

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À  
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES

PAR  
MAMADOU LAMINE CISSÉ

DE NOUVEAUX ESTIMATEURS SEMI-PARAMÉTRIQUES DE L'INDICE DE  
DÉPENDANCE EXTRÊME DE QUEUE

FÉVRIER 2020

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

## AVANT-PROPOS

Le choix d'étudier à l'Université du Québec à Trois Rivières est lié à plusieurs motifs. D'abord cette université offre des spécialisations dans le domaine de la statistique qui collent parfaitement avec mes compétences académiques, telles que l'application des méthodes classiques, l'analyse de données, les techniques de modélisations, afin de mieux expliquer et appréhender les problèmes mathématiques d'ordre pratique. Ensuite, ce département de mathématiques et informatique appliquées offre à l'étudiant les meilleures conditions de travail en planifiant son projet de recherche dans une problématique plus globale et en le préparant à l'exécution de sa propre recherche.

Je tiens à remercier sincèrement Monsieur Jean-François Quessy, mon directeur de recherche, pour sa confiance, sa patience et son dévouement. Le partage de son inébranlable expertise en statistique et en recherche m'a permis d'affiner autant ma connaissance que ma rigueur de travail. Il m'a transmis la passion de la recherche mathématique, m'a encouragé et soutenu financièrement durant mon séjour. Qu'il soit assuré de ma sincère reconnaissance et de ma profonde sympathie.

Enfin, je veux remercier les professeurs Fadel Touré et Mhamed Mesfioui, du Département de mathématiques et d'informatique de l'UQTR, pour avoir gracieusement accepté de lire et de commenter ce travail de recherche. Leurs remarques judicieuses ont permis d'améliorer la version finale de mon travail.

Le financement durant mes études a été fourni par un octroi individuel à M. Jean-François Quessy dans le cadre du programme des Subventions à la Découverte du Conseil de Recherche en Sciences Naturelle et en Génie (CRSNG) du Canada.

# Table des matières

Avant-propos	ii
Table des matières	iii
Liste des tableaux	vi
Table des figures	vii
<b>1 Mise en contexte et motivation</b>	<b>1</b>
1.1 Perspective historique . . . . .	1
1.2 Choix d'une mesure de dépendance . . . . .	2
1.3 Utilisation de la copule de Fisher pour modéliser la dépendance . . . . .	3
1.4 Importance de mesurer la dépendance caudale . . . . .	4
1.5 Objectifs et structure du mémoire . . . . .	5
<b>2 Les indices de dépendance caudale</b>	<b>6</b>
2.1 Les mesures de dépendance caudale $\lambda_L$ et $\lambda_U$ . . . . .	6
2.2 Lois bivariées et copules . . . . .	8
2.2.1 Introduction . . . . .	8
2.2.2 Théorème de Sklar . . . . .	8
2.2.3 Importance pour la modélisation . . . . .	9
2.3 Les mesures $\lambda_L$ et $\lambda_U$ s'écrivent comme des fonctionnelles de la copule .	10
2.4 Indices de dépendance caudale pour quelques familles de copules . . . . .	11
2.4.1 Cas d'une symétrie radiale . . . . .	11
2.4.2 Copule d'indépendance et bornes de Fréchet . . . . .	12

2.4.3	Copule Normale . . . . .	13
2.4.4	Copule de Student . . . . .	14
2.4.5	Copules Archimédiennes . . . . .	15
2.4.6	Copules à valeurs extrêmes . . . . .	17
<b>3</b>	<b>Revue des méthodes d'estimation de <math>\lambda_U</math> et regard critique</b>	<b>20</b>
3.1	Préliminaires . . . . .	20
3.1.1	Mise en contexte . . . . .	20
3.1.2	Copule empirique de Deheuvels (1979) . . . . .	22
3.1.3	Copule empirique à base de rangs . . . . .	23
3.2	Quelques estimateurs de dépendance caudale . . . . .	23
3.2.1	Estimateur de Capéreaa <i>et coll.</i> (1997) . . . . .	24
3.2.2	Estimateur de Schmidt et Stadtmuller (2006) . . . . .	24
3.2.3	Estimateur de Coles <i>et coll.</i> (1999) . . . . .	25
3.2.4	Estimateur de Joe <i>et coll.</i> (1992) . . . . .	25
3.3	Limites et faiblesses des estimateurs existants . . . . .	26
<b>4</b>	<b>De nouveaux estimateurs des coefficients de dépendance caudale</b>	<b>29</b>
4.1	La famille des copules Fisher . . . . .	29
4.1.1	Famille des copules Khi-deux . . . . .	29
4.1.2	Construction d'une copule Fisher . . . . .	30
4.1.3	Coefficients de dépendance caudale de la copule Fisher . . . . .	32
4.2	Description des nouveaux estimateurs de l'indice de dépendance caudale	35
4.2.1	Idée générale . . . . .	35
4.2.2	Estimation de $\rho$ par inversion du tau de Kendall . . . . .	35
4.2.3	Estimation de $(\rho, \nu)$ par maximum de vraisemblance . . . . .	36
4.2.4	Estimation de $(\rho, \nu)$ par une méthode en deux étapes . . . . .	38
4.2.5	Estimation de $\lambda_U$ : degré de liberté fixé . . . . .	38
4.2.6	Estimation de $\lambda_U$ : degré de liberté inconnu . . . . .	39
4.3	Étude de la performance des estimateurs de $\lambda_U$ . . . . .	40
4.3.1	Description générale de l'étude de simulations . . . . .	40

4.3.2	Données générées de la copule Fisher . . . . .	41
4.3.3	Données générées de la copule de Gumbel . . . . .	42
4.3.4	Données générées de la copule Khi-deux . . . . .	43
4.3.5	Commentaires globaux sur les simulations et quelques recommandations . . . . .	43
4.3.6	Tableaux des résultats de simulations . . . . .	45
4.4	Illustration sur de vraies données . . . . .	49
4.4.1	Présentation des données . . . . .	49
4.4.2	Analyse statistique . . . . .	50
	<b>Conclusion</b>	<b>53</b>
	<b>Bibliographie</b>	<b>54</b>
<b>A</b>	<b>Éléments complémentaires</b>	<b>60</b>
A.1	Mesurer la dépendance . . . . .	60
A.1.1	Inconvénients de la corrélation . . . . .	60
A.1.2	Rho de Spearman . . . . .	61
A.1.3	Tau de Kendall . . . . .	63
A.1.4	Mesures de concordances . . . . .	66
A.2	Estimation par noyau dans un contexte de copules . . . . .	69
A.2.1	Estimation de la copule . . . . .	69
A.2.2	Estimation de la densité d'une copule . . . . .	72
A.3	Méthode du maximum vraisemblance canonique . . . . .	74

# Liste des tableaux

2.1	Caractéristiques de trois copules Archimédiennes . . . . .	16
2.2	Caractéristiques de quelques copules à valeurs extrêmes . . . . .	19
4.1	Erreur quadratique moyenne de sept estimateurs de $\lambda_U$ , estimée à l'aide de 1 000 réplicats, sous des données simulées de la copule Fisher . . . . .	45
4.2	Biais de sept estimateurs de $\lambda_U$ , estimé à l'aide de 1 000 réplicats, sous des données simulées de la copule Fisher . . . . .	46
4.3	Erreur quadratique moyenne de sept estimateurs de $\lambda_U$ , estimée à l'aide de 1 000 réplicats, sous des données simulées de la copule de Gumbel . . . . .	47
4.4	Biais de sept estimateurs de $\lambda_U$ , estimé à l'aide de 1 000 réplicats, sous des données simulées de la copule de Gumbel . . . . .	47
4.5	Erreur quadratique moyenne de sept estimateurs de $\lambda_U$ , estimée à l'aide de 1 000 réplicats, sous des données simulées de la copule Khi-deux . . . . .	48
4.6	Biais de sept estimateurs de $\lambda_U$ , estimé à l'aide de 1 000 réplicats, sous des données simulées de la copule Khi-deux . . . . .	48
4.7	Résultats de l'estimation des coefficients de dépendance caudale $\lambda_U$ (partie supérieure du tableau) et $\lambda_L$ (partie inférieure du tableau) pour quelques paires de variables du jeu de données de l' <i>Australian Institute of Sport</i> . . . . .	52

# Table des figures

4.1	Histogrammes (sur la diagonale), nuages de points des données brutes (triangle supérieur) et nuages de points des pseudo-observations (triangle inférieur) pour les paires $(X_1, X_2, X_3)$ du jeu de données de l' <i>Australian Institute of Sport</i> . . . . .	50
4.2	Histogrammes (sur la diagonale), nuage de points des données brutes (triangle supérieur) et nuage de points des pseudo-observations (triangle inférieur) pour la paire $(X_4, X_5)$ du jeu de données de l' <i>Australian Institute of Sport</i> . . . . .	51

# Chapitre 1

## Mise en contexte et motivation

### 1.1 Perspective historique

Mesurer la mesure de dépendance entre des variables aléatoires est maintenant une pratique très répandue. Une panoplie de mesures de dépendance est disponibles notamment : le coefficient de corrélation de Pearson, le tau de Kendall et le rho de Spearman. Néanmoins, ces derniers, bien que faciles à calculer et à interpréter, ne peuvent mettre en exergue toutes les formes de dépendance possibles. C'est ainsi qu'un autre moyen est mis en place pour remédier à ce manquement. En effet, la fonction copule offre l'avantage de modéliser complètement la dépendance entre deux ou plusieurs variables aléatoires.

La notion de “copule” a été introduite dans les années 50. C'est ainsi que dans les années 70, avec le développement de la théorie des processus empiriques, des propriétés et appellations nouvelles sur les copules apparaissent, incluant les “représentations uniformes” de Kimeldorf and Sampson [34], la “fonction de dépendance” de Galambos [23] et Deheuvels [16], ainsi que la “forme standard” de Cook and Johnson [14].

Issue du mot latin "copulae", qui signifie lien, alliance, liaison ou union, la copule apparaît aussi sous d'autres connotations dans des études faites par Fréchet [21], Féron [19] et Dall'Aglio [15] sur l'étude des tables de contingence et des lois multidimensionnelles à structures marginales fixées. Mais c'est grâce à Sklar [51] que le terme copule apparaît pour la première fois avec l'étude sur la théorie des lois multidimensionnelles.

De manière explicite, la notion de copule met en exergue la liaison entre deux variables  $(X, Y)$  par la relation

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

avec  $H$  est la fonction de répartition conjointe des variables aléatoires  $X$  et  $Y$  et  $F$  et  $G$  sont les fonctions de répartition marginales. Un des avantages des copules est que ces dernières peuvent être des lois marginales différentes sans impacter la structure de dépendance. Cette flexibilité facilite l'application des copules dans plusieurs domaines notamment l'hydrologie avec les travaux de Salvadori et al. [42], les sciences actuarielles avec les travaux de Frees and Valdez [22], ainsi qu'en finance avec les études de Cherubini et al. [12] et McNeil et al. [36].

## 1.2 Choix d'une mesure de dépendance

Étant donné que le coefficient de corrélation et la fonction copule sont tous des indicateurs permettant de mesurer la dépendance entre des variables aléatoires, ce coefficient présente des limites que la copule permet de surmonter :

- (i) Pour les distributions à queues lourdes où les variances peuvent être infinies, le coefficient de corrélation est indéfini ;
- (ii) Pour les valeurs extrêmes, le coefficient de corrélation peut être constant tandis que la structure de dépendance peut être différente ;

- (iii) Pour les variables non gaussiennes, le coefficient de corrélation et sa caractérisation ne permettent pas de capter la structure de dépendance ;
- (iv) Le coefficient de corrélation ne permet pas de capter l'ensemble des aspects de la structure de dépendance.

Modéliser la dépendance par des indicateurs statistiques est une chose, la modéliser à l'aide d'une fonction de dépendance en est une autre. Ainsi, la copule est adéquate pour faire ce travail, car elle permet de

- (1) modéliser les propriétés des structures de dépendance ;
- (2) mesurer la dépendance pour les distributions à queues lourdes ;
- (3) construire des modèles de distributions multidimensionnelles, notamment des modèles non gaussiens.

### 1.3 Utilisation de la copule de Fisher pour modéliser la dépendance

Les familles de copules multivariées qui existent présentement présentent des limites pour la modélisation. Par exemple, la copule normale se limite aux cas où la structure de dépendance est obligatoirement radialement symétrique et lorsque la dépendance caudale est nulle. Quant à elle, la copule de Student permet de la dépendance caudale, mais elle doit être radialement symétrique, à l'instar de la copule normale. En ce qui concerne les copules Khi-deux étudiées par Bårdossy [7] et Quessy [39], elles sont asymétriques, sauf que leurs coefficients de dépendance caudale sont nuls. Il existe quelques alternatives, notamment les constructions à base de vignes popularisées entre autres par Aas et al. [1] ; ces modèles ne seront toutefois pas considérés dans la suite.

À l'égard des limitations listées ci-dessus, un modèle intéressant est la famille des copules de Fisher introduite par Favre et al. [17]. En effet, les copules de Fisher affichent les propriétés suivantes :

- Elles permettent la modélisation en haute dimension ;
- La dépendance caudale supérieure dans ces modèles est non nulle ;
- Ces copules possèdent la propriété d'asymétrie radiale ;
- La structure de dépendance par paire permet d'interpréter facilement les résultats d'une modélisation statistique, en particulier dans des applications comme la statistique spatiale.

## 1.4 Importance de mesurer la dépendance caudale

Les indices de dépendance caudale mesurent la propension qu'ont deux variables aléatoires à prendre simultanément des valeurs semblables (simultanément grandes ou petites). D'une certaine manière, c'est la probabilité conjointe d'apparition d'événements extrêmes. Pour une paire de variables aléatoires  $(X, Y)$  avec les distributions marginales  $F$  et  $G$ , le coefficient de dépendance caudale supérieure est la probabilité conditionnelle

$$\lambda_U = \lambda_U(X, Y) = \lim_{u \uparrow 1} \mathbb{P} \{X > F^{-1}(u) | Y > G^{-1}(u)\}. \quad (1.1)$$

On dit que  $X$  et  $Y$  sont asymptotiquement dépendantes supérieurement si  $\lambda_U > 0$ . Sinon, c'est-à-dire si  $\lambda_U = 0$ , on dit que  $X$  et  $Y$  sont asymptotiquement indépendantes. Dans la même lignée, le coefficient de dépendance caudale inférieur est défini par

$$\lambda_L = \lambda_L(X, Y) = \lim_{u \downarrow 0} \mathbb{P} \{X \leq F^{-1}(u) | Y \leq G^{-1}(u)\}. \quad (1.2)$$

Les indices  $\lambda_U$  et  $\lambda_L$  sont abondamment utilisés dans les sciences appliquées, notamment en hydrologie et en climatologie, où les événements extrêmes en deux endroits plus ou moins rapprochés (comme les pluies abondantes, les sécheresses) ont une forte tendance à se produire en même temps.

## 1.5 Objectifs et structure du mémoire

Tel que clairement démontré et argumenté par Serinaldi et al. [49], les estimateurs de  $\lambda_U$  et  $\lambda_L$  proposés jusqu'ici dans la littérature scientifiques sont généralement extrêmement biaisés. Cela pose donc un problème évident lorsque vient le temps de quantifier le niveau de dépendance extrême lors de l'observation d'un phénomène physique.

L'objectif principal de ce mémoire est de combler cette lacune en proposant de nouveaux estimateurs semi-paramétriques fondés sur l'expression obtenue par Favre et al. [17] pour l'indice de dépendance caudale sous l'hypothèse d'une structure de dépendance appartenant à la famille des copules de Fisher. Leur efficacité en termes de biais et d'erreur quadratique moyenne sera ensuite étudiée à l'aide de simulations, non seulement sous des dépendances de type Fisher, mais aussi sous une panoplie de modèles de copules afin de montrer que leur usage peut être étendu à l'extérieur de l'hypothèse d'une copule de Fisher.

Le reste du mémoire est organisé comme suit. Au Chapitre 2, les différents concepts qui seront utilisés tout au long de ce travail seront définis, incluant une revue des principaux résultats théoriques pertinents ; quelques familles de copules parmi les plus utilisées en pratique, de même que leurs propriétés fondamentales, seront également décrites. Au Chapitre 3, les différentes méthodes d'estimation des paramètres d'une copule seront décrites ; ceci permettra de définir les principaux estimateurs du coefficient de dépendance caudale utilisés dans la pratique, et de discuter de leurs défauts. Au Chapitre 4, les nouveaux estimateurs des coefficients de dépendance caudale sont introduits ; ceux-ci sont basés sur le coefficient de dépendance caudale supérieure de la copule de Fisher. Deux cas sont considérés, à savoir lorsque le degré de liberté est connu et lorsque celui-ci est supposé inconnu. La performance de ces estimateurs est étudiée à l'aide de simulations de type Monte-Carlo. Quelques explications complémentaires concernant la modélisation de la dépendance ont été placées en annexe.

# Chapitre 2

## Les indices de dépendance caudale

### 2.1 Les mesures de dépendance caudale $\lambda_L$ et $\lambda_U$

La notion de dépendance caudale donne une idée plus pertinente sur la survenue des risques simultanés au niveau des queues de distribution. Ceci permet d'étudier l'occurrence concomitante des valeurs extrêmes. Contrairement aux mesures de concordance appliquées sur toute la distribution, la dépendance caudale est locale. Cet indice peut s'interpréter différemment (Malevergne et al. [35]). En prenant l'exemple de deux événements aléatoires  $X$  et  $Y$ , leur coefficient de dépendance indiquerait dans ce cas la probabilité que les deux événements se produisent simultanément. Si  $X$  et  $Y$  désignent les volatilités de deux marchés distincts, leur coefficient de dépendance est la probabilité que les deux marchés soient affectés ensemble d'une forte volatilité.

Soient  $X$  et  $Y$ , deux variables aléatoires dont les marges sont respectivement  $F$  et  $G$ . Alors l'indice de dépendance caudale supérieure est défini par

$$\lambda_U = \lambda_U(X, Y) = \lim_{u \uparrow 1} \mathbb{P} \{ X > F^{-1}(u) | Y > G^{-1}(u) \}.$$

Une définition semblable concerne la dépendance au niveau des queues inférieures, à savoir l'indice de dépendance caudale inférieure dont la définition est

$$\lambda_L = \lambda_L(X, Y) = \lim_{u \downarrow 0} \mathbb{P} \{ X \leq F^{-1}(u) | Y \leq G^{-1}(u) \}.$$

Ces deux indices de dépendance sont intimement liés. En effet, on montre que  $\lambda_L(X, Y) = \lambda_U(-X, -Y)$ . Pour le montrer, on remarque d'abord que la fonction de répartition de  $-X$  est  $\tilde{F}(x) = 1 - F(-x)$ , alors que celle de  $-Y$  est  $\tilde{G}(y) = 1 - G(-y)$ . Ainsi, comme  $\tilde{F}^{-1}(u) = -F^{-1}(1 - u)$  et  $\tilde{G}^{-1}(u) = -G^{-1}(1 - u)$ , alors

$$\begin{aligned} \lambda_U(-X, -Y) &= \lim_{u \uparrow 1} \mathbb{P} \left\{ -X > \tilde{F}^{-1}(u) \mid -Y > \tilde{G}^{-1}(u) \right\} \\ &= \lim_{u \uparrow 1} \mathbb{P} \left\{ -X > -F^{-1}(1 - u) \mid -Y > -G^{-1}(1 - u) \right\} \\ &= \lim_{u \uparrow 1} \mathbb{P} \left\{ X < F^{-1}(1 - u) \mid Y < G^{-1}(1 - u) \right\} \\ &= \lim_{u \downarrow 0} \mathbb{P} \left\{ X < F^{-1}(u) \mid Y < G^{-1}(u) \right\} \\ &= \lambda_L(X, Y). \end{aligned}$$

Il y'a plusieurs lois de probabilités bidimensionnelles qui sont symétriques au sens où la loi de  $(X, Y)$  est la même que celle de  $(-X, -Y)$ . Par exemple, les lois normale et Student possèdent cette propriété. On déduit alors facilement que  $\lambda_L = \lambda_U$ . Une propriété importante des indices  $\lambda_U$  et  $\lambda_L$  est qu'ils dépendent seulement de la copule de la distribution conjointe de  $(X, Y)$ . Ce lien sera établi à la Section 2.3. Juste avant, il convient de définir formellement la notion de copule; la section suivante y est consacrée.

## 2.2 Lois bivariées et copules

### 2.2.1 Introduction

Introduit en 1959 par Abe Sklar, le concept de copule est appréhendé comme solution d'un problème de probabilité. Très peu utilisé, les travaux de Kimeldorf and Sampson [34] sur la dépendance et les recherches de Deheuvels [16] ont suscité un grand intérêt. L'article "*The joy of copulas : Bivariate distributions with uniform marginals*" de Genest and MacKay [24] constitue un fondement important dans la recherche sur la théorie des copules. Suite à cela, C. Genest et ses coauteurs (R. J. MacKay, L.-P. Rivest, P. Capéraà, A. L. Fougères, K. Ghoudi, B. Rémillard, entre autres) ont entrepris de nombreuses recherches statistiques. Ainsi devenu un outil de base, les copules servent maintenant à modéliser des fonctions multivariées, notamment dans le domaine de la finance, de l'assurance et de l'environnement.

### 2.2.2 Théorème de Sklar

Le point de départ de l'étude des copules est un théorème important démontré par Sklar en 1959. Ce résultat permet d'isoler la dépendance lorsqu'on étudie un couple de variables aléatoires  $(X, Y)$ . Il est donc d'une grande importance pour la modélisation et la vérification d'hypothèses concernant des phénomènes bivariés.

**Théorème 2.1.** *Soit  $H$ , une fonction de répartition bivariée de marges continues  $F$  et  $G$ . Alors il existe une unique fonction  $C : [0, 1]^2 \rightarrow [0, 1]$  telle que*

$$H(x, y) = C \{F(x), G(y)\}, \quad \text{pour tout } (x, y) \in \mathbb{R}^2.$$

*La fonction  $C$  s'appelle la copule de  $H$ .*

Pour un couple  $(X, Y)$  de loi  $H$  et de marges  $F$  et  $G$ , la copule  $C$  correspond à la loi conjointe de  $U = F(X)$  et de  $V = G(Y)$ . Ce résultat d'obtient en calculant

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

À noter que l'on a utilisé le fait que comme  $F$  et  $G$  sont continues,  $F\{F^{-1}(u)\} = u$  et  $G\{G^{-1}(v)\} = v$ . Par conséquent,  $C$  est une fonction de répartition sur  $[0, 1]^2$  dont les marges sont uniformes, car on sait que  $U = F(X)$  et  $V = G(Y)$  sont uniformément distribuées sur  $[0, 1]$ .

D'un point de vue strictement mathématique, une fonction  $C : [0, 1]^2 \rightarrow [0, 1]$  doit satisfaire les propriétés  $\mathcal{P}_1$  et  $\mathcal{P}_2$  suivantes pour être qualifiée de copule :

( $\mathcal{P}_1$ ) Pour tout  $u \in [0, 1]$ ,  $C(u, 1) = C(1, u) = 0$  et  $C(u, 0) = C(0, u) = 0$ ;

( $\mathcal{P}_2$ ) Pour tout  $u_1 < u_2$  et  $v_1 < v_2$ ,

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

### 2.2.3 Importance pour la modélisation

Le Théorème de Sklar met en lumière le fait que la loi d'un couple aléatoire  $(X, Y)$  est composée des éléments suivants :

- Le comportement marginal de  $X$ , représenté par  $F$ ;
- Le comportement marginal de  $Y$ , représenté par  $G$ ;
- La dépendance entre  $X$  et  $Y$ , représentée par  $C$ .

Cette constatation est d'une grande utilité pour la modélisation. En effet, ce résultat permet, d'une part, de choisir des modèles univariés adéquats pour les lois marginales

$F$  et  $G$ . D'autre part, la sélection d'un modèle de copule approprié peut se faire indépendamment de ce choix des marges.

## 2.3 Les mesures $\lambda_L$ et $\lambda_U$ s'écrivent comme des fonctionnelles de la copule

De la définition de  $\lambda_U$  donnée à l'Équation (1.1), on déduit que l'indice de dépendance caudale supérieur s'écrit uniquement en fonction de la copule de  $(X, Y)$ . En effet, puisque  $(U, V) \sim C$ , où  $U = F(X)$  et  $V = G(Y)$ , on a

$$\begin{aligned}
 \lambda_U &= \lim_{u \uparrow 1} \mathbb{P} \{X > F^{-1}(u) | Y > G^{-1}(u)\} \\
 &= \lim_{u \uparrow 1} \frac{\mathbb{P} \{X > F^{-1}(u), Y > G^{-1}(u)\}}{\mathbb{P} \{Y > G^{-1}(u)\}} \\
 &= \lim_{u \uparrow 1} \frac{\mathbb{P} \{F(X) > u, G(Y) > u\}}{\mathbb{P} \{G(Y) > u\}} \\
 &= \lim_{u \uparrow 1} \frac{\mathbb{P} \{U > u, V > u\}}{\mathbb{P} \{V > u\}} \\
 &= \lim_{u \uparrow 1} \frac{\bar{C}(u, u)}{1 - u} \\
 &= \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u}. \tag{2.1}
 \end{aligned}$$

Similairement, en partant de la définition de  $\lambda_L$  à l'Équation (1.2), on montre que

$$\lambda_L = \lim_{u \downarrow 0} \frac{C(u, u)}{u}. \tag{2.2}$$

Dans la suite, on écrira  $\lambda_L(C)$  et  $\lambda_U(C)$  afin de mettre l'emphasis sur le fait que les indices de dépendance caudale dépendent uniquement de la copule.

Il est intéressant de noter que les indices de dépendance caudale  $\lambda_L(C)$  et  $\lambda_U(C)$  sont des fonctions croissantes de la dépendance. Plus spécifiquement, on dit qu'une copule

$C_2$  est plus concordante (génère plus de dépendance) qu'une copule  $C_1$  si et seulement si  $C_1(u, v) \leq C_2(u, v)$  pour tout  $(u, v) \in [0, 1]^2$ . À partir des expressions (2.1) et (2.2), on obtient alors facilement que

$$\lambda_U(C_1) \leq \lambda_U(C_2) \quad \text{et} \quad \lambda_L(C_1) \leq \lambda_L(C_2).$$

Autrement dit, la valeur des indices de dépendance caudale augmente à mesure que le niveau de dépendance augmente.

## 2.4 Indices de dépendance caudale pour quelques familles de copules

### 2.4.1 Cas d'une symétrie radiale

On dit qu'une copule possède la propriété de symétrie radiale si

$$C(u, v) = u + v - 1 + C(1 - u, 1 - v).$$

Dans ce cas, les indices de dépendance caudale sont égaux, c'est-à-dire que  $\lambda_L(C) = \lambda_U(C)$ . En effet, par un calcul direct basé sur les Équations (2.1) et (2.2),

$$\begin{aligned}
 \lambda_L(C) &= \lim_{u \downarrow 0} \frac{C(u, u)}{u} \\
 &= \lim_{u \downarrow 0} \frac{u + u - 1 + C(1 - u, 1 - u)}{u} \\
 &= \lim_{u \downarrow 0} \frac{2u - 1 + C(1 - u, 1 - u)}{u} \\
 &= \lim_{v \uparrow 1} \frac{2(1 - v) - 1 + C(v, v)}{1 - v} \\
 &= \lim_{v \uparrow 1} \frac{1 - 2v + C(v, v)}{1 - v} \\
 &= \lambda_U(C).
 \end{aligned}$$

### 2.4.2 Copule d'indépendance et bornes de Fréchet

La copule qui correspond à l'indépendance entre deux variables est  $\Pi(u, v) = uv$ . On a alors facilement que

$$\lambda_L(\Pi) = \lim_{u \downarrow 0} \frac{u^2}{u} = 0.$$

Comme cette copule possède la propriété de symétrie radiale, on déduit immédiatement que  $\lambda_U(\Pi) = 0$ . Les bornes inférieure et supérieure de Fréchet sont les copules définies par

$$M(u, v) = \min(u, v) \quad \text{et} \quad W(u, v) = \max(u + v - 1, 0).$$

Elles sont telles que pour toute copule  $C$ , on a

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

La copule  $W$  correspond à la dépendance négative parfaite, alors que  $M$  correspond à la dépendance positive parfaite. On peut montrer que ces deux copules possèdent la

propriété de symétrie radiale. De plus, par des calculs directs, on montre facilement que  $\lambda_L(W) = 0$  et  $\lambda_L(M) = 1$ . Ainsi,  $\lambda_L(W) = \lambda_U(W) = 0$  et  $\lambda_L(M) = \lambda_U(M) = 1$ .

### 2.4.3 Copule Normale

Par une application du Théorème de Sklar, il est facile de construire un modèle dont la dépendance est héritée du modèle normal. Notons d'abord que la fonction de répartition de la loi normale standard est

$$H_\rho(x, y) = \int_{-\infty}^x \int_{-\infty}^y \phi_\rho(s, t) dt ds,$$

où  $\rho$  est le coefficient de corrélation. Comme les deux marges dans ce cas sont celles de la loi Normale standard univariée notée  $\Phi$ , alors la copule normale est

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \phi_\rho(s, t) dt ds.$$

On peut donc construire une loi bivariée dont la dépendance, *i.e.* la copule, est de type Normal et dont les marges sont  $F$  et  $G$ . Si on choisit des marges normales, c'est-à-dire

$$F(x) = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right) \quad \text{et} \quad G(y) = \Phi\left(\frac{y - \mu_Y}{\sigma_Y}\right),$$

alors on retrouve le modèle traditionnel normal de moyennes  $\mu_X, \mu_Y$ , de variances  $\sigma_X^2, \sigma_Y^2$ , et de corrélation  $\rho$ . La copule Normale possède la propriété de symétrie radiale. Cependant, elle ne présente pas de dépendance caudale, sauf lorsque  $\rho = 1$ . Ainsi, on a  $\lambda_L(C_\rho) = \lambda_U(C_\rho) = 0$  pour  $\rho \in [-1, 1)$ , alors que  $\lambda_L(C_\rho) = \lambda_U(C_\rho) = 1$  si  $\rho = 1$ .

### 2.4.4 Copule de Student

Un vecteur aléatoire  $(X, Y)$  est distribué selon la loi de Student à  $\nu \in \mathbb{N}$  degrés de liberté et de corrélation  $\rho \in [-1, 1]$  s'il admet la représentation stochastique

$$(X, Y) = \left( \frac{Z_1}{\sqrt{W/\nu}}, \frac{Z_2}{\sqrt{W/\nu}} \right),$$

où  $(Z_1, Z_2)$  est de loi Normale standard de coefficient de corrélation  $\rho$  et  $W$  suit la loi Khi-deux univariée à  $\nu$  degrés de liberté. Dans ce cas, la densité de  $(X, Y)$  est

$$h_{\rho, \nu}^T(x, y) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \pi \nu \rho} \left( 1 + \frac{x^2 + y^2 - 2\rho xy}{\nu} \right)^{-(\nu+2)/2},$$

où  $\Gamma$  est la fonction gamma. La copule de Student est alors définie implicitement par

$$C_{\rho, \nu}^T(u, v) = \int_{-\infty}^{F_\nu^{-1}(u)} \int_{-\infty}^{F_\nu^{-1}(v)} h_{\rho, \nu}^T(x, y) dy dx,$$

où  $F_\nu$  est la fonction de répartition de la distribution de Student univariée à  $\nu$  degrés de liberté. Cette copule est radialement symétrique. La densité de la copule de Student peut être exprimée via

$$c_{\rho, \nu}^T(u, v) = \frac{h_{\rho, \nu}^T\{F_\nu^{-1}(u), F_\nu^{-1}(v)\}}{f_\nu\{F_\nu^{-1}(u)\} f_\nu\{F_\nu^{-1}(v)\}}, \quad (2.3)$$

où  $f_\nu = F_\nu'$  est la densité de la Student à  $\nu$  degrés de liberté. De plus, on a

$$\lambda_L(C_{\rho, \nu}^T) = \lambda_U(C_{\rho, \nu}^T) = 2 F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right).$$

Contrairement à la copule Normale, la copule de Student possède donc de la dépendance caudale, c'est-à-dire que  $\lambda_L(C_{\rho, \nu}^T) = \lambda_U(C_{\rho, \nu}^T) > 0$ . En particulier, lorsque  $\rho = 1$ ,

$$\lambda_L(C_{\rho, \nu}^T) = \lambda_U(C_{\rho, \nu}^T) = 2 F_{\nu+1}(0) = 1.$$

La copule de Student tend vers la copule Normale lorsque  $\nu \rightarrow \infty$ . Il n'est donc pas surprenant de constater que si  $\rho \neq 1$ , alors

$$\lim_{\nu \rightarrow \infty} \lambda_L(C_{\rho, \nu}^T) = \lim_{\nu \rightarrow \infty} 2F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right) = \lim_{\kappa \rightarrow -\infty} 2F_{\nu+1}(\kappa) = 0.$$

### 2.4.5 Copules Archimédiennes

Du fait de la facilité avec laquelle on peut les construire, les copules Archimédiennes sont beaucoup utilisées. On retrouve dans cette famille les copules Clayton, Frank et Gumbel. De façon générale, une copule est Archimédienne si elle admet une représentation de la forme

$$C(u, v) = \begin{cases} \varphi^{-1} \{ \varphi(u) + \varphi(v) \}, & \text{si } \varphi(u) + \varphi(v) \leq 0; \\ 0 & \text{sinon,} \end{cases}$$

où  $\varphi : [0, 1] \rightarrow \mathbb{R}$  est une fonction décroissante et convexe telle que  $\varphi(1) = 0$ . On appelle  $\varphi$  la fonction génératrice.

Puisque  $C(u, u) = \varphi^{-1} \{ 2\varphi(u) \}$ , on déduit facilement les formules pour les indices de dépendance caudale d'une copule Archimédienne. En posant  $v = \varphi(u)$ , c'est-à-dire que  $u = \varphi^{-1}(v)$ , alors

$$\lambda_L = \lim_{u \downarrow 0} \frac{\varphi^{-1} \{ 2\varphi(u) \}}{u} = \lim_{v \downarrow \infty} \frac{2\varphi^{-1}(2v)}{\varphi^{-1}(v)} = \lim_{v \downarrow \infty} \frac{2\varphi'(2v)}{\varphi'(v)}.$$

De même,

$$\lambda_U = \lim_{u \downarrow 1} \frac{1 - 2u + \varphi^{-1} \{ 2\varphi(u) \}}{1 - u} = \lim_{u \downarrow 0} \frac{1 - 2u + \varphi^{-1}(v) + \varphi^{-1}(2v)}{1 - \varphi^{-1}(v)}$$

Un cas particulier est la copule Clayton dont le générateur paramétrique est donné

pour  $\theta > 0$  par

$$\varphi_\theta(t) = \frac{t^{-\theta} - 1}{\theta}.$$

La copule de Clayton s'écrit donc

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

La copule de Frank est générée pour  $\theta \in \mathbb{R} \setminus \{0\}$  par

$$\varphi_\theta(t) = -\ln \left( 1 + \frac{e^{\theta t} - 1}{e^\theta - 1} \right).$$

La forme de la copule associée est alors

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{\theta u} - 1)(e^{\theta v} - 1)}{e^\theta - 1} \right\}.$$

Un autre modèle de dépendance populaire est la copule de Gumbel, qui possède le générateur  $\varphi_\theta(t) = |\ln t|^\theta$  pour  $\theta \in [0, 1]$  et dont la forme est

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

Le Tableau 2.1 fait une synthèse de ces trois copules, incluant les valeurs de  $\lambda_L$  et  $\lambda_U$ .

TABLE 2.1 – Caractéristiques de trois copules Archimédiennes

Copule	$\varphi_\theta(u)$	$C_\theta(u, v)$	$\lambda_L$	$\lambda_U$
Clayton	$\frac{t^{-\theta} - 1}{\theta}$	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$2^{-1/\theta}$	0
Frank	$-\ln \left( 1 + \frac{e^{\theta t} - 1}{e^\theta - 1} \right)$	$-\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{\theta u} - 1)(e^{\theta v} - 1)}{e^\theta - 1} \right\}$	0	0
Gumbel	$ \ln t ^\theta$	$\exp \left[ - \{  \ln u ^\theta +  \ln v ^\theta \}^{1/\theta} \right]$	0	$2 - 2^{1-\theta}$

### 2.4.6 Copules à valeurs extrêmes

Une copule extrême est la fonction de dépendance extraite d'une loi bivariée extrême. Leur construction se fera à partir d'une caractérisation. D'abord, une copule  $C$  est max-stable si pour tout  $r > 0$ ,

$$\{C(u^{1/r}, v^{1/r})\}^r = C(u, v).$$

On peut montrer que  $C$  est extrême si et seulement si elle est max-stable. Ceci permet de déduire que les copules à valeurs extrêmes s'écrivent toujours sous la forme

$$C_A(u, v) = \exp \left\{ \ln uv A \left( \frac{\ln u}{\ln uv} \right) \right\},$$

où  $A : [0, 1] \rightarrow [1/2, 1]$  s'appelle la fonction de dépendance extrême. Il s'agit de la forme générale des copules à valeurs extrêmes. Il s'ensuit alors que

$$\begin{aligned} \ln C_A(u, u) &= \ln uu A \left( \frac{\ln u}{\ln uu} \right) \\ &= \ln u^2 A \left( \frac{\ln u}{\ln u^2} \right) \\ &= 2A \left( \frac{1}{2} \right) \ln u \\ &= \ln u^{2A(1/2)}. \end{aligned}$$

On a donc  $C_A(u, u) = u^{2A(1/2)}$ , ce qui permet de montrer que le coefficient de dépendance caudale inférieure d'une copule à valeurs extrêmes s'exprime par

$$\lambda_L(C_A) = \lim_{u \uparrow 0} \frac{u^{2A(\frac{1}{2})}}{u} = \lim_{u \uparrow 0} u^{2A(\frac{1}{2})-1}.$$

On aboutit alors à

$$\lambda_L(C_A) = \begin{cases} 1, & \text{si } A(1/2) = 1/2; \\ 0, & \text{sinon} \end{cases}.$$

Pour le coefficient de dépendance caudale supérieur, on a

$$\begin{aligned}\lambda_U(C_A) &= \lim_{u \uparrow 1} \frac{1 - 2u + u^{2A(1/2)}}{1 - u} \\ &= \lim_{u \uparrow 1} 2 - 2A(1/2)u^{2A(1/2)-1} \\ &= 2 \left\{ 1 - A \left( \frac{1}{2} \right) \right\}.\end{aligned}$$

Lorsque  $A(t) = 1$ , on retrouve la copule d'indépendance car alors,  $C_A(u, v) = uv$ ; on retrouve donc  $\lambda_L(C_A) = 0$  et  $\lambda_U(C_A) = 1$ . Quand  $A(t) = \max(t, 1 - t)$ , la copule associée est la borne supérieure de Fréchet-Hoeffding. Puisqu'alors,  $A(1/2) = 1/2$ , il s'ensuit que  $\lambda_L(C_A) = \lambda_U(C_A) = 1$ .

La copule Archimédienne de Gumbel—Hougaard appartient également à la familles des valeurs extrêmes. Sa fonction de dépendance extrême est définie par

$$A_\theta(t) = \{t^{1/(1-\theta)} + (1-t)^{1/(1-\theta)}\}^{1-\theta}, \quad \theta \in [0, 1].$$

On a alors

$$A \left( \frac{1}{2} \right) = \left\{ \left( \frac{1}{2} \right)^{1/(1-\theta)} + \left( 1 - \frac{1}{2} \right)^{1/(1-\theta)} \right\}^{1-\theta} = \frac{2^{1-\theta}}{2}.$$

Si  $\theta < 1$ , on a donc  $\lambda_L(C_A) = 0$ , alors que  $\lambda_L(C_A) = 1$  si  $\theta = 1$ . On a aussi

$$\lambda_U(C_A) = 2 \left( 1 - A \left( \frac{1}{2} \right) \right) = 2 - 2^{1-\theta}.$$

Pour la copule de Galambos, on a

$$A_\theta \left( \frac{1}{2} \right) = 1 - 2^{-(1+\theta)/\theta}, \quad \theta \in [0, 1].$$

On obtient alors  $\lambda_L(C_A) = 0$  et  $\lambda_U(C_A) = 2^{-1/\theta}$ . Le Tableau 2.2 offre une synthèse de quelques copules à valeurs extrêmes.

TABLE 2.2 – Caractéristiques de quelques copules à valeurs extrêmes

Copule	$A(1/2)$	$C_A(u, u)$	$\lambda_L(C_A)$	$\lambda_U(C_A)$
Indépendance	1	$u^2$	0	1
Fréchet-Hoeffding	1/2	$u$	1	1
Gumbel-Hougaard	$2^{1-\theta}/2$	$u^{2^{1-\theta}}$	0	$2 - 2^{1-\theta}$
Galambos	$1 - 2^{-(1+\theta)/\theta}$	$u^{2+1/\theta}$	0	$2^{-1/\theta}$

# Chapitre 3

## Revue des méthodes d'estimation de $\lambda_U$ et regard critique

### 3.1 Préliminaires

#### 3.1.1 Mise en contexte

Il existe essentiellement trois façons d'estimer une copule ; il s'agit d'approches

- paramétriques ;
- semi-paramétriques ;
- non-paramétriques.

La méthode paramétrique nécessite de spécifier à la fois la structure de dépendance, *i.e.* la copule, ainsi que les distributions marginales. Dans ce cas, on obtient typiquement les estimateurs des paramètres à l'aide d'un maximum de vraisemblance basé sur la loi multidimensionnelle complète. Cette approche est généralement fastidieuse en regard de la complexité des calculs, car il faut estimer simultanément beaucoup de

paramètres, à savoir ceux associés à la copule et ceux des lois marginales.

Pour remédier aux limites de la méthode paramétrique basée sur la vraisemblance complète, Joe and Hu [31] ont proposé une méthode par vraisemblance en deux étapes. Cette idée est relativement simple à mettre en œuvre. La première étape consiste à estimer les paramètres des lois marginales. Dans la deuxième étape, ces estimateurs sont injectés dans la vraisemblance et les estimateurs de la copule sont obtenus conditionnellement à ces valeurs estimées.

Les deux approches précédemment décrites entrent dans la catégorie des méthodes paramétriques. Ainsi, celles-ci montrent le désavantage de dépendre des hypothèses sur les distributions marginales. Pour relâcher substantiellement ces hypothèses, Shih and Louis [50] et Genest et al. [25] ont développé une méthode semi-paramétrique qui consiste à estimer les distributions marginales par le biais des distributions empiriques. Ensuite, conditionnellement à ces estimations entièrement non-paramétriques, les paramètres de la copule sont estimés par maximum de vraisemblance. Les méthodes d'estimation des paramètres de type paramétriques et semi-paramétriques ont été étudiées par Kim et al. [33] dans une étude de simulation exhaustive.

L'approche non-paramétrique consiste à estimer la copule d'une population sans que ne soit émise la moindre hypothèse sur sa forme. La façon habituelle de procéder consiste à calculer la copule empirique. La première suggestion dans ce sens est attribuable à Deheuvels [16], qui introduisit la *fonction de dépendance* dans un cadre de tests d'indépendance. La version la plus utilisée est toutefois due à Fermanian et al. [18]; voir également Segers [46] pour les propriétés asymptotiques de la copule empirique. Une manière alternative d'estimer une copule est d'utiliser une méthode de lissage par noyau. Cette idée a été élaborée, par exemple, par Gijbels and Mielniczuk [28], Chen and Huang [10] et Omelka et al. [37]. Cette méthode est intéressante dans la mesure où elle permet de minimiser le biais aux frontières.

### 3.1.2 Copule empirique de Deheuvels (1979)

Connue sous le nom de fonction de distribution empirique de dépendance, Deheuvels [16] a introduit une estimation de la copule afin de construire des tests non paramétriques pour l'indépendance. Pour la décrire, soit un échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  issu d'une distribution conjointe  $H$  à marges continues  $F$  et  $G$ . Dans un premier temps, pour  $(x, y) \in \mathbb{R}^2$ , on définit la fonction de répartition conjointe empirique par

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

On déduit ensuite de  $H_n$  les fonctions de répartition empiriques marginales, à savoir

$$F_n(x) = H_n(x, \infty) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(X_i \leq x) \quad \text{et} \quad G_n(y) = H_n(\infty, y) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(Y_i \leq y).$$

On rappelle que l'unique copule  $C$  associée à  $H$  s'obtient en faisant

$$C(u, v) = H \{F^{-1}(u), G^{-1}(v)\}.$$

De là, Deheuvels [16] a proposé une estimation non-paramétrique de la copule, à savoir

$$C_n^*(u, v) = H_n \{F_n^{-1}(u), G_n^{-1}(v)\}. \quad (3.1)$$

De la définition même de  $H_n$ , on a explicitement

$$C_n^*(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{I} \{X_i \leq F_n^{-1}(u), Y_i \leq G_n^{-1}(v)\}. \quad (3.2)$$

### 3.1.3 Copule empirique à base de rangs

La copule empirique  $C_n^*$  définie à l'Équation (3.1) est cependant peu employée en pratique. La raison principale est la difficulté à obtenir les inverses  $F_n^{-1}$  et  $G_n^{-1}$  des fonctions étagées  $F_n$  et  $G_n$ . Même s'il est faux de dire que  $X_i \leq F_n^{-1}(u)$  est vrai si et seulement si  $F_n(X_i) \leq u$ , cette affirmation est *approximativement* vraie. Ainsi, en partant de l'Équation (3.2),

$$C_n^*(u, v) \approx C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{F_n(X_i) \leq u, G_n(Y_i) \leq v\}.$$

La fonction empirique  $C_n$  est la version la plus couramment utilisée de copule empirique. Elle a été utilisée dans de nombreux contextes, par exemple par Genest et al. [26] pour tester l'adéquation.

Il est intéressant de remarquer que  $n F_n(X_i)$  correspond au rang  $R_i$  de l'observation  $X_i$  dans l'échantillon  $X_1, \dots, X_n$ . De même,  $n G_n(Y_i)$  est le rang  $S_i$  de  $Y_i$  parmi  $Y_1, \dots, Y_n$ . On peut alors écrire

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\left(\frac{R_i}{n} \leq u, \frac{S_i}{n} \leq v\right).$$

Ceci met en lumière le fait que la copule empirique  $C_n$ , de même que les méthodes statistiques qui en découlent, sont à base de techniques de rangs.

## 3.2 Quelques estimateurs de dépendance caudale

Dans la suite, soient les paires de pseudos-observations  $(\widehat{U}_1, \widehat{V}_1), \dots, (\widehat{U}_n, \widehat{V}_n)$ , où pour chaque  $i \in \{1, \dots, n\}$ , on a défini  $\widehat{U}_i = F_n(X_i) = R_i/n$  et  $\widehat{V}_i = G_n(Y_i) = S_i/n$ .

### 3.2.1 Estimateur de Capéreaa et coll. (1997)

L'estimateur de Capéreaa et al. [9] est donné par

$$\widehat{\lambda}_U^{\text{CFG}} = 2 - 2 \left\{ \frac{1}{n} \sum_{i=1}^n \ln \left( \frac{\sqrt{\ln \widehat{U}_i \ln \widehat{V}_i}}{-\ln(\widehat{U}_i^2 \vee \widehat{V}_i^2)} \right) \right\},$$

où  $a \vee b$  est le maximum entre deux réels  $a, b \in \mathbb{R}$ .

### 3.2.2 Estimateur de Schmidt et Stadtmüller (2006)

Pour  $(U, V) \sim C$ , soit  $\bar{C}(u, v) = \mathbb{P}(U > u, V > v)$ . Comme  $\bar{C}(u, v) = 1 - u - v + C(u, v)$ , on peut alors écrire

$$\lambda_U = \lim_{u \uparrow 1} \frac{\bar{C}(u, u)}{1 - u}.$$

Les chercheurs Schmidt and Stadtmüller [44] se sont inspirés de cette représentation pour suggérer l'estimateur

$$\begin{aligned} \widehat{\lambda}_U^{\text{SS}} &= \frac{\bar{C}_n \{1 - (k/n), 1 - (k/n)\}}{1 - \{1 - (k/n)\}} \\ &= \frac{1}{k} \sum_{i=1}^n \mathbb{I} \left( \widehat{U}_i > 1 - \frac{k}{n}, \widehat{V}_i > 1 - \frac{k}{n} \right) \\ &= \frac{1}{k} \sum_{i=1}^n \mathbb{I} \left( 1 - \widehat{U}_i < \frac{k}{n}, 1 - \widehat{V}_i < \frac{k}{n} \right). \end{aligned}$$

Cet estimateur nécessite de choisir une valeur seuil  $k \in \mathbb{N}$ , où  $k \ll n$ . On peut, à cette fin, utiliser l'algorithme heuristique de recherche de plateau tel que suggéré par Schmidt [43] (voir aussi Frahm et al. [20]).

### 3.2.3 Estimateur de Coles *et coll.* (1999)

En partant de la définition de  $\lambda_U$ , on peut déduire que

$$\lambda_U = 2 - \lim_{u \uparrow 1} \frac{\ln C(u, u)}{\ln u}.$$

En effet,

$$\begin{aligned} \lambda_U &= \lim_{u \uparrow 1} \frac{1 + 2u + C(u, u)}{1 - u} \\ &= \lim_{u \uparrow 1} \frac{2(1 - u) - 1 + C(u, u)}{1 - u} \\ &= 2 - \lim_{u \uparrow 1} \frac{1 - C(u, u)}{1 - u} \\ &= 2 - \lim_{u \uparrow 1} \frac{dC(u, u)/du}{du} \\ &= 2 - \lim_{u \uparrow 1} \frac{d \ln C(u, u)/du}{d \ln u/du} \\ &= 2 - \lim_{u \uparrow 1} \frac{\ln C(u, u)}{\ln u}. \end{aligned}$$

En partant de cette représentation pour  $\lambda_U$ , Coles *et al.* [13] ont proposé l'estimateur

$$\widehat{\lambda}_U^{\text{Coles}} = 2 - \frac{\ln C_n\{1 - (k/n), 1 - (k/n)\}}{\ln\{1 - (k/n)\}}.$$

### 3.2.4 Estimateur de Joe *et coll.* (1992)

On note que

$$\lambda_U = \lim_{u \uparrow 1} \frac{1 + 2u + C(u, u)}{1 - u} = \lim_{u \uparrow 1} \frac{2(1 - u) - 1 + C(u, u)}{1 - u} = 2 - \lim_{u \uparrow 1} \frac{1 - C(u, u)}{1 - u}.$$

Tel que suggéré par Joe et al. [32], on déduit alors l'estimateur

$$\begin{aligned}\widehat{\lambda}_U^{\text{Joe}} &= 2 - \frac{1 - C_n\{1 - (k/n), 1 - (k/n)\}}{1 - \{1 - (k/n)\}} \\ &= 2 - \frac{n}{k} \left\{ 1 - \frac{1}{n} \sum_{i=1}^n \mathbb{I} \left( \widehat{U}_i \leq 1 - \frac{k}{n}, \widehat{V}_i \leq 1 - \frac{k}{n} \right) \right\} \\ &= 2 - \frac{n}{k} + \frac{1}{k} \sum_{i=1}^n \mathbb{I} \left( 1 - \widehat{U}_i \geq \frac{k}{n}, 1 - \widehat{V}_i \geq \frac{k}{n} \right).\end{aligned}$$

En fait,  $\widehat{\lambda}_U^{\text{Joe}} = \widehat{\lambda}_U^{\text{SS}}$ , et donc  $\widehat{\lambda}_U^{\text{Joe}}$  ne sera plus considéré dans la suite de ce travail.

### 3.3 Limites et faiblesses des estimateurs existants

L'analyse fréquentielle a beaucoup évolué avec l'introduction d'outils statistiques plus maniables comme les copules. Plusieurs études hydrologiques ont conduit à une application extensive de modèles. On peut citer les travaux de Ghizzoni et al. [27] sur l'étude de l'occurrence d'événements extrêmes et les travaux de Bardossy and Pegram [4] sur les problèmes d'interpolation spatiale et de simulation. Néanmoins ce corpus littéraire est plus axé sur l'application potentielle au détriment de la nature réelle des variables disponibles. A cela s'ajoutent les lacunes liées à la fiabilité des méthodes utilisées ou appliquées à des séries chronologiques très courtes. Bardossy and Pegram [5] quant à eux ont mis l'accent sur certains concepts de base à savoir : la nature des relations entre les variables hydrologiques et la fiabilité de l'inférence multivariée appliquée sur des séries chronologiques typiquement courtes. Ainsi il s'agit d'appliquer des concepts à priori univariés vers le cadre multivarié.

Le coefficient de dépendance caudale donné par  $\lambda_U$  donne une idée de la tendance de l'occurrence des événements extrêmes pour un certain seuil quantile fixé. De ce fait, pour mieux cerner la manière dont les événements extrêmes se comportent, il est important de l'évaluer afin de pouvoir choisir le meilleur modèle. Ainsi dans la

littérature les estimateurs les plus utilisés sont au nombre de quatre à savoir :

- L'estimateur  $\widehat{\lambda}_U^{\text{SS}}$  (Schmidt and Stadtmüller [44]) présente parfois une valeur élevée de la variance ;
- L'estimateur  $\widehat{\lambda}_U^{\text{Coles}}$  basé sur la sécante et est défini comme étant la pente de la sécante le long de la diagonale de la copule proche du point (1,1) mais peut être biaisé et montre une très faible variance ;
- L'estimateur  $\widehat{\lambda}_U^{\text{CFG}}$  pose l'hypothèse que la fonction de copule empirique se rapproche d'une copule à valeur extrême mais pourrait être biaisé et montrer de très faible variance ;
- L'estimateur  $\widehat{\lambda}_U^{\text{Coles}}$  est la contrepartie non paramétrique de l'estimateur proposé par Coles et al. [13].

Poulin et al. [38] et Serinaldi [47] ont été les pionniers de l'usage du coefficient de dépendance caudale dans les études hydrologiques. Ainsi ils ont réalisé une étude de cas spécifique pour choisir le meilleur estimateur. Pour cela ils ont retenu les estimateurs de Coles-Heffernan-Tawn noté  $\widehat{\lambda}_U^{\text{Coles}}$  (Coles et al. [13]) et de Capéraà-Fougères-Genest noté  $\widehat{\lambda}_U^{\text{CFG}}$  (Capéraà et al. [9]). Serinaldi [47] a construit un diagramme de diagnostic dans l'objectif de choisir le modèle adéquat. Pour cela, il a mesuré le lien entre  $\lambda_U$  et le tau de Kendall. La fiabilité de son outil est démontrée en réalisant des simulations qui font ressentir le biais de  $\widehat{\lambda}_U^{\text{CFG}}$  et l'incertitude de  $\widehat{\lambda}_U^{\text{SS}}$  (Schmidt and Stadtmüller [44]). Cependant une évaluation approfondie n'a pas été faite et les mises en gardes de Schmidt [43] et Frahm et al. [20] ne sont pas fidèlement rapportées.

Les hydrologues se sont malheureusement basés beaucoup sur les travaux traitant de  $\lambda_U$  dans le même domaine, au détriment de la littérature originelle. Cela a débouché sur un usage plutôt irrationnel de  $\widehat{\lambda}_U^{\text{CFG}}$  et des autres estimateurs de  $\lambda_U$ . Dans cette même lancée, Serinaldi and Kilsby [48] ont nuancé les recommandations de AghaKouchak et al. [2] sur l'utilisation des  $\lambda_U$  comme outils d'aide à l'élaboration de politiques ainsi qu'à la validation de modèles numériques. Ils ont en effet prouvé que les champs de précipitations simulés par la structure de dépendance spatiale méta-gaussienne présentent des estimations de  $\lambda_U$  similaires à celles des champs de précipitations observés.

Serinaldi et al. [49] se sont focalisés seulement sur deux estimateurs parmi les quatre à savoir : l'estimateur  $\widehat{\lambda}_U^{SS}$  du fait qu'il ne nécessite aucune hypothèse de distribution et l'estimateur  $\widehat{\lambda}_U^{CFG}$  en raison de sa popularité liée à sa formule de forme fermée. Ainsi ils ont pu montrer que les estimateurs  $\widehat{\lambda}_U^{SS}$  et  $\widehat{\lambda}_U^{CFG}$  sont généralement biaisés et donnent des valeurs  $\lambda_U$  fortement liées à la corrélation globale même si la structure de dépendance sous-jacente (vraie) a une dépendance caudale nulle. En plus, l'incertitude est élevée même si les échantillons d'expériences sont de tailles grandes. De ce fait, ces estimateurs devraient être utilisés avec beaucoup de prudence car leur interprétation automatique peut donner des résultats erronés.

# Chapitre 4

## De nouveaux estimateurs des coefficients de dépendance caudale

### 4.1 La famille des copules Fisher

#### 4.1.1 Famille des copules Khi-deux

Soit  $(Z_1, Z_2)$ , une paire de variables aléatoires de distribution Normale standard (*i.e.* de moyennes nulles et de variances unitaires) de coefficient de corrélation  $\rho \in [-1, 1]$ . Alors pour un certain paramètre de décentralité  $a \in \mathbb{R}$ , la copule Khi-deux, notée  $C_{\rho, a}^X$ , est définie comme la structure de dépendance du couple

$$(Y_1, Y_2) = ((Z_1 + a)^2, (Z_2 + a)^2).$$

Cette famille de modèles de copules a d'abord été introduite par Bårdossy [7], alors que Quessy et al. [40] ont étudié ses propriétés. Le nom de cette copule découle du fait que les marges du couple  $(Y_1, Y_2)$  qui sert de base à la construction sont des Khi-deux

décentrées à 1 degré de liberté. Parmi les nombreuses propriétés identifiées par Quessy et al. [40], on note que la copule Khi-deux de paramètre de décentralité  $a = -r < 0$  est la même que lorsque  $a = r$ . On peut donc supposer, sans perte de généralité, que  $a \geq 0$ . En outre, la copule normale apparaît dans le cas particulier où  $a \rightarrow \infty$ . Sinon, c'est-à-dire lorsque  $a \in [0, \infty)$ , les copules Khi-deux sont radialement asymétriques. Cependant, les coefficients de dépendance caudale inférieure et supérieure sont nulles pour les copules Khi-deux, c'est-à-dire que

$$\lambda_L(C_{\rho,a}^X) = \lambda_U(C_{\rho,a}^X) = 0.$$

La seule exception survient quand  $|\rho| = 1$ , dans lequel cas  $\lambda_L(C_{\rho,a}^X) = \lambda_U(C_{\rho,a}^X) = 1$ .

### 4.1.2 Construction d'une copule Fisher

Suivant une idée similaire à celle menant à la copule Khi-deux, la famille des copules Fisher sera construite. Pour ce faire, soit un vecteur aléatoire  $(X_1, X_2)$  distribué selon la loi de Student à  $\nu \in \mathbb{N}$  degrés de liberté et de paramètre  $\rho \in [-1, 1]$ . Alors la copule de Fisher, notée  $C_{\rho,\nu}^F$ , est la structure de dépendance extraite de la loi de

$$(Y_1, Y_2) = (X_1^2, X_2^2).$$

Cette famille de copules a été introduite par Favre et al. [17] dans le cas  $d$ -dimensionnel. Son nom origine du fait que les marges du couple  $(Y_1, Y_2)$  qui sert de base à la construction sont Fisher à 1 degré de liberté. Puisque  $\tilde{F}_\nu(x) = \mathbb{P}(X_j^2 \leq x) = 2F_\nu(\sqrt{x}) - 1$ , la copule de Fisher correspond à la loi conjointe de  $(\tilde{F}_\nu(X_1^2), \tilde{F}_\nu(X_2^2)) = (2F_\nu(|X_1|) -$

$1, 2F_\nu(|X_2|) - 1$ ). La copule Fisher en  $d = 2$  dimensions peut alors s'exprimer par

$$\begin{aligned}
 C_{\rho, \nu}^F(u, v) &= \mathbb{P} \left\{ 2F_\nu(|X_1|) - 1 \leq u, 2F_\nu(|X_2|) - 1 \leq v \right\} \\
 &= \mathbb{P} \left\{ |X_1| \leq F_\nu^{-1} \left( \frac{1+u}{2} \right), |X_2| \leq F_\nu^{-1} \left( \frac{1+v}{2} \right) \right\} \\
 &= \mathbb{P} \left\{ -F_\nu^{-1} \left( \frac{1+u}{2} \right) \leq X_1 \leq F_\nu^{-1} \left( \frac{1+u}{2} \right), \right. \\
 &\quad \left. -F_\nu^{-1} \left( \frac{1+v}{2} \right) \leq X_2 \leq F_\nu^{-1} \left( \frac{1+v}{2} \right) \right\} \\
 &= \mathbb{P} \left\{ F_\nu^{-1} \left( \frac{1-u}{2} \right) \leq X_1 \leq F_\nu^{-1} \left( \frac{1+u}{2} \right), \right. \\
 &\quad \left. F_\nu^{-1} \left( \frac{1-v}{2} \right) \leq X_2 \leq F_\nu^{-1} \left( \frac{1+v}{2} \right) \right\} \\
 &= \mathbb{P} \left\{ \frac{1-u}{2} \leq F_\nu(X_1) \leq \frac{1+u}{2}, \frac{1-v}{2} \leq F_\nu(X_2) \leq \frac{1+v}{2} \right\}. \quad (4.1)
 \end{aligned}$$

Pour obtenir cette chaîne d'égalités, on a utilisé le fait que comme  $F_\nu$  est symétrique autour de zéro, alors  $-F_\nu^{-1}(x) = F_\nu^{-1}(1-x)$ . Étant donné que par construction,  $(X_1, X_2)$  suit une distribution de Student, le vecteur  $(U, V) = (F_\nu(X_1), F_\nu(X_2))$  est distribué selon la copule de Student. Par conséquent, on peut écrire

$$\begin{aligned}
 C_{\rho, \nu}^F(u, v) &= \mathbb{P} \left( \frac{1-u}{2} \leq U \leq \frac{1+u}{2}, \frac{1-v}{2} \leq V \leq \frac{1+v}{2} \right) \\
 &= \mathbb{P} \left( U \leq \frac{1+u}{2}, V \leq \frac{1+v}{2} \right) + \mathbb{P} \left( U \leq \frac{1-u}{2}, V \leq \frac{1-v}{2} \right) \\
 &\quad - \mathbb{P} \left( U \leq \frac{1-u}{2}, V \leq \frac{1+v}{2} \right) - \mathbb{P} \left( U \leq \frac{1+u}{2}, V \leq \frac{1-v}{2} \right) \\
 &= C_{\rho, \nu}^T \left( \frac{1+u}{2}, \frac{1+v}{2} \right) + C_{\rho, \nu}^T \left( \frac{1-u}{2}, \frac{1-v}{2} \right) \\
 &\quad - C_{\rho, \nu}^T \left( \frac{1-u}{2}, \frac{1+v}{2} \right) - C_{\rho, \nu}^T \left( \frac{1+u}{2}, \frac{1-v}{2} \right),
 \end{aligned}$$

où  $C_{\rho,\nu}^T$  est la copule de Student. La densité correspondante est donc

$$\begin{aligned} c_{\rho,\nu}^F(u, v) &= \frac{\partial^2}{\partial u \partial v} C_{\rho,\nu}^F(u, v) \\ &= \frac{1}{4} \left\{ c_{\rho,\nu}^T \left( \frac{1+u}{2}, \frac{1+v}{2} \right) + c_{\rho,\nu}^T \left( \frac{1-u}{2}, \frac{1-v}{2} \right) \right. \\ &\quad \left. + c_{\rho,\nu}^T \left( \frac{1-u}{2}, \frac{1+v}{2} \right) + c_{\rho,\nu}^T \left( \frac{1+u}{2}, \frac{1-v}{2} \right) \right\}, \end{aligned} \quad (4.2)$$

où  $c_{\rho,\nu}^T$  est la densité de la copule de Student.

### 4.1.3 Coefficients de dépendance caudale de la copule Fisher

Tel que démontré par Favre et al. [17], la copule Fisher ne possède pas de dépendance caudale inférieure, c'est-à-dire que  $\lambda_L(C_{\rho,\nu}^F) = 0$ . Par contre, ces auteurs montrent que le coefficient de dépendance caudale supérieur est positif. Pour le montrer, soit un vecteur aléatoire  $(X_1, X_2)$  distribué selon la loi de Student à  $\nu \in \mathbb{N}$  degrés de liberté et de paramètre  $\rho \in [-1, 1]$ . En posant  $F_1(u) = F_2(u) = 2F_\nu(\sqrt{u}) - 1$ , on a

$$\begin{aligned} \lambda_U(C_{\rho,\nu}^F) &= \lim_{u \uparrow 1} \mathbb{P} \{ X_1^2 > F_1^{-1}(u) - 1 \mid X_2^2 > F_1^{-1}(u) - 1 \} \\ &= \lim_{u \uparrow 1} \mathbb{P} \left\{ X_1^2 > \left( F_\nu^{-1} \left( \frac{1+u}{2} \right) \right)^2 \mid X_2^2 > \left( F_\nu^{-1} \left( \frac{1+u}{2} \right) \right)^2 \right\} \\ &= \lim_{s \uparrow 1} \mathbb{P} \left\{ X_1^2 > (F_\nu^{-1}(s))^2 \mid X_2^2 > (F_\nu^{-1}(s))^2 \right\} \\ &= \lim_{s \uparrow 1} \mathbb{P} \{ |X_1| > F_\nu^{-1}(s) \mid |X_2| > F_\nu^{-1}(s) \} \\ &= \lim_{s \uparrow 1} \mathbb{P} \{ X_1 > F_\nu^{-1}(s) \mid |X_2| > F_\nu^{-1}(s) \} \\ &\quad + \mathbb{P} \{ -X_1 > F_\nu^{-1}(s) \mid |X_2| > F_\nu^{-1}(s) \}. \end{aligned} \quad (4.3)$$

Pour le premier terme dans la dernière expression, on a

$$\begin{aligned}
 A &= \mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid |X_2| > F_\nu^{-1}(s)\} \\
 &= \frac{\mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid |X_2| > F_\nu^{-1}(s)\}}{\mathbb{P}\{|X_2| > F_\nu^{-1}(s)\}} \\
 &= \frac{\mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid X_2 > F_\nu^{-1}(s)\} + \mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid -X_2 > F_\nu^{-1}(s)\}}{\mathbb{P}\{|X_2| > F_\nu^{-1}(s)\}} \\
 &= \frac{\mathbb{P}\{X_2 > F_\nu^{-1}(s)\}}{\mathbb{P}\{|X_2| > F_\nu^{-1}(s)\}} \mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid X_2 > F_\nu^{-1}(s)\} \\
 &\quad + \frac{\mathbb{P}\{-X_2 > F_\nu^{-1}(s)\}}{\mathbb{P}\{|X_2| > F_\nu^{-1}(s)\}} \mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid -X_2 > F_\nu^{-1}(s)\} \\
 &= \frac{1}{2} [\mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid X_2 > F_\nu^{-1}(s)\} \\
 &\quad + \mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid -X_2 > F_\nu^{-1}(s)\}].
 \end{aligned}$$

où la dernière égalité découle du fait que pour tout  $a > 0$ ,

$$\frac{\mathbb{P}\{X_2 > a\}}{\mathbb{P}\{|X_2| > a\}} = \frac{1 - F_\nu(a)}{2\{1 - F_\nu(a)\}} = \frac{1}{2}.$$

Pour le deuxième terme, on peut écrire

$$\begin{aligned}
 B &= \mathbb{P}\{-X_1 > F_\nu^{-1}(s) \mid |X_2| > F_\nu^{-1}(s)\} \\
 &= \frac{1}{2} [\mathbb{P}\{-X_1 > F_\nu^{-1}(s) \mid X_2 > F_\nu^{-1}(s)\} \\
 &\quad + \mathbb{P}\{-X_1 > F_\nu^{-1}(s) \mid -X_2 > F_\nu^{-1}(s)\}].
 \end{aligned}$$

Ces deux expressions dans l'Équation (4.3) donnent

$$\begin{aligned}
 \lambda_U(C_{\rho,\nu}^F) &= A + B & (4.4) \\
 &= \frac{1}{2} \lim_{s \uparrow 1} [\mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid X_2 > F_\nu^{-1}(s)\} \\
 &\quad + \mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid -X_2 > F_\nu^{-1}(s)\}] \\
 &\quad + \frac{1}{2} \lim_{s \uparrow 1} [\mathbb{P}\{-X_1 > F_\nu^{-1}(s) \mid X_2 > F_\nu^{-1}(s)\} \\
 &\quad + \mathbb{P}\{-X_1 > F_\nu^{-1}(s) \mid -X_2 > F_\nu^{-1}(s)\}] \\
 &= \lim_{s \uparrow 1} [\mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid X_2 > F_\nu^{-1}(s)\} \\
 &\quad + \mathbb{P}\{X_1 > F_\nu^{-1}(s) \mid -X_2 > F_\nu^{-1}(s)\}] \\
 &= \lambda_U(C_{\rho,\nu}^T) + \lambda_U(C_{-\rho,\nu}^T) & (4.5)
 \end{aligned}$$

Or, tel qu'on l'a vu, pour  $F_{\nu+1}$  qui est la fonction de répartition de la loi de Student univariée à  $\nu + 1$  degrés de liberté, on a

$$\lambda_U(C_{\rho,\nu}^T) = 2F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right). \quad (4.6)$$

On a donc l'expression

$$\lambda_U(C_{\rho,\nu}^F) = 2 \left\{ F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right) + F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1+\rho)}{1-\rho}} \right) \right\}. \quad (4.7)$$

La copule Fisher possède ainsi de la dépendance caudale supérieure, car  $\lambda_U(C_{\rho,\nu}^F) > 0$ .

Une exception survient à la limite lorsque  $\nu \rightarrow \infty$ , car alors

$$\lim_{\nu \rightarrow \infty} \lambda_U(C_{\rho,\nu}^F) = 0.$$

Cette constatation coïncide avec le fait que  $\nu \rightarrow \infty$  correspond à la copule Khi-deux.

## 4.2 Description des nouveaux estimateurs de l'indice de dépendance caudale

### 4.2.1 Idée générale

Soient  $(X_1, Y_1), \dots, (X_n, Y_n)$ , des paires aléatoires indépendantes provenant d'une loi conjointe  $H$  dont les marges sont continues. L'objectif de cette section est de proposer des estimateurs du coefficient de dépendance caudale  $\lambda_U$  qui vont concurrencer ceux décrits aux Chapitre 3. Pour ce faire, on va considérer que la vraie copule de  $H$  peut raisonnablement être considérée comme faisant partie de la famille des copules de Fisher. Dans ce cas, on sait que

$$\lambda_U(C_{\rho, \nu}^F) = 2 \left\{ F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right) + F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1+\rho)}{1-\rho}} \right) \right\}.$$

Partant de là, un estimateur semi-paramétrique de  $\lambda_U$  s'obtient en remplaçant les paramètres inconnus de  $\lambda_U(C_{\rho, \nu}^F)$  par des estimateurs basés sur les observations  $(X_1, Y_1), \dots, (X_n, Y_n)$ . Dans la suite, deux situations seront considérées, à savoir les cas où (i) le degré de liberté  $\nu$  est fixé, et (ii) le degré de liberté  $\nu$  est inconnu.

### 4.2.2 Estimation de $\rho$ par inversion du tau de Kendall

Soit une famille de copules paramétriques  $\{C_\theta; \theta \in \Theta \subseteq \mathbb{R}\}$ . Dans ce cas, le tau de Kendall est généralement une fonction  $\kappa : \Theta \rightarrow [-1, 1]$  de la forme

$$\kappa(\theta) = 4 \int_{[0,1]^2} C_\theta(u, v) dC_\theta(u, v) - 1.$$

La méthode d'inversion du tau de Kendall est similaire à une méthode des moments qui consiste à estimer  $\theta$  par le nombre  $\theta_n^{\text{ITK}}$  tel que

$$\kappa(\theta_n^{\text{ITK}}) = \tau_n,$$

où  $\tau_n$  est le tau de Kendall empirique. Spécifiquement, pour un échantillon de paires  $(X_1, Y_1), \dots, (X_n, Y_n)$ ,

$$\tau_n = -1 + \frac{4}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{I}\{(X_i - X_j)(Y_i - Y_j) > 0\}.$$

Dans le cas de la copule Fisher, le tau de Kendall s'exprime via la fonction

$$\kappa_\nu(\rho) = 4 \int_{[0,1]^2} C_{\rho,\nu}^{\text{F}}(u, v) dC_{\rho,\nu}^{\text{F}}(u, v) - 1.$$

Cependant, la fonction  $\kappa_\nu$  n'a pas de formulation explicite, ce qui empêche de déduire une expression pour l'estimateur de  $\rho$  en faisant  $\rho_n^{\text{ITK}} = \kappa_\nu^{-1}(\tau_n)$ . Pour cette raison, l'estimateur  $\rho_n^{\text{ITK}}$  s'obtient par des méthodes numériques.

### 4.2.3 Estimation de $(\rho, \nu)$ par maximum de vraisemblance

Soit un échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  i.i.d. de loi conjointe  $H$  et de marges continues  $F$  et  $G$ . Dans ce cas, on sait que la densité de  $H$  peut s'écrire sous la forme

$$h(x, y) = c\{F(x), G(y)\} f(x) g(y).$$

Supposons que la copule  $C$  appartient à une certaine famille paramétrique  $\{C_\theta; \theta \in \Theta\}$ . De même, assumons que  $F \in \{F_\alpha; \alpha \in A\}$  et  $G \in \{G_\beta; \beta \in B\}$ . Afin d'estimer le vecteur des paramètres  $(\theta, \alpha, \beta)$ , la façon classique consiste à maximiser la fonction

de log-vraisemblance, c'est-à-dire

$$\begin{aligned} L(\theta, \alpha, \beta) &= \sum_{i=1}^n \ln [c_{\theta} \{F_{\alpha}(X_i), G_{\beta}(Y_i)\} f_{\alpha}(X_i) g_{\beta}(Y_i)] \\ &= \sum_{i=1}^n \ln c_{\theta} \{F_{\alpha}(X_i), G_{\beta}(Y_i)\} + \sum_{i=1}^n \ln f_{\alpha}(X_i) + \sum_{i=1}^n \ln g_{\beta}(Y_i). \end{aligned}$$

Ainsi, l'estimateur à maximum de vraisemblance de  $(\theta, \alpha, \beta)$  est le vecteur

$$(\theta_n^{\text{MV}}, \alpha^{\text{MV}}, \beta^{\text{MV}}) = \arg \max_{(\theta, \alpha, \beta) \in \Theta \times A \times B} L(\theta, \alpha, \beta).$$

L'inconvénient avec cette méthode est que le temps de calcul peut s'avérer très long, surtout lorsque la dimension des paramètres est élevée. Cela est dû en bonne partie au fait qu'il faut conjointement estimer les paramètres des lois marginales et les paramètres de la structure de dépendance.

L'estimateur de pseudo-vraisemblance élaboré par Genest et al. [25] est un moyen efficace pour l'estimation du paramètre d'une copule; il s'agit d'une alternative à l'inversion du tau de Kendall qui ne souffre pas de la contrainte que  $\theta$  soit unidimensionnel. Cette méthode d'estimation semi-paramétrique est basée sur les rangs des observations. Spécifiquement, on estime d'abord les marges univariées à l'aide des fonctions de répartition empiriques, à savoir

$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(X_i \leq x) \quad \text{et} \quad G_n(y) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(Y_i \leq y).$$

Ensuite, on définit la fonction de pseudo-vraisemblance, à savoir

$$\tilde{L}(\theta) = \sum_{i=1}^n \ln c_{\theta} \{F_n(X_i), G_n(Y_i)\},$$

de sorte que l'estimateur à pseudo-maximum de vraisemblance de  $\theta$  est

$$\theta_n^{\text{PMV}} = \arg \max_{\theta \in \Theta} \tilde{L}(\theta).$$

Pour adapter cette méthodologie à la copule Fisher  $C_{\rho,\nu}^F$ , soit d'abord  $A_{20} = \{1, \dots, 20\}$ , l'ensemble des vingt premiers entiers. Alors l'estimateur à pseudo-maximum de vraisemblance de  $(\rho, \nu)$  est donné par

$$(\rho_n^{\text{PMV}}, \nu_n^{\text{PMV}}) = \arg \max_{\rho \in (-1,1), \nu \in A_{20}} \sum_{i=1}^n \ln c_{\rho,\nu}^F \{F_n(X_i), G_n(Y_i)\}.$$

#### 4.2.4 Estimation de $(\rho, \nu)$ par une méthode en deux étapes

Afin de rendre le calcul de l'estimateur  $(\rho_n^{\text{PMV}}, \nu_n^{\text{PMV}})$  plus efficace, on suggère de procéder en deux étapes. La première consiste à fixer  $\nu \in \mathbb{N}$  et à déduire l'estimateur de  $\rho$  par inversion du tau de Kendall conditionnellement à  $\nu$  : on note alors cette valeur  $\rho_n(\nu) = \kappa_\nu^{-1}(\tau_n)$ . On obtient ensuite l'estimateur de  $\nu$  en maximisant

$$\sum_{i=1}^n \ln c_{\rho_n(\nu), \nu}^F \{F_n(X_i), G_n(Y_i)\}$$

pour  $\nu \in A_{20}$ , c'est-à-dire que

$$\nu_n^{\text{PMV2}} = \arg \max_{\nu \in A_{20}} \sum_{i=1}^n \ln c_{\rho_n(\nu), \nu}^F \{F_n(X_i), G_n(Y_i)\}.$$

Enfin, l'estimateur de  $\rho$  correspond à la valeur optimale de  $\nu$ , à savoir que

$$\rho_n^{\text{PMV2}} = \rho_n(\nu_n^{\text{PMV2}}).$$

#### 4.2.5 Estimation de $\lambda_U$ : degré de liberté fixé

Les estimateurs proposés pour  $\lambda_U$  sont très simples. Il s'agira, en effet, de supposer que la dépendance des paires observées peut raisonnablement se modéliser par une copule Fisher. Sous cette hypothèse, on sait que l'indice de dépendance caudale supérieure

prend la forme explicite

$$\lambda_U = 2 \left\{ F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}} \right) + F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1+\rho)}{1-\rho}} \right) \right\}.$$

Si on suppose que le degré de liberté est fixe, alors  $\rho$  s'estime par inversion du tau de Kendall, c'est-à-dire que  $\rho_n^{\text{ITK}} = \kappa_\nu^{-1}(\tau_n)$ . L'estimateur proposé pour  $\lambda_U$  est alors

$$\widehat{\lambda}_U^{\text{ITK}}(\nu) = 2 \left\{ F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1-\rho_n^{\text{ITK}})}{1+\rho_n^{\text{ITK}}}} \right) + F_{\nu+1} \left( -\sqrt{\frac{(\nu+1)(1+\rho_n^{\text{ITK}})}{1-\rho_n^{\text{ITK}}}} \right) \right\}.$$

#### 4.2.6 Estimation de $\lambda_U$ : degré de liberté inconnu

Deux estimateurs de  $(\rho, \nu)$  ont été décrits lorsque le degré de liberté de la copule Fisher est inconnu. Il s'agit de l'estimateur à pseudo-maximum de vraisemblance  $(\rho_n^{\text{PMV}}, \nu_n^{\text{PMV}})$  et de sa version en deux étapes, à savoir  $(\rho_n^{\text{PMV2}}, \nu_n^{\text{PMV2}})$ . Ces deux méthodes d'estimation induisent donc deux estimateurs de  $\lambda_U$ , c'est-à-dire

$$\begin{aligned} \widehat{\lambda}_U^{\text{PMV}} &= 2F_{\nu_n^{\text{PMV}}+1} \left( -\sqrt{\frac{(\nu_n^{\text{PMV}}+1)(1-\rho_n^{\text{PMV}})}{1+\rho_n^{\text{PMV}}}} \right) \\ &\quad + 2F_{\nu_n^{\text{PMV}}+1} \left( -\sqrt{\frac{(\nu_n^{\text{PMV}}+1)(1+\rho_n^{\text{PMV}})}{1-\rho_n^{\text{PMV}}}} \right) \end{aligned}$$

et

$$\begin{aligned} \widehat{\lambda}_U^{\text{PMV2}} &= 2F_{\nu_n^{\text{PMV2}}+1} \left( -\sqrt{\frac{(\nu_n^{\text{PMV2}}+1)(1-\rho_n^{\text{PMV2}})}{1+\rho_n^{\text{PMV2}}}} \right) \\ &\quad + 2F_{\nu_n^{\text{PMV2}}+1} \left( -\sqrt{\frac{(\nu_n^{\text{PMV2}}+1)(1+\rho_n^{\text{PMV2}})}{1-\rho_n^{\text{PMV2}}}} \right). \end{aligned}$$

## 4.3 Étude de la performance des estimateurs de $\lambda_U$

### 4.3.1 Description générale de l'étude de simulations

Cette section étudie l'efficacité de quelques estimateurs de  $\lambda_U$  proposés dans ce travail, et la compare aux estimateurs existants. On considérera ainsi :

- (i) Les estimateurs  $\widehat{\lambda}_U^{\text{CFG}}$ ,  $\widehat{\lambda}_U^{\text{SS}}$  et  $\widehat{\lambda}_U^{\text{Coles}}$ , qui sont abondamment utilisés en pratique, ainsi que dans la littérature scientifique. On choisit dans la suite pour le calcul de ces estimateurs, une valeur seuil  $k = n/10$  avec  $n$  la taille de l'échantillon ;
- (ii) Les nouveaux estimateurs  $\widehat{\lambda}_U^{\text{ITK}}(1)$ ,  $\widehat{\lambda}_U^{\text{ITK}}(5)$  et  $\widehat{\lambda}_U^{\text{ITK}}(10)$ , qui imposent de fixer la valeur  $\nu$  du degré de liberté ;
- (iii) Le nouvel estimateur  $\widehat{\lambda}_U^{\text{PMV}^2}$ , qui ne nécessite pas de fixer la valeur de  $\nu$ .

L'efficacité de ces sept estimateurs sera évaluée selon leur erreur quadratique moyenne (EQM) et leur Biais (B). On rappelle que l'erreur quadratique moyenne et le biais d'un estimateur  $\widehat{\lambda}_U$  de  $\lambda_U$  sont définis par

$$\text{EQM}(\widehat{\lambda}_U) = \text{E} \left\{ \left( \widehat{\lambda}_U - \lambda_U \right)^2 \right\} \quad \text{et} \quad \text{B}(\widehat{\lambda}_U) = \text{E} \left( \widehat{\lambda}_U - \lambda_U \right).$$

Cependant, il est difficile, voire impossible, de calculer  $\text{EQM}(\widehat{\lambda}_U)$  et  $\text{B}(\widehat{\lambda}_U)$  explicitement ; on procèdera donc par simulations. Ainsi, sous un modèle de copules donné, disons  $C$ , et pour une certaine taille d'échantillons, on générera  $M$  échantillons indépendants  $(X_1^{(1)}, Y_1^{(1)}), \dots, (X_n^{(1)}, Y_n^{(1)}), \dots, (X_1^{(M)}, Y_1^{(M)}), \dots, (X_n^{(M)}, Y_n^{(M)})$ , chacun composé d'observations i.i.d.  $C$  ; on obtient alors les estimateurs  $\widehat{\lambda}_U^{(1)}, \dots, \widehat{\lambda}_U^{(M)}$ . On estime

alors  $\text{EQM}(\widehat{\lambda}_U)$  et  $\text{B}(\widehat{\lambda}_U)$  avec, respectivement,

$$\widehat{\text{EQM}}(\widehat{\lambda}_U) = \frac{1}{M} \sum_{m=1}^M \left( \widehat{\lambda}_U^{(1)} - \bar{\lambda}_U \right)^2 \quad \text{et} \quad \widehat{\text{B}}(\widehat{\lambda}_U) = \frac{1}{M} \sum_{m=1}^M \left( \widehat{\lambda}_U^{(1)} - \bar{\lambda}_U \right),$$

où  $\bar{\lambda}_U = (\widehat{\lambda}_U^{(1)} + \dots + \widehat{\lambda}_U^{(M)})/M$ . On considérera le comportement des estimateurs de  $\lambda_U$  sous trois types de structures de dépendance, à savoir Fisher, Gumbel et Khi-deux. Afin d'alléger le texte, tous les tableaux des résultats ont été déplacés à la sous-section 4.3.6 située à la fin de cette section.

### 4.3.2 Données générées de la copule Fisher

Les Tableaux 4.1 et 4.2 concernent respectivement l'erreur quadratique moyenne et le biais qui sont estimés à partir de  $n = 75$  et  $n = 150$  échantillons de la copule Fisher. Ces derniers ont été paramétrés en termes du tau de Kendall pour des valeurs de  $\tau \in \{1/4, 1/2, 3/4\}$  correspondant à des niveaux de dépendance faibles, modérés et élevés. Comme on pouvait s'y attendre, les valeurs de l'erreur quadratique moyenne et du biais tendent à diminuer à mesure que la taille de l'échantillon augmente.

Quand le degré de liberté est faible, les erreurs quadratiques moyennes et le biais diminuent significativement à mesure que la valeur du tau de Kendall augmente. Dans le cas où le degré de liberté est élevé, toutefois, le tau de Kendall n'a pas d'impact notable sur la performance des estimateurs. Pour les estimateurs  $\widehat{\lambda}_U^{\text{ITK}(1)}$  et  $\widehat{\lambda}_U^{\text{ITK}(5)}$ , on constate que leur EQM et leur biais diminuent significativement lorsque la valeur du degré de liberté augmente, contrairement à  $\widehat{\lambda}_U^{\text{ITK}(10)}$ . Globalement, ces estimateurs sont plus précis que les estimateurs non-paramétriques  $\widehat{\lambda}_U^{\text{CFG}}$ ,  $\widehat{\lambda}_U^{\text{SS}}$  et  $\widehat{\lambda}_U^{\text{Coles}}$ .

Parmi les autres observations que l'on peut tirer des Tableaux 4.1 et 4.2, on note que

(1) Dans le cas où le tau de Kendall est faible ( $\tau = 1/4$ ), de même que le degré de liberté ( $\nu = 1$ ), les estimateurs  $\hat{\lambda}_U^{\text{CFG}}$ ,  $\hat{\lambda}^{\text{ITK}}(5)$  et  $\hat{\lambda}^{\text{ITK}}(10)$  et  $\hat{\lambda}^{\text{PMV2}}(10)$  sont les plus précis ; ceci est particulièrement notable pour  $\hat{\lambda}^{\text{ITK}}(10)$  et  $\hat{\lambda}^{\text{PMV2}}(10)$ .

(2) Quand  $\tau$  est élevé ( $\tau = 3/4$ ) et que le degré de liberté est faible ( $\nu = 1$ ), les estimateurs  $\hat{\lambda}_U^{\text{CFG}}$ ,  $\hat{\lambda}^{\text{ITK}}(1)$ ,  $\hat{\lambda}^{\text{ITK}}(5)$ ,  $\hat{\lambda}^{\text{ITK}}(10)$  et  $\hat{\lambda}^{\text{PMV2}}(10)$  sont les meilleurs ; ceci est particulièrement vrai pour  $\hat{\lambda}_U^{\text{CFG}}$  et  $\hat{\lambda}^{\text{ITK}}(10)$ .

(3) Dans le cas où le tau de Kendall est faible ( $\tau = 1/4$ ) et le degré de liberté est élevé ( $\nu = 10$ ), les estimateurs  $\hat{\lambda}^{\text{ITK}}(5)$  et  $\hat{\lambda}^{\text{ITK}}(10)$  et  $\hat{\lambda}^{\text{PMV2}}(10)$  sont les plus précis, mais surtout  $\hat{\lambda}^{\text{ITK}}(5)$  et  $\hat{\lambda}^{\text{ITK}}(10)$ .

(4) Quand  $\tau$  est élevé ( $\tau = 3/4$ ) et que le degré de liberté est élevé ( $\nu = 10$ ), la meilleure précision est obtenue par l'estimateur  $\hat{\lambda}^{\text{PMV2}}(10)$ .

### 4.3.3 Données générées de la copule de Gumbel

Les Tableaux 4.3 et 4.4 montrent respectivement les valeurs de l'erreur quadratique moyenne (EQM) et du biais (B) des mêmes sept estimateurs de la sous-section 4.3.2, mais cette fois dans le cas où les observations proviennent de la copule de Gumbel. On remarque d'emblée que, tel qu'attendu, les EQM et les biais tendent à diminuer à mesure que la taille de l'échantillon augmente.

Dans le cas où la valeur du tau de Kendall est élevée, les estimateurs  $\hat{\lambda}_U^{\text{ITK}}(10)$ ,  $\hat{\lambda}_U^{\text{ITK}}(5)$ ,  $\hat{\lambda}_U^{\text{ITK}}(10)$  et  $\hat{\lambda}_U^{\text{CFG}}$  sont meilleurs que les estimateurs  $\hat{\lambda}_U^{\text{SS}}$  et  $\hat{\lambda}_U^{\text{Coles}}$ . Les estimateurs  $\hat{\lambda}_U^{\text{CFG}}$  et  $\hat{\lambda}_U^{\text{ITK}}(10)$  sont les plus performants lorsque la la valeur de  $\tau$  est faible ( $\tau = 1/4$ ).

#### 4.3.4 Données générées de la copule Khi-deux

Les Tableaux 4.5 et 4.6 montrent respectivement les valeurs de l'erreur quadratique moyenne (EQM) et du biais (B) des sept estimateurs sous des structures de dépendance de type Khi-deux. À l'instar des études effectuées aux sous-section 4.3.2 et 4.3.3, les erreurs quadratiques moyennes et les biais diminuent lorsque la taille de l'échantillon augmente. Pour des valeurs faibles du tau de Kendall, *i.e.* quand  $\tau = 1/4$ , les estimateurs  $\hat{\lambda}_U^{\text{ITK}}(1)$ ,  $\hat{\lambda}_U^{\text{ITK}}(5)$  et  $\hat{\lambda}_U^{\text{PMV}^2}$  donnent les meilleures précisions. Toutefois, si la valeur du tau de Kendall est élevée ( $\tau = 3/4$ ), aucun des estimateurs n'est précis.

#### 4.3.5 Commentaires globaux sur les simulations et quelques recommandations

Tel qu'attendu, les valeurs de l'erreur quadratique moyenne (EQM) et du biais (B) des sept estimateurs sous des structures de dépendance de type Fisher, Gumbel et Khi-deux diminuent lorsque la taille de l'échantillon augmente. Globalement, on peut émettre les commentaires suivants :

(i) Sous de la dépendance de type Fisher, les nouveaux estimateurs sont précis quel que soit le niveau de dépendance et le degré de liberté ; cette constatation n'est pas surprenante, dans la mesure où les estimateurs proposés sont développés sous l'hypothèse d'une structure de dépendance Fisher ;

(ii) Sous de la dépendance extrême de type Gumbel, qui comporte un degré élevé de dépendance caudale supérieure, la performance des estimateurs dépend du niveau de dépendance. Ainsi, lorsque la valeur du tau de Kendall est élevée, les nouveaux estimateurs performant bien.

(iii) Sous de la dépendance de type Khi-deux, qui ne comporte pas de dépendance caudale supérieure, les estimateurs sont généralement assez biaisés, surtout quand le niveau de dépendance est élevé. Ce comportement en l'absence de dépendance caudale des estimateurs non-paramétriques a été bien documenté par Serinaldi et al. [49].

Si une recommandation générale devait être émise, on pourrait suggérer l'usage de l'estimateur  $\widehat{\lambda}_U^{\text{PMV}^2}$  lorsqu'on peut assurer d'une dépendance caudale significative. Cet estimateur sera d'autant plus efficace que le niveau de dépendance est élevé ; ce dernier aspect peut s'apprécier facilement en calculant la valeur du tau de Kendall empirique. Par ailleurs,  $\widehat{\lambda}_U^{\text{PMV}^2}$  a l'avantage de ne pas faire d'hypothèse sur la valeur du degré de liberté, contrairement à l'estimateur  $\widehat{\lambda}_U^{\text{TK}}(\nu)$ , qui nécessite de choisir quelque peu arbitrairement la valeur de  $\nu$ .

### 4.3.6 Tableaux des résultats de simulations

TABLE 4.1 – Erreur quadratique moyenne de sept estimateurs de  $\lambda_U$ , estimée à l'aide de 1 000 réplicats, sous des données simulées de la copule Fisher

$n$	$\nu$	$\tau$	$\hat{\lambda}_U^{\text{CFG}}$	$\hat{\lambda}_U^{\text{SS}}$	$\hat{\lambda}_U^{\text{Coles}}$	$\hat{\lambda}_U^{\text{ITK}(1)}$	$\hat{\lambda}_U^{\text{ITK}(5)}$	$\hat{\lambda}_U^{\text{ITK}(10)}$	$\hat{\lambda}_U^{\text{PMV2}}$
75	1	0.25	5.0	12.3	11.1	5.2	4.7	2.5	4.1
		0.50	1.3	5.3	5.2	2.8	2.2	0.8	2.5
		0.75	0.3	2.1	2.3	0.7	0.5	0.2	0.6
	5	0.25	6.5	10.3	8.6	2.6	2.3	3.8	2.5
		0.50	7.4	11.8	10.9	7.4	5.5	5.3	4.6
		0.75	3.0	5.3	5.1	5.0	3.5	2.7	2.5
	10	0.25	11.6	15.2	12.5	1.8	1.3	7.5	3.3
		0.50	17.5	21.3	19.4	16.7	12.2	14.4	8.6
		0.75	8.3	10.7	10.3	11.6	8.9	7.8	5.6
150	1	0.25	4.0	10.5	8.9	4.5	4.3	1.8	3.6
		0.50	1.0	4.2	3.8	1.5	1.0	0.4	1.5
		0.75	0.2	1.3	1.3	0.7	0.5	0.1	0.5
	5	0.25	5.3	9.4	7.3	2.1	2.0	3.1	2.3
		0.50	6.9	10.4	9.1	6.8	4.6	5.2	4.6
		0.75	2.9	4.6	4.3	4.9	3.4	2.6	2.2
	10	0.25	9.9	14.8	11.6	0.8	0.4	6.8	3.3
		0.50	16.4	20.5	18.4	16.2	11.1	14.0	8.5
		0.75	8.2	9.7	9.1	11.6	8.9	7.8	5.1

TABLE 4.2 – Biais de sept estimateurs de  $\lambda_U$ , estimé à l'aide de 1 000 répliqués, sous des données simulées de la copule Fisher

$n$	$\nu$	$\tau$	$\widehat{\lambda}_U^{\text{CFG}}$	$\widehat{\lambda}_U^{\text{SS}}$	$\widehat{\lambda}_U^{\text{Coles}}$	$\widehat{\lambda}_U^{\text{ITK}(1)}$	$\widehat{\lambda}_U^{\text{ITK}(5)}$	$\widehat{\lambda}_U^{\text{ITK}(10)}$	$\widehat{\lambda}_U^{\text{PMV2}}$
75	1	0.25	2.1	3.0	2.6	-0.7	-0.5	1.2	-0.5
		0.50	0.9	1.5	1.3	0.4	0.4	0.3	0.1
		0.75	0.4	0.7	0.5	0.8	0.6	0.1	0.5
	5	0.25	2.4	2.6	2.0	-0.8	-0.9	1.5	0.2
		0.50	2.6	3.0	2.7	2.3	1.8	2.2	1.4
		0.75	1.7	1.9	1.7	2.2	1.8	1.6	1.3
	10	0.25	3.3	3.4	2.8	0.1	-0.1	2.5	1.0
		0.50	4.1	4.3	3.9	3.8	3.2	3.7	2.5
		0.75	2.9	3.0	2.9	3.4	3.0	2.8	2.1
150	1	0.25	1.9	3.0	2.6	-1.2	-1.0	1.1	-0.5
		0.50	0.8	1.7	1.4	0.5	0.5	0.3	0.4
		0.75	0.4	0.8	0.7	0.8	0.7	0.1	0.7
	5	0.25	2.2	2.7	2.2	-1.1	-1.3	1.5	0.4
		0.50	2.6	3.0	2.7	2.4	1.9	2.2	1.7
		0.75	1.7	1.9	1.8	2.2	1.8	1.6	1.2
	10	0.25	3.1	3.6	3.1	-0.2	-0.4	2.5	1.2
		0.50	4.0	4.4	4.1	3.9	3.2	3.7	2.6
		0.75	2.9	3.0	2.9	3.4	3.0	2.8	2.1

TABLE 4.3 – Erreur quadratique moyenne de sept estimateurs de  $\lambda_U$ , estimée à l'aide de 1 000 réplicats, sous des données simulées de la copule de Gumbel

$n$	$\tau$	$\hat{\lambda}_U^{\text{CFG}}$	$\hat{\lambda}_U^{\text{SS}}$	$\hat{\lambda}_U^{\text{Coles}}$	$\hat{\lambda}_U^{\text{ITK}(1)}$	$\hat{\lambda}_U^{\text{ITK}(5)}$	$\hat{\lambda}_U^{\text{ITK}(10)}$	$\hat{\lambda}_U^{\text{PMV2}}$
75	0.25	0.8	4.1	7.6	8.7	9.7	1.3	4.1
	0.50	0.5	8.1	12.1	2.0	7.8	0.5	4.9
	0.75	0.3	8.6	11.1	0.4	0.3	0.1	1.1
150	0.25	0.5	2.5	5.3	9.4	10.0	0.7	3.9
	0.50	0.4	6.4	9.8	0.6	7.2	0.3	4.0
	0.75	0.3	6.4	8.4	0.3	0.2	0.1	0.4

TABLE 4.4 – Biais de sept estimateurs de  $\lambda_U$ , estimé à l'aide de 1 000 réplicats, sous des données simulées de la copule de Gumbel

$n$	$\tau$	$\hat{\lambda}_U^{\text{CFG}}$	$\hat{\lambda}_U^{\text{SS}}$	$\hat{\lambda}_U^{\text{Coles}}$	$\hat{\lambda}_U^{\text{ITK}(1)}$	$\hat{\lambda}_U^{\text{ITK}(5)}$	$\hat{\lambda}_U^{\text{ITK}(10)}$	$\hat{\lambda}_U^{\text{PMV2}}$
75	0.25	0.1	-1.1	-2.0	-2.7	-3.1	-0.3	-14.7
	0.50	-0.3	-2.2	-2.8	-0.1	-2.0	-0.2	-15.1
	0.75	-0.4	-2.3	-2.6	0.5	-0.3	-0.1	-4.7
150	0.25	-0.2	-1.0	-1.9	-3.0	-3.2	-0.2	-15.9
	0.50	-0.5	-2.2	-2.8	0.1	-2.2	-0.2	-14.6
	0.75	-0.5	-2.2	-2.5	0.5	-0.3	-0.1	-1.7

TABLE 4.5 – Erreur quadratique moyenne de sept estimateurs de  $\lambda_U$ , estimée à l'aide de 1 000 réplicats, sous des données simulées de la copule Khi-deux

$n$	$\tau$	$\widehat{\lambda}_U^{\text{CFG}}$	$\widehat{\lambda}_U^{\text{SS}}$	$\widehat{\lambda}_U^{\text{Coles}}$	$\widehat{\lambda}_U^{\text{ITK}(1)}$	$\widehat{\lambda}_U^{\text{ITK}(5)}$	$\widehat{\lambda}_U^{\text{ITK}(10)}$	$\widehat{\lambda}_U^{\text{PMV}^2}$
75	0.25	14.2	18.5	15.3	2.0	1.1	9.5	3.7
	0.50	36.7	41.3	38.2	35.3	26.3	32.7	18.3
	0.75	65.4	67.0	65.1	74.7	66.4	64.5	51.3
150	0.25	12.4	16.9	13.2	0.6	0.2	9.2	3.1
	0.50	35.5	39.8	36.5	35.7	25.4	32.6	17.2
	0.75	64.8	67.7	65.9	74.6	66.2	64.3	48.1

TABLE 4.6 – Biais de sept estimateurs de  $\lambda_U$ , estimé à l'aide de 1 000 réplicats, sous des données simulées de la copule Khi-deux

$n$	$\tau$	$\widehat{\lambda}_U^{\text{CFG}}$	$\widehat{\lambda}_U^{\text{SS}}$	$\widehat{\lambda}_U^{\text{Coles}}$	$\widehat{\lambda}_U^{\text{ITK}(1)}$	$\widehat{\lambda}_U^{\text{ITK}(5)}$	$\widehat{\lambda}_U^{\text{ITK}(10)}$	$\widehat{\lambda}_U^{\text{PMV}^2}$
75	0.25	3.7	3.9	3.3	0.6	0.3	2.9	13.0
	0.50	6.0	6.2	5.8	5.8	4.9	5.7	40.0
	0.75	8.1	8.1	7.9	8.6	8.1	8.0	70.9
150	0.25	3.5	3.9	3.3	0.2	0.1	2.9	13.1
	0.50	5.9	6.2	5.9	5.9	4.9	5.7	39.4
	0.75	8.0	8.2	8.1	8.6	8.1	8.0	68.9

## 4.4 Illustration sur de vraies données

### 4.4.1 Présentation des données

Les méthodes d'adéquation présentées dans ce travail seront appliquées à des données provenant de l'*Australian Institute of Sport*. Il s'agit de treize mesures biométriques prises sur  $n = 202$  athlètes australiens (100 femmes, 102 hommes). Ces données ont été analysées dans l'article de Telford and Cunningham [53] publié dans la revue *Medicine and science in sports and exercise*. Dans le contexte de nouveaux tests d'adéquation pour les copules, elles ont également été considérées par Bahraoui et al. [3].

Pour l'analyse subséquente, on se concentrera sur cinq des treize variables, à savoir

- $X_1$ , le nombre de globules rouges dans le sang ;
- $X_2$ , le taux d'hématocrite, qui constitue le pourcentage du volume sanguin total occupé par les globules rouges ;
- $X_3$ , la concentration d'hémoglobine, mesurée en grammes par décilitre de sang ; l'hémoglobine assure le transport de l'oxygène ;
- $X_4$ , la masse maigre, mesurée en kilogrammes ;
- $X_5$ , la taille de l'athlète en centimètres.

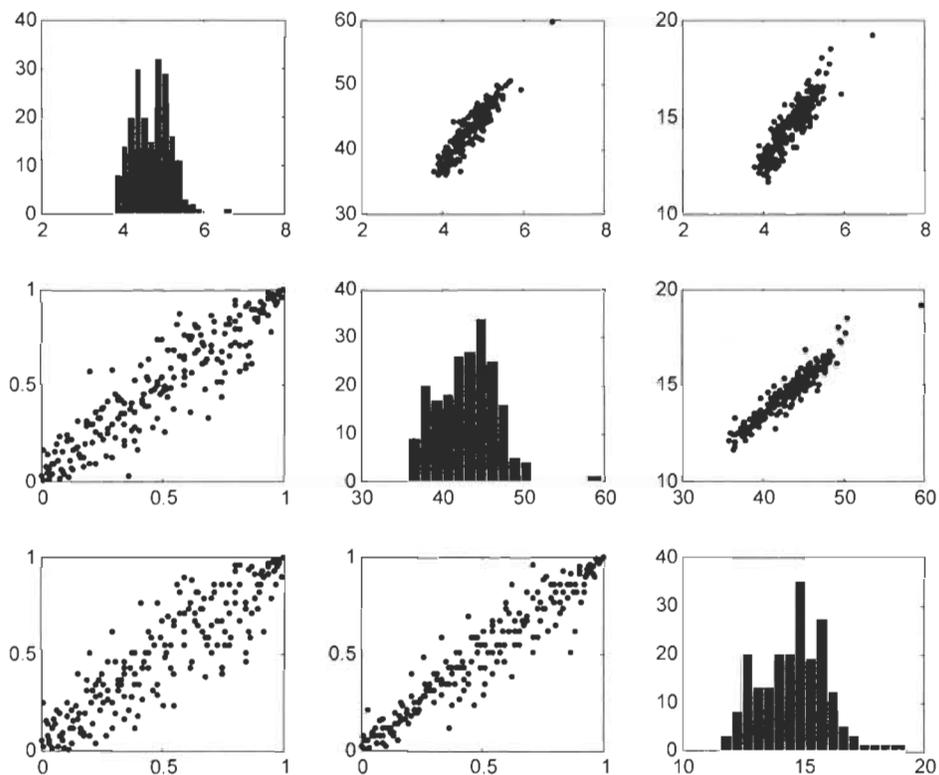


FIGURE 4.1 – Histogrammes (sur la diagonale), nuages de points des données brutes (triangle supérieur) et nuages de points des pseudo-observations (triangle inférieur) pour les paires  $(X_1, X_2, X_3)$  du jeu de données de l’*Australian Institute of Sport*

#### 4.4.2 Analyse statistique

Dans la suite, on procédera à l’analyse bidimensionnelle des paires de variables  $(X_1, X_2)$ ,  $(X_2, X_3)$ ,  $(X_1, X_3)$  et  $(X_4, X_5)$ . Les histogrammes, les nuages de points des données brutes, de même que les nuages de points des pseudo-observations, se retrouvent à la Figure 4.1 pour le triplet  $(X_1, X_2, X_3)$ . La Figure 4.2 montre les données pour la paire

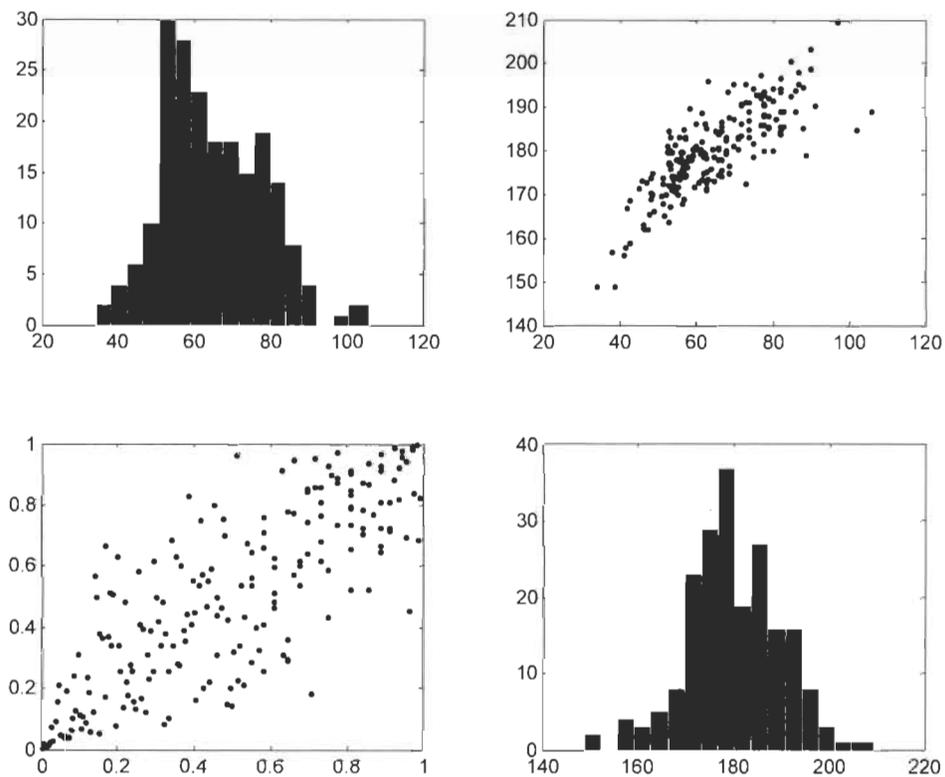


FIGURE 4.2 – Histogrammes (sur la diagonale), nuage de points des données brutes (triangle supérieur) et nuage de points des pseudo-observations (triangle inférieur) pour la paire  $(X_4, X_5)$  du jeu de données de l'*Australian Institute of Sport*

$(X_4, X_5)$ . L'examen des nuages de points des pseudo-observations révèle des niveaux de dépendance élevés, notamment du point de vue des queues supérieures. On rappelle que l'indice de dépendance caudale supérieur  $\lambda_U$  mesure justement la propension qu'ont deux variables aléatoires à prendre simultanément des valeurs très élevées.

On considérera également l'estimation de l'indice de dépendance caudale inférieur,

c'est-à-dire  $\lambda_L$ . Pour des paires observées  $(X_1, Y_1), \dots, (X_n, y_n)$ , on obtient facilement l'estimateur  $\hat{\lambda}_L$  en prenant  $\hat{\lambda}_U$  calculé sur  $(-X_1, -Y_1), \dots, (-X_n, -Y_n)$ . Les résultats de ces analyses se retrouvent au Tableau 4.7. À noter que pour les estimateurs non-paramétriques  $\hat{\lambda}_U^{CFG}$ ,  $\hat{\lambda}_U^{SS}$  et  $\hat{\lambda}_U^{Coles}$ , on a pris la valeur  $k = 20$ , ce qui fait que la portion des données qui entrent dans l'estimation est d'environ 20%.

Globalement, ces résultats suggèrent la présence significative de dépendance caudale supérieure et inférieure pour les quatre paires étudiées. De façon générale, les valeurs des estimations sont relativement concordantes. À la lumière des simulations effectuées précédemment, on peut considérer la valeur de  $\hat{\lambda}_U^{PMV2}$  comme fiable. On voit que pour les paires du triplet  $(X_1, X_2, X_3)$ , les indices  $\lambda_U$  et  $\lambda_L$  sont assez semblables, ce qui suggère une structure de dépendance en symétrie radiale.

TABLE 4.7 – Résultats de l'estimation des coefficients de dépendance caudale  $\lambda_U$  (partie supérieure du tableau) et  $\lambda_L$  (partie inférieure du tableau) pour quelques paires de variables du jeu de données de l'*Australian Institute of Sport*

Paire	$\hat{\lambda}_U^{CFG}$	$\hat{\lambda}_U^{SS}$	$\hat{\lambda}_U^{Coles}$	$\hat{\lambda}_U^{ITK(1)}$	$\hat{\lambda}_U^{ITK(5)}$	$\hat{\lambda}_U^{ITK(10)}$	$\hat{\lambda}_U^{PMV2}$
$(X_1, X_2)$	0,7872	0,8000	0,7869	0,8498	0,7846	0,7089	0,6725
$(X_1, X_3)$	0,7362	0,7000	0,6785	0,8016	0,7326	0,6425	0,5739
$(X_2, X_3)$	0,8451	0,7500	0,7328	0,8881	0,8307	0,7704	0,8459
$(X_4, X_5)$	0,6445	0,5000	0,4580	0,7210	0,6584	0,5493	0,6102
$(X_1, X_2)$	0,7644	0,6500	0,6238	0,8498	0,7846	0,7089	0,7215
$(X_1, X_3)$	0,7200	0,6000	0,5689	0,8016	0,7326	0,6425	0,6583
$(X_2, X_3)$	0,8337	0,7500	0,7328	0,8881	0,8307	0,7704	0,8307
$(X_4, X_5)$	0,6856	0,6500	0,6238	0,7210	0,6584	0,5493	0,4079

# Conclusion

L'indice de dépendance caudale mesure la propension qu'ont deux variables aléatoires à prendre simultanément des valeurs très élevées. Ainsi les estimateurs de  $\lambda_U$  et  $\lambda_L$  proposés jusqu'ici dans la littérature scientifiques sont généralement extrêmement biaisés. Ce travail consistait à pallier à cette lacune en proposant de nouveaux estimateurs semi-paramétriques fondés sur l'expression obtenue par Favre et al. [17] pour l'indice de dépendance caudale. Cette dernière est construite sous l'hypothèse d'une structure de dépendance appartenant à la famille des copules de Fisher.

Par ailleurs, l'efficacité des nouveaux estimateurs en termes de biais et d'erreur quadratique moyenne est étudiée à l'aide de simulations. Pour cela, on a d'abord procédé à une étude comparative du biais et de l'erreur quadratique moyenne entre les estimateurs non-paramétriques  $\hat{\lambda}_U^{\text{CFG}}$ ,  $\hat{\lambda}_U^{\text{SS}}$ ,  $\hat{\lambda}_U^{\text{Coles}}$  et les estimateurs proposés, à savoir  $\hat{\lambda}_U^{\text{ITK}}(1)$ ,  $\hat{\lambda}_U^{\text{ITK}}(5)$ ,  $\hat{\lambda}_U^{\text{ITK}}(10)$  et  $\hat{\lambda}_U^{\text{PMV}^2}$ . Cette comparaison a été faite sous des données simulées de la copule de Fisher, mais aussi sous des structures de dépendance modélisées par la copule de Gumbel et Khi-deux. Une modélisation à l'aide de ces estimateurs a été effectuée sur des données provenant de l'*Australian Institute of Sport*.

Les simulations ont montré que dans la plupart des cas, les estimateurs proposés produisent de bonnes estimations de l'indice de dépendance caudale supérieur. Notamment, l'estimateur  $\hat{\lambda}_U^{\text{PMV}^2}$ , qui ne nécessite ni le choix d'aucun paramètre de lissage, ni le choix d'un degré de liberté, est particulièrement précis lorsque les données présentent un niveau de dépendance élevé. On peut donc affirmer que ce mémoire a atteint son objectif principal, à savoir une estimation fiable de l'indice  $\lambda_U$ .

# Bibliographie

- [1] K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance Math. Econom.*, 44(2) :182–198, 2009. ISSN 0167-6687.
- [2] A. Aghakouchak, S. Sellars, and S. Sorooshian. Methods of tail dependence estimation. in : Aghakouchak a, easterling d, hsu k, schubert s, sorooshian s (eds) extremes in a changing climate. *Water science and technology library*, 65 : 163–179, 2013.
- [3] T. Bahraoui, T. Bouezmarni, and J.-F. Quessy. A family of goodness-of-fit tests for copulas based on characteristic functions. *Scand. J. Stat.*, 45(2) :301–323, 2018.
- [4] A. Bardossy and G. Pegram. Interpolation of precipitation under topographic influence at different time scales. *Water Resour Res.*, 49(8) :4545–4565, 2013.
- [5] A. Bardossy and G. Pegram. The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs. *Water Resour Res.*, 46(6) :3423–3442, 2013.
- [6] C. B. Borkowf. Computing the nonnull asymptotic variance and the asymptotic relative efficiency of spearman’s rank correlation. *Comput. Statist. Data Anal.*, 39(3) :271–286, 2002. ISSN 0167-9473.
- [7] A. Bardossy. Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11) :1–12, 2006.

- [8] B. M. Brown and S. X. Chen. Beta-Bernstein smoothing for regression curves with compact support. *Scand. J. Statist.*, 26(1) :47–59, 1999. ISSN 0303-6898.
- [9] P. Capéraà, A.-L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3) :567–577, 1997. ISSN 0006-3444.
- [10] F. Chen and Y. R. Huang. Some results for classic Lucas-Fibonacci sequences. *Commun. Appl. Math. Comput.*, 21(1) :116–120, 2007. ISSN 1006-6330.
- [11] S. X. Chen. Beta kernel estimators for density functions. *Comput. Statist. Data Anal.*, 31(2) :131–145, 1999. ISSN 0167-9473.
- [12] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula methods in finance*. Wiley, Chichester, 2004.
- [13] R. Coles, C. S. Calude, P. H. Hertling, and B. Khoussainov. Degree-theoretic aspects of computably enumerable reals. In *Models and computability (Leeds, 1997)*, volume 259 of *London Math. Soc. Lecture Note Ser.*, pages 23–39. Cambridge Univ. Press, Cambridge, 1999.
- [14] R. D. Cook and M. E. Johnson. A family of distributions for modelling nonelliptically symmetric multivariate data. *J. Roy. Statist. Soc. Ser. B*, 43(2) :210–218, 1981. ISSN 0035-9246.
- [15] G. Dall’Aglia. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuola Norm. Sup. Pisa (3)*, 10 :35–74, 1956.
- [16] P. Deheuvels. La fonction de dépendance empirique et ses propriétés. un test non paramétrique d’indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, 65 :274–292, 1979.
- [17] A.-C. Favre, J.-F. Quessy, and M.-H. Toupin. The new family of Fischer copulas to model upper tail dependence and radial asymmetry : properties and application to high-dimensional rainfall data. *Environmetrics*, 29(3) :2494, 17, 2018.

- [18] J.-D. Fermanian, D. Radulović, and M. H. Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10 :847–860, 2004.
- [19] R. Féron. Sur les tableaux de corrélation dont les marges sont données. Cas de l'espace a trois dimensions. *Publ. Inst. Statist. Univ. Paris*, 5 :3–12, 1956.
- [20] G. Frahm, M. Junker, and R. Schmidt. Estimating the tail-dependence coefficient : properties and pitfalls. *Insurance Math. Econom.*, 37(1) :80–100, 2005. ISSN 0167-6687.
- [21] M. Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon. Sect. A. (3)*, 14 :53–77, 1951.
- [22] E. W. Frees and E. A. Valdez. Understanding relationships using copulas. *N. Am. Actuar. J.*, 2(1) :1–25, 1998. ISSN 1092-0277.
- [23] J. Galambos. *The asymptotic theory of extreme order statistics*. John Wiley & Sons, New York-Chichester-Brisbane, 1978. ISBN 0-471-02148-2. Wiley Series in Probability and Mathematical Statistics.
- [24] C. Genest and J. MacKay. The joy of copulas : bivariate distributions with uniform marginals. *Amer. Statist.*, 40(4) :280–283, 1986. ISSN 0003-1305.
- [25] C. Genest, K. Ghoudi, and L.-P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3) :543–552, 1995. ISSN 0006-3444. doi : 10.1093/biomet/82.3.543.
- [26] C. Genest, B. Rémillard, and D. Beaudoin. Goodness-of-fit tests for copulas : a review and a power study. *Insurance Math. Econom.*, 44(2) :199–213, 2009. ISSN 0167-6687. doi : 10.1016/j.insmatheco.2007.10.005. URL <http://dx.doi.org/10.1016/j.insmatheco.2007.10.005>.
- [27] T. Ghizzoni, G. Roth, and R. Rudari. Multivariate skew-t approach to the design of accumulation risk scenarios for the flooding hazard. *Adv Water Resour.*, 33 (10) :1243–1255, 2010.

- [28] I. Gijbels and J. Mielniczuk. Estimating the density of a copula function. *Comm. Statist. Theory Methods*, 19(2) :445–464, 1990. ISSN 0361-0926. doi : 10.1080/03610929008830212.
- [29] F. Harrell and C. E. Davis. A new distribution-free quantile estimator. *Biometrika*, 69(3) :635–640, 1982. ISSN 0006-3444.
- [30] P. Hominal and P. Deheuvels. Estimation non paramétrique de la densité compte tenu d'informations sur le support. *Rev. Statist. Appl.*, 27(3) :47–68, 1979. ISSN 0035-175X.
- [31] H. Joe and T. Hu. Multivariate distributions from mixtures of max-infinitely divisible distributions. *J. Multivariate Anal.*, 57(2) :240–265, 1996. ISSN 0047-259X. doi : 10.1006/jmva.1996.0032.
- [32] H. Joe, R. Smith, and I. Weissman. Bivariate threshold methods for extremes. volume 54(1), page 171–183. *J R Stat Soc Ser B (Methodol)*, 1992.
- [33] G. Kim, M. Silvapulle, and P. Silvapulle. Comparison of semiparametric and parametric methods for estimating copulas. *Comput. Statist. Data Anal.*, 51(6) : 2836–2850, 2007. ISSN 0167-9473. doi : 10.1016/j.csda.2006.10.009.
- [34] G. Kimeldorf and A. Sampson. Uniform representations of bivariate distributions. *Comm. Statist.*, 4(7) :617–627, 1975.
- [35] Y. Malevergne, V. Pisarenko, and D. Sornette. Empirical distributions of stock returns : between the stretched exponential and the power law ? *Quant. Finance*, 5(4) :379–401, 2005. ISSN 1469-7688.
- [36] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management : Concepts, techniques and tools*. Princeton University Press, Princeton, NJ, 2005.
- [37] M. Omelka, I. Gijbels, and N. Veraverbeke. Improved kernel estimation of copulas : weak convergence and goodness-of-fit testing. *Ann. Statist.*, 37(5B) : 3023–3058, 2009. ISSN 0090-5364.

- [38] A. Poulin, D. Huard, A. Favre, and S. Pugin. Importance of tail dependence in bivariate frequency analysis. *J Hydrol Eng.*, 12(4) :394–403, 2007.
- [39] J.-F. Quessy. A general framework for testing homogeneity hypotheses about copulas. *Electron. J. Stat.*, 10(1) :1064–1097, 2016. ISSN 1935-7524.
- [40] J.-F. Quessy, L.-P. Rivest, and M.-H. Toupin. On the family of multivariate chi-square copulas. *J. Multivariate Anal.*, 152 :40–60, 2016.
- [41] F. H. Ruymgaart and M. C. A. Zuijlen. Asymptotic normality of multivariate linear rank statistics in the non-i.i.d. 6(3) :588–602, 1978.
- [42] G. Salvadori, C. de Michele, N. T. Kottegoda, and R. Rosso. *Extremes in nature : An approach using copulas*. Springer, New York, 2007.
- [43] A. Schmidt. A multi-mesh finite element method for phase-field simulations. In *Interface and transport dynamics*, volume 32 of *Lect. Notes Comput. Sci. Eng.*, pages 208–217. Springer, Berlin, 2003.
- [44] R. Schmidt and U. Stadtmüller. Non-parametric estimation of tail dependence. *Scand. J. Statist.*, 33(2) :307–335, 2006. ISSN 0303-6898.
- [45] P. Schuster. Effects of finite population size and other stochastic phenomena in molecular evolution. In *Complex systems—operational approaches in neurobiology, physics, and computers (Schloss Elmau, 1985)*, volume 31 of *Springer Ser. Synergetics*, pages 16–35. Springer, Berlin, 1985.
- [46] J. Segers. Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, 18(3) :764–782, 2012.
- [47] F. Serinaldi. Analysis of inter-gauge dependence by kendall’s, upper tail dependence coefficient, and 2-copulas with application to rainfall fields. *Stoch Environ Res Risk Assess*, 22(6) :671–688, 2007.
- [48] F. Serinaldi and C. Kilsby. Simulating daily rainfall fields over large areas for collective risk estimation. *J Hydrol*, 512 :285–302, 2014.

- [49] F. Serinaldi, A. Bardossy, and C. G. Kilsby. Upper tail dependence in rainfall extremes : would we know it if we saw it? *Stoch. Environ. Res. Risk Assess.*, 2015. doi : 10.1007/s00477-014-0946-8.
- [50] J. H. Shih and T. A. Louis. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4) :1384–1399, 1995. ISSN 0006-341X. doi : 10.2307/2533269. URL <http://dx.doi.org/10.2307/2533269>.
- [51] A. Sklar. Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8 :229–231, 1959.
- [52] C. Spearman. The proof and measurement of association between two things. *Am. J. Psychol.*, 15 :72–101, 1904.
- [53] R. D. Telford and R. B. Cunningham. Sex, sport, and body-size dependency of hematology in highly trained athletes. *Medicine and science in sports and exercise*, 23(7) :788–794, 1991.
- [54] N. Vaschide and A. Binet. Corrélation des épreuves physiques. *In : L'année psychologique*, 4 :142–172, 1897.

# Annexe A

## Éléments complémentaires

### A.1 Mesurer la dépendance

#### A.1.1 Inconvénients de la corrélation

La corrélation entre deux variables aléatoires  $X$  et  $Y$  est définie pour  $\rho_X = E(X)$ ,  $\rho_Y = E(Y)$ ,  $\sigma_X^2 = Var(X)$  et  $\sigma_Y^2 = Var(Y)$  par

$$\rho_{X,Y} = \frac{E\{(X - \rho_X)(Y - \rho_Y)\}}{\sigma_X \sigma_Y}.$$

Lorsque  $X$  et  $Y$  sont indépendantes, on a

$$\rho_{X,Y} = \frac{E\{(X - \rho_X)\} E\{(Y - \rho_Y)\}}{\sigma_X \sigma_Y} = 0.$$

Néanmoins le contraire n'est pas toujours vérifié. En effet, il se peut que  $\rho_{X,Y} = 0$ , mais que  $X$  et  $Y$  soient dépendantes. Une mesure de dépendance ne devrait pas dépendre des comportements individuels des variables. Malheureusement, ce n'est pas le cas avec corrélation.

Il y'a deux problèmes liés à l'utilisation  $\rho_{X,Y}$  pour mesurer la force de la dépendance :

- Sauf dans certains cas particuliers (ex. Normalité bivariée),  $\rho_{X,Y} = 0$  n'implique pas l'indépendance,
- La valeur de  $\rho_{X,Y}$  n'est pas invariante sous des transformations des marges.

## A.1.2 Rho de Spearman

### Corrélation de rangs

Plutôt que de mesurer la corrélation à partir de  $(X_1, Y_1), \dots, (X_n, Y_n)$ , Vaschide and Binet [54] et Spearman [52] ont eu l'idée d'utiliser les rangs. Spécifiquement, notons :

- $R_i$ , le rang de  $X_i$  parmi  $X_1, \dots, X_n$  ;
- $S_i$ , le rang de  $Y_i$  parmi  $Y_1, \dots, Y_n$ .

On définit la corrélation de rangs par la corrélation calculée à partir de l'échantillon  $(R_1, S_1), \dots, (R_n, S_n)$ . Ainsi, on définit la corrélation de rangs par

$$r_S = \rho((R_i, S_i), \dots, (R_n, S_n)).$$

On appelle aussi  $r_S$  le *rho* de Spearman. On montre que

$$r_S = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - S_i).$$

Les rangs sont invariants sous une transformation monotone croissante des observations. Autrement dit, le rang de  $X_i$  est le même que le rang de  $r(X_i)$ , où  $r$  est une fonction croissante arbitraire. Par conséquent,  $r_S$  est invariante sous des transformations monotones croissantes des observations marginales. On note que

- $R_i = S_i$ , sous de la dépendance positive parfaite ;
- $R_i = n + 1 - S_i$ , sous de la dépendance négative parfaite.

Par conséquent, on montre que

- $r_S = 1$  sous de la dépendance positive parfaite;
- $r_S = -1$  sous de la dépendance négative parfaite.

### Versión théorique

Soient les fonctions de répartition empiriques

$$F_n(X) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(X_i \leq x)} \quad \text{et} \quad G_n(Y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(Y_i \leq y)},$$

on remarque que

$$R_i = nF_n(X_i) \quad \text{et} \quad S_i = nG_n(Y_i).$$

Puisque pour  $n \rightarrow \infty$ , on a intuitivement que  $F_n \rightarrow F$  et  $G_n \rightarrow G$ , on conclut de manière heuristique que

$$r_S \rightarrow \rho_S = \frac{E\{F(X)G(Y)\} - E\{F(X)\}E\{G(Y)\}}{\sqrt{\text{Var}\{F(X)\}\text{Var}\{G(Y)\}}}.$$

En posant  $U = F(X)$  et  $V = G(Y)$ , on peut écrire

$$\rho_S = \frac{E\{UV\} - E\{U\}E\{V\}}{\sqrt{\text{Var}\{U\}\text{Var}\{V\}}},$$

puisque  $(U, V) \sim C$ , où  $C$  est la copule de  $(X, Y)$ ,  $\rho_S = \rho_S(C)$  dépend uniquement de la copule de  $(X, Y)$ .

### Monotonie en fonction de C

On dit qu'une copule  $C_1$  est moins concordante qu'une copule  $C_2$ , notée  $C_1 \prec C_2$ , si

$$C_1(u, v) \leq C_2(u, v) \quad \text{pour tout} \quad (u, v) \in [0, 1]^2.$$

La relation  $\prec$  induit un ordre de concordance partiel sur l'ensemble des copules.

### Normalité asymptotique

Ruymgaart and Zuijlen [41] ont démontré la normalité asymptotique de la version standardisée de  $r_S$ . Spécifiquement, ils ont obtenu

$$\sqrt{n} \{r_S - \rho_S(C)\} \sim \mathcal{N}(0, \sigma_\rho^2).$$

Ici, le symbole  $\sim$  veut dire convergence en loi. Borkowf [6] a obtenu une formule pour la variance asymptotique  $\sigma^2$  en fonction de la copule  $C$ . Soient les hypothèses d'indépendance

$$H_0 : \rho_S = 0 \quad \text{et} \quad H_1 : \rho_S \neq 0.$$

Puisque sous l'indépendance,  $\sigma_\rho^2 = 1$ , on pourrait rejeter  $H_0$  quand  $|\sqrt{n}r_S| > Z_{\alpha/2}$  où  $Z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ . Également, un intervalle de confiance pour  $\rho_S$  est :

$$I = r_S \pm Z_{\alpha/2} \frac{\hat{\sigma}_\rho}{\sqrt{n}},$$

où  $\hat{\sigma}_\rho$  est une estimation de  $\sigma_\rho$ .

### A.1.3 Tau de Kendall

#### Concordance et version empirique

Une façon intéressante d'envisager la dépendance est de considérer la notion de concordance. Pour la définir, soient deux couples indépendantes  $(X_1, Y_1)$  et  $(X_2, Y_2)$  de loi

$H$ . On dit que ces couples sont concordants si

$$X_1 < X_2, Y_1 < Y_2 \quad \text{ou} \quad X_1 > X_2, Y_1 > Y_2.$$

À l'opposé, on dit qu'elles sont discordants si

$$X_1 < X_2, Y_1 > Y_2 \quad \text{ou} \quad X_1 > X_2, Y_1 < Y_2.$$

Soit un échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  i.i.d  $H$ . Définissons, parmi les  $n(n-1)/2$  paires possibles

- $C_n$ , le nombre de concordances ;
- $D_n$ , le nombre de discordances.

Le tau de Kendall empirique est défini par

$$\tau_n = \frac{C_n - D_n}{\frac{n}{2}}.$$

Cette définition est basée sur la constatation que

- $C_n = \frac{n}{2}$  et  $D_n = 0$  en présence de dépendance positive parfaite ;
- $C_n = 0$  et  $D_n = \frac{n}{2}$  en présence de dépendance négative parfaite ;
- $C_n \simeq D_n$  sous l'indépendance.

### Version théorique

On montre par un calcul direct que  $\tau_n$  est sans biais pour

$$\tau = P(C) - P(D).$$

Ici, pour une loi bivariée  $H$

- $P(C)$  est la probabilité de concordance ;
- $P(D)$  est la probabilité de discordance.

Le tau de Kendall théorique d'un couple  $(X, Y)$  ne dépend que de la copule sous-jacente  $C$ .

### Monotonie en fonction de $C$

Si  $C_1 \prec C_2$ , alors on a

$$E_{C_1}(C_1) \leq E_{C_2}(C_2),$$

une conséquence directe est que

$$\tau(C_1) \leq \tau(C_2),$$

et il s'ensuit que pour toute copule  $C$

$$-1 \leq \tau(C) \leq 1.$$

### Normalité asymptotique

Dans un cadre général, soit des observations  $X_1, \dots, X_n$  où  $X_1 \in \mathbb{R}^d$ . Pour une fonction  $\ell_1 : \mathbb{R}^d \rightarrow \mathbb{R}$ , une  $U$ -statistique d'ordre 1 est définie par

$$U_{n_1} = \binom{n}{1}^{-1} \sum_{i=1}^n \ell_1(X_i).$$

Pour une fonction  $\ell_2 : \mathbb{R}^d * \mathbb{R}^d \rightarrow \mathbb{R}$ , une  $U$ -statistique d'ordre 2 est définie par

$$U_{n_2} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \ell_2(X_i, X_j).$$

### A.1.4 Mesures de concordances

#### Opérateur de concordance

Soient  $(U_1, V_1) \sim C_1$  et  $(U_2, V_2) \sim C_2$ , on définit l'opérateur de concordance  $Q$  par la probabilité de concordance entre  $(U_1, V_1)$  et  $(U_2, V_2)$ . Spécifiquement

$$Q(C_1, C_2) = P(U_1 < U_2, V_1 < V_2) + P(U_1 > U_2, V_1 > V_2).$$

On montre que

$$Q(C_1, C_2) = E_{C_2}(C_1) + E_{C_1}(C_2).$$

On voit que si  $C_1 \prec C_1^*$  et  $C_2 \prec C_2^*$ , alors

$$Q(C_1, C_2) \leq Q(C_1^*, C_2^*).$$

#### Quelques mesures de concordance

Le tau de Kendall et le rho de Spearman sont des mesures de concordance car

$$\tau = Q(C, C) - 1 \quad \text{et} \quad \rho = 6Q(C, \pi) - 3.$$

Il existe aussi d'autres mesures de concordance telles que

- le coefficient de Gini  $\gamma(C) = 2\{Q(C, M) - Q(C, W) - 1\}$ ,
- le Spearman's footrule  $\tilde{\rho}_S(C) = 3Q(C, M) - 2$ .

### Concordance multivariée

Pour des vecteurs aléatoires  $U_1 = (U_{11}, \dots, U_{1d}) \sim C_1$  et  $U_2 = (U_{21}, \dots, U_{2d}) \sim C_2$ , on définit l'opérateur de concordance multivarié par

$$Q(C_1, C_2) = P(U_1 < U_2) + P(U_1 > U_2).$$

On a que

$$Q(C_1, C_2) = E_{C_2}(C_1) + E_{C_1}(C_2).$$

Afin de généraliser le tau de Kendall au cas multivarié, soit

$$\tau_d(C) = AQ(C, C) + B.$$

On cherchera des valeurs  $A, B \in \mathbb{R}$  de telle sorte que

$$\tau_d(\Pi) = 0 \quad \text{et} \quad \tau_d(M) = 1.$$

On déduit alors le tau de Kendall multivarié par

$$\tau_d(C) = \frac{2^{d-1}Q(C, C) - 1}{2^{d-1} - 1}.$$

On procède de la même façon pour généraliser le rho de Spearman en considérant

$$\rho_d(C) = AQ(C, C) + B.$$

Les mêmes conditions que celles imposées à  $\tau_d(C)$  amènent

$$\rho_d(C) = \left( \frac{d+1}{2^d - d - 1} \right) \{2^{d-1}Q(C, \Pi) - 1\}.$$

### Versions empiriques

Soit  $X_1, X_2 \sim H$ , où  $C$  est la copule de  $H$ . Alors il est facile de montrer qu'un estimateur sans biais pour  $Q(C, C)$  est

$$\mathbb{1}_{(X_1 < X_2)} + \mathbb{1}_{(X_1 > X_2)}.$$

Cette constatation permet de construire un estimateur de  $\tau_d$  à l'aide d'une  $U$ -statistique d'ordre 2. Pour ce faire, soit un échantillon  $X_1, X_2$  d'une population de loi  $H$  et de copule  $C$ . On estime alors  $Q(C, C)$  avec

$$Q_n = \left(\frac{n}{2}\right)^{-1} \sum_{1 \leq i < j \leq n} \{ \mathbb{1}_{(X_i < X_j)} + \mathbb{1}_{(X_i > X_j)} \}.$$

Une version empirique de  $\tau_d(C)$  est donc

$$\tau_{n,d}(C) = \frac{2^{d-1} Q_n - 1}{2^{d-1} - 1}.$$

En définissant  $l(X_1, X_2) = \mathbb{1}_{(X_1 < X_2)} + \mathbb{1}_{(X_1 > X_2)}$ , on a la représentation

$$Q_n = \left(\frac{n}{2}\right)^{-1} \sum_{1 \leq i < j \leq n} l(X_i, X_j).$$

Par un résultat sur les  $U$ -statistiques d'ordre 2, on déduit que pour  $\sigma_Q^2 = \text{Var} \{H(X) + \bar{H}(X)\}$ , alors

$$\sqrt{n} \{Q_n - Q(C, C)\} \rightsquigarrow N(0, 4\sigma_Q^2).$$

Ensuite, il s'agit de remarquer que

$$\sqrt{n} \{\tau_{n,d} - \tau_d(C)\} \left(\frac{2^{d-1}}{2^{d-1} - 1}\right)^2 \sqrt{n} \{Q_n - Q(C, C)\}.$$

On déduit alors que  $\sqrt{n} \{\tau_{n,d} - \tau_d(C)\} \rightsquigarrow N(0, 4\sigma_Q^2)$ , où

$$\sigma_\tau^2 = \left( \frac{2^d}{2^{d-1} - 1} \right) \text{Var} \{H(X) + \bar{H}(X)\}.$$

## A.2 Estimation par noyau dans un contexte de couples

### A.2.1 Estimation de la copule

Une version lisse notée  $C_n^{(SE)}$  est proposée par les auteurs Fermanian et al. [18].

$$C_n^{(SE)}(u, v) = \hat{H}_n \left( \hat{F}_n^{-1}(u), \hat{G}_n^{-1}(v) \right),$$

où les quantités  $\hat{H}_n$ ,  $\hat{F}_n$  et  $\hat{G}_n$  sont

$$\hat{H}_n(x, y) = \frac{1}{n} \sum_{i=1}^n K_n(x - X_i, y - Y_i)$$

et

$$\hat{F}_n(x) = \hat{H}_n(x, +\infty), \quad \hat{G}_n(y) = \hat{H}_n(+\infty, x)$$

avec

$$K_n(x, y) = K \left( \frac{x}{b_n}, \frac{y}{b_n} \right), \quad K(x, y) = \int_{-\infty}^x \int_{-\infty}^y k(s, t) ds dt$$

où  $k(s, t)$  est la fonction densité du noyau bidimensionnelle, et  $K_n$  est un paramètre de lissage qui tend vers 0 quand  $n$  converge vers l'infini. Cet estimateur connaît un problème de biais lié à l'intervalle d'étude sur  $[0, 1]^2$ . L'estimateur  $C_n^{(SE)}$  n'est pas convergent et pour remédier à cela d'autres méthodes d'estimation semi-paramétriques sont proposées notamment : l'estimateur par transformation des données établi par Omelka et al. [37] et l'estimateur local linéaire proposé par Chen and Huang [10] afin

de faire face aux biais près des frontières du carré unité.

### Estimateur par transformation des données

Étant donnée de l'invariance de la fonction copule, une transformation des données originales pour réduire le biais aux bornes de l'estimateur à noyau est nécessaire. Soient  $X'_1 = T_1(X_1)$  et  $X'_2 = T_2(X_2)$  une transformation des données originales  $X_1$  et  $X_2$  avec  $T_1$  et  $T_2$  des fonctions croissantes. Néanmoins le choix des fonctions  $T_1$  et  $T_2$  n'est pas facile raison pour laquelle les auteurs Omelka et al. [37] ont mis en place la procédure suivante :

— construction de pseudo-observations uniformes

$$\widehat{U}_i^{(E)} = \frac{n}{n+1} F_n(X_i) \quad \text{et} \quad \widehat{V}_i^{(E)} = \frac{n}{n+1} C_n(Y_i);$$

— soit  $\phi$  une fonction de distribution de données, posons les deux fonctions

$$\widehat{S}_i = \phi^{-1} U_i^{(E)} \quad \text{et} \quad \widehat{T}_i = \phi^{-1} V_i^{(E)};$$

— on utilise les observations transformées sur l'estimateur  $C_n^{(SE)}$

$$C_n^{(SE)}(u, v) = \frac{1}{n} \sum_{i=1}^n K \left( \frac{\phi^{-1}(u) - \widehat{S}_i}{h_n} \right) \times K \left( \frac{\phi^{-1}(v) - \widehat{T}_i}{h_n} \right)$$

avec  $h_n$  paramètres de lissages.

Cette approche est faite pour deux observations, mais elle peut s'étendre sur plusieurs transformations en considérant différentes  $\phi$  et différents noyaux  $K$ .

### Estimateur local linéaire

Chen and Huang [10] ont proposé un estimateur local linéaire basé sur la procédure suivante

— L'estimation des fonctions marginales

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{b_{n_1}}\right)$$

et

$$\widehat{G}_n(y) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{y - Y_i}{b_{n_2}}\right),$$

avec  $K$  l'intégrale d'une fonction à noyau  $k$  symétrique délimitée sur  $[-1, 1]$ ,  $b_{n_1}$  et  $b_{n_2}$  des paramètres de lissage.

— Les pseudo-observations

$$\widehat{U}_i = \widehat{F}_n(X_i) \quad \text{et} \quad \widehat{V}_i = \widehat{G}_n(Y_i)$$

sont des données transformées pour les non observées  $F(X_i)$  et  $G(Y_i)$ .

— L'estimateur local linéaire de la copule proposée est

$$\widehat{C}_n^{(LL)}(u, v) = \frac{1}{n} \sum_{i=1}^n K_{u, h_n}\left(\frac{u - \widehat{U}_i}{h_n}\right) \times K_{v, h_n}\left(\frac{v - \widehat{V}_i}{h_n}\right),$$

avec

$$K_{u, h_n}(x) = \int_{-\infty}^x k_{u, h_n}(s) ds.$$

Selon Chen and Huang [10] le choix optimal des constantes de lissage  $h = h_n$  peut minimiser la variance et l'erreur quadratique moyenne de l'estimation de la copule. Cependant l'expérience de Monte Carlo montre que les paramètres de lissage  $b_{n_1}$  et  $b_{n_2}$  impactent moins sur le calcul de l'estimateur  $\widehat{C}_n^{(LL)}$ . L'inconvénient de cet estimateur  $\widehat{C}_n^{(LL)}$  est que pour de nombreuses familles de copules, le biais est seulement d'ordre  $O(h_n)$  comparé à  $O(h_n^2)$  qui est l'ordre du biais à l'intérieur. Le problème est causé par l'absence de limites des dérivées partielles du second ordre pour de nombreuses familles de copules. Omelka et al. [37] proposent l'estimateur  $\widehat{C}_n^{(LLS)}$  qui est une version

de celui de  $\widehat{C}_n^{(LL)}$  donné par

$$\widehat{C}_n^{(LLS)}(u, v) = \frac{1}{n} \sum_{i=1}^n K_{u, h_n} \left( \frac{u - \widehat{U}_i}{b(u)h_n} \right) \times K_{v, h_n} \left( \frac{v - \widehat{V}_i}{b(v)h_n} \right),$$

avec  $b(u) = \min(\sqrt{u}, \sqrt{1-u})$  et  $h(n)$  substitué par  $b(u)h_n$  réduisant la valeur du paramètre de lissage au niveau des coins de l'unité carrée.

## A.2.2 Estimation de la densité d'une copule

Nous avons deux méthodes d'estimation de la densité de copule notamment la méthode de l'image miroir et la méthode à noyau bêta.

### Méthode de l'image miroir

Élaborée par les auteurs Gijbels and Mielniczuk [28], cette approche s'occupe de la correction du biais aux frontières par le biais de l'augmentation des données obtenues en réfléchissant les données originales par rapport aux bords et les coins du carré de l'unité. Développée par Hominal and Deheuvels [30] et Schuster [45], cette méthode consiste à ajouter certaines masses manquantes en reflétant l'échantillon par rapport aux bords. L'estimateur par image miroir de la copule est donné par

$$\begin{aligned} \widehat{C}_n^{(MR)}(u, v) &= \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^9 \left[ K \left( \frac{u - \widehat{U}_i^l}{h_n} \right) - K \left( \frac{\widehat{U}_i^l}{h_n} \right) \right] \\ &\quad \times \left[ K \left( \frac{v - \widehat{V}_i^l}{h_n} \right) - K \left( \frac{\widehat{V}_i^l}{h_n} \right) \right]. \end{aligned}$$

Le problème avec cet estimateur est le même que celui de l'estimateur linéaire local car tous les deux connaissent un problème de biais au niveau des bords. Ainsi pour remédier à ce problème, un nouveau estimateur  $\widehat{C}_n^{(MRS)}$  est proposé par Omelka et al.

[37].

$$\begin{aligned} \widehat{C}_n^{(MRS)}(u, v) &= \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^9 \left[ K \left( \frac{u - \widehat{U}_i^l}{b_u h_n} \right) - K \left( \frac{\widehat{U}_i^l}{b_u h_n} \right) \right] \\ &\quad \times \left[ K \left( \frac{v - \widehat{V}_i^l}{b_v h_n} \right) - K \left( \frac{\widehat{V}_i^l}{b_v h_n} \right) \right]. \end{aligned}$$

### Méthode à noyau Bêta

Introduite par Brown and Chen [8] et Chen [11], la méthode à noyau Bêta consiste à faire une estimation non paramétrique des courbes de régression et des densités univariées à support compact. Harrell and Davis [29] et Chen [11] ont mis en place un nouveau estimateur pour éliminer le biais au bord de l'estimateur à noyau standard. Pour cela cet estimateur est une fonction de densité avec support compact connu  $[0, 1]$  défini comme suite

$$\widehat{f}_{h_n}(x) = \frac{1}{n} \sum_{i=1}^n K \left( X_i, \frac{x}{h_n} + 1, \frac{1-x}{h_n} + 1 \right),$$

où  $K(\cdot, \alpha, \beta)$  est la densité de la distribution bêta avec paramètres  $\alpha$  et  $\beta$ .

$$K(x, \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad x \in [0, 1],$$

où

$$B(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

L'inconvénient avec cette méthode c'est que le choix du paramètre de lissage  $h_n$  peut être difficile. Néanmoins cette approche comporte deux avantages

- le noyau bêta correspond parfois au support compact de l'objet à estimer,
- le noyau bêta est flexible quand on s'éloigne des bords.

Par conséquent, les estimateurs à noyau bêta sont naturellement sans biais aux bords.

Cet estimateur à noyau bêta est donné par l'expression de la densité  $\widehat{c}_{h_n}$  suivante

$$\widehat{c}_{h_n}(u, v) = \frac{1}{n} \sum_{i=1}^n K \left( X_i, \frac{u}{h_n} + 1, \frac{1-u}{h_n} + 1 \right) \times K \left( Y_i, \frac{v}{h_n} + 1, \frac{1-v}{h_n} + 1 \right).$$

### A.3 Méthode du maximum vraisemblance canonique

Contrairement à la méthode de vraisemblance à deux étapes, la méthode canonique proposée par Genest et al. [25] ne fait aucune hypothèse sur les distributions marginales. Pour cela on commence par construire les fonctions de distributions ci-dessous, en utilisant les fonctions de distributions empiriques

$$F_n(x_i) = \frac{1}{n+1} \sum_{j=1}^n 1_{(x_j \leq x_i)} \quad \text{et} \quad G_n(y_i) = \frac{1}{n+1} \sum_{j=1}^n 1_{(y_j \leq y_i)}.$$

L'estimation du paramètre de la famille de la copule paramétrique est donné par

$$\widehat{\alpha} = \arg \max_{\alpha} \sum_{i=1}^n \log c(F_n(x_i), G_n(y_i); \alpha).$$