

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
BABACAR NDOYE

ÉTUDE DE LA PUISSANCE DE TESTS DE SYMÉTRIE RADIALE POUR
COPULES MULTIDIMENSIONNELLES SOUS DE LA DÉPENDANCE DE TYPE
FISHER

Mai 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

L'Université du Québec à Trois Rivières, fondée en 1969, fait partie du réseau des universités de Québec. Elle offre plusieurs spécialités parmi lesquelles le programme de maîtrise en mathématiques et informatique appliquées. La formation comprend un volet théorique et un volet pratique. Après la formation, les élèves doivent être des acteurs de développement affectés auprès des ministères, des organismes publics et/ou privés, des grandes entreprises etc. En tenant compte de cet objectif, les enseignements dispensés sont d'un niveau théorique assez poussé. Le passage à l'aspect pratique se fait de façon alternée avec des travaux de recherche encadrés. C'est dans ce cadre que ce présent mémoire a été rédigé. Il s'agira non seulement d'appliquer les connaissances théoriques acquises, mais de les approfondir en explorant de nouvelles pistes.

Plusieurs personnes ont, directement ou indirectement, contribué à la réalisation de ce mémoire de recherche. Je remercie plus particulièrement mon encadreur, le professeur Jean-François Quessy, sans qui ce mémoire n'aurait assurément jamais vu le jour ; en outre, les bourses qu'il m'a octroyées m'ont permis de mener à terme ce travail. Des remerciements sont aussi dus à l'Institut des Sciences Mathématiques du Québec pour m'avoir offert une bourse pendant deux années consécutives, ainsi qu'au gouvernement du Québec pour m'avoir accordé une bourse d'exemption des frais de scolarité.

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 du Canada.

Enfin, je dédie ce mémoire à ma maman Awa Leye Ndiaye, à mon papa Feu Mar Ndoye, ainsi qu'à mes frères et soeurs, pour leur soutien indéfectible.

Table des matières

Avant-propos	ii
Table des matières	iii
Liste des tableaux	vi
Table des figures	vii
1 Mise en contexte, motivation et structure du mémoire	1
1.1 Mise en contexte	1
1.2 Mesure de la dépendance et copules	2
1.3 Utilisation de la copule de Fisher pour modéliser l'asymétrie radiale	3
1.4 Objectifs et structure du mémoire	4
2 Structures de dépendance, symétrie radiale et copules de Fisher	5
2.1 Qu'est-ce qu'une copule : quelques définitions et propriétés	5
2.1.1 Définition analytique d'une copule et propriétés	6
2.1.2 Théorème de Sklar	6
2.1.3 Propriété d'invariance d'une copule	7
2.1.4 Extension multidimensionnelle	8
2.2 Le concept de symétrie radiale	9
2.3 Quelques modèles de copules	11
2.3.1 Copule d'indépendance	11
2.3.2 Copule Normale	12
2.3.3 Copules Archimédiennes	13

2.3.4	Copules de Plackett	16
2.3.5	Copules de Student	17
2.3.6	Copules Khi-deux	19
2.3.7	Copules de Fisher	19
3	Extension d-dimensionnelle des tests de Genest et Nešlehová	22
3.1	Tests de Genest et Nešlehová pour la symétrie bivariée	22
3.1.1	Copule empirique bivariée et copule de survie empirique	23
3.1.2	Convergence faible de la copule empirique	24
3.1.3	Version moltiplicateur du processus de copule	25
3.1.4	Statistiques de test et leurs versions moltiplicateur	26
3.1.5	Algorithme pour performer les tests et calculer leurs p-valeurs .	27
3.2	Versions d -dimensionnelles des tests de Genest & Nešlehová (2014) .	29
3.2.1	Copule empirique multivariée et sa version survie	29
3.2.2	Convergence faible de la copule empirique multivariée	30
3.2.3	Versions moltiplicateurs de \mathbb{C}_n et \mathbb{C}_n^* quand $d \geq 2$	32
3.2.4	Statistiques de test et leurs versions moltiplicateur	33
3.2.5	Algorithme pour performer les tests et calculer leurs p-valeurs .	34
4	Étude de puissance des tests de symétrie radiale	36
4.1	Procédure de test de Bahraoui & Quessy (2017)	37
4.2	Étude de simulations dans le cas $d = 2$	39
4.2.1	Plan de simulation	39
4.2.2	Résultats et commentaires	40
4.3	Étude de simulations dans le cas $d > 2$	42
4.3.1	Plan de simulation	42
4.3.2	Résultats et commentaires	44
4.4	Analyse de la symétrie radiale de données multidimensionnelles	45
Conclusion		49
Bibliographie		50

A Formules explicites pour les statistiques de test	53
A.1 Cas bidimensionnel	53
A.1.1 Formule explicite pour R_n	53
A.1.2 Formule explicite pour S_n	55
A.1.3 Formule explicite pour T_n	55
A.2 Cas multidimensionnel	56
A.2.1 Formule explicite pour R_n	56
A.2.2 Formule explicite pour S_n	56
A.2.3 Formule explicite pour T_n	57
B Programmes Matlab utilisés pour les simulations	58
B.1 Test de symétrie radiale basé sur la fonction empirique	58
B.2 Estimation de la puissance des tests	60

