mathbf{I}^2} (C(u, v) - uv) dudP_k(v).$$

On peut aussi l'écrire de la manière suivante

$$\delta_k = \int_{\mathbf{I}^2} uP_k(v)dC(u, v),$$

où  $P_k(v)$  est le polynôme de Legendre déplacé (PLD), tel que

$$P_k(v) := \sum_{l=0}^k p_{k,l} v^l,$$

où  $p_{k,l} = [(-1)^{k+l} (k+l)!] / [(l!)^2 (k-l)!]$ .

Les trois premiers PLD sont :  $P_0(v) = 1$ ,  $P_1(v) = 2v - 1$ ,  $P_2(v) = 6v^2 - 6v + 1$ .

**Exemple 1.8.3** *Estimer les deux paramètres de chacune des deux copules Marshall-Olkin (MO) et Farlie-Gumbel-Morgenstern (FGM). Pour cela on suit les étapes suivante :*

a. *Calculer les deux premiers BLM de chaque copule.*

$$\text{- Copule MO : } \begin{cases} \delta_1 = \frac{ab}{4a + 4b - 2ab}, \\ \delta_2 = \frac{ab}{6a + 4b - 2ab} - \frac{3a^2b}{2(2a + 2b - ab)(3a + 2b - ab)}. \end{cases}$$

$$\text{- Copule FGM : } \begin{cases} \delta_1 = \frac{a}{18} + \frac{b}{72}, \\ \delta_2 = \frac{b}{120}. \end{cases}$$

b. *Estimer  $\delta_1$  et  $\delta_2$ .*

c. *Résoudre un système d'équation à deux inconnues pour obtenir les estimateurs de  $a$  et  $b$ .*

Pour plus de détail sur la représentation des BLM par la copule et la méthode d'estimation des paramètres de cette dernière, voir Brahim et al. [12]<sup>4</sup>.

### Estimation non-paramétrique

Si le modèle paramétrique de la copule est mal spécifié, l'approche paramétrique et semi-paramétrique conduisent généralement à un mauvais ajustement aux données. Dans ce cas les méthodes d'estimation non paramétriques peuvent être une bonne alternative.

<sup>4</sup>Brahimi, B., Chebana, F., Necir, A., 2015. Copula representation of bivariate L-moments : A new estimation method for multiparameter two-dimensional copula models. *Statistics* **49** (3), 497-521.

- **Copule empirique** : La plupart des méthodes non-paramétriques sont basées sur la copule empirique voir la section 1.4.7. Celle-ci se base sur le rang des observations pour extraire ensuite la structure de dépendance. On peut notamment l'utiliser pour fournir des estimateurs non-paramétriques de la mesure d'association, comme le tau de Kendall et le rho de Spearman.

**Théorème 1.8.1** Soient  $\hat{C}$  et  $\hat{c}$  la copule empirique et sa fonction de densité respectivement de l'échantillon  $\{(x_i, y_i)\}_{1 \leq i \leq n}$ . Le tau de Kendall et le rho de Spearman sont estimés par les formules suivantes

$$\hat{\tau} = \frac{2n}{n-1} \sum_{j=2}^n \sum_{k=2}^n \sum_{p=1}^{j-1} \sum_{q=1}^{k-1} \left( \hat{c} \left( \frac{j}{n}, \frac{k}{n} \right) \hat{c} \left( \frac{p}{n}, \frac{q}{n} \right) - \hat{c} \left( \frac{j}{n}, \frac{q}{n} \right) \hat{c} \left( \frac{p}{n}, \frac{k}{n} \right) \right).$$

$$\hat{\rho} = \frac{12}{n^2 - 1} \sum_{j=1}^n \sum_{k=1}^n \left( \hat{C} \left( \frac{j}{n}, \frac{k}{n} \right) - \frac{jk}{n^2} \right).$$

Pour la démonstration de ce théorème, voir Nelsen [66] ou Deheuvels [24].

Cette copule satisfait plusieurs propriétés qu'on peut montrer, comme sa convergence asymptotique vers la vraie copule  $C$ .

- **Procédure de GENEST et RIVEST** : Cette méthode a été proposée par Genest et Rivest [42] pour sélectionner une copule appartenant à la famille Archimédienne. Sa méthodologie se base directement sur le générateur de la copule.

Soit  $X$  un vecteur de  $n$  vas,  $C$  la copule associée au générateur  $\phi$  et  $K$  la fonction de Kendall définie en 1.9. Un estimateur non paramétrique de  $K$  est alors donné par

$$\hat{K}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{T_i \leq t\}}, \quad (1.9)$$

où  $T_i = \frac{1}{n-1} \sum_{j=1}^n \mathbf{1}_{\{X_j \leq x_i, Y_j \leq y_i\}}$ . Comme son nom l'indique la fonction  $K$  est liée au tau de Kendall du fait que

$$\tau = 4 \int_{\mathbf{I}} (1 - K(t)) dt - 1.$$

Pour plus de détail de cette méthode voir Genest et Rivest [42] et Denuit et al. [26].

**Remarque 1.8.1** *Il existe d'autres méthodes d'estimation de la copule qu'on n'a pas présentées dans cette partie telles que :*

- Omnibus par Genest et Werker [44];
- Estimation à noyau par Fermanian et Scaillet [35];
- Bench-mark;
- Estimation bayésienne;
- Estimation à distance minimale,... ect.

Dans notre application, on se concentre sur la méthode d'inversion des mesures d'association (semi-paramétrique) pour estimer les paramètres de la copule, où on va comparer la robustesse de ces différentes méthodes d'inversion. Pour cela on va la confronter à la méthode des BLM.

### 3. Validation de modèle

Une méthode graphique permettant de valider la copule ajustée. Elle consiste à simuler des réalisations du modèle ajusté et à comparer la forme du nuage de points simulés avec celle du nuage de points observés.

De manière plus formelle, on peut faire des tests d'adéquation. Une implémentation rigoureuse des tests d'adéquation conduit à des temps de calculs importants.

La sélection de la meilleure copule doit s'appuyer sur ces tests statistique. Par exemple le test du chi-deux utilisé dans le cadre de l'ajustement d'une distribution paramétrique à une distribution empirique constitue à cet égard un outil intéressant pour le choix d'une copule. Habituellement mis en œuvre pour ajuster une distribution univariée, il offre l'avantage de s'adapter facilement au cas des distributions multivariées. Il est vrai qu'il nécessite de procéder à un découpage en classes qui réduit sa puissance et que la manière de choisir ces classes peut être critiquée. Néanmoins, sa facilité d'utilisation en fait de lui un outil de décision très intéressant pour sélectionner la meilleure copule. En plus du test du chi-deux, il existe plusieurs d'autres types de tests, comme le test de Cramer-Von-Mises, test de Kolomogorov-Smirnov...etc.

## Outils graphiques d'ajustement

Pour l'évaluation de l'ajustement d'une copule à la structure de dépendance d'un échantillon, on peut utiliser les outils graphiques suivants :

- Comparaison des densités empiriques et théoriques en trois dimensions ou sous forme de courbes de niveaux,
- Dépendogramme,
- Kendall plot ou K-plot (voir Genest et Boies [45]).

Ces graphes permettent notamment d'observer si la dépendance de queue est présentée dans les données et si elle est bien modélisée par la copule choisie.

- **Dépendogramme** : Le dépendogramme représente la structure de dépendance sous la forme du nuage de points des marges uniformes  $(U, V)$  extraites de l'échantillon, ou encore de simulation d'une copule théorique. On précise que le couple  $(U, V)$  issu de l'échantillon forme ce qu'on appellera la copule empirique. Cette dernière est simplement définie par les statistiques de rang issues de l'échantillon. On comparera le dépendogramme de cette copule empirique à celui d'autres copules théoriques estimées sur l'échantillon.

Le dépendogramme permet d'observer aussi le caractère plus ou moins simultané des réalisations issues de l'échantillon. Dans les queues plus précisément, il sera utile d'analyser si la simultanéité est forte et donc s'il est nécessaire de calibrer sur notre échantillon une copule avec une dépendance de queue.

- **Kendall plot** : Le Kendall plot va permettre une comparaison plus directe entre la copule empirique et la copule théorique. L'algorithme de construction d'un Kendall plot est le suivant :

1. Calculer pour chaque paire de rangs d'observation de l'échantillon de taille  $n$   $\{(u_i, v_i)\}_{1 \leq i \leq n}$  la fonction  $H_i^{emp}$ , tel que

$$H_i^{emp} = \frac{1}{n-1} \text{card} \{u_j \leq u_i, v_j \leq v_i : i \neq j\}.$$

Il faut ensuite les ordonner pour obtenir  $H_{(1)}^{emp} \leq \dots \leq H_{(n)}^{emp}$ . On obtient ainsi la partie empirique du K-plot issue de l'échantillon.

2. Comparer la partie empirique du K-plot issue de l'échantillon, à celle correspondant à la copule théorique qu'on estime sur l'échantillon.

Pour cela, il faut déterminer  $H_{i:n}^{theo} = \mathbb{E} \left[ H_{(i)}^{theo} \right]$ , avec

$$H_i^{theo} = \frac{1}{n-1} \text{card} \{ u_j^{theo} \leq u_i^{theo}, v_j^{theo} \leq v_i^{theo} : i \neq j \},$$

pour chaque couple  $(u_i^{theo}, v_i^{theo})$  issu de la copule théorique. On détermine ces couples à l'aide des simulations de la copule théorique.

- a) On simule  $n$  couples de la copule théorique  $m$  fois.
- b) Ainsi, on détermine  $H_i^{theo}$  pour chaque  $m^{ième}$  simulation, que l'on ordonne ensuite.
- c) Ensuite, on calcule l'espérance des  $m$   $H_{(i)}^{theo}$  pour chaque  $i$ .
- d) Enfin, on obtient  $H_{i:n}^{theo}$ .

3. Grapher les paires  $(H_i^{emp}, H_{i:n}^{theo})$  pour obtenir le K-plot.

Plus le Kendall plot se rapproche d'une droite, plus l'ajustement entre la structure de dépendance de l'échantillon et la copule estimée sur ce même échantillon est bon.

Prenons comme exemple de comparaison la copule de Student de corrélation 0.5 et de degré de liberté 3 avec les copules suivantes : produit, Gaussienne et Clayton. Le dépendogramme ci-dessous représente les quatre Kendall plots des quatre copules choisies.

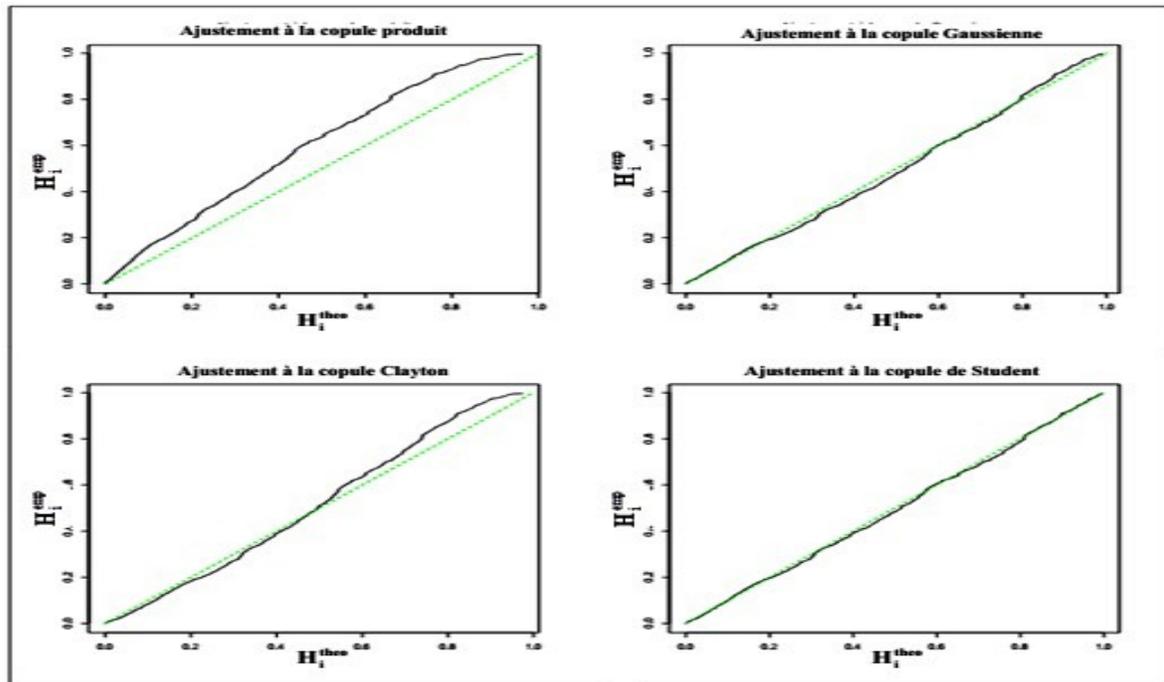


FIG. 1.11 – Kendall plots comparant un échantillon de données simulées à partir d’une copule Student avec  $\rho = 0.5$  et  $\nu = 3$  à plusieurs copules estimées sur ce même échantillon[5].

### Test du chi-deux

On propose ci-après une méthode statistique de sélection des copules basées sur un ajustement bivarié entre la copule théorique et la copule empirique effectuée à partir de la statistique du chi-deux. L’approche est inspirée d’un article de Hurlimann [54] dans lequel le test du chi-deux est utilisé dans une version modifiée de la méthode IFM. Le processus de sélection d’une copule est décrit ci-après.

#### 1. Choix d’une copule et estimation du paramètre de celle-ci :

On choisit une copule, on estime son paramètre à partir de la méthode semi-paramétrique afin de s’affranchir d’une éventuelle erreur de spécification des lois marginales. On note  $C_{\hat{\theta}}$  la copule théorique (paramétrique) obtenue.

#### 2. Calcul de la copule empirique :

On définit un treillis d’ordre  $n$  adapté au nombre d’observations et aux contraintes de temps de calcul puis on calcule la copule empirique bivariée sur ce treillis en utilisant la formule 1.6.

### 3. Choix d'un nombre d'intervalles pour découper les uniformes univariées :

On se donne un nombre  $l$  afin de discrétiser chacune des lois uniformes à partir de  $l$  intervalles et de construire un tableau de contingence bivarié des effectifs.  $l$  et  $n$  doivent être choisis ensemble afin d'obtenir aisément les effectifs empiriques à partir de la copule empirique. Sans à priori sur la distribution empirique bivariée, on découpe chaque uniforme sur des intervalles identiques.

On note  $]u_0, u_1], ]u_1, u_2], \dots, ]u_{l-1}, u_l]$  et  $]v_0, v_1], ]v_1, v_2], \dots, ]v_{l-1}, v_l]$  les  $l$  intervalles retenus pour représenter respectivement  $U$  et  $V$ .

### 4. Calcul des effectifs et construction du tableau de contingence bivarié :

L'objet de cette étape est de calculer les effectifs empiriques et théoriques sur chacun des  $l^2$  intervalles bidimensionnels  $]u_{k-1}, u_k] \times ]v_{m-1}, v_m]$  avec  $k, m = 1, \dots, l$ .

On calcule les probabilités  $p_{k,m}$  d'appartenance à l'intervalle  $]u_{k-1}, u_k] \times ]v_{m-1}, v_m]$

pour  $k, m = 1, \dots, l$  selon la formule suivante

$$\boxed{p_{k,m} = C(u_k, v_m) - C(u_{k-1}, v_m) - C(u_k, v_{m-1}) + C(u_{k-1}, v_{m-1})}, \quad (1.10)$$

en choisissant comme copule  $C$  la copule empirique  $\hat{C}$  (respectivement la copule  $C_{\hat{\theta}}$ ) pour obtenir le tableau de contingence empirique (respectivement théorique) utile au test d'adéquation du chi-deux.

**Preuve.** (démonstration de la formule 1.10)

$$\begin{aligned} p_{k,m} &= \mathbb{P}[(U, V) \in ]u_{k-1}, u_k] \times ]v_{m-1}, v_m]] \\ &= \mathbb{P}[U \leq u_k \text{ et } V \leq v_m, U > u_{k-1} \text{ et } V > v_{m-1}] \\ &= \mathbb{P}[U \leq u_k \text{ et } V \leq v_m] - \mathbb{P}[U \leq u_k \text{ et } V \leq v_m, U \leq u_{k-1} \text{ ou } V \leq v_{m-1}] \\ &= \mathbb{P}[U \leq u_k \text{ et } V \leq v_m] - \mathbb{P}[U \leq u_{k-1} \text{ et } V \leq v_m \text{ ou } U \leq u_k \text{ et } V \leq v_{m-1}] \\ &= C(u_k, v_m) - (\mathbb{P}[U \leq u_{k-1} \text{ et } V \leq v_m] + \mathbb{P}[U \leq u_k \text{ et } V \leq v_{m-1}] \\ &\quad - \mathbb{P}[U \leq u_{k-1} \text{ et } V \leq v_{m-1}]) \\ &= C(u_k, v_m) - C(u_{k-1}, v_m) - C(u_k, v_{m-1}) + C(u_{k-1}, v_{m-1}). \end{aligned}$$

CQFD. ■

Les effectifs  $f_{k,m}$  sont calculés en multipliant les probabilités par le nombre total d'observations. On obtient alors un tableau de contingence bivarié donnant les effectifs empiriques issus de la copule empirique (respectivement les effectifs théoriques issus de la copule théorique). Il a la forme suivante

$U \setminus V$	$]v_0, v_1]$	$]v_1, v_2]$	...	$]v_{l-1}, v_l]$
$]u_0, u_1]$	$f_{1,1}$	$f_{1,2}$	...	$f_{1,l}$
$]u_1, u_2]$	$f_{2,1}$	$f_{2,2}$	...	$f_{2,l}$
...	...	...	...	...
$]u_{l-1}, u_l]$	$f_{l,1}$	$f_{l,2}$	...	$f_{l,l}$

TAB. 1.5 – Tableau de contingence donnant les effectifs empiriques issus de la copule empirique (respectivement les effectifs théoriques issus de la copule théorique).

### 5. Application du critère de Cochran et regroupement en classes :

Dans la pratique, les effectifs théoriques de certaines cases peuvent être très faibles voir proches de zéro. Il faut alors procéder à un regroupement des  $l^2$  intervalles initiaux en un nombre  $R$  plus faible, de classes permettant de respecter le critère de Cochran qui recommande d'avoir des effectifs théoriques  $\geq 1\%$  du nombre total d'observations dans chaque classe, et  $\geq 5\%$  du nombre total d'observations dans au moins 80% des classes.

Ce regroupement est aussi à effectuer sur le tableau des effectifs empiriques.

### 6. Calcul de la statistique bivariée du chi-deux et sélection de la copule :

Après avoir défini le regroupement en  $R$  classes vérifiant le critère de Cochran, on calcule la statistique du chi-deux observée notée  $\chi_{obs}^2$  comme suit

$$\chi_{obs}^2 = \sum_{r=1}^R \frac{(n_r^{emp} - n_r^{theo})^2}{n_r^{theo}},$$

où  $n_r^{theo}$  et  $n_r^{emp}$  représentent respectivement les effectifs théoriques et empiriques dans la classe  $r$  avec  $r = 1, \dots, R$ . Cette statistique suit une loi du chi-deux à  $(R - \omega - 1)$  degré de liberté où  $\omega$  désigne le nombre de paramètres estimés de la copule.

Plus  $\chi_{obs}^2$  sera grand, plus le désaccord sera important. La coincidence sera parfaite si  $\chi_{obs}^2 = 0$ .

On se donne ensuite un seuil critique  $\alpha$  pour la zone de rejet et on calcule la p-valeur

$\hat{\alpha} = \mathbb{P}(\chi_{R-\omega-1}^2 > \chi_{obs}^2)$ . On rejette l'adéquation de la copule théorique si  $\hat{\alpha} < \alpha$ . Il faut préciser pour les praticiens que la p-valeur exacte  $\hat{\alpha}$  se situe en fait dans l'intervalle suivant :

$$\mathbb{P}(\chi_{R-1}^2 > \chi_{obs}^2) \leq \hat{\alpha} \leq \mathbb{P}(\chi_{R-\omega-1}^2 > \chi_{obs}^2).$$

Cet encadrement étant de faible amplitude lorsque  $\omega$  est petit vis à vis de  $R$ , le seuil critique  $\alpha$  sera le plus souvent en dehors de cet encadrement et on conclura sans ambiguïté (pour plus de détail voir Besson et Partrat [7]). Parmi les copules acceptées, on sélectionne alors celle qui maximise la p-valeur.

### Test de Cramér-von Mises

On suppose qu'on cherche à tester  $C \in \mathcal{C}$  (l'hypothèse nulle), où  $\mathcal{C}$  est une famille de copules (l'hypothèse alternative étant  $C \notin \mathcal{C}$ ). Si la famille  $\mathcal{C}$  est une famille théorique, Fermanian [36] ou Genest et Remillard [47] ont suggéré d'utiliser le test de Cramér-von Mises, avec comme statistique

$$T = n \iint_{\mathbf{I}^2} \left[ \hat{C}(u, v) - C_{\hat{\theta}}(u, v) \right]^2 dC_{\hat{\theta}}(u, v).$$

Une approche un peu plus simple est d'utiliser la fonction de Kendall 1.5. Si on suppose que cette fonction appartient à une famille paramétrique, alors la statistique précédente peut s'écrire

$$T = n \iint_{\mathbf{I}^2} \left[ \hat{K}(t) - K_{\hat{\theta}}(t) \right]^2 dK_{\hat{\theta}}(t),$$

comme suggéré par Genest et Rivest [42]. On rejette l'hypothèse nulle si la valeur de cet indicateur  $T$  est supérieure au seuil critique  $\alpha$ .

### 1.8.2 Simulation de la copule

Souhaiter simuler une copule bivariée  $C$  signifie simuler les arguments  $u$  et  $v$  de cette fonction tirés d'un vecteur aléatoire de loi uniforme  $(U, V)$  de distribution jointe  $H$ . Ceci permet alors de déterminer la simulation d'un vecteur aléatoire  $(X, Y)$  dont la structure de dépendance est définie par  $C$  et dont les marges particulières sont  $F$  et  $G$ . Pour cela on utilise la transformation 1.2.

Toute la difficulté est de simuler des nombres aléatoires issus de  $C$ . Par exemple la simulation de la copule produit est relativement simple : on simule des réalisations de loi uniforme sur  $\mathbf{I}$  de manière indépendante. Mais pour la majorité des copules, cela s'avère plus complexe. On propose alors trois méthodes différentes selon la forme de la copule.

### Méthode des distributions

Dans le cas où la distribution jointe  $H$  est plus facile à simuler que la copule  $C$ , on utilise la méthode des distributions. Simuler les réalisations du vecteur aléatoire  $(U, V)$  revient à :

1. Simuler les réalisations du vecteur aléatoire  $(X, Y)$  de distribution  $H$ .
2. Appliquer la transformation  $(F(X), G(Y))$ .

C'est le cas par exemple d'une copule Gaussienne et de la copule de Student.

### Méthode analytique

Une méthode analytique est une méthode de simulation spécifique à chaque copule.

#### Exemple 1.8.4 (Copule de Clayton)

On choisit de présenter ici la méthode analytique de simulation d'une copule de Clayton proposée par Devroye [27]. Cette méthode revient à :

1. Simuler deux vas uniformes  $s$  et  $t$  à partir desquelles on détermine  $x = -\ln(s)$  et  $y = -\ln(t)$ .

2. Simuler une va de loi Gamma de paramètres 1 et  $\theta$  :  $z \sim \Gamma(1, \theta)$ <sup>5</sup>.

3. Déterminer les réalisations de la copule Clayton à partir des expressions suivantes :

$$u = \left(1 + \frac{x}{z}\right)^{-\theta} \quad \text{et} \quad v = \left(1 + \frac{y}{z}\right)^{-\theta}.$$

---

<sup>5</sup>Une va  $X$  suit une loi Gamma de paramètres  $\mu, \theta \in \mathbb{R}_+^*$  si elle admet pour densité de probabilité la fonction :

$$f_X(x) = \begin{cases} \frac{\theta^\mu}{\Gamma(\mu)} x^{\mu-1} \exp\{-\theta x\} & \text{si } 0 < x; \\ 0 & \text{si } x \leq 0, \end{cases}$$

$$\text{avec } \Gamma(\mu) = \int_0^{+\infty} x^{\mu-1} \exp\{-x\} dx.$$

**Exemple 1.8.5** (Simulation de la copule MO)

D'après Devroye [27], il existe deux méthodes :

**Algorithme 1 :**

**Étape 1.** Générer trois variables indépendantes uniformes  $p, q$  et  $r$  sur  $\mathbf{I}$ .

**Étape 2.** Soient  $x = \min\left(\frac{-\ln p}{\lambda_1}, \frac{-\ln r}{\lambda_{12}}\right)$  et  $y = \min\left(\frac{-\ln q}{\lambda_2}, \frac{-\ln r}{\lambda_{12}}\right)$ , avec  $\lambda_1, \lambda_2$  et  $\lambda_{12} \geq 0$ .

**Étape 3.** La paire souhaitée  $(u, v)$  est obtenue à partir de  $u = \exp\{-(\lambda_1 + \lambda_{12})x\}$  et de  $v = \exp\{-(\lambda_2 + \lambda_{12})y\}$ .

**Algorithme 2 :**

**Étape 1.** Générer des variables indépendantes exponentielles  $Z_1 \sim \mathcal{E}(\lambda_1)$ ,  $Z_2 \sim \mathcal{E}(\lambda_2)$  et

$Z_{12} \sim \mathcal{E}(\lambda_{12})$ , avec  $\lambda_1, \lambda_2$  et  $\lambda_{12} \geq 0$ .

**Étape 2.** Soient  $U = 1 - \exp\{-(\lambda_1 + \lambda_{12})\max(Z_1, Z_{12})\}$  et  $V = 1 - \exp\{-(\lambda_2 + \lambda_{12})\max(Z_2, Z_{12})\}$ .

**Étape 3.** Soient  $a = \frac{\lambda_{12}}{\lambda_1 + \lambda_{12}}$  et  $b = \frac{\lambda_{12}}{\lambda_2 + \lambda_{12}}$ .

Dans la figure (1.12), on présente les nuages de points de la copule MO pour un échantillon de taille  $n = 500$ , selon les trois cas de dépendance.

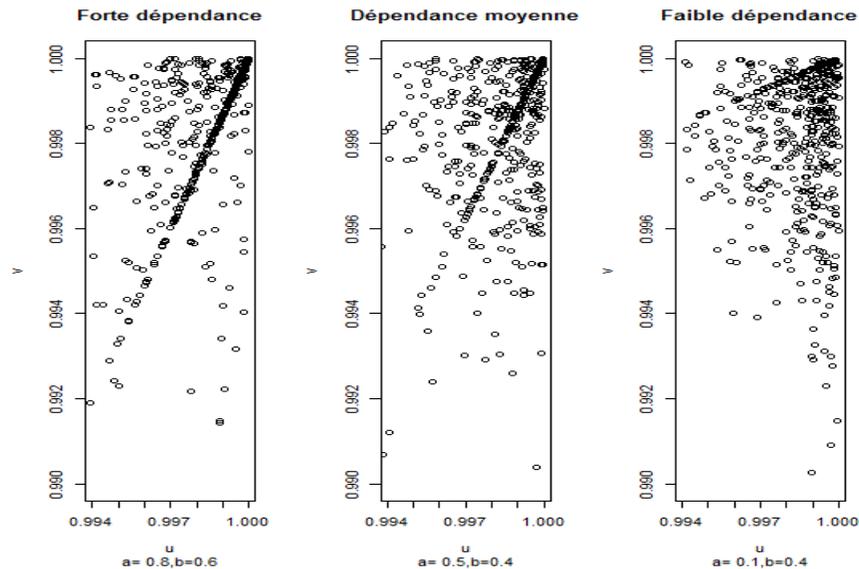


FIG. 1.12 – Nuages de points de la copule MO selon les trois cas de dépendance.

La composante singulière ( $u^a = v^b$ ) est clairement visible sur les figures. On remarque aussi dans les trois figures, qu'il y a une forte dépendance sur la queue supérieure.

## Méthode des distributions conditionnelles

Cette méthode consiste à simuler directement la copule  $C$ , dont l'algorithme est le suivant :

1. Simuler deux variables uniformes  $v_1$  et  $v_2$ .
2. Imposer  $u_1 = v_1$ .
3. Déterminer  $u_2$  à l'aide de la distribution conditionnelle de la copule  $C$  sachant  $u_1$ . Cette dernière se définit de la manière suivante :

$$\begin{aligned}
 C_{U_2|U_1}(u_1, u_2) &= \mathbb{P}[U_2 \leq u_2 | U_1 = u_1] \\
 &= \lim_{\Delta u \rightarrow 0^+} \frac{\mathbb{P}[U_2 \leq u_2, u_1 \leq U_1 \leq u_1 + \Delta u]}{\mathbb{P}[u_1 \leq U_1 \leq u_1 + \Delta u]} \\
 &= \lim_{\Delta u \rightarrow 0^+} \frac{\mathbb{P}[U_1 \leq u_1 + \Delta u, U_2 \leq u_2] - \mathbb{P}[U_2 \leq u_2, U_1 \leq u_1]}{\Delta u - 0} \\
 &= \lim_{\Delta u \rightarrow 0^+} \frac{C(u_1 + \Delta u, u_2) - C(u_1, u_2)}{\Delta u} \\
 &= \partial C(u_1, u_2) / \partial u_1 \\
 &= d_{U_1} C(u_1, u_2).
 \end{aligned}$$

4. Pour déterminer  $u_2$ , il nous reste à :

- Inverser la distribution conditionnelle

$$C_{U_2|U_1}^{-1}(u; u_1) = \{u_2 : C_{U_2|U_1}(u_1, u_2) = u\}.$$

- Remplacer  $u$  par  $v_2$ . Ainsi, on a la forme générale de l'inverse de la distribution conditionnelle suivante

$$u_2 = C_{U_2|U_1}^{-1}(v_2; u_1).$$

C'est le cas par exemple de la copule de Frank et la copule de Gumbel.

**Exemple 1.8.6** (Copule de Frank)

Dans le cas de la copule de Frank, la méthode des distributions conditionnelles peut être mise en œuvre simplement. L'expression de cette copule est

$$C(u_1, u_2) = -\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right); \theta \in \mathbb{R}^*.$$

On en déduit que

$$\begin{aligned} C_{U_2|U_1}(u_1, u_2) &= \partial C(u_1, u_2) / \partial u_1 \\ &= [e^{-\theta u_1}(e^{-\theta u_2} - 1)] / [e^{-\theta} - 1 + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)]. \end{aligned}$$

On peut alors inverser cette équation (en résolvant l'équation en  $u_2$ ,  $C_{U_2|U_1}(u_1, u_2) = u$  pour obtenir

$$C_{U_2|U_1}^{-1}(u; u_1) = -\frac{1}{\theta} \ln \left( 1 + \frac{u(e^{-\theta} - 1)}{u + (1 - u)e^{-\theta u_1}} \right).$$

Donc

$$u_2 = C_{U_2|U_1}^{-1}(v_2; u_1) = -\frac{1}{\theta} \ln \left( 1 + \frac{v_2(e^{-\theta} - 1)}{v_2 + (1 - v_2)e^{-\theta u_1}} \right).$$

La simulation peut donc être mise en œuvre simplement.

**Exemple 1.8.7** (Copule de Gumbel)

L'expression de cette copule est

$$C(u_1, u_2) = \exp \left\{ - [(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta} \right\}; \theta \geq 1.$$

On en déduit que

$$\begin{aligned} C_{U_2|U_1}(u_1, u_2) &= \partial \exp \left\{ - [(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta} \right\} / \partial u_1 \\ &= \frac{1}{u_1} \left[ 1 + \left( \frac{\ln u_2}{\ln u_1} \right)^\theta \right]^{-1+1/\theta} \exp \left\{ - [(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta} \right\}. \end{aligned}$$

L'inversion de cette expression n'est pas aisée, et on aura ici recours à des méthodes numériques.

**Exemple 1.8.8** (Simulation de la copule de FGM)

D'après Nelsen ([66], p. 41), on suit les étapes suivantes :

**Étape 1.** Générer deux variables indépendantes uniformes  $u$  et  $t$  sur  $\mathbf{I}$ .

**Étape 2.** Soit  $v = c_u^{[-1]}(t)$  où  $c_u^{[-1]}$  est l'inverse de  $c_u$ , tel que

$$c_u(v) = \frac{dC(u, v)}{du}.$$

**Étape 3.** La paire désirée est  $(u, v)$ .

Dans la figure (1.13), on présente les nuages de points de la copule FGM pour un échantillon de taille  $n=1000$ , selon les trois cas de dépendance.

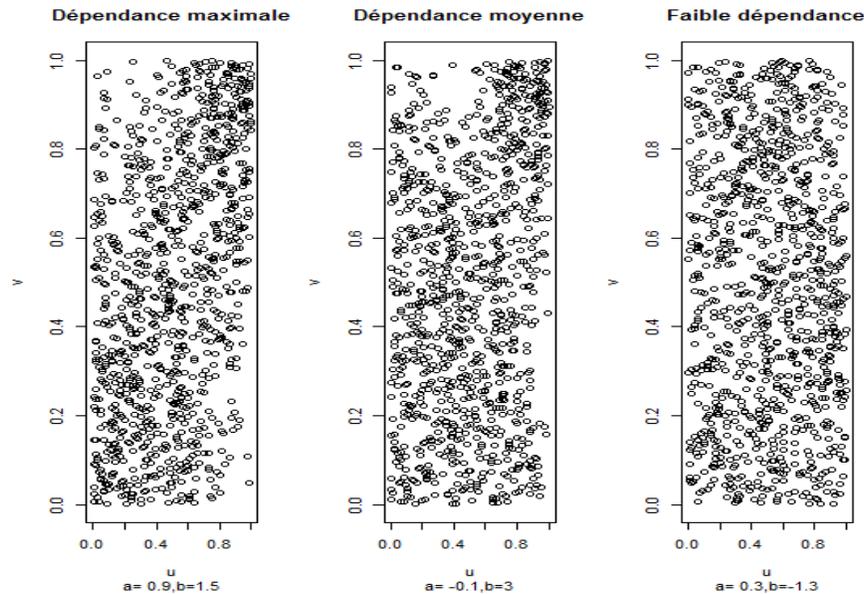


FIG. 1.13 – Nuages de points de la copule FGM selon les trois cas de dépendance.

D'après ce graphe on remarque qu'il n'y a pas de forte dépendance car cette famille ne permet pas la modélisation des dépendances élevées étant donné que le coefficient de corrélation est limité à  $[-0.33, 0.33]$ . Cela dit on peut clairement remarquer la différence entre la dépendance moyenne et faible.

**Remarque 1.8.2** Les packages que fournit le logiciel statistique R présentent le gros avantage de donner rapidement du code documenté, et donc réutilisable par quelqu'un d'autre que ses

*concepteurs. Pour cela, on fait usage de plusieurs packages :*

- **copula** : *générer les valeurs de la copule.*
- **gumbel** : *générer les valeurs de la copule de Gumbel.*
- **BB** : *résoudre des systèmes d'équations non linéaires.*
- **asbio** : *calcul du beta de Blomqvist empirique.*
- **lmomco** : *estimer les BLM.*

# Chapitre 2

## Mesures de dépendance et de concordance

Le concept de dépendance ou d'association a été largement étudié en probabilités et en statistiques, notamment dans le cas de deux variables. De manière informelle, deux variables  $X$  et  $Y$  sont dites concordantes lorsque ces dernières prennent de fortes (resp. faibles) valeurs simultanément. Différentes mesures de concordance ont été proposées dans la littérature.

Lorsqu'on évoque la dépendance, le premier mot qui vient généralement à l'esprit est celui du coefficient de corrélation linéaire. Ce dernier également connu sous le nom de coefficient de corrélation de Pearson, est en effet le plus couramment utilisé. Il est tout à fait approprié lorsqu'on étudie des distributions normales ou de Student multivariées, mais celui-ci perd de son intérêt si le modèle est différent.

### 2.1 Fonction de concordance

Une mesure de dépendance doit en fait nous permettre de nous faire une idée sur la structure de dépendance entre deux variables, ceci et exprimé à l'aide d'un seul nombre.

### 2.1.1 Corrélation linéaire de Pearson

Soient  $X$  et  $Y$  deux vas. La covariance de  $X$  et  $Y$  est le nombre

$$\text{Cov}(X, Y) := \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

**Définition 2.1.1** Soient  $X$  et  $Y$  deux vas ayant des variances finies  $\mathbb{V}(X)$  et  $\mathbb{V}(Y)$  respectivement. Le coefficient de corrélation entre  $X$  et  $Y$  se calcule par la formule suivante

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}}; \rho \in [-1, 1],$$

où  $\mathbb{V}(Z) = \mathbb{E}(Z - \mathbb{E}(Z))^2$  pour  $Z = X$  et  $Y$ .

Il est important de rappeler que la dépendance et la corrélation sont des notions différentes. En effet, si  $X$  et  $Y$  sont des variables indépendantes elles sont non corrélées mais la réciproque est fautive sauf dans le cas où les variables sont gaussiennes car la dépendance est alors entièrement caractérisée par le coefficient de corrélation.

On peut donner le contre-exemple suivant :

Soient  $X \rightsquigarrow \mathcal{N}(0, 1)$  et  $Y = X^2$ . Les deux variables sont non-corrélées :  $\text{Cov}(X, Y) = \mathbb{E}(X^3) = 0$  mais présentent un lien de dépendance évident.

### 2.1.2 Concept de concordance

Soit  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  un échantillon de  $n$  observations d'un couple  $(X, Y)$ . Il existe  $C_n^2 = n!/2(n-2)!$  paires de distributions distinctes de couples  $(x_i, y_i)$  et  $(x_j, y_j)$  qui sont dites concordantes ou discordantes si :

- **Concordantes** :  $(x_i - x_j)(y_i - y_j) > 0$  i.e.  $(x_i < x_j \text{ et } y_i < y_j)$  ou  $(x_i > x_j \text{ et } y_i > y_j)$ .
- **Discordantes** :  $(x_i - x_j)(y_i - y_j) < 0$  i.e.  $(x_i < x_j \text{ et } y_i > y_j)$  ou  $(x_i > x_j \text{ et } y_i < y_j)$ .

**Définition 2.1.2 (Fonction de concordance)** La fonction de concordance est la différence entre la probabilité de concordance et celle de discordance entre deux couples  $(X_1, Y_1)$  et  $(X_2, Y_2)$ . Elle est donnée par

$$Q := \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]. \quad (2.1)$$

Comme les vas sont continues, donc

$$\begin{aligned} Q &= \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - (1 - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0]) \\ &= 2\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1. \end{aligned}$$

Avec :  $\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] = \mathbb{P}[X_1 > X_2, Y_1 > Y_2] + \mathbb{P}[X_1 < X_2, Y_1 < Y_2]$ .

### 2.1.3 Propriétés de la fonction de concordance

Le résultat suivant permet d'établir une relation entre la fonction de concordance  $Q$  et les copules associées aux deux couples.

**Théorème 2.1.1** Soient  $(X_1, Y_1)$  et  $(X_2, Y_2)$  deux couples de vas indépendantes de distributions jointes  $H_1$  et  $H_2$  avec des marges communes  $F$  et  $G$  respectivement (i.e.  $F$  pour  $X_1$  et  $X_2$ ,  $G$  pour  $Y_1$  et  $Y_2$ ). Soient  $C_1$  et  $C_2$  les copules associées aux distributions  $H_1$  et  $H_2$  respectivement.

On a

$$Q = Q(C_1, C_2) := 4 \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) - 1. \quad (2.2)$$

On peut montrer que les deux formules 2.1 et 2.2 sont équivalentes.

**Preuve.** Pour commencer on calcule d'abord  $\mathbb{P}[X_1 > X_2, Y_1 > Y_2]$  puis

$\mathbb{P}[X_1 < X_2, Y_1 < Y_2]$  comme suit

$$\begin{aligned}
1. \mathbb{P}[X_1 > X_2, Y_1 > Y_2] &= \mathbb{P}[X_2 < X_1, Y_2 < Y_1] \\
&= \iint_{\mathbb{R}^2} \mathbb{P}[X_2 < x, Y_2 < y] dH_1(x, y) \\
&= \iint_{\mathbb{R}^2} \mathbb{P}[X_2 < x, Y_2 < y] dC_1(F(x), G(y)) \\
&= \iint_{\mathbb{R}^2} C_2(F(x), G(y)) dC_1(F(x), G(y)) \\
&= \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v).
\end{aligned}$$

De façon similaire on a :

$$\begin{aligned}
2. \mathbb{P}[X_1 < X_2, Y_1 < Y_2] &= \mathbb{P}[X_2 > X_1, Y_2 > Y_1] \\
&= \iint_{\mathbb{R}^2} \mathbb{P}[X_2 > x, Y_2 > y] dH_1(x, y) \\
&= \iint_{\mathbb{R}^2} (1 - \mathbb{P}[X_2 \leq x, Y_2 \leq y]) dH_1(x, y) \\
&= \iint_{\mathbb{R}^2} (1 - H_2(x, y)) dC_1(F(x), G(y)) \\
&= \iint_{\mathbb{R}^2} \bar{H}_2(x, y) dH_1(x, y) \\
&= \iint_{\mathbf{I}^2} [1 - u - v + C_2(u, v)] dC_1(u, v) \\
&= 1 - \frac{1}{2} - \frac{1}{2} + \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) \\
&= \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v).
\end{aligned}$$

CQFD. ■

**Preuve.** En regroupant ces deux résultats on en déduit que :

$$\begin{aligned}
Q &= 2\mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1 \\
&= 2(\mathbb{P}[X_1 < X_2, Y_1 < Y_2] + \mathbb{P}[X_1 > X_2, Y_1 > Y_2]) - 1 \\
&= 2 \left( \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) + \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) \right) - 1 \\
&= 4 \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) - 1.
\end{aligned}$$

CQFD. ■

Le résultat suivant résume les propositions essentielles de la fonction  $Q$ .

**Proposition 2.1.1** *Soient  $C_1, C_2$  et  $Q$  donnés par le théorème 2.1.1, alors*

- i)  $Q$  est symétrique :  $Q(C_1, C_2) = Q(C_2, C_1)$ .
- ii)  $Q$  conserve l'ordre : si  $C_1 < \hat{C}_1$  et  $C_2 < \hat{C}_2$  alors  $Q(C_1, C_2) < Q(\hat{C}_1, \hat{C}_2)$ .
- iii)  $Q$  est invariante par rapport à la survie :  $Q(C_1, C_2) = Q(\hat{C}_1, \hat{C}_2)$ .

**Preuve.** i) La première proposition est évidente.

$$\text{ii) } Q(\hat{C}_1, \hat{C}_2) - Q(C_1, C_2) = 4 \left[ \iint_{\mathbf{I}^2} \hat{C}_2(u, v) d\hat{C}_1(u, v) - \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) \right].$$

Comme  $C_1 < \hat{C}_1$  cela implique que  $\partial^2 C_1 / \partial u \partial v < \partial^2 \hat{C}_1 / \partial u \partial v$ , car  $C_1$  et  $\hat{C}_1$  sont 2-croissantes.

$$\text{On pose : } \begin{cases} \partial^2 C_1 / \partial u \partial v = dC_1(u, v), \\ \partial^2 \hat{C}_1 / \partial u \partial v = d\hat{C}_1(u, v). \end{cases}$$

Donc  $dC_1(u, v) < d\hat{C}_1(u, v)$ .

$$\text{Alors : } \iint_{\mathbf{I}^2} \hat{C}_2(u, v) d\hat{C}_1(u, v) > \iint_{\mathbf{I}^2} \hat{C}_2(u, v) dC_1(u, v) > \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v).$$

$$\implies \iint_{\mathbf{I}^2} \hat{C}_2(u, v) d\hat{C}_1(u, v) - \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) > 0.$$

$$\implies 4 \left[ \iint_{\mathbf{I}^2} \hat{C}_2(u, v) d\hat{C}_1(u, v) - \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) \right] > 0.$$

$$\boxed{Q(\hat{C}_1, \hat{C}_2) - Q(C_1, C_2) > 0.}$$

iii) Pour toute copule  $C$ , on a  $\partial^2 C / \partial u \partial v = dC(u, v)$  et d'après la copule de survie dans la section 1.5.1, on a  $\hat{C}(u, v) = 1 - u - v + C(u, v)$ . Il s'en suit que :  $d\hat{C}(u, v) = dC(u, v)$ , alors

$$\begin{aligned} Q(\hat{C}_1, \hat{C}_2) &= 4 \iint_{\mathbf{I}^2} \hat{C}_2(u, v) d\hat{C}_1(u, v) - 1 \\ &= 4 \iint_{\mathbf{I}^2} [1 - u - v + C_2(u, v)] dC_1(u, v) - 1 \\ &= 4 \iint_{\mathbf{I}^2} (1 - u - v) dC_1(u, v) + 4 \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) - 1. \end{aligned}$$

$$\text{Or : } \iint_{\mathbf{I}^2} (1 - u - v) dC_1(u, v) = 1 - \frac{1}{2} - \frac{1}{2} = 0.$$

Cela implique que

$$Q(\hat{C}_1, \hat{C}_2) = 4 \iint_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) - 1 = Q(C_1, C_2).$$

CQFD. ■

**Remarque 2.1.1** (*Fonctions de dépendance des copules usuelles*)

*Les résultats suivants donnent les fonctions de dépendance des copules usuelles  $W$ ,  $\Pi$  et  $M$  pris deux à deux.*

$$Q(W, W) = Q(W, M) = -1; \quad Q(\Pi, \Pi) = 0; \quad Q(\Pi, M) = 1/3; \quad Q(W, \Pi) = -1/3; \quad Q(M, M) = 1.$$

$$\text{Pour toute copule } C \text{ on a : } 0 \leq Q(C, \Pi) \leq 1/3; \quad -1 \leq Q(C, W) \leq 1/3; \quad -1 \leq Q(C, M) \leq 1.$$

**Preuve. I)** Le support de la copule minimum  $M$  est l'ensemble  $\mathcal{L}_M = \{u, v \in \mathbf{I} : u = v\}$ .

Si  $g$  est une fonction intégrable dans le domaine  $\mathbf{I}^2$ , alors

$$\boxed{\iint_{\mathbf{I}^2} g(u, v) dM(u, v) = \int_{\mathbf{I}} g(u, u) du.}$$

Donc :  $M(u, v) = u$ ;  $W(u, v) = 2u - 1$ ;  $\Pi(u, v) = u^2$ . Alors

$$- Q(M, M) : 4 \int_{\mathbf{I}} u du - 1 = 1.$$

$$- Q(M, \Pi) : 4 \int_{\mathbf{I}} u^2 du - 1 = 1/3.$$

$$- Q(M, W) := 4 \int_{\mathbf{I}} (2u - 1) du - 1 = -1.$$

**II)** De même le support de la copule maximum  $W$  est l'ensemble  $\mathcal{L}_W = \{u, v \in \mathbf{I} : v = 1 - u\}$ ,

alors

$$\boxed{\iint_{\mathbf{I}^2} g(u, v) dW(u, v) = \int_{\mathbf{I}} g(u, 1 - u) du.}$$

Donc :  $M(u, v) = \min(u, 1 - u)$ ;  $W(u, v) = 0$ ;  $\Pi(u, v) = u(1 - u)$ . Alors

$$- Q(W, W) := 4 \int_{\mathbf{I}} 0 du - 1 = -1.$$

$$- Q(W, \Pi) : 4 \int_{\mathbf{I}} u(1 - u) du - 1 = -1/3.$$

**III)** Enfin puisque  $d\Pi(u, v) = dudv$ , alors :  $Q(\Pi, \Pi) = 4 \iint_{\mathbf{I}^2} uv dudv - 1 = 0$ .

CQFD. ■

**Résultat.** Pour toute copule quelconque  $C$  on a :

$$\left\{ \begin{array}{l} W \leq C \leq \Pi \implies -1 \leq Q(C, W) \leq -1/3, \\ \Pi \leq C \leq M \implies 0 \leq Q(C, \Pi) \leq 1/3. \\ W \leq C \leq M \implies -1 \leq Q(C, M) \leq 1. \end{array} \right.$$

## 2.2 Mesures de concordance

Plus classiquement, les mesures d'association couramment utilisées d'après D'Agostino et Stephens [20], Joe [57] et Genest et Favre [46], pour la caractérisation de la dépendance sont : le coefficient de corrélation linéaire  $\varrho$  de Pearson ; le coefficient de corrélation des rangs  $\rho$  de Spearman et le coefficient de corrélation des rangs  $\tau$  de Kendall. Le coefficient de corrélation de Pearson mesure la dépendance linéaire entre deux vas. Pour l'étude d'une dépendance non linéaire, les deux autres coefficients sont plus adaptés puisqu'ils ne dépendent que du rang de chaque observation par opposition au coefficient de corrélation de Pearson qui utilise les réalisations des variables. En outre,  $\rho$  et  $\tau$  sont invariants par transformation strictement croissante des vas  $X$  et  $Y$ .

Une mesure d'association  $\kappa$  entre deux vas  $X$  et  $Y$  de copule  $C$  est une mesure de concordance si elle vérifie les propriétés suivantes :

- $\kappa$  est définie pour tout couple  $(X, Y)$  de va continues ;
- $-1 \leq \kappa_{X,-X} \leq \kappa_{X,Y} \leq \kappa_{X,X} = 1$ ;
- Si  $Y$  est presque sûrement une fonction croissante de  $X$ , alors  $\kappa_{X,Y} = \kappa_M = 1$ ;
- Si  $Y$  est presque sûrement une fonction décroissante de  $X$ , alors  $\kappa_{X,Y} = \kappa_W = -1$ ;
- Si  $X$  et  $Y$  sont indépendantes, alors  $\kappa_{X,Y} = 0$ ;
- $\kappa_{X,Y} = \kappa_{Y,X}$ ;
- $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$ ;
- Si  $\alpha$  et  $\beta$  sont des fonctions strictement croissantes, alors  $\kappa_{\alpha(X),\beta(Y)} = \kappa_{X,Y}$ ;
- Si  $C_1$  et  $C_2$  sont deux copules telles que  $C_1 < C_2$  alors on a :  $\kappa_{C_1} < \kappa_{C_2}$ ;
- Si  $\{(X_n, Y_n)\}_{n \geq 1}$  est une suite de vas dont la copule est  $C_n$ , et si  $\{C_n\}_{n \geq 1}$  converge vers  $C$  lorsque  $n$  tend vers l'infini alors  $\kappa_{C_n}$  converge vers  $\kappa_C$ .

### 2.2.1 Tau de Kendall

Le tau de Kendall et le rho de Spearman sont deux mesures de concordance bien connues en statistique. Elles offrent l'avantage de s'exprimer simplement en fonction de la copule associée au couple de vas.

On peut exprimer le tau de Kendall de différentes manières, soit en fonction des observations, de la fonction de concordance et même en fonction de la copule associée.

Le tau de Kendall de la série exprimé en fonction des observations s'écrit sous la forme suivante :

**Définition 2.2.1 (Version échantillon)** Soit une série de  $n$  observations  $\{(x_i, y_i)\}_{1 \leq i \leq n}$  d'un couple  $(X, Y)$ . Le tau de Kendall est défini par

$$\tau := \frac{2}{n(n-1)}(n_c - n_d); \tau \in [-1, 1], \quad (2.3)$$

$$\text{où } \begin{cases} n_c : \text{nombre de paires concordantes,} \\ n_d : \text{nombre de paires discordantes,} \\ n : \text{nombre total de paires.} \end{cases}$$

Sachant que  $n_c - n_d = \sum_{i=1}^{j-1} \sum_{j=2}^n \text{sign}\{(x_i - x_j)(y_i - y_j)\}$ .

$$\text{En posant } z_{ij} = (x_i - x_j)(y_i - y_j), \text{ on a } \text{sign}\{z_{ij}\} = \begin{cases} 1 & \text{if } z_{ij} > 0, \\ -1 & \text{if } z_{ij} < 0. \end{cases}$$

Le tau de Kendall n'est autre que la différence entre la probabilité de concordance et celle de discordance.

**Définition 2.2.2 (Version population)** Soient  $(X_1, Y_1)$  et  $(X_2, Y_2)$  deux couples de vecteurs continus iid de fd jointe  $H$ . Le tau de Kendall est donné par

$$\tau_{XY} := \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Son expression en terme de la copule associée est la suivante.

**Définition 2.2.3** Soit  $(X, Y)$  un couple de vas continues de copule  $C$ . Si les couples sont identiquement distribués i.e.  $H_1 = H_2 = H$ , le tau de Kendall a pour expression

$$\tau_C = Q(C, C) := 4 \iint_{\mathbf{I}^2} C(u, v) dC(u, v) - 1. \quad (2.4)$$

Comme les couples sont uniformément distribués sur  $\mathbf{I}$ , alors

$$\iint_{\mathbf{I}^2} C(u, v) dC(u, v) = \mathbb{E}[C(u, v)].$$

Ainsi, on peut écrire

$$\tau_C = 4\mathbb{E}[C(u, v)] - 1.$$

**Remarque 2.2.1** Le tau de Kendall de  $X$  et  $Y$  d'une copule Archimédienne s'obtient simplement à l'aide du générateur de la copule, selon la formule ci-dessous

$$\tau_C = 1 + 4 \int_{\mathbf{I}} \frac{\phi(u)}{\phi'(u)} du.$$

## 2.2.2 Rho de Spearman

Le coefficient de corrélation rho de Spearman représente l'une des mesures les plus connues pour quantifier le degré d'association entre deux vas. La valeur de ce coefficient dénotée par  $\rho$  est équivalente au coefficient de corrélation de Pearson. Il a été développé par Spearman [76]. Le rho de Spearman de deux vas  $X$  et  $Y$  est égal au coefficient de corrélation entre les variables  $F(X)$  et  $G(Y)$  distribuées selon la loi uniforme sur l'intervalle  $\mathbf{I}$  tel que

$$\rho(X, Y) = \rho(F(X), G(Y)).$$

**Définition 2.2.4 (Version échantillon)** Soit une série de  $n$  observations  $\{(x_i, y_i)\}_{1 \leq i \leq n}$  d'un couple  $(X, Y)$ . On définit le rho de Spearman par

$$\rho := 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n D_i^2, \quad (2.5)$$

où  $D_i = R_{X_i} - R_{Y_i}$ ,  $R_{X_i}$  et  $R_{Y_i}$  les rangs des observations de  $X$  et  $Y$  respectivement.

**Définition 2.2.5 (Version population)** Soient  $(X_1, Y_1)$ ,  $(X_2, Y_2)$  et  $(X_3, Y_3)$  trois vecteurs aléatoires indépendants de même loi  $H$ . La version population du rho de Spearman est définie comme étant proportionnelle à la différence de la probabilité de concordance et celle de discordance des couples aléatoires  $(X_1, Y_1)$ ,  $(X_2, Y_3)$ . Elle s'écrit comme suit

$$\rho_{XY} := 3([\mathbb{P}(X_1 - X_2)(Y_1 - Y_3) > 0] - [\mathbb{P}(X_1 - X_2)(Y_1 - Y_3) < 0]).$$

La distribution de  $(X_2, Y_3)$  étant  $\Pi$  (car les variables  $X_2$  et  $Y_3$  sont indépendantes) alors d'après ce qui précède on a le théorème suivant.

**Théorème 2.2.1** (Expression de rho de Spearman en terme de copule)

$$\rho_C = 3Q(C, \Pi) = 3 \times \left( 4 \iint_{\mathbf{I}^2} C(u, v) d\Pi(u, v) - 1 \right) = 12 \iint_{\mathbf{I}^2} C(u, v) d(uv) - 3.$$

Alors

$$\rho_C := 12 \iint_{\mathbf{I}^2} C(u, v) dudv - 3; \rho \in [-1, 1]. \quad (2.6)$$

**Preuve.**  $F(X) = U$  et  $G(Y) = V$  sont deux variables qui suivent la loi  $\mathcal{U}_{\mathbf{I}}$ . Par conséquent leur espérance et leur variance sont respectivement égales à  $1/2$  et  $1/12$ .

Par la suite, on a :

$$\begin{aligned}
\rho(U, V) &= \text{Cov}(U, V) / \sqrt{\mathbb{V}(U)} \sqrt{\mathbb{V}(V)} \\
&= 12 \text{Cov}(U, V) \\
&= 12 (\mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V)) \\
&= 12 \left( \iint_{\mathbf{I}^2} uv dC(u, v) - \frac{1}{4} \right) \\
&= 12 \iint_{\mathbf{I}^2} uv dC(u, v) - 3. \\
&= \rho_C
\end{aligned}$$

CQFD. ■

Le tableau (2.1) présente le tau de Kendall et le rho de Spearman de quelques familles classiques.

Copule	$\tau_C$	$\rho_C$
Clayton	$\theta / (\theta + 2)$	/
Gumbel	$(\theta - 1) / \theta$	/
Frank	$1 - [4(1 - D_1(\theta)) / \theta]$	$1 - [12(D_1(\theta) - D_2(\theta)) / \theta]$
AMH	$[\theta(3\theta - 2) - 2(1 - \theta)^2 \ln(1 - \theta)] / 3\theta^2$	$[12(1 + \theta)L_1(1 - \theta) - L_2] / \theta^2$
Gaussienne	$[2 \arcsin(\sigma)] / \pi$	$[6 \arcsin(\sigma/2)] / \pi$
FGM	$2\theta/9$	$\theta/9$

TAB. 2.1 – Le tau de Kendall et le rho de Spearman de quelques familles classiques.

Avec

$$D_k(\theta) : \text{Fonction de Debye est égale à } \frac{k}{\theta^k} \int_0^\theta \frac{t^k}{\exp(t) - 1} dt.$$

$$L_1(z) : \text{Fonction dilogarithm est égale à } \int_1^z \frac{\ln t}{1 - t} dt.$$

$$L_2 = 24(1 - \theta) \ln(1 - \theta) - 3\theta(\theta + 12).$$

## 2.3 Autres mesures de corrélation

Parmi les autres mesures intéressantes ne dépendant que des rangs, on pourra introduire l'indice de beta de Blomqvist  $\beta$  et le gamma de Gini  $\gamma$ .

### 2.3.1 Beta de Blomqvist

Appelé aussi le coefficient de corrélation de la médiane noté  $\beta$ , a été proposé par Blomqvist [9].

**Définition 2.3.1** Soit une série de  $n$  observations  $\{(x_i, y_i)\}_{1 \leq i \leq n}$  d'un couple  $(X, Y)$ . On définit le beta de Blomqvist par

$$\beta := \frac{n_1 - n_2}{n_1 + n_2}; \beta \in [-1, 1], \quad (2.7)$$

où

$n_1$  : nombre de points situés soit dans le quadrant inférieur gauche ( $Q_3$ ) ou dans le quadrant supérieur droit ( $Q_1$ );

$n_2$  : nombre de points situés soit dans le quadrant supérieur gauche ( $Q_2$ ) ou dans le quadrant inférieur droit ( $Q_4$ ).

$$\left\{ \begin{array}{l} Q_1 := \{(x, y) \in \mathbf{I}^2; \frac{1}{2} < x < 1, \frac{1}{2} < y < 1\}, \\ Q_2 := \{(x, y) \in \mathbf{I}^2; 0 < x < \frac{1}{2}, \frac{1}{2} < y < 1\}, \\ Q_3 := \{(x, y) \in \mathbf{I}^2; 0 < x < \frac{1}{2}, 0 < y < \frac{1}{2}\}, \\ Q_4 := \{(x, y) \in \mathbf{I}^2; \frac{1}{2} < x < 1, 0 < y < \frac{1}{2}\}. \end{array} \right.$$

Donc  $Q_1, Q_3 : (x - \frac{1}{2})(y - \frac{1}{2}) > 0$  et  $Q_2, Q_4 : (x - \frac{1}{2})(y - \frac{1}{2}) < 0$ .

Tout point sur l'axe  $x$  ou  $y$  est dit "n'appartient à aucun quadrant". Pour mieux comprendre voir la figure (2.1).

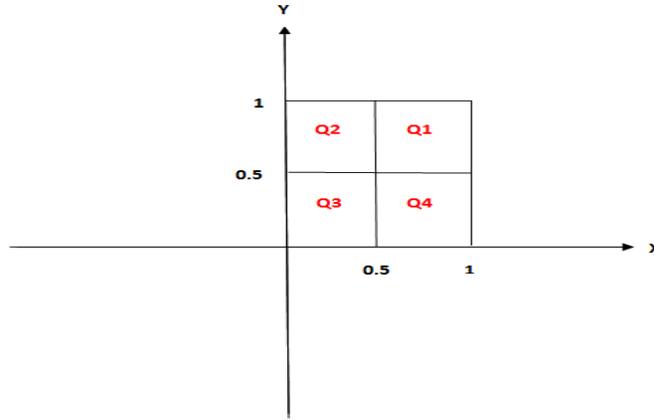


FIG. 2.1 – Les quatre quadrants.

**Définition 2.3.2 (Version population)** Soient  $\tilde{x}$  et  $\tilde{y}$  deux valeurs médianes de  $X$  et  $Y$  respectivement, le coefficient beta de Blomqvist est donné par

$$\beta_{XY} := \mathbb{P}\{(X - \tilde{x})(Y - \tilde{y}) > 0\} - \mathbb{P}\{(X - \tilde{x})(Y - \tilde{y}) < 0\}. \quad (2.8)$$

L'indice beta de Blomqvist peut aussi être exprimé en terme de la copule associée  $C$ .

$$\beta_C := 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1. \quad (2.9)$$

**Preuve.** On montre que les deux formules 2.8 et 2.9 sont équivalentes.

Soient  $X$  et  $Y$  deux vas de marges  $F$  et  $G$  respectivement et de fd jointe  $H$ , on a

$$\begin{aligned} \beta_{X,Y} &= 2[\mathbb{P}(X \leq \tilde{x}, Y \leq \tilde{y}) + \mathbb{P}(X > \tilde{x}, Y > \tilde{y})] - 1 \\ &= 2[H(\tilde{x}, \tilde{y}) + (1 - F(\tilde{x}) - G(\tilde{y}) + H(\tilde{x}, \tilde{y}))] - 1 \\ &= 4H(\tilde{x}, \tilde{y}) - 1; \text{ car } F(\tilde{x}) = G(\tilde{y}) = \frac{1}{2} \\ &= 4C(F(\tilde{x}), G(\tilde{y})) - 1 \\ &= 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1 \\ &= \beta_C. \end{aligned}$$

CQFD. ■

### 2.3.2 Gamma de Gini

Parmi les mesures d'association, il existe une autre mesure que Corrado Gini a proposée [66] nommée l'indice de Gini notée par  $\gamma$ .

**Définition 2.3.3 (Version échantillon)** *Le coefficient gamma de Gini s'écrit en fonction des rangs des observations de  $X$  et  $Y$  respectivement. Il est défini comme suit*

$$\gamma := \frac{1}{\lfloor n^2/2 \rfloor} \left( \sum_{i=1}^n |R_{X_i} + R_{Y_i} - (n+1)| - \sum_{i=1}^n |R_{X_i} - R_{Y_i}| \right); \gamma \in [-1, 1], \quad (2.10)$$

où  $\lfloor t \rfloor$  désigne la partie entière de  $t$ . Le gamma de Gini peut aussi s'écrire en terme de la copule  $C$ .

$$\gamma_C := 2 \iint_{\mathbf{I}^2} (|u+v-1| - |u-v|) dC(u, v),$$

ou bien

$$\gamma_C := 4 \left( \int_{\mathbf{I}} C(u, 1-u) du - \int_{\mathbf{I}} (u - C(u, u)) du \right). \quad (2.11)$$

#### Lien entre les différentes mesures de concordance

Généralement, les valeurs du  $\tau$  et du  $\rho$  pour une copule  $C$  sont différentes. D'après Daniels [22] certaines relations lient les deux mesures de concordance comme suit

$$-1 \leq 3\tau - 2\rho \leq 1. \quad (2.12)$$

**Théorème 2.3.1** *D'après Durbin et Stuart [31], Soit  $X$  et  $Y$  deux vas, alors*

$$\frac{1+\rho}{2} \geq \left( \frac{1+\tau}{2} \right)^2, \quad \frac{1-\rho}{2} \geq \left( \frac{1-\tau}{2} \right)^2. \quad (2.13)$$

En combinant 2.12 et 2.13, on obtient :

$$\left\{ \begin{array}{ll} \frac{3\tau-1}{2} \leq \rho \leq \frac{1+2\tau-\tau^2}{2} & \text{si } \tau \geq 0, \\ \frac{-1+2\tau+\tau^2}{2} \leq \rho \leq \frac{1+3\tau}{2} & \text{si } \tau < 0. \end{array} \right.$$

On a aussi :

$$\begin{cases} \frac{1}{4}(1 + \beta)^2 - 1 \leq \tau \leq 1 - \frac{1}{4}(1 - \beta)^2, \\ \frac{3}{16}(1 + \beta)^3 - 1 \leq \rho \leq 1 - \frac{3}{16}(1 - \beta)^3, \\ \frac{3}{8}(1 + \beta)^2 - 1 \leq \gamma \leq 1 - \frac{3}{8}(1 - \beta)^2. \end{cases}$$

On prends comme exemple deux types de copules : la copule de Marshall-olkin (MO) et la copule de Farlie-Gumbel-Morgenstern (FGM), où on va calculer leurs quatre mesures de dépendance dans la partie suivante.

## 2.4 Copule de Marshall-Olkin

Les copules sont un outil très populaire et largement appliquées dans divers domaines. Par exemple :

- Pour effectuer des tests de robustesse dans des situations où les évènements extrêmes peuvent se produire.
- Dans la modélisation flexible de la structure de dépendance pour les portefeuilles de dimensions élevées.
- Dans l'évaluation des risques financiers et l'analyse actuarielle.
- Dans la formulation de base de données pour l'analyse de la fiabilité.
- Dans diverses études de simulation multidimensionnelles dans les applications d'ingénierie.
- Dans le climat et la recherche de la modélisation météorologique.
- Dans diverses études de performance basées sur la simulation.

On a choisi d'utiliser comme copule bivariée à deux paramètres; la copule de MO où on va étudier certaines de ses propriétés. Cette copule décrit une distribution exponentielle bivariée, pour étudier le système complexe dans lesquels les deux composants ne sont pas indépendants et sont soumis à des chocs qui sont fatales, pour une des deux composantes ou pour les deux en même temps.

Soient  $X$  et  $Y$  les durées de vie de deux composantes dans un système. On suppose que les chocs sur les deux composantes suivent trois processus de Poisson indépendant avec des paramètres

$\lambda_1, \lambda_2$  et  $\lambda_{12} \geq 0$ , selon que le choc n'affecte que la composante 1, seulement la composante 2 ou les deux composantes simultanément. Par conséquent, la fonction de survie est

$$\forall x, y \geq 0 : \bar{H}(x, y) = \mathbb{P}(X > x, Y > y) = \exp[-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)],$$

avec  $\max(x, y) = x + y - \min(x, y)$ .

Donc :

$$\begin{aligned} \bar{H}(x, y) &= \exp[-(\lambda_1 + \lambda_{12})x - (\lambda_2 + \lambda_{12})y + \lambda_{12} \min(x, y)] \\ &= \bar{F}(x)\bar{G}(y) \min[\exp(\lambda_{12}x), \exp(\lambda_{12}y)], \end{aligned}$$

où  $\bar{F}$  est  $\bar{G}$  sont les fonctions de survie des marginales de  $X$  et  $Y$  respectivement. Elle sont égales à

$$\bar{F}(x) = \exp[-(\lambda_1 + \lambda_{12})x] \text{ et } \bar{G}(y) = \exp[-(\lambda_2 + \lambda_{12})y].$$

Il convient alors de rappeler que les vas  $X$  et  $Y$  sont obtenues à partir des vas indépendantes et exponentiellement distribuées  $Z_1, Z_2$  et  $Z_{12}$  avec les paramètres  $\lambda_1, \lambda_2$  et  $\lambda_{12}$  respectivement, par

$$X = \min\{Z_1, Z_{12}\} \text{ et } Y = \min\{Z_2, Z_{12}\}.$$

On pose  $a = \lambda_{12}/\lambda_1 + \lambda_{12}, b = \lambda_{12}/\lambda_2 + \lambda_{12}, \bar{F}(x) = u$  et  $\bar{G}(y) = v$ . La copule de survie  $\hat{C}_{a,b}$  associée à la fonction de survie  $\bar{H}(x, y)$  est

$$\hat{C}_{a,b}(u, v) = uv \min(u^{-a}, v^{-b}) = \min(u^{1-a}v, uv^{1-b}); 0 < a, b < 1.$$

La copule de survie de la distribution exponentielle bivariée, conduit à une famille de copules à deux paramètres

$$C_{a,b}(u, v) = \begin{cases} u^{1-a}v; & u^a \geq v^b, \\ uv^{1-b}; & u^a \leq v^b. \end{cases}$$

Lorsque  $a = b$  correspondant au cas où  $\lambda_1 = \lambda_2$ , la copule est connue sous le nom de Generalized Cuadras-Augé. Autrement dit,  $X$  et  $Y$  sont échangeables. Pour certaines lectures pertinentes intéressantes voir Hutchinson et Lai [55], Embrechts [32] et Nelsen [66].

### 2.4.1 Propriétés de la copule

On va présenter et prouver certaines propriétés de la copule  $C_{a,b}$  à l'égard des distributions de probabilité, des mesures de dépendance et des dépendances de queue.

1. **Distribution de probabilité** : Contrairement aux copules présentées jusqu'à maintenant, cette copule a à la fois une continuité absolue et une composante singulière. Pour la composante absolument continue, on peut calculer la densité de la copule  $C_{a,b}$  comme suit

$$c_{a,b}(u, v) = \begin{cases} (1-a)u^{-a}; & u^a > v^b, \\ (1-b)v^{-b}; & u^a < v^b. \end{cases}$$

La figure (2.2) représente la copule de MO. À gauche on a le graphe de la densité de la copule, à droite on a une simulation de 2000 vas avec  $a = 0.25$  et  $b = 0.75$ , ou bien avec  $\lambda_1 = 1.1$ ,  $\lambda_2 = 0.2$  et  $\lambda_{12} = 0.6$ .

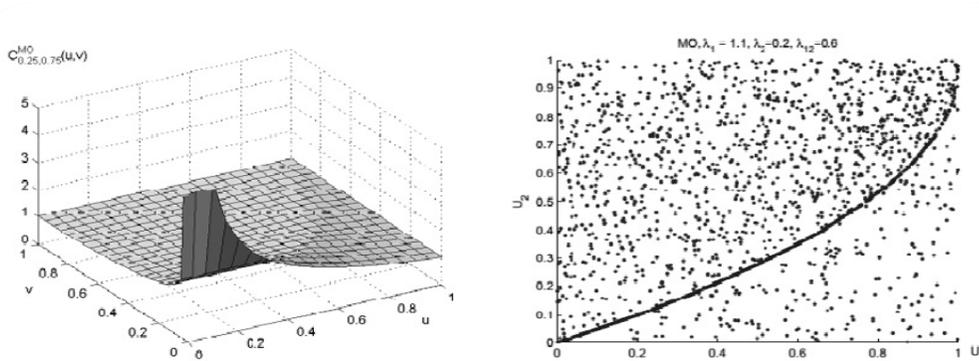


FIG. 2.2 – Copule de MO. Gauche : Densité de la copule. Droite : Simulation de 2000 vas avec  $a = 0.25$  et  $b = 0.75$ .

La composante singulière distribue la masse que sur la courbe  $u^a = v^b$ . Plus précisément, on peut montrer que <sup>1</sup>

$$\mathbb{P}(U = V^{b/a}) = \frac{ab}{a + b - ab} = \frac{\lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}}.$$

<sup>1</sup>Voir Nelsen (1999), Section 3.1.1.

Cet effet devient visible à partir des données simulés présentées dans la figure (2.2) (à droite). La partie discontinue des résultats de la copule dans les points de données se trouvent souvent sur le répertoire de la courbe  $u = v^{b/a}$ , pour plus de détail voir Embrechts et al. [32].

**2. Mesures de dépendance :** D'après Hoeffding [51], Sklar [75] et Schweizer et Wolff [72], la plupart des mesures d'association sont d'échelle-invariant par rapport aux transformations strictement croissantes des variables, comme le  $\tau$  de Kendall et le  $\rho$  de Spearman (voir la section 2.2). On présente quelques résultats de la copule MO liés aux mesures de dépendance.

- Tau de Kendall :  $\tau_{C_{a,b}} = 1 - 4 \iint_{\mathbf{I}^2} \frac{\partial C_{a,b}(u,v)}{\partial u} \frac{\partial C_{a,b}(u,v)}{\partial v} dudv$ , alors

$$\tau_{C_{a,b}} = \frac{ab}{a+b-ab}.$$

- Rho de Spearman :  $\rho_{C_{a,b}} = 12 \iint_{\mathbf{I}^2} uv dC_{a,b}(u,v) - 3$ , alors

$$\rho_{C_{a,b}} = \frac{3ab}{2a+2b-ab}.$$

Pour les preuves de  $\tau_{C_{a,b}}$  et de  $\rho_{C_{a,b}}$  voir Nelsen ([66], p : 165-168).

- Beta de Blomqvist :  $\beta_{C_{a,b}} = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1$ , alors

$$\beta_{C_{a,b}} = 4 \times 2^{\min(a,b)-2} - 1.$$

Bien évidemment

$$\begin{aligned} C_{a,b}(u, u) &= \min(u^{2-a}, u^{2-b}) \\ &= u^{\max(2-a, 2-b)}, \text{ car } 0 < a, b < 1 \text{ et } u \in \mathbf{I} \\ &= u^{2-\min(a,b)}. \end{aligned}$$

Donc

$$\begin{aligned} \beta_{C_{a,b}} &= 4 \left(\frac{1}{2}\right)^{2-\min(a,b)} - 1 \\ &= 4 \times 2^{\min(a,b)-2} - 1. \end{aligned}$$

- Gamma de Gini :  $\gamma_{C_{a,b}} = 4 \left( \int_{\mathbf{I}} C_{a,b}(u, 1-u) du - \int_{\mathbf{I}} (u - C_{a,b}(u, u)) du \right)$ , alors

$$\boxed{\gamma_{C_{a,b}} = \frac{4}{3 - \min(a, b)} - 2 + 4 \times D,}$$

avec

$$D = \frac{4}{3 - \min(a, b)} - 2 + 4 \left[ \frac{1}{(3-a)} u_0^{(3-a)} + \frac{1}{(3-b)} (1-u_0)^{(3-b)} - \frac{1}{(2-a)} u_0^{(2-a)} - \frac{1}{(2-b)} (1-u_0)^{(2-b)} + \frac{1}{(2-a)} + \frac{1}{(2-b)} - \frac{1}{(3-a)} - \frac{1}{(3-b)} \right],$$

et  $u_0$  est la solution de  $(1-u)^b = u^a$  i.e.  $u^{a/b} + u - 1 = 0$ .

**Preuve.** Le gamma de Gini est égale à :  $4 \int_{\mathbf{I}} C_{a,b}(u, 1-u) du - 2 + 4 \int_{\mathbf{I}} C_{a,b}(u, u) du$ .

$$\text{Par conséquent : } 4 \int_{\mathbf{I}} C_{a,b}(u, u) du = 4 \int_{\mathbf{I}} u^{2-\min(a,b)} du = \frac{4}{3 - \min(a, b)}.$$

$$\begin{aligned} \text{On calcule maintenant : } C_{a,b}(u, 1-u) &= \min((1-u)u^{1-a}, u(1-u)^{1-b}) \\ &= u^{1-a}(1-u)^{1-b} \min((1-u)^b, u^a). \end{aligned}$$

Avec les conditions sur  $u$ ,  $a$  et  $b$  la solution  $u_0$  est unique, alors

$$\min((1-u)^b, u^a) = \begin{cases} u^a & ; u_0 \geq 0, \\ (1-u)^b & ; u_0 < 0. \end{cases} = \begin{cases} u^a & ; u \leq u_0, \\ (1-u)^b & ; u > u_0. \end{cases}$$

$$\text{Donc : } C_{a,b}(u, 1-u) = \begin{cases} u(1-u)^{1-b} & ; u \leq u_0, \\ (1-u)u^{1-a} & ; u > u_0. \end{cases}$$

$$\begin{aligned} \text{Enfin : } \int_{\mathbf{I}} C_{a,b}(u, 1-u) du &= \int_{[0, u_0]} u(1-u)^{1-b} du + \int_{[u_0, 1]} (1-u)u^{1-a} du \\ &= \frac{1}{(3-a)} u_0^{(3-a)} + \frac{1}{(3-b)} (1-u_0)^{(3-b)} - \frac{1}{(2-a)} u_0^{(2-a)} - \frac{1}{(2-b)} (1-u_0)^{(2-b)} + \frac{1}{(2-a)} \\ &\quad + \frac{1}{(2-b)} - \frac{1}{(3-a)} - \frac{1}{(3-b)} = D. \end{aligned}$$

$$\text{On aboutit au faite que : } \gamma_{C_{a,b}} = \frac{4}{3 - \min(a, b)} - 2 + 4 \times D.$$

CQFD. ■

**Remarque 2.4.1** Comme  $u_0$  est une fonction de  $a/b$ . On peut également obtenir une idée sur  $u_0(a/b)$  en traçant son inverse  $a/b = \frac{\log(1-u)}{\log(u)}$ .

**Corollaire 2.4.1** Dans le cas où  $a = b$  et  $u_0 = 1/2$ , les mesures de dépendance de la copule MO à un paramètre  $a$  sont

$\tau_{C_{a,b}}$	$\rho_{C_{a,b}}$	$\beta_{C_{a,b}}$	$\gamma_{C_{a,b}}$
$a/(2-a)$	$3a/(4-a)$	$2^a - 1$	$\frac{2(2+3a-a^2) - (4-a)2^a}{(3-a)(2-a)}$

La relation entre le paramètre de la copule MO et les quatre mesures de concordance est représentée graphiquement dans la figure (2.3).

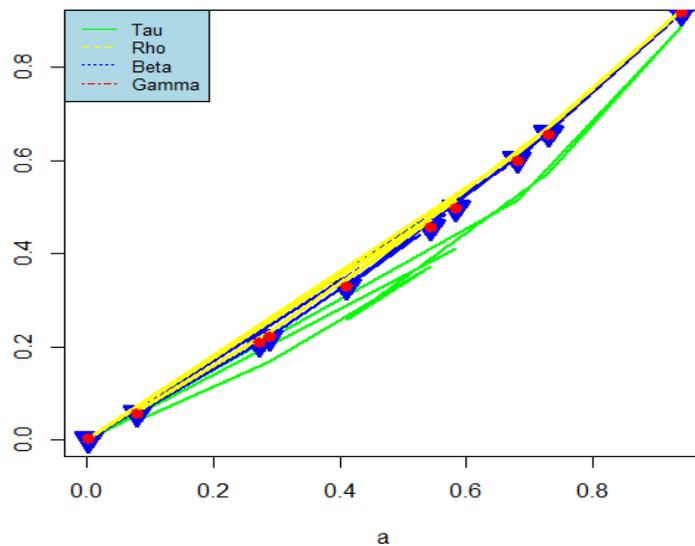


FIG. 2.3 – Relation entre le paramètre de la copule et les quatre mesures de concordance.

**3. Dépendance de queue :** Le concept de la dépendance de queue de la copule  $C_{a,b}$  est de mesurer la dépendance entre les variables  $u$  et  $v$  dans le quadrant supérieur droit et le quadrant inférieur gauche dans  $\mathbf{I}^2$ . La copule de MO n'a pas de dépendance de queue inférieure ( $\lambda_L = 0$ ) par contre elle a une dépendance de queue supérieure ( $\lambda_U = \min(a, b)$ ,  $\lambda_U \in ]0, 1[$ ).

**Preuve.** A partir de la définition des dépendances de queue dans la section 1.1.3 on a :

- Dépendance de queue inférieure :

$$\begin{aligned}\lambda_L &= \lim_{u \rightarrow 0^+} C_{a,b}(u, u)/u \\ &= \lim_{u \rightarrow 0^+} u^{2-\min(a,b)}/u \\ &= 0, \text{ car } 1 - \min(a, b) > 0.\end{aligned}$$

- Dépendance de queue supérieure :

$$\begin{aligned}\lambda_U &= \lim_{u \rightarrow 1^-} [1 - 2u + C_{a,b}(u, u)] / [1 - u] \\ &= 2 - \lim_{u \rightarrow 1^-} [1 - C_{a,b}(u, u)] / [1 - u] \\ &= 2 - \lim_{u \rightarrow 1^-} [1 - u^{2-\min(a,b)}] / [1 - u] \\ &= 2 - \lim_{u \rightarrow 1^-} \frac{-[2 - \min(a, b)] u^{1-\min(a,b)}}{-1}, \text{ par la règle d'Hopital} \\ &= \min(a, b).\end{aligned}$$

CQFD. ■

**Remarque 2.4.2** La copule MO est une copule de valeurs extrêmes, sa fonction de dépendance Archimax  $A_{a,b}(t) = \max\{1 - at, 1 - b(1 - t)\}$ .

## 2.5 Copule de Farlie-Gumbel-Morgenstern

Une des familles paramétriques les plus populaires des copules est la famille de FGM. Elle a été d'abord étudiée par Morgenstern [65] puis par Farlie [34] et Gumbel [50]. Kim et Sungur [60] ont utilisé cette famille comme modèle impliquant la censure. Chalabian et Dunnington [14] et Blischke et Prabhakar Murthy [8] ont aussi évoqué son utilisation dans le domaine médicale, en ingénierie, assurance et même en hydrologie,... Elle est intéressante en raison de sa simplicité. Cependant, elle est restrictive parce que cette copule n'est utile que lorsque la dépendance entre les deux marginales est de grandeur modeste. La copule de FGM est définie par

$$C_a(u, v) = uv[1 + a(1 - u)(1 - v)]; a \in [-1, 1],$$

si  $a = 0$  alors  $C_a(u, v) = \Pi$ . Sa densité est donnée par

$$c_a(u, v) = [1 + a(2u - 1)(2v - 1)].$$

Durante et al. [30] ont montré une limitation bien connue de cette famille car elle ne permet pas la modélisation des dépendances élevées étant donné que le coefficient de corrélation est limité à  $\frac{a}{3}$  i.e. que le  $\rho \in [-0.33, 0.33]$ .

Un certain nombre d'extensions de la distribution FGM ont été développées principalement pour augmenter la valeur maximale du coefficient de corrélation. La plupart d'entre elles sont les copules de type polynôme (copules qui sont exprimées en termes de polynômes en  $u$  et  $v$ ) pour plus de détail voir Huang et Kotz [53].

### 2.5.1 Itération des distributions de FGM

Afin d'augmenter le nombre de paramètres de cette famille, Johnson et Kotz [59] ont introduit la  $(r - 1)$ <sup>ième</sup> itération où le paramètre  $a$  devient de dimension  $r$  i.e.  $a = (a_1, \dots, a_r)$  :

$$C_a(u, v) = uv + \sum_{j=1}^r a_j (uv)^{[j/2]+1} [(1-u)(1-v)]^{[(j+1)/2]}; -1 \leq a_j \leq 1,$$

où  $[z]$  représente le plus grand nombre entier inférieur ou égal à  $z$ . Ce processus peut bien entendu être répété.

Par exemple, pour une itération de la famille de FGM, on a le modèle de la copule à deux paramètres suivants

$$C_{a_1, a_2}(u, v) = uv [1 + a_1(1-u)(1-v) + a_2 uv(1-u)(1-v)].$$

Pour tout ce qui suit concernant la copule de FGM à deux paramètres on pose,  $a_1 = a$  et  $a_2 = b$ .

La fd  $C_{a,b}$  est donnée par

$$C_{a,b}(u, v) = uv [1 + a(1-u)(1-v) + buv(1-u)(1-v)], \quad (2.14)$$

où les combinaisons valides de  $a$  et  $b$  sont  $-1 \leq a \leq 1$  et  $-1-a \leq b \leq \frac{1}{2}(3-a+\sqrt{9-6a-3a^2})$ .

Pour plus de détails sur la distribution de FGM, se référer à Lin [63], Huang et Kotz [52].

## 2.5.2 Propriétés de la copule

On va présenter et prouver certaines propriétés de la copule de FGM à l'égard des distributions de probabilité, des mesures de dépendance et des dépendances de queue.

1. **Distribution de probabilité** : La densité de la copule  $C_{a,b}$  est définie comme suit

$$c_{a,b}(u, v) = (1 + a(1 - 2u)(1 - 2v) + buv(2 - 3u)(2 - 3v)).$$

La figure (2.4) représente la copule de FGM. À gauche la Densité de la copule, et à droite simulation de 2000 vas avec  $a = 0.5$  et  $b = 1.2$ .

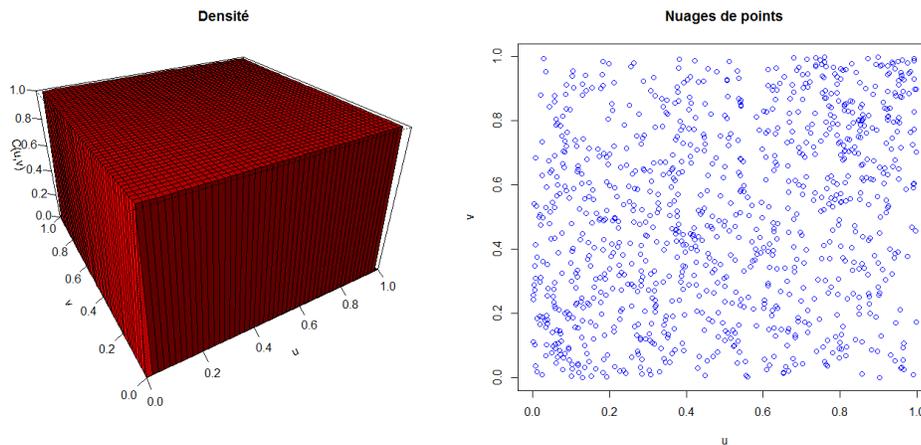


FIG. 2.4 – Copule de FGM. Gauche : Densité de la copule. Droite : Simulation de 2000 vas avec  $a = 0.5$  et  $b = 1.2$ .

Il est clair que cette copule n'a pas de dépendance de queue, ni à droite ni à gauche.

2. **Mesures de dépendance** : Les quatre mesures de dépendance de la copule sont les suivantes

- Tau de Kendall :  $\tau_{C_{a,b}} = 1 - 4 \iint_{\mathbf{I}^2} \frac{\partial}{\partial u} C_{a,b}(u, v) \frac{\partial}{\partial v} C_{a,b}(u, v) dudv$ , alors

$$\tau_{C_{a,b}} = \frac{2a}{9} + \frac{b}{18} + \frac{ab}{450}.$$

- Rho de Spearman :  $\rho_{C_{a,b}} = 12 \iint_{\mathbf{I}^2} uv dC_{a,b}(u, v) - 3$ , alors

$$\rho_{C_{a,b}} = \frac{a}{3} + \frac{b}{12}.$$

- Beta de Blomqvist :  $\beta_{C_{a,b}} = 4C_{a,b}\left(\frac{1}{2}, \frac{1}{2}\right) - 1$ , alors

$$\beta_{C_{a,b}} = \frac{a}{4} + \frac{b}{16}.$$

- Gamma de Gini :  $\gamma_{C_{a,b}} = 4 \left[ \int_{\mathbf{I}} C_{a,b}(u, 1-u) du - \int_{\mathbf{I}} (u - C_{a,b}(u, u)) du \right]$ , alors

$$\gamma_{C_{a,b}} = \frac{4a}{15} + \frac{b}{15}.$$

3. **Dépendance de queue** : La copule de FGM n'a ni de dépendance de queues inférieure ni supérieure ( $\lambda_L = \lambda_U = 0$ ).

**Preuve.** A partir de la définition des dépendances de queue dans la section 1.1.3 on a :

- Dépendance de queue inférieure :

$$\begin{aligned} \lambda_L &= \lim_{u \rightarrow 0^+} C_{a,b}(u, u)/u \\ &= \lim_{u \rightarrow 0^+} (u^2 [1 + (1-u)^2 (a + bu^2)]) / u \\ &= 0. \end{aligned}$$

- Dépendance de queue supérieure :

$$\begin{aligned}\lambda_U &= \lim_{u \rightarrow 1^-} (1 - 2u + C(u, u)) / (1 - u) \\ &= \lim_{u \rightarrow 1^-} (1 - 2u + u^2 [1 + (1 - u)^2 (a + bu^2)]) / (1 - u) \\ &= \lim_{u \rightarrow 1^-} \frac{-2 + 2u + 2au(1 - u)(1 - 2u) + 2bu(1 - u)(2 - 3u)}{-1}, \text{ par la règle d'Hopital} \\ &= 0.\end{aligned}$$

CQFD. ■

## Chapitre 3

# Association measures and estimation of copula parameters

On va appliquer la méthode d'inversion des mesures d'association, avec plusieurs combinaisons de deux parmi les quatre mesures les plus populaires, pour estimer les paramètres des copules dans le cas de distributions bivariées. On va réaliser une étude de simulation sur deux exemples, à savoir les copules à deux paramètres de Farlie-Gumbel-Morgenstern et de Marshall-Olkin, pour faire des comparaisons entre les estimateurs, en matière du biais (*bias*) et de la racine de l'erreur quadratique moyenne (*rmse*), qui sont définis comme suit :

$$bias = \frac{1}{r} \sum_{j=1}^r (\hat{\theta}_j - \theta), \quad rmse = \left( \frac{1}{r} \sum_{j=1}^r (\hat{\theta}_j - \theta)^2 \right)^{1/2},$$

où  $\theta$  est la valeur exacte (à estimer),  $\hat{\theta}$  est la valeur approchée (l'estimateur) et  $r$  est le nombre de réplifications. L'estimateur le plus efficace est simplement celui qui a l'*rmse* la plus petite <sup>1</sup>.

### Abstract

We apply the inversion method of estimation, with several combinations of two among the four most popular association measures, to estimate the parameters of copulas in the case of bivariate distributions. We carry out a simulation study with two examples, namely Farlie-

---

<sup>1</sup>Benelmir, I., Meraghni, D., 2016. Association measures and estimation of copula parameters. *Journal Afrika Statistika* **11** (1), 933-942.

Gumbel-Morgenstern and Marshall-Olkin two-parameter copulas to make comparisons between the obtained estimators, with respect to bias and root of the mean squared error.

### 3.1 Introduction

Dependence relations between random variables (rv's) are one of the most important issues that got a great deal of interest in probability and statistics. Such dependence is appropriately modelled by a very useful tool for handling multivariate distributions with given univariate marginals, known as copula. As mentioned by Fisher [38], the copulas are of interest to statisticians for two main reasons : first, they represent a way of studying scale-free measures of dependence and second, they may be considered as a starting point for constructing families of multivariate distributions. The concept of copulas has become very useful in real life applications as diverse as risk management, reliability, survival analysis, finance, actuarial and medical sciences. An exhaustive list of copula applications can be found in, for instance, Balakrishnan and Lai [3], pages 55-58, with full details. The notion of copula is used by Brahim and al. [10] to analyze the distortion risk measures of the sum of two or more insurance losses, where the dependence structure is a very significant factor. A copula is a mean of linking a multivariate distribution function (df) with its margins. Indeed, if  $X^{(1)}, \dots, X^{(d)}$  are  $d \geq 2$  rv's with joint df  $F$  and margins  $F_i$ ,  $i = 1, \dots, d$ , then according to Sklar's Theorem [75], there exists a copula  $C : [0, 1]^d \rightarrow [0, 1]$  such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

In other words, the copula  $C$  is the joint df of the  $(0, 1)$ -uniform rv's  $U_i := F_i(X^{(i)})$ ,  $i = 1, \dots, d$ . It is defined on  $[0, 1]^d$  by

$$C(u_1, \dots, u_d) = F(F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d)),$$

where  $F_i^{\leftarrow}(s) := \inf\{x \in \mathbb{R} \setminus F_i(x) \geq s\}$ ,  $0 \leq s \leq 1$ , denotes the generalized inverse or quantile function of df  $F_i$ . Note that if all the margins  $F_i$  are continuous, then  $C$  is unique. In the sequel and for the sake of simplicity, we will restrict ourselves to the two-dimensional case, i.e. we take

$d = 2$ . For a full description of copulas and comprehensive details on their properties, we refer the reader to the textbooks of Cherubini and al [17], Joe [57] and Nelsen [66]

In the process of modelling the dependence between two (ore more) rv's, we should look for a copula with appropriate parametric form. For each specific problem, two of the fundamental characteristics required in our search are flexibility and analytical simplicity. This leads to the availability of various families of copulas, among which we may cite, in addition of the trivial ones (product, maximum and minimum), the elliptical copulas (Gauss, Student,...), Archimediens copulas (Gumbel-Hougaard, Clayton, Frank,...) and extreme value copulas. Any copula  $C$  is delimited by what is called Hoeffding and Fréchet bounds generally denoted by  $W$  and  $M$ , which are trivial copulas known as minimum and maximum copulas respectively. That is, for  $0 \leq u, v \leq 1$ , we have

$$W(u, v) \leq C(u, v) \leq M(u, v),$$

where  $W(u, v) := \max(u + v - 1, 0)$  and  $M(u, v) := \min(u, v)$ .

The central issue in statistical modelling is the estimation of the parameters upon which the probability distribution depends. In the case of copulas, there is a variety of estimation procedures available in the literature. Depending on the situation, one may consider parametric, semi-parametric or non-parametric copula inference methods based on independent and identically distributed observations of random vectors with dependent components. The forms of the joint distribution and its marginals play a crucial role in choosing the right estimation approach. For this matter, the authors of Choros and al. [18] provide a detailed survey that may be summarized into :

- Parametric model : exact maximum likelihood methods, inference from likelihoods for margins and inversion method of association measures.
- Semi-parametric model : maximum pseudo-likelihood and canonical maximum likelihood.
- Non-parametric model : empirical copula processes.

We must emphasize that, in this paper, we don't directly deal with the estimation of the copulas themselves, but we are rather concerned with estimating their parameters. In multivariate

statistical analysis, the inference on copula parameters represents a major topic that has got a great deal of interest from several authors. For an overview of the different estimation methods, see the introductions of Brahim and Necir [11] and Brahim and al [12] (and the references therein) where the authors recently applied the notions of moments and bivariate L-moments (BLM's) of copulas to provide new parameter estimators which they compared to the already existing ones. However, a close look at their comparison results shows that the conclusions are not always against the inversion method. Indeed, there are instances where this latter produces estimators of better performance, mainly from the root of the mean squared error (rmse) perspective (this will be confirmed later on in Section 3.3). This suggests that one should be more or at least equally attracted by this method of copula parameters estimation. In addition to this, we think that the inversion approach still has the advantage because it is based on the concordance coefficients which are familiar quantities, expressed as simple functions of the copula and whose empirical counterparts are implemented in most statistical software such as **R**, for simulations and application needs. All this motivated us to focus on the inversion method and to look for the combination of association measures that yield the most accurate parameter estimates. Another reason of our interest in this method is the question raised in the introduction of Brahim and Necir [11] with respect to the choice of the appropriate association measures with which the estimation is to be made. Thus, this work may be seen as an attempt to answer that question. Finally, note that, in our study, we consider the bivariate case with two particular classes of copulas, namely Farlie-Gumbel-Morgenstern (FGM) and Marshall-Olkin (MO) two-parameter copulas, that we briefly describe below.

### 3.1.1 Marshall-Olkin copula

This copula, also known as generalized Cuadras-Augé copula, is mainly used in reliability, finance, insurance. . . It originates from a concrete model assumption which can easily be used to simulate pseudo rv's. It may be recalled that rv's  $X$  and  $Y$  with a MO copula are obtained from independent and exponentially distributed rv's  $Z_1$ ,  $Z_2$  and  $Z_{12}$ , with respective parameters

$\lambda_1$ ,  $\lambda_2$  and  $\lambda_{12}$ , by  $X := \min(Z_1, Z_{12})$  and  $Y := \min(Z_2, Z_{12})$ . The respective df's of  $X$  and  $Y$

$$F(x) = 1 - \exp\{-(\lambda_1 + \lambda_{12})x\} \text{ and } G(y) = 1 - \exp\{-(\lambda_2 + \lambda_{12})y\},$$

are readily obtained. Therefore, MO copula with two parameters  $a$  and  $b$  is given by

$$C_{a,b}^{MO}(u, v) = \min\{vu^{1-a}, uv^{1-b}\} = \begin{cases} vu^{1-a}, & u^a \geq v^b, \\ uv^{1-b}, & u^a \leq v^b, \end{cases}$$

where  $a := \lambda_{12}/(\lambda_1 + \lambda_{12})$  and  $b := \lambda_{12}/(\lambda_2 + \lambda_{12})$  meaning that  $0 \leq a, b \leq 1$ . This copula has the particularity to possess an absolutely continuous component and a singular one. Note that when  $a = b$ , we have  $\lambda_1 = \lambda_2$  and so  $X$  and  $Y$  are exchangeable. For more details on MO copulas, we refer to Embrechts and al. [32] and Marshall and Olkin [64].

### 3.1.2 Farlie-Gumbel-Morgenstern copula

This is one of the most popular parametric families of copulas, which is particularly applied in insurance, hydrology, the health field, . . . The FGM family with one parameter  $a$  is defined by

$$C_a^{FGM}(u, v) = uv [1 + a(1 - u)(1 - v)], \quad -1 \leq a \leq 1.$$

The corresponding correlation coefficient is equal to  $a/3$ , which clearly lies in the interval  $[-1/3, 1/3]$ . In other words, the maximum degree of dependence that an FGM copula can describe does not exceed one third (in absolute value). This represents a limitation to this family as it does not allow the modeling of high dependences.

**Iterated FGM copula.** For some integer  $r \geq 1$ , Johnson and Kotz [59] introduced the  $(r - 1)$ -iterated FGM family with  $r$ -dimensional parameter  $\theta = (\theta_1, \dots, \theta_r)$ , where  $|a| \leq 1$  such us

$$C_{\theta}^{FGM}(u, v) = uv + \sum_{j=1}^r \theta_j (uv)^{[j/2]+1} (\bar{u}\bar{v})^{[j/2+1/2]},$$

where  $\bar{w} := 1 - w$  and  $[t]$  denotes the greatest integer less than or equal to  $t$ . For  $r = 2$  and  $\theta = (a, b)$ , we get the one-iterated FGM family as follows

$$C_{a,b}^{FGM}(u, v) = uv[1 + (1 - u)(1 - v)(a + buv)],$$

where the valid combinations of  $a$  and  $b$  are  $-1 \leq a \leq 1$  and  $-1 - a \leq b \leq (3 - a + \sqrt{9 - 6a - 3a^2})/2$ . For further details on FGM distributions, we refer to Bekrizadeh and al. [4], Huang and Kotz [52] and Lin [63]

## 3.2 Association measures

Four of the most common non-parametric measures of association between the components of a continuous random vector  $(X, Y)$  are Kendall's tau  $\tau$ , Spearman's rho  $\rho$ , Gini's gamma  $\gamma$  and Blomqvist's beta  $\beta$  (also known as the medial correlation coefficient). These measures, which only depend on the copula  $C$  pertaining to the pair  $(X, Y)$ , are respectively equal to

$$\tau = 1 - 4 \iint \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v) dudv, \quad \rho = 12 \iint uvC(u, v) dudv - 3, \quad (3.1)$$

$$\gamma = 4 \left[ \int C(u, 1 - u) du - \int (u - C(u, u)) du \right] \quad \text{and} \quad \beta = 4 \times C\left(\frac{1}{2}, \frac{1}{2}\right) - 1. \quad (3.2)$$

It is noteworthy that the aforementioned coefficients lie between  $-1$  and  $1$ . For a further discussion of their properties, see, for instance, Nelsen [66] Applying the formulas (3.1) and (3.2) to MO and FGM copulas, with two parameters  $a$  and  $b$ , yields the values that we summarize in Table(3.1), where

$$\lambda := \frac{u_0^{3-a}}{3-a} + \frac{(1-u_0)^{3-b}}{3-ab} - \frac{u_0^{2-a}}{2-a} - \frac{(1-u_0)^{2-b}}{2-b} + \frac{1}{2-a} + \frac{1}{2-b} - \frac{1}{3-a} - \frac{1}{3-b},$$

with  $u_0$  being solution of  $u^a - (1 - u)^b = 0$ .

	MO copula	FGM copula
Kendall's $\tau$	$ab / (a + b - ab)$	$(100a + 25b + ab) / 450$
Spearman's $\rho$	$3ab / (2a + 2b - ab)$	$(4a + b) / 12$
Gini's $\gamma$	$4\lambda - 2 + \frac{4}{3 - \min(a, b)}$	$(4a + b) / 15$
Blomqvist's $\beta$	$2^{\min(a, b)} - 1$	$(4a + b) / 16$

TAB. 3.1 – Association measures of FGM and MO copulas with two parameters.

### 3.2.1 Kendall's tau

Kendall's tau is defined in terms of concordance as follows. A sample of size  $n \geq 2$  is drawn from a random vector  $(X, Y)$ , then there are  $\binom{n}{2}$  distinct pairs  $(x_i, y_i)$  and  $(x_j, y_j)$  in this sample. If  $(x_i - x_j)(y_i - y_j) > 0$  the pairs are said to be concordant, otherwise they are discordant. Let  $n_c$  be the number of concordant pairs and  $n_d$  that of discordant ones. Then the sample Kendall's tau is defined as

$$\tau_{emp} := \frac{2}{n(n-1)}(n_c - n_d).$$

Note that  $n_c - n_d = \sum_{j=2}^n \sum_{i=1}^{j-1} \text{sign} \{(x_i - x_j)(y_i - y_j)\}$ , where  $\text{sign} \{z\} = 1$  if  $z > 0$  and  $-1$  if  $z < 0$ . If  $(x_i - x_j)(y_i - y_j) = 0$  the pair is neither concordant nor discordant.

### 3.2.2 Spearman's rho

The empirical Spearman's rho is defined as the correlation coefficient of the ranks of  $X$  and  $Y$ . It is equal to

$$\rho_{emp} := 1 - \frac{6 \sum_{i=1}^n (R_{x_i} - R_{y_i})^2}{n(n^2 - 1)},$$

where  $R_{x_i}$  and  $R_{y_i}$  are the ranks of  $X$  and  $Y$  respectively.

### 3.2.3 Gini's gamma

The sample Gini's gamma is defined by

$$\gamma_{emp} := \frac{1}{\lfloor n^2/2 \rfloor} \left\{ \sum_{i=1}^n |R_{x_i} + R_{y_i} - (n+1)| - \sum_{i=1}^n |R_{x_i} - R_{y_i}| \right\}.$$

### 3.2.4 Blomqvist's beta

The empirical Blomqvist's beta is defined in terms of the four quadrants as follows

$$\beta_{emp} := \frac{n_1 - n_2}{n_1 + n_2},$$

where  $n_1$  is the number of points located in either the upper right quadrant ( $Q_1$ ) or the lower left quadrant ( $Q_3$ ) and  $n_2$  is the number of points located in either the upper left quadrant ( $Q_2$ ) or the lower right quadrant ( $Q_4$ ). If we set  $\mathbf{I} := [0, 1]$ , then we have

$$Q_1 := \{(x, y) \in \mathbf{I}^2; \frac{1}{2} < x < 1, \frac{1}{2} < y < 1\}, \quad Q_2 := \{(x, y) \in \mathbf{I}^2; 0 < x < \frac{1}{2}, \frac{1}{2} < y < 1\},$$

$$Q_3 := \{(x, y) \in \mathbf{I}^2; 0 < x < \frac{1}{2}, 0 < y < \frac{1}{2}\}, \quad Q_4 := \{(x, y) \in \mathbf{I}^2; \frac{1}{2} < x < 1, 0 < y < \frac{1}{2}\}.$$

In other words,  $Q_1$  and  $Q_3$  are defined by the inequality  $(x - \frac{1}{2})(y - \frac{1}{2}) > 0$ , whereas  $Q_2$  and  $Q_4$  satisfy the inequality  $(x - \frac{1}{2})(y - \frac{1}{2}) < 0$ .

## 3.3 Simulation comparative study

This study is intended to evaluate and compare, with respect to bias and rmse, the copula parameter estimators obtained by the inversion procedure using all possible combinations of two association measures chosen among Kendall's tau, Spearman's rho, Gini's gamma and Blomqvist's beta. To this end, we make use of several packages of the statistical software **R**. Copula values are generated by means of the package **copula** and parameter estimates are calculated using the package **BB**, which permits to solve systems of nonlinear equations. Also, the package **asbio** was needed for the computation of the empirical Blomqvist's beta. On the other hand, for comparison purposes as well, we use the package **lmomco** to do the computations relative to BLM based estimators. This method of copula parameter estimation was very recently introduced in article of Brahim and al. [12], where the authors provide complete details on the notion of BLM's of copulas.

The choice of the parameters  $a$  and  $b$  is made in such a way that we cover the three dependence

cases, namely weak, moderate and strong. For each copula, we generate 1000 samples of size  $n = 100$ , and for every combination of two association measures we determine the estimate values of  $a$  and  $b$  and then we compute their biases and rmse's. We do likewise with the BLM estimator. Our overall results, summarized in Tables (3.2), (3.3) and (3.4) for MO copula and Tables (3.5), (3.6) and (3.7) for FGM copula, are taken as the empirical means of the results obtained through all repetitions. The same simulation procedure is repeated for a second sample size  $n = 500$ . Finally, it is noteworthy that when we speak of strong dependence for FGM copula, we mean the maximum degree of dependence that it can cover. With this regard, we carried out a very large number of simulations and found that the absolute value of this top level does not exceed 0.5 for any association coefficient.

On the light of these results, one may draw two overall conclusions. First, regardless of the combination of association measures and the dependence level, the estimation for MO copula is by far more accurate compared to that related to FGM copula. Second, the first parameter of both copulas is generally more precisely estimated than the second one. We also have some remarks that are specific to each one of the copulas :

- MO copula : for the second parameter, the best estimates at any level of dependence, are those based on Kendall's tau and Spearman's rho. However, as far as the first parameter is concerned, the conclusion is not as clear cut. Indeed, the inversion of Kendall's tau and Spearman's rho does not always yield the best estimate values. For instance, in the cases of strong and weak dependences, the estimators of the first parameter built via Kendall's tau and Blomqvist's beta seem to be the most accurate.

- FGM copula : for both parameters, Kendall's tau and Spearman's rho based estimators perform quite well in all three dependence cases. Moreover, they are better than those obtained through the other combinations of association measures which can often be of extremely poor reliability. For this reason, there must be no hesitation in picking the combination of Kendall's tau and Spearman's rho, when dealing with the estimation of the parameters of FGM copula by the inversion method.

Finally, the last (seperate) line of each table in the Appendix indicates that the first parameter

estimator based on the appropriate couple of association measures outclasses that obtained by means of the BLM's. This means that one should not neglect nor underrate the inversion method when it comes to estimate copula parameters.

**Acknowledgements.** We are grateful to the reviewer for his pertinent comments which allowed us to improve our work.

### 3.4 Appendix

We summarize the results of the simulations performed in Section 4.3. The first three tables consist in those related to MO copula whereas the other three concern FGM copula.

	$a = 0.8$		$b = 0.6$		$a = 0.8$		$b = 0.6$	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\tau - \rho$	-0.036	0.154	0.050	0.182	-0.030	0.125	0.038	0.117
$\tau - \beta$	0.030	0.049	0.220	0.226	0.026	0.036	0.219	0.222
$\tau - \gamma$	-0.404	0.406	1.876	1.883	-0.404	0.404	1.883	1.884
$\rho - \beta$	0.045	0.065	0.243	0.248	0.042	0.052	0.242	0.244
$\rho - \gamma$	-0.405	0.408	1.874	1.882	-0.404	0.405	1.882	1.884
$\beta - \gamma$	0.200	0.200	2.515	2.516	0.200	0.200	2.516	2.516
<i>BLM</i>	-0.766	0.767	-0.101	0.108	-0.759	0.762	-0.086	0.090

TAB. 3.2 – Estimation biases and rmse's for MO copula parameters under strong dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel).

	$a = 0.5$		$b = 0.4$		$a = 0.5$		$b = 0.4$	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\tau - \rho$	0.001	0.091	0.019	0.180	-0.005	0.065	0.013	0.097
$\tau - \beta$	0.064	0.098	0.406	0.480	0.068	0.107	0.408	0.457
$\tau - \gamma$	-0.252	0.258	1.540	1.569	-0.251	0.252	1.565	1.570
$\rho - \beta$	0.073	0.114	0.363	0.448	0.076	0.111	0.388	0.442
$\rho - \gamma$	-0.252	0.260	1.538	1.570	-0.251	0.253	1.565	1.570
$\beta - \gamma$	2.124	2.545	1.486	1.808	2.315	2.706	1.358	1.676
<i>BLM</i>	-0.474	0.475	0.006	0.024	-0.467	0.468	0.005	0.012

TAB. 3.3 – Estimation biases and rmse's for MO copula parameters under moderate dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel).

	$a = 0.1$		$b = 0.4$		$a = 0.1$		$b = 0.4$	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\tau - \rho$	0.015	0.070	0.050	0.234	-0.001	0.034	0.010	0.168
$\tau - \beta$	0.027	0.050	0.465	1.196	0.008	0.018	0.384	1.367
$\tau - \gamma$	-0.016	0.045	0.762	1.230	-0.020	0.032	0.829	1.099
$\rho - \beta$	0.051	0.091	0.766	1.551	0.023	0.052	0.616	1.968
$\rho - \gamma$	-0.011	0.046	0.727	1.216	-0.019	0.032	0.973	1.339
$\beta - \gamma$	1.446	1.455	0.662	0.956	1.47	1.47	0.60	0.60
<i>BLM</i>	-0.076	0.079	0.021	0.033	-0.073	0.077	0.012	0.014

TAB. 3.4 – Estimation biases and rmse’s for MO copula parameters under weak dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel).

	$a = 0.9$		$b = 1.5$		$a = 0.9$		$b = 1.5$	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\tau - \rho$	-0.054	0.311	0.178	1.525	-0.034	0.199	0.143	0.963
$\tau - \beta$	-0.576	1.296	2.406	5.918	-0.377	0.918	1.581	4.040
$\tau - \gamma$	-0.147	0.588	0.577	2.569	-0.022	0.286	0.098	1.173
$\rho - \beta$	0.605	10.572	-2.489	43.021	-0.169	4.131	0.680	16.812
$\rho - \gamma$	0.203	6.493	-0.874	26.498	-0.427	10.359	1.712	41.452
$\beta - \gamma$	0.491	7.131	-2.031	29.341	0.168	3.778	-0.664	15.322
<i>BLM</i>	-0.008	0.449	0.007	1.417	-0.003	0.202	0.027	0.615

TAB. 3.5 – Estimation biases and rmse’s for FGM copula parameters under strong dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel).

	$a = -0.1$		$b = 3$		$a = -0.1$		$b = 3$	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\tau - \rho$	0.046	0.413	-0.206	1.634	0.044	0.216	-0.160	0.911
$\tau - \beta$	-0.283	0.965	1.236	4.797	-0.165	0.652	0.733	3.039
$\tau - \gamma$	-0.055	0.509	0.225	2.400	0.017	0.313	-0.045	1.346
$\rho - \beta$	0.431	11.626	-1.753	47.316	-0.120	5.007	0.495	20.383
$\rho - \gamma$	0.241	9.323	-0.999	37.942	0.279	6.956	-1.101	27.879
$\beta - \gamma$	0.282	8.495	-1.154	34.885	0.132	3.742	-0.512	15.354
<i>BLM</i>	0.002	0.497	-0.020	1.574	-0.002	0.267	0.026	0.675

TAB. 3.6 – Estimation biases and rmse’s for FGM copula parameters under moderate dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel).

	$a = 0.3$		$b = -1.3$		$a = 0.3$		$b = -1.3$	
	bias	rmse	bias	rmse	bias	rmse	bias	rmse
$\tau - \rho$	0.016	0.362	-0.034	1.650	-0.013	0.214	0.072	0.737
$\tau - \beta$	-0.610	1.406	2.635	6.413	-0.374	0.925	1.585	4.048
$\tau - \gamma$	-0.248	0.761	1.059	3.404	0.000	0.334	0.029	1.342
$\rho - \beta$	-0.136	11.350	0.571	46.235	-0.119	4.748	0.501	19.348
$\rho - \gamma$	-0.326	7.955	1.331	32.423	0.237	5.224	-0.929	20.965
$\beta - \gamma$	-0.342	9.199	1.409	37.595	-0.029	5.280	0.145	21.349
<i>BLM</i>	0.011	0.504	-0.023	1.582	-0.001	0.217	0.023	0.665

TAB. 3.7 – Estimation biases and rmse’s for FGM copula parameters under weak dependence, based on 1000 samples of sizes 100 (left panel) and 500 (right panel).

# Conclusion générale

La copule est un concept de modélisation de la structure de dépendance. Contrairement aux différentes mesures de dépendance, la relation entre les variables est bien expliquée par la copule. Autrement dit, au lieu de résumer la structure de la dépendance par un seul nombre, par exemple le coefficient de corrélation, on peut utiliser la fonction copule qui décrit au mieux cette dépendance.

Dans cette thèse on a abordé la notion des copules ainsi que les différentes propriétés qui lui sont associées. Au début, on a présenté quelques familles de copules paramétriques les plus utilisées et leurs propriétés fondamentales à savoir : les copules elliptiques, les copules Archimédiennes,...etc. Puis, on a exprimé les probabilités au terme de celles-ci, expliqué comment les construire, et enfin évoquer leurs domaines d'applications.

On a parlé de la mesure de dépendance régulièrement utilisée qui est la corrélation linéaire de Bravais-Pearson. Cet indicateur est performant lorsque la relation de dépendance est linéaire et l'univers considéré gaussien, ce cadre d'analyse est malheureusement rare dans certains domaines d'applications. Pour remédier à cela, On a recours à d'autres indicateurs de dépendance se fondant sur les discordances et concordances observées sur un échantillon. Alors, on a utilisé des coefficients de corrélation non linéaires et non paramétriques, comme le tau de Kendall, le rho de Spearman ou encore le gamma de Gini et le beta de Blomqvist qui sont de bons indicateurs globaux de la dépendance entre variables. On a étudié les différentes propriétés des deux copules (MO et FGM) utilisées dans notre application, à savoir la dépendance de queue et les quatre mesures de concordances.

On a présenté trois méthodes pour estimer la copule. La première méthode est paramétrique, i.e. elle impose un modèle paramétrique pour la copule et pour les distributions marginales.

Une comparaison entre les deux techniques de cette méthode, a montré que la méthode "IFM" aurait présenter l'avantage de reposer sur des calculs plus légers que ceux générés par la méthode "EML". La seconde approche est semi-paramétrique, i.e. elle suppose un modèle paramétrique pour la copule et non-paramétrique pour les distributions marginales. La troisième méthode d'estimation de la copule est non-paramétrique, celle-ci ne fait pas de restriction sur les modèles. On a présenté les divers outils pour la sélection de la meilleure copule basée sur les outils graphiques d'ajustement, les tests statistiques et les fonctions qui permettent de faire le tri parmi les copules et de choisir celle qui conviendra à notre série de données. La sélection des copules représentant au mieux les structures de dépendance, constitue un axe d'investigation supplémentaire. Les tests d'ajustement comme par exemple le test d'adéquation du chi-deux est utile car il donne au critère de décision un fondement statistique et peut être utilisé dans un cadre multivarié. Cependant, il nécessite un découpage en classes qui lui fait perdre de la puissance. La recherche de tests multidimensionnels plus puissants appliquée aux copules, pourrait offrir de nouvelles perspectives d'application.

Le travail s'achève dans la partie application qui est basée sur les résultats obtenus par des techniques de simulation. Le but est de comparer la robustesse des estimateurs (obtenus par la méthode d'inversion des mesures d'association) des paramètres de la copule en terme du bias et de l'rmse, quand les distributions marginales sont inconnues. Pour cela, on a confronter l'utilisation de ce processus d'inversion à la méthode BLM.

On a aboutit à deux conclusions générales. **Tout d'abord**, quelle que soit la combinaison de deux mesures d'association et le niveau de dépendance, l'estimation pour la copule de MO est de loin plus précise par rapport à celle liée à la copule de FGM. **D'autre part**, l'estimation du premier paramètre des deux copules est généralement plus précise que la seconde. On a également quelques remarques qui sont spécifiques à chacune des deux copules :

- Copule de MO : les meilleures estimations du second paramètre de la copule, à tout niveau de dépendance, sont celles fondées sur le tau de Kendall et le rho de Spearman. Cependant, en ce qui concerne le premier paramètre, la conclusion n'est pas aussi claire. En effet, l'inversion du tau de Kendall et du rho de Spearman ne produit pas souvent les meilleures

estimateurs. Par exemple, dans les cas de forte et de faible dépendance, les estimateurs du premier paramètre construits via le tau de Kendall et le beta de Blomqvist semblent être les plus précis.

- Copule de FGM : pour les deux paramètres, les estimations basées sur le tau de Kendall et rho de Spearman performant assez bien dans les trois cas de dépendance. De plus, elles sont meilleures que celles obtenues par les autres combinaisons de mesures d'association qui peuvent souvent être de très mauvaise fiabilité. Pour cette raison, il ne doit y avoir aucune hésitation à choisir la combinaison du tau de Kendall et du rho de Spearman, lorsqu'on traite l'estimation des paramètres de la copule FGM par la procédure d'inversion.

Enfin, la dernière ligne séparée de chaque tableau en Annexe 3.4, indique que l'estimateur du premier paramètre basé sur le couple approprié de mesures d'association surclasse celui obtenu par la méthode des BLM. Cela signifie qu'il ne faut ni négliger ni sous-estimer la méthode d'inversion lorsqu'on vient à estimer les paramètres de la copule.

On espère que ce travail a permis d'appréhender précisément les tenants et les aboutissants de la modélisation de la structure de dépendance d'un échantillon à travers les copules. Malgré sa relative simplicité mathématique, l'outil copule n'est pas toujours utilisé dans toutes ses potentialités. En finance, par exemple, de nombreux travaux sont encore à réaliser pour appliquer cet objet rigoureusement et de manière adéquate dans des domaines aussi variés que le pricing de produits structurés ou encore la gestion des risques multiples.

# Annexe : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de cette thèse sont expliquées ci-dessous :

<u>Symbole</u>	<u>Signification</u>
$va$	Variable aléatoire.
i.i.d.	indépendantes et identiquement distribuées.
fd	Fonction de distribution.
$\mathbb{P}(A)$	Probabilité d'un évènement $A$ .
$X, Y$	Vas réelles.
$U, V$	Vas qui suivent la loi uniforme $\mathcal{U}$ .
$\mathbb{E}[X]$	Espérance mathématique de la va $X$ .
$\mathbb{V}[X]$	Variance mathématique de la va $X$ .
$\sigma_X$	Ecart type de la va $X$ .
$Cov(X, Y)$	Covariance mathématique du couple $(X, Y)$ .
$F, G$	fd des lois marginales.
$H$	Distribution jointe.
$f, g$	Densités des lois marginales.
$h$	Densité de distribution jointe.
$\tilde{x}$	Médianes de $X$ .

<u>Symbole</u>	<u>Signification</u>
$\mathbb{R}$	Ensemble des nombres réels.
$\mathbf{I}$	Intervalle $[0, 1]$ .
$C$	Distribution de la copule.
$c$	Densité de la copule.
$\Phi(x)$	fd de la va $x$ qui suit la loi normale.
$T_\nu$	Loi de Student à $\nu$ degré de liberté.
$\phi$	Fonction génératrice de la copule archimédienne.
$\phi^{-1}$	Fonction quantile de $\phi$ .
$A(t)$	Fonction de dépendance.
$C_{\phi,A}$	Copule Archimax.
$M_{X,n}$	Maximum des vas $(X_i)_{1 \leq i \leq n}$ .
$C^n$	Copule de valeurs extrêmes.
$\hat{C}_n$	Copule empirique.
$\kappa$	Mesure d'association.
$Q$	Fonction de concordance.
$\tau$	Tau de Kendall.
$\rho$	Rho de Spearman.
$\beta$	Beta de Blomqvist.
$\gamma$	Gamma de Gini.
$\lambda$	Paramètre de la loi exponentielle.
$\Pi$	Copule produit.
$W$	Copule maximum.
$M$	Copule minimum.
$F_n$	fd empirique.
$R_{X_i}$	Rang des observations de $X$ .
$D_i$	Différence entre les rangs des observations de $X$ et $Y$ .

<u>Symbole</u>	<u>Signification</u>
i.e.	c'est-à-dire.
e.g.	par exemple.
$\rho$	Coefficients de corrélation.linéaire.
$C_{\rho}^{\Phi}$	Copule normale.
$C_{\rho,v}^t$	Copule de Student.
$F_n$	fd empirique.
$F^{-1}$	Fonction quantile de $F$ .
$K$	Fonction de Kendall.
$n_c, n_d$	Nombre de paires concordantes, discordantes.
$L(\theta)$	Fonction log-vraisemblance de $\theta$
$\overset{*}{C}$	Co-copule.
$\acute{C}$	Copule mixte.
$\tilde{C}$	Copule duale.
$\bar{C}$	Fonction de survie.
$\hat{C}$	Copule de survie.
$\lambda_U, \lambda_L$	Dépendance de queue supérieure, inférieure.
$\theta, a, b$	Paramètres de la copule.
$\mathcal{C}$	Famille de copules.
$\Theta$	Espace des paramètres.
$P$	Polynôme de legendre déplacé.
$[t]$	Partie entière de $t$ .
$\alpha, \beta$	Fonctions strictement croissantes.

# Bibliographie

- [1] Abbas, A. E., 2006. Entropy methods for joint distributions in decision analysis. *IEEE Transactions on engineering management* **53**, 146–159.
- [2] Acciolya, R. D. E., Chiyoshi, F. Y. 2004. Modeling dependence with copulas : A useful tool for field development decision process. *Journal of petroleum science and engineering* **44**, 83–91.
- [3] Balakrishnan, N., Lai, CD., 2009. Continuous bivariate distributions. Springer, New York.
- [4] Bekrizadeh, H., Parham, GH.A., Zadkarmi, M.R., 2012. The new generalization of Farlie-GumbelMorgenstern copula. *App. Math. Scie* **6**, 3527-3533.
- [5] Belguise, O., 2001. Tempêtes : étude des dépendances entre les branches auto et incendie avec la théorie des copules. Mémoire DUAS troisième année. *Actuariat, Univ. Louis. Pasteur. Strasbourg. France.*
- [6] Benelmir, I., Meraghni, D., 2016. Association measures and estimation of copula parameters. *Afr. Stat.* **11** (1), 933-942.
- [7] Besson, J. L., Partrat, C., 2005. Assurance non-vie - Modélisation, simulation, *Economica*.
- [8] Blischke, W. R., Prabhakar Murthy, D.N., 2000. Reliability modeling, prediction, and optimization, Wiley, New York.
- [9] Blomqvist, N., 1950. On a Measure of Dependence Between two Random Variables. *Ann. Math. Statist.* **21**, 593-600.
- [10] Brahim, B., Meraghni, D., Necir, A., 2010. Distortion risk measures for sums of dependent losses. *Journal Afrika Statistika* **5** (9), 260-267.

- [11] Brahim, B., Necir, A., 2012. A semiparametric estimation of copula models based on the method of moments. *Statistical Methodology* **9** (4), 467-477.
- [12] Brahim, B., Chebana, F., Necir, A., 2015. Copula representation of bivariate L-moments : A new estimation method for multiparameter two-dimensional copula models. *Statistics* **49** (3), 497-521.
- [13] Capéraà, P., Fougères, A. L., Genest, C., 2000. Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis* **72**, 30–49.
- [14] Chalabian, J., Dunnington, G., 1998. Do our current assessments assure competency in clinical breast evaluation results? *The American Journal of Surgery* **175** (6), 497- 502.
- [15] Chen, X. H., Fan, Y. Q., 2007. A model selection test for bivariate failure-time data. *Econometric theory* **23**, 414–4393.
- [16] Chen, S. X., Huang, T. M., 2007. Nonparametric estimation of copula functions for dependence modelling. *La Revue Canadienne de statistique* **35**, 265–282.
- [17] Cherubini, U., Luciano, E., Vecchiato, W., 2004, Copula methods in finance. John Wiley & Sons. 310.
- [18] Choros, B., Ibragimov, R., Permiakova, E., 2010. Copula theory and its applications. *Springer*. **198**, 77-91.
- [19] Cook, R. D., Johnson, M. E., 1981. A family of distributions for modeling non-elliptically symmetric multivariate data. *J. Roy. Statist. Soc, Ser. B* **43**, 210–218.
- [20] D’Agostino, R. B., Stephens, M. A., 1986. Goodness-of-Fit Techniques. New York : Marcel Dekker Inc.
- [21] Dall’Aglia, G., 1956. Sur les extrémités des doubles moments de la fonction de distribution. *Annales de l’école normale supérieure de Pise, la classe des sciences*, **10** (3), 35–74.
- [22] Daniels, HE., 1950. Rank correlation and population models. *J. Roy. Statist. Soc. Ser. B* **12**, 171-181.
- [23] Darlap, P., Mayr, B., 2006. Group aspects of regulatory reform in the insurance sector. *Geneva papers on risk and insurance : Issues and Practice* **31**, 96–123.

- [24] Deheuvels, P., 1979. La fonction de dépendance empirique et ses propriétés. Un test non-paramétrique d'indépendance. *Académie Royale de Belgique Bulletin de la Classe des Sciences*, **65** (5), 274–292.
- [25] Deheuvels, P., 1940. La fonction de dépendance empirique et ses propriétés. Académie Royale de Belgique (Bulletin de la Classe des Sciences) **65**, 274– 292.
- [26] Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R., 2005. Actuarial theory for dependent risks measures, Wiley, England.
- [27] Devroye, L., 1986. Non-Uniform Random Variate Generation. *Springer*, New York.
- [28] Dobric, J., Schmid, F., 2005. Testing goodness of fit for parametric families of copulas : Application to financial data. *Communications in statistics-simulation and computation* **34**, 1053–1068.
- [29] Dobric, J., Schmid, F., 2007. A goodness of fit test for copulas based on rosenblatt's transformation. *Comput. Stat. Data Anal.* **51** (9), 4633–4642.
- [30] Durante, F., Quesada-Molina, J. J., Sempì, C., 2006. A generalization of the archimedean class of bivariate copulas. *Ann. Inst. Statist. Math.* **59**, 487-498.
- [31] Durbin, J., Stuart, A., 1951. Inversions and Rank Correlation Coefficients. *Journal of the Royal Statistical Society.* **13** (2), 303-309.
- [32] Embrechts, P., Lindskog, F., Mcneil, A., 2003. Modeling dependence with copulas and applications to risk management, in : S. Rachev (Ed.), Handbook of Heavy Tailed Distributions in Finance. *Elsevier*, New York, 329-384.
- [33] Escarela, G., Mena, R. H., Castillo-Morales, A., 2006. A flexible class of parametric transition regression models based on copulas : Application to poliomyelitis incidence. *Statistical methods in medical Research* **15**, 593–609.
- [34] Farlie, D. J. G., 1960. The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**, 307–323.
- [35] Fermanian, J. D., Scaillet, O., 2002. Non parametric estimation of copulas for time series. *Journal of risks* **5**, 25-54.

- [36] Fermanian, J. D., 2005. Goodness of fit tests for copulas. *Journal of Multivariate Analysis*, **95** (1), 119–152.
- [37] Féron, R. 1956. Sur les tableaux de corrélation dont les marges sont données, cas de l'espace à trois dimensions. *Publ. Inst. Statist. Univ. Paris* 5, 3–12.
- [38] Fisher, N. I., 1997. Copulas. *Encyclopedia of Statistical Sciences* **1**, 159-163.
- [39] Fréchet, M., 1951. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, Sect. A* **9**, 53–77.
- [40] Galambos, J., 1978. The Asymptotic theory of extreme order statistics. John Wiley and Sons, USA.
- [41] Genest, C., MacKay, J., 1986. The joy of copulas : Bivariate distributions with uniform marginals. *The American Statistician* **40**, 280–285.
- [42] Genest, C., Rivest, L., 1993. Statistical inference procedures for bivariate Archimedean copulas. *Journal of American Statistical Association*, **88**, 423.
- [43] Genest, C., Ghoudi, K., Rivest, L. P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552.
- [44] Genest, C., Werker, B. J. M., 2002. Conditions of asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models, in C.M. Cuadras and J.A. Rodríguez Lallena (Eds.). *Springer*. New York, 103-112.
- [45] Genest, Ch., Boies, J. C., 2003. Detecting Dependence with Kendall Plots. American Statistical Association.
- [46] Genest, C., Favre, A. C., 2007. Everything you always wanted to know about copula modelling but were afraid to ask. *J. Hydrol. Eng*, **12** (4),347-368.
- [47] Genest, C., Rémillard, B., 2008. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'institut Henri Poincaré : Probabilités et statistiques* **44**, 1096–1127.
- [48] Geoffroy, J. 1958. Contributions à la théorie des valeurs extrêmes. *Publ. Inst. Statist. Univ. Paris* **7**, 37–121.

- [49] Gijbels, I., Mielniczuk, J. 1990. Estimating the density of a copula function. *Communications in statistics-theory and methods* **19**, 445–464.
- [50] Gumbel, E. J., 1960. Bivariate exponential distributions. *Journal of American Statistical Association* **55**, 698–707.
- [51] Hoeffding, W., 1940. Masstabinvariance korrelationsmasse, *Schriften des Mathematischen Instituts für Angewandte Mathematik der Universität Berlin*, **5** (3), 179-233.
- [52] Huang, JS., Kotz, S., 1984. On some generalized Farlie-Gumbel-Morgenstern distribution. *Biometrika* **71**, 633-636.
- [53] Huang, J. S., Kotz, S., 1999. Modifications of the Farlie–Gumbel–Morgenstern distributions : A tough hill to climb. *Metrika* **49**, 307–323.
- [54] Hurlimann, W., 2001. Fitting bivariate cumulative returns with copulas. Working paper, Winterthur Life and Pensions.
- [55] Hutchinson, T. P., Lai, C. D., 1990. Continuous bivariate distributions emphasizing applications. *Rumsby Scientific Publishing*. Adelaide, Australia.
- [56] Joe, H., Xu, J., 1996. The estimation method of inference functions for margins for multivariate models. Technical Report, *Dep. Statist.* University of British Columbia, 166.
- [57] Joe, H., 1997. Multivariate models and dependence concepts. *Chapman & Hall*, London, 73.
- [58] Joe, H. 2005. Asymptotic efficiency of the two-stage estimation method for copulabased models. *Journal of Multivariate Analysis* **94**, 401–419.
- [59] Johnson, N. L., Kotz, S., 1977. On some generalized Farlie-Gumbel-Morgenstern distributions-II regression, correlation and further generalizations. *Comm Stat.* **6**, 485-496.
- [60] Kim, J. M., Sungur, E. A., 2004. New class of bivariate copulas, In Proceedings for the Spring Conference of Korean Statistical Society, 207–212.
- [61] Kim, G., Silvapulle, J., Silvapulle, P., 2007. Comparison of semi-parametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis* **51**, 2836–2850.

- [62] Kimeldorf, G., Sampson, A., 1975. One-parameter families of bivariate distributions with fixed marginals. *Comm. Statist.* **4**, 293–301.
- [63] Lin, G. D., 1987. Relationships between two extensions of Farlie-Gumbel-Morgenstern distribution. *Ann. Inst. Statist. Math.* **39**, 129-140.
- [64] Marshall, A. W., Olkin, I., 1967. A Generalized Bivariate Exponential Distribution. *J. Appl. Prob.* **4**, 291-302.
- [65] Morgenstern, D., 1956. Simple examples of two-dimensional distributions. *Statist. Math. Doc.* **8**, 234–235.
- [66] Nelsen, R. B., 2006. An Introduction to copula. Springer Verlag, New York.
- [67] Oakes, D. 1982. A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Ser. B* **44**, 414–422.
- [68] Oakes, D. 1986. Semiparametric inference in a model for association in bivariate survival data. *Biometrika* **73**, 353–361.
- [69] Purcaru, O., 2003. Semi-parametric Archimedean copula modelling in actuarial science. *Insurance mathematics and economics* **33**, 419–420.
- [70] Salvadori, G., De Michele, C., 2007. On the use of copulas in hydrology : Theory and practice. *Journal of hydrologic engineering* **12**, 369–380.
- [71] Sancetta, A., Satchell, S. 2004. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Economic Theory* **20**, 535–562.
- [72] Schweizer, B., Wolff, E. F., 1981. On nonparametric measures of dependence for random variables. *Ann. Statist.* **9**, 879-885.
- [73] Shih, J. H., Louis, T. A., 1995.. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* (51), 1384–1399.
- [74] Sibuya, M., 1960. Bivariate extreme statistics. *Annals of the institute of statistical mathematics* **11**, 195–210.
- [75] Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. *Inst. Statist. Univ. Paris.* **8**, 229-231.

- 
- [76] Spearman, C., 1904. The proof and measurement of association between two things. *Am J Psychol.* **15**, 72–101.
- [77] Srinivas, S., Menon, D., Prasad, A. M., 2006. Multivariate simulation and multimodal dependence modeling of vehicle axle weights with copulas. *Journal of transportation engineering* **132**, 945–955.
- [78] Tiago de Olivera, J., 1958. Extremal distributions. *Rev. Fac. Ciencias. Univ. Lisboa, A* **8**, 299–310.
- [79] Van Der Hoek, J., Sherris, M., 2006. A flexible approach to multivariate risk modelling with a new class of copulas. *Insurance mathematics and economics.* **39**, 398–399.
- [80] Venter, G., 2001. Tails of copulas. *23ème ASTIN Colloquium*, Washington D. C.
- [81] Viswanathan, B., Manatunga, A. K., 2001. Diagnostic plots for assessing the frailty distribution in multivariate survival data. *Lifetime Data Analysis* **7**, 143–155.
- [82] Wist, H. T., Myrhaug, D., Rue, H., 2004. Statistical properties of successive wave heights and successive wave periods. *Applied ocean research* **26**, 114–136.