Liste des tableaux

4.1	Pourcentages de rejet estimés à partir de 1 000 échantillons de la copule de Fisher à $\nu = 1$ degré de liberté (partie du haut : $n = 125$; partie du bas : $n = 250$)	42
4.2	Pourcentages de rejet estimés à partir de 1 000 échantillons de la copule de Fisher à $\nu = 4$ degrés de liberté (partie du haut : $n = 125$; partie du bas : $n = 250$)	43
4.3	Pourcentages de rejet estimés à partir de 1 000 échantillons de la copule de Fisher à $\nu = 7$ degrés de liberté (partie du haut : $n = 125$; partie du bas : $n = 250$)	44
4.4	Pourcentages de rejet estimés à partir de 1 000 échantillons des copules Normale, Fisher et Student à $\nu = 1$ degré de liberté et à $d \in \{3, 5\}$ (partie du haut : $n = 125$; partie du bas : $n = 250$)	45
4.5	Pourcentages de rejet estimés à partir de 1 000 échantillons des copules Normale, Fisher et Student à $\nu = 4$ degrés de liberté et à $d \in \{3, 5\}$ (partie du haut : $n = 125$; partie du bas : $n = 250$)	46
4.6	Pourcentages de rejet estimés à partir de 1 000 échantillons des copules Normale, Fisher et Student à $\nu = 7$ degrés de liberté et à $d \in \{3, 5\}$ (partie du haut : $n = 125$; partie du bas : $n = 250$)	47

Table des figures

2.1	Nuage de points de 2000 paires simulées à partir de la copule d'indépendance	12
2.2	Nuage de points de 1 000 paires simulées à partir de la copule Normale avec un tau de Kendall de 0,6	13
2.3	Nuage de points de 1000 paires simulées à partir de la copule de Clayton avec un tau de Kendall de 0,6	15
2.4	Nuage de points de 1000 paires simulées à partir de la copule de Gumbel avec un tau de Kendall de 0,6	16
2.5	Nuage de points de 1000 paires simulées à partir de la copule de Frank avec un tau de Kendall de 0,6	16
2.6	Nuage de points de 1000 paires simulées à partir de la copule de Plackett avec un tau de Kendall de 0,6	17
2.7	Nuage de points de 1000 paires simulées à partir de la copule de Student à 1 degré de liberté avec un tau de Kendall de 0,6	18
2.8	Nuage de points de 1000 paires simulées à partir de la copule Khi-deux avec $a = 0$ et un tau de Kendall de 0,6	20
2.9	Nuage de points de 1000 paires simulées à partir de la copule de Fisher à 1 degré de liberté avec un tau de Kendall de 0,6	21
4.1	Nuage de points 3000 paires simulées des copules de Fisher à 1, 4 et 7 degrés de liberté (respectivement de haut en bas) pour trois niveaux de dépendance (0,25 à gauche, 0,5 au milieu et 0,75 à droite)	41

4.2 Graphiques pour les données sur les habitudes alimentaires de $n = 737$ femmes aux États-Unis ; sur la diagonale, les histogrammes ; dans le triangle supérieur, les nuages de points des données brutes ; sur le triangle inférieur, les nuages de points des rangs standardisés	48
---	----

Chapitre 1

Mise en contexte, motivation et structure du mémoire

1.1 Mise en contexte

Supposons que l'on observe un phénomène qui met en scène deux ou plusieurs variables aléatoires. Une façon porteuse de modéliser leur interdépendance consiste à utiliser les copules. En effet, ces dernières permettent de caractériser complètement la structure de dépendance dans un vecteur de variables aléatoires. Beaucoup de travaux se sont attardés au développement de modèles de copules, ainsi qu'à l'inférence statistique pour ceux-ci. On se contentera ici de citer les excellentes monographies de Nelsen (2006), Mai & Scherer (2012) et Joe (2015), qui permettent d'avoir une vue d'ensemble sur les résultats mathématiques de base les plus importants.

Une caractéristique importante dans l'étude de la dépendance est de savoir si oui ou non, la force d'attraction entre deux ou plusieurs variables aléatoires est de nature symétrique ou asymétrique. Il existe plusieurs notions de symétrie en statistique. D'un

intérêt particulier pour ce travail est la symétrie radiale. En effet, on peut se demander si la population à partir de laquelle des données proviennent possède une structure de dépendance qui possède la symétrie radiale.

Pour modéliser la dépendance avec des modèles paramétriques de copules, un statisticien peut être intéressé à savoir si la copule est symétrique ou non afin de le guider vers les bonnes familles de modèles à considérer. Par exemple, si on établit que la structure de dépendance est symétrique, on pourra privilégier les copules normale, Student, Laplace, Frank et Plackett ; dans le cas contraire, on se tournerait vers des modèles comme les copules Khi-deux, Fisher et Clayton, ou encore un modèle dans la famille des copules à valeurs extrêmes. Ces modèles seront décrits en détails ultérieurement.

Récemment, il y a eu un intérêt pour le développement de tests statistiques formels afin de vérifier si la copule d'une population possède la symétrie radiale :

- (i) Genest & Nešlehová (2014) ont proposé des tests basés sur la copule empirique, mais dans le cas bivarié seulement ;
- (ii) Bahraoui & Quessy (2017) ont développé des tests basés sur la fonction caractéristique de copules, valides à la fois dans le cas bivarié et multidimensionnel.

1.2 Mesure de la dépendance et copules

Les différentes théories et applications des méthodes statistiques occupent une place prépondérante dans le domaine de la science. Elles sont principalement utilisées pour décrire et interpréter un jeu de données. Les variables portant sur l'analyse de ces données peuvent être sous différentes formes : qualitatives et quantitatives. Dans chacun des cas, l'étude de la liaison entre des variables a toujours été une préoccupation primordiale des chercheurs. Dès lors, par le biais des travaux de nombreux auteurs, différentes mesures de dépendance ont été élaborées, parmi lesquelles le coefficient

de corrélation linéaire. Ce dernier permet de quantifier le degré de liaison entre deux variables aléatoires quantitatives X et Y .

Formellement, le coefficient de corrélation de Pearson associé à la paire (X, Y) est

$$\rho_{(X,Y)}^{\text{Pe}} = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

En particulier, $\rho_{(X,Y)}^{\text{Pe}} = 0$ quand les variables X et Y sont indépendantes. L'implication inverse est généralement fausse, car il se peut que $\rho_{(X,Y)}^{\text{Pe}} = 0$ même si X et Y ne sont pas indépendantes.

Le coefficient $\rho_{(X,Y)}^{\text{Pe}}$ affiche d'autres limites. Notamment, il est généralement incapable de bien saisir le degré de dépendance entre deux variables dont le lien n'est pas linéaire. Dans ce contexte, d'autres mesures de dépendance ont été proposées. Les plus connues sont certainement le tau de Kendall et le coefficient de corrélation des rangs de Spearman. Quoi qu'il en soit, ces mesures ne donnent qu'une indication sommaire de la structure de dépendance qui unit deux, ou plusieurs, variables aléatoires. C'est pourquoi la notion de copule a été introduite pour modéliser de façon générale plusieurs variables aléatoires par une fonction de dépendance.

1.3 Utilisation de la copule de Fisher pour modéliser l'asymétrie radiale

La famille des copules de Fisher construite par Favre et al. (2018) comporte plusieurs avantages du point de vue de la modélisation de la dépendance. D'une part, elle permet une modélisation pour des dimensions arbitraires, tout en possédant un indice de dépendance extrême de queue non-nul. De surcroît, contrairement à plusieurs modèles de copules multidimensionnelles, la copule de Fisher possède la propriété d'asymétrie

radiale ; ce comportement est abondamment présent dans les données hydrologiques et climatologiques. Les données de précipitations conjointes, par exemple, montrent généralement des asymétries de la queue et une dépendance de la queue supérieure.

1.4 Objectifs et structure du mémoire

L'objectif principal de ce travail est d'étudier la puissance de plusieurs tests de symétrie radiale pour copules multidimensionnelles sous de la dépendance de type Fisher. De manière spécifique, il y a trois objectifs poursuivis par ce mémoire :

- (1) Développer une extension d -dimensionnelle aux tests de Genest & Nešlehová (2014) basés sur la copule empirique ;
- (2) Étudier la puissance des tests de symétrie radiale de Bahraoui & Quessy (2017) et Genest & Nešlehová (2014) dans le cas bivarié sous des alternatives de Fisher ;
- (3) Comparer nos nouveaux tests basés sur la copule empirique avec ceux de Bahraoui & Quessy (2017).

Le reste du mémoire est organisé comme suit. Le Chapitre 2 présente les structures de dépendance, la notion de symétrie radiale et les copules de Fisher. À cet effet, quelques définitions et propriétés des copules seront données ; ensuite, la notion de symétrie radiale sera étudiée et quelques exemples de modèles de copules symétriques et asymétriques seront présentés. Le Chapitre 3 développe l'extension d -dimensionnelle des tests de Genest & Nešlehová (2014). Le Chapitre 4 présente les résultats d'une étude de puissance des tests de symétrie radiale sous des contre-hypothèses de copules de Fisher. Le cœur du mémoire se termine par une brève conclusion. Quelques éléments techniques concernant le calcul de statistiques de tests sont dans l'Annexe A, alors que les codes Matlab utilisés pour les simulations ont été placés dans l'Annexe B.

Chapitre 2

Structures de dépendance, symétrie radiale et copules de Fisher

2.1 Qu'est-ce qu'une copule : quelques définitions et propriétés

Le concept de copule vient étymologiquement du mot latin *copulae* qui signifie lien, alliance, union ou liaison. Son origine remonte vers les années 1950 au cours du développement de la théorie des espaces métriques probabilistes, principalement en relation avec l'étude des familles d'opérations binaires sur les fonctions de répartition dans l'espace probabilisé. Toutefois, Sklar (1959) est le premier à introduire cette notion de *copule* dans son article intitulé *Fonctions de répartition à n dimensions et leurs marges*. Par la suite, plusieurs auteurs se sont intéressés à la théorie des copules, qui est devenue un outil de base dans la modélisation des distributions multivariées.

2.1.1 Définition analytique d'une copule et propriétés

Une copule à deux dimensions est une fonction $C : [0, 1]^2 \rightarrow [0, 1]$ possédant les propriétés suivantes (voir Nelsen (2006) pour plus de détails) :

- i) $C(u, 0) = 0$ et $C(0, v) = 0$ pour tout $u, v \in [0, 1]$;
- ii) $C(u, 1) = u$ et $C(1, v) = v$ pour tout $u, v \in [0, 1]$;
- iii) Pour tout $u_1, u_2, v_1, v_2 \in [0, 1]$ tels que $u_1 \leq u_2$ et $v_1 \leq v_2$, on a

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

La fonction définie par $C(u, v) = uv$ satisfait aux trois conditions précédentes : c'est donc une copule. D'abord, $C(u, 0) = u \times 0 = 0$ et $C(0, v) = 0 \times v = 0$, ce qui montre (i). Ensuite, $C(u, 1) = u \times 1 = u$ et $C(1, v) = 1 \times v = v$, ce qui montre (ii). Enfin, (iii) est satisfaite car pour $u_1 \leq u_2$ et $v_1 \leq v_2$, on a

$$\begin{aligned} C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) &= u_2 v_2 - u_2 v_1 - u_1 v_2 + u_1 v_1 \\ &= (u_2 - u_1)(v_2 - v_1) \\ &\geq 0. \end{aligned}$$

2.1.2 Théorème de Sklar

Le Théorème de Sklar (1959), énoncé dans la suite, est fondamental pour la modélisation de plusieurs variables aléatoires. Il permet de voir que la loi de (X, Y) est composée de la dépendance entre X et Y , ainsi que des comportements marginaux de X et de Y , représentés respectivement par F et G . À l'inverse, il permet aussi de construire un modèle avec une copule C désirée pour la dépendance, et possédant les marges univariées F et G voulues.

Théorème 2.1. Soit le couple (X, Y) de loi H dont les marges sont $F(x) = H(x, \infty)$

et $G(y) = H(\infty, y)$. Alors le Théorème de Sklar (1959) assure qu'il existe une copule $C : [0, 1]^2 \rightarrow [0, 1]$ telle que pour tout $(x, y) \in \mathbb{R}^2$,

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

Si les marges F et G sont continues, alors la copule C est unique.

La copule C correspond à la loi conjointe de $U = F(X)$ et de $V = G(Y)$. En effet,

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

2.1.3 Propriété d'invariance d'une copule

Une des propriétés fondamentales de la copule C est son invariance. En effet, lors d'une transformation monotone croissante, la copule reliant les distributions marginales est invariante. Pour démontrer cette propriété, posons $\tilde{X} = \vartheta(X)$ et $\tilde{Y} = \gamma(Y)$, où ϑ et γ sont deux fonctions monotones croissantes. On a alors

$$\begin{aligned} \tilde{H}(x, y) &= \mathbb{P}\{\vartheta(X) \leq x, \gamma(Y) \leq y\} \\ &= \mathbb{P}\{X \leq \vartheta^{-1}(x), Y \leq \gamma^{-1}(y)\} \\ &= H\{\vartheta^{-1}(x), \gamma^{-1}(y)\} \\ &= C\{F \circ \vartheta^{-1}(x), G \circ \gamma^{-1}(y)\}. \end{aligned}$$

Étant donné que les lois marginales de \tilde{X} et \tilde{Y} sont respectivement $F \circ \vartheta^{-1}$ et $G \circ \gamma^{-1}$, la copule de \tilde{H} est C . Cependant, si ϑ est monotone décroissante, la copule de (\tilde{X}, \tilde{Y})

n'est toutefois pas la même que celle de (X, Y) . Dans ce cas, on a plutôt

$$\begin{aligned}\tilde{H}(x, y) &= \mathbb{P}\{\vartheta(X) \leq x, \gamma(Y) \leq y\} \\ &= \mathbb{P}\{X > \vartheta^{-1}(x), Y > \gamma^{-1}(y)\} \\ &= G \circ \gamma^{-1}(y) - H\{\vartheta^{-1}(x), \gamma^{-1}(y)\} \\ &= G \circ \gamma^{-1}(y) - C\{F \circ \vartheta^{-1}(x), G \circ \gamma^{-1}(y)\}.\end{aligned}$$

Par conséquent, la copule de \tilde{H} est $\tilde{C}(u, v) = v - C(1 - u, v)$.

2.1.4 Extension multidimensionnelle

Le Théorème de Sklar se généralise au cadre multidimensionnel. Pour le décrire, soit un vecteur aléatoire à d dimensions $\mathbf{X} = (X_1, \dots, X_d)$ de loi H et de marges continues F_1, \dots, F_d . Alors pour tout $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, il existe une unique copule $C : [0, 1]^d \rightarrow [0, 1]$ telle que

$$H(\mathbf{x}) = C\{F_1(x_1), \dots, F_d(x_d)\}. \quad (2.1)$$

À l'instar du cas bivarié, la copule C correspond à la loi conjointe de $\mathbf{U} = (U_1, \dots, U_d)$, où $U_\ell = F_\ell(X_\ell)$ pour chaque $\ell \in \{1, \dots, d\}$. Ce théorème joue un rôle primordial dans la mesure où il met en exergue la structure de dépendance entre les composantes de \mathbf{X} , représentées par C , et les distributions marginales F_1, \dots, F_d .

En posant $u_\ell = F_\ell(x_\ell)$ dans l'Équation (2.1), on peut extraire l'unique copule d'une loi d -dimensionnelle continue H . On obtient alors

$$C(\mathbf{u}) = H\{F^{-1}(u_1), \dots, F^{-1}(u_d)\}.$$

Comme dans le cas bivarié, la notion d'invariance de la copule C peut être généralisée

dans le cadre multidimensionnel. En effet, soient des fonctions monotones croissantes $\alpha_1, \dots, \alpha_d : \mathbb{R} \rightarrow \mathbb{R}$ et posons $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_d)$, où $\tilde{X}_\ell = \alpha_\ell(X_\ell)$. On a alors

$$\begin{aligned}\tilde{H}(\mathbf{x}) &= \mathbb{P}\{\alpha_1(X_1) \leq x_1, \dots, \alpha_d(X_d) \leq x_d\} \\ &= \mathbb{P}\{X_1 \leq \alpha_{-1}^{-1}(x_1), \dots, X_d \leq \alpha_d^{-1}(x_d)\} \\ &= H\{\alpha_1^{-1}(x_1), \dots, \alpha_d^{-1}(x_d)\} \\ &= C\{F_1 \circ \alpha_1^{-1}(x_1), \dots, F_d \circ \alpha_d^{-1}(x_d)\}.\end{aligned}$$

On peut déduire que la copule de \tilde{H} est C , car les lois marginales de $\tilde{X}_1, \dots, \tilde{X}_d$ sont respectivement $F_1 \circ \alpha_1^{-1}, \dots, F_d \circ \alpha_d^{-1}$.

2.2 Le concept de symétrie radiale

Soit X , une variable aléatoire. La symétrie de X autour d'un point $c \in \mathbb{R}$ signifie que $X - c$ et $c - X$ ont la même loi. Cela se traduit par $\mathbb{P}(X - c \leq x) = \mathbb{P}(c - X \leq x)$, ce qui équivaut à $F(c + x) = 1 - F(c - x)$.

Dans le cas d'une loi bivariée, la notion de symétrie apparaît sous plusieurs formes : marginale, radiale et jointe. Ainsi, pour un couple aléatoire (X, Y) de loi H et de marges F et G , on définit trois types de symétrie autour de $(r_1, r_2) \in \mathbb{R}^2$:

- Symétrie marginale : F est symétrique autour de r_1 et G autour de r_2 ;
- Symétrie jointe : $(X - r_1, Y - r_2)$, $(X - r_1, r_2 - Y)$, $(r_1 - X, Y - r_2)$ et $(r_1 - X, r_2 - Y)$ ont la même loi ;
- Symétrie radiale : $(X - r_1, Y - r_2)$ et $(r_1 - X, r_2 - Y)$ ont la même loi.

Dans ce travail, on s'attardera plus particulièrement à la notion de symétrie radiale. On peut montrer que si (X, Y) est un couple aléatoire de loi H et de marges continues F et G qui sont symétriques, alors on a la symétrie radiale si et seulement si l'unique

copule C de H est telle que $C = C^*$, où

$$C^*(u, v) = u + v - 1 + C(1 - u, 1 - v)$$

est la copule de survie de C . En effet, comme $X - r_1$ a la même loi que $r_1 - X$ et que $Y - r_2$ a la même loi que $r_2 - Y$, alors $\mathbb{P}(X - r_1 \leq x, Y - r_2 \leq y) = \mathbb{P}(r_1 - X \leq x, r_2 - Y \leq y)$. Cette égalité se traduit par

$$H(x + r_1, y + r_2) = \bar{H}(r_1 - x, r_2 - y), \quad (2.2)$$

où $\bar{H}(x, y) = \mathbb{P}(X > x, Y > y)$ est la fonction de survie de H . Ensuite, en posant $\bar{F}(x) = 1 - F(x)$ et $\bar{G}(y) = 1 - G(y)$, on remarque que

$$\begin{aligned} \bar{H}(x, y) &= 1 - \mathbb{P}(X \leq x) - \mathbb{P}(Y \leq y) + \mathbb{P}(X \leq x, Y \leq y) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C\{1 - \bar{F}(x), 1 - \bar{G}(y)\} \\ &= C^*\{\bar{F}(x), \bar{G}(y)\}. \end{aligned}$$

On peut donc écrire l'Équation (2.2) sous la forme

$$C\{F(x + r_1), G(y + r_2)\} = C^*\{\bar{F}(r_1 - x), \bar{G}(r_2 - y)\}.$$

Comme X et Y sont symétriques, respectivement, autour de r_1 et de r_2 , on a $F(x + r_1) = \bar{F}(r_1 - x)$ et $G(y + r_2) = \bar{G}(r_2 - y)$, ce qui permet de conclure que $C = C^*$.

Cette notion se généralise au cas d'un vecteur $\mathbf{X} \in \mathbb{R}^d$. On dit que \mathbf{X} possède la propriété de symétrie radiale autour de $\mathbf{r} = (r_1, \dots, r_d) \in \mathbb{R}^d$ si et seulement si

- $F_\ell(x_\ell + r_\ell) = \bar{F}_\ell(r_\ell - x_\ell)$, c'est-à-dire que X_ℓ est symétrique autour de r_ℓ ;
- $C = C^*$, où C^* est la copule de survie de C , qui correspond à la loi de $\mathbf{1} - \mathbf{U} = (1 - U_1, \dots, 1 - U_d)$ pour $\mathbf{U} \sim C$.

Cette deuxième propriété veut dire que la copule C possède la propriété de symétrie radiale au niveau de la dépendance. On a alors $\mathbb{P}(\mathbf{1} - \mathbf{U} \leq \mathbf{u}) = \mathbb{P}(\mathbf{U} \leq \mathbf{u})$.

2.3 Quelques modèles de copules

2.3.1 Copule d'indépendance

On sait que deux variables aléatoires $X \sim F$ et $Y \sim G$ sont indépendantes si et seulement si $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \times \mathbb{P}(Y \leq y) = F(x) \times G(y)$ pour tout $(x, y) \in \mathbb{R}^2$. D'après le Théorème de Sklar, si les marges F et G sont continues, alors la copule correspondante est

$$C_{\Pi}(u, v) = F \circ F^{-1}(u) \times G \circ G^{-1}(v) = uv.$$

La fonction C_{Π} est appelée la copule d'indépendance. Dans le cas d -dimensionnel, des variables X_1, \dots, X_d de lois F_1, \dots, F_d sont indépendantes si et seulement si, pour tout $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$\mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \prod_{\ell=1}^d F_{\ell}(x_{\ell}).$$

On déduit alors que la copule d'indépendance multidimensionnelle est donnée pour $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$ par $C_{\Pi}(\mathbf{u}) = u_1 \times \dots \times u_d$. Cette copule possède la propriété de symétrie radiale. En effet, si $\mathbf{U} = (U_1, \dots, U_d) \sim C_{\Pi}$, alors

$$\begin{aligned} \mathbb{P}(\mathbf{1} - \mathbf{U} \leq \mathbf{u}) &= \mathbb{P}(\mathbf{U} \geq \mathbf{1} - \mathbf{u}) \\ &= \prod_{\ell=1}^d \mathbb{P}(U_{\ell} \geq 1 - u_{\ell}) \\ &= u_1 \times \dots \times u_d \\ &= \mathbb{P}(\mathbf{U} \leq \mathbf{u}). \end{aligned}$$

La Figure 2.1 montre le nuage de points de 2000 paires simulées de la copule C_{Π} .

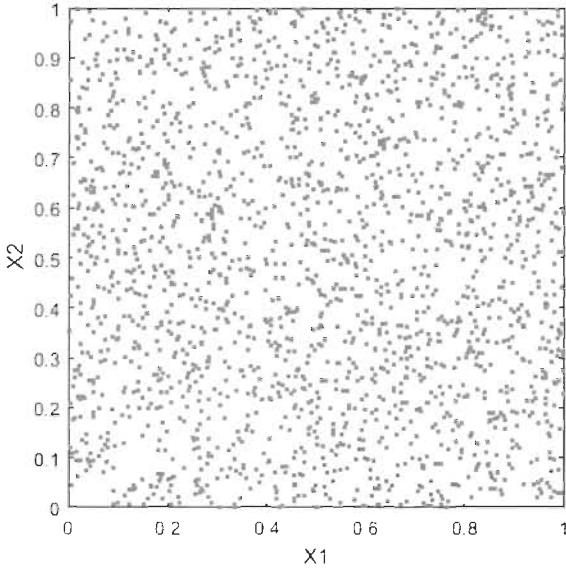


FIGURE 2.1 – Nuage de points de 2000 paires simulées à partir de la copule d’indépendance

2.3.2 Copule Normale

Une copule appartient à la famille des copules Normales si elle est extraite d’une distribution gaussienne multivariée. Les copules Normales sont très utilisées pour la construction d’un modèle dont la structure de dépendance est héritée du modèle normal et dont les lois marginales sont telles que désirées. Rappelons que la fonction de répartition de la loi Normale standard s’écrit implicitement

$$\Phi_{\Sigma}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \phi_{\Sigma}(s_1, \dots, s_d) ds_1 \cdots ds_d,$$

où ϕ_{Σ} est la densité de la loi Normale d -dimensionnelle de matrice de corrélation Σ . Étant donné que les marges sont toutes Φ , la copule Normale peut alors s’écrire

$$C_{\Sigma}^N(\mathbf{u}) = \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_d)} \phi_{\Sigma}(s_1, \dots, s_d) ds_1 \cdots ds_d.$$

La copule Normale possède la propriété de symétrie radiale. Parmi les nombreux travaux qui se sont intéressés à la copule Normale, on retrouve entre autres Klaassen

& Wellner (1997), qui traitent de l'inférence optimale dans ces modèles, ainsi que Meyer (2013), qui énonce plusieurs propriétés analytiques intéressantes.

La Figure 2.2 montre un nuage de points de 1000 paires simulées de la copule C_{Σ}^N lorsque le niveau de dépendance, tel que mesuré par le tau de Kendall, est 0,6.

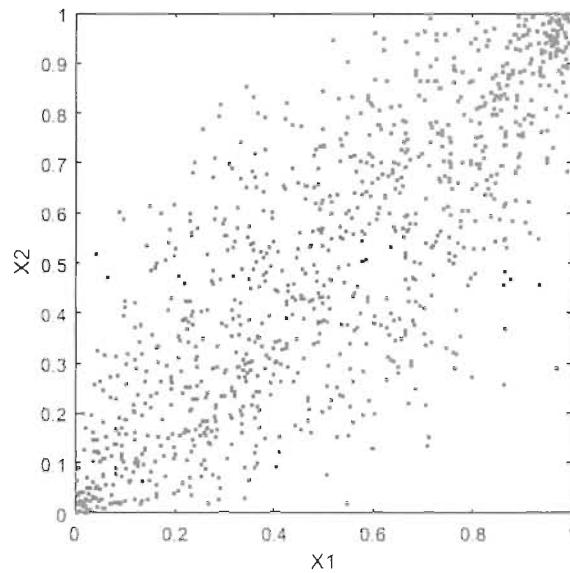


FIGURE 2.2 – Nuage de points de 1 000 paires simulées à partir de la copule Normale avec un tau de Kendall de 0,6

2.3.3 Copules Archimédiennes

Les copules Archimédiennes constituent une classe très importante de copules, principalement en raison de la facilité avec laquelle on peut les construire, du nombre important de familles qui appartiennent à celles-ci et des propriétés intéressantes qu'elles possèdent (voir Nelsen (2006)). Considérons une fonction $\varphi : [0, 1] \rightarrow [0, \infty)$ qui est décroissante, convexe et telle que $\varphi(1) = 0$. Selon Genest & MacKay (1986), une copule est Archimédienne si elle admet une écriture de la forme

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

La copule C_φ possède les propriétés suivantes :

- (i) C_φ est symétrique au sens où $C_\varphi(u, v) = C_\varphi(v, u)$;
- (ii) C_φ est associative, c'est-à-dire que pour tout $u, v, w \in [0, 1]$,

$$C_\varphi \{u, C_\varphi(v, w)\} = C_\varphi \{C_\varphi(u, v), w\}.$$

Dans la version multidimensionnelle, on suppose que le générateur φ est complètement monotone. En d'autres termes, pour tout $j \in \{1, \dots, d\}$,

$$(-1)^j \frac{d^j}{dt^j} \varphi^{-1}(t) > 0.$$

La copule Archimédienne d -dimensionnelle de générateur φ est alors définie par

$$C_\varphi(\mathbf{u}) = \varphi^{-1} \left\{ \sum_{l=1}^d \varphi(u_l) \right\}.$$

La copule de Clayton est Archimédienne dont le générateur est défini pour $\theta > 0$ par

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

Dans le cas $d = 2$, sa copule s'écrit sous la forme

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

La copule de Gumbel est aussi Archimédienne ; son générateur est défini pour $\theta \in [0, 1)$ par $\varphi_\theta(t) = |\ln t|^{1/(1-\theta)}$, ce qui permet de montrer que sa copule bivariée est

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

Enfin, la copule de Frank est générée par

$$\varphi_\theta(t) = \ln \left(\frac{1 - e^{-\theta}}{1 - e^{-\theta t}} \right), \quad \theta \in \mathbb{R} \setminus \{0\}.$$

Sa copule bivariée est

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

Les copules de Clayton et de Gumbel sont asymétriques du point de vu radial, alors que la copule de Frank possède la propriété de symétrie radiale. En fait, dans la famille des copules Archimédiennes, seule la copule de Frank est symétrique. On peut le constater en examinant les Figures 2.3, 2.4 et 2.5, qui montrent respectivement des nuages de points de 1000 paires simulées des copules Clayton, Gumbel et Frank pour un niveau de dépendance, tel que mesuré par le tau de Kendall, de 0,6.

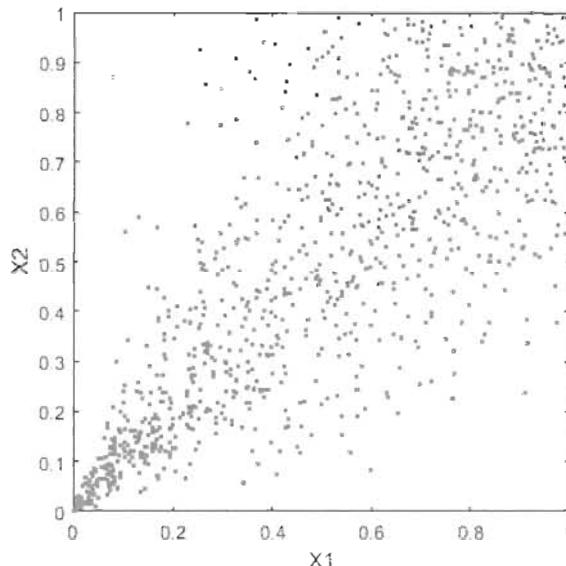


FIGURE 2.3 – Nuage de points de 1000 paires simulées à partir de la copule de Clayton avec un tau de Kendall de 0,6

Parmi un grand nombre de travaux portant sur les copules Archimédiennes, on mentionnera Nelsen (1997) pour les propriétés de mesures de dépendance, McNeil & Nešlehová (2009) qui découvrent une représentation alternative et intéressante de ces modèles, ainsi que Genest et al. (2011), qui mettent au point des techniques d’inférence.

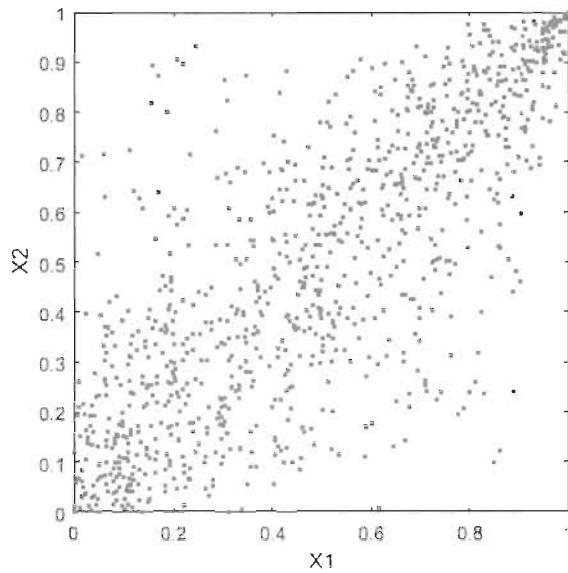


FIGURE 2.4 – Nuage de points de 1000 paires simulées à partir de la copule de Gumbel avec un tau de Kendall de 0,6

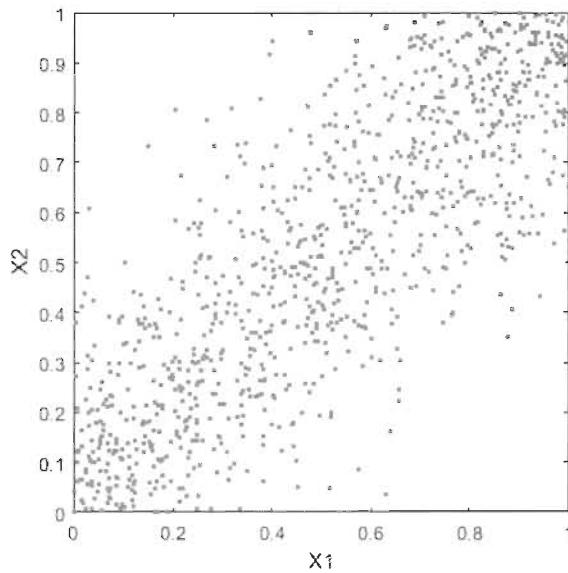


FIGURE 2.5 – Nuage de points de 1000 paires simulées à partir de la copule de Frank avec un tau de Kendall de 0,6

2.3.4 Copules de Plackett

La copule de Plackett, issue d'un modèle décrit initialement par Plackett (1965), s'écrit pour $\theta \geq 1$ par

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

où $L_\theta(u, v) = 1 + (\theta - 1)(u + v)$. Cette copule possède la propriété de symétrie radiale. Une extension multidimensionnelle de cette copule a été proposée par Molenberghs & Lesaffre (1994), mais son usage demeure limité à cause des nombreuses contraintes sur ses paramètres. La Figure 2.6 montre un nuage de points de 1000 paires simulées de la copule bivariée C_θ^{Pl} dont le tau de Kendall est fixé à $\tau_C = 0,6$.

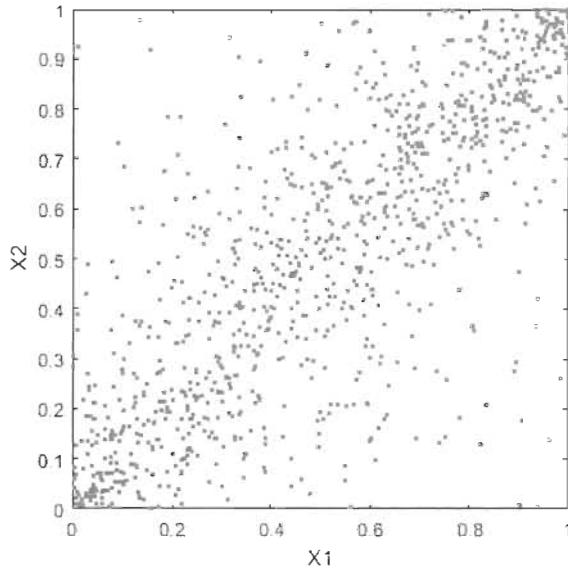


FIGURE 2.6 – Nuage de points de 1000 paires simulées à partir de la copule de Plackett avec un tau de Kendall de 0,6

2.3.5 Copules de Student

Un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$ suit une distribution de Student à d dimensions avec une matrice de corrélation Σ et $\nu \in \mathbb{N}$ degrés de liberté s'il admet la représentation

$$X = \frac{\mathbf{Z}}{\sqrt{W/\nu}},$$

où $\mathbf{Z} = (Z_1, \dots, Z_d) \sim \Phi_\Sigma$ et W , qui est indépendante de \mathbf{Z} , est distribuée selon la loi Khi-deux à ν degrés de liberté. La densité de \mathbf{X} est alors donnée pour tout

$\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ par

$$h_{\Sigma, \nu}^T(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{d/2}} |\Sigma|^{-1/2} \left(1 + \frac{\mathbf{x}\Sigma^{-1}\mathbf{x}^\top}{\nu}\right)^{-(\nu+d)/2},$$

où $\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx$ est la fonction gamma. La copule de Student est alors

$$C_{\Sigma, \nu}^T(\mathbf{u}) = \int_{-\infty}^{F_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{F_\nu^{-1}(u_d)} h_{\Sigma, \nu}^T(x_1, \dots, x_d) dx_d \cdots dx_1,$$

où F_ν est la fonction de répartition de la Student univariée à ν degrés de liberté. La copule de Student a la propriété de symétrie radiale. Ainsi, les queues inférieure et supérieure de la densité de la copule de Student sont identiques. Cette caractéristique des structures de dépendance de Student peut être limitative dans la pratique, en particulier dans le cas des données de précipitations. La Figure 2.7 montre un nuage de 1000 paires simulées de la copule de Student $C_{\Sigma, \nu}^T$ avec un tau de Kendall de 0,6.

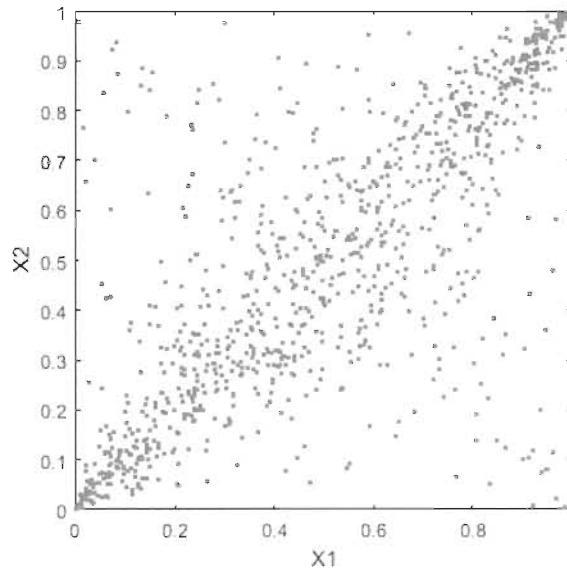


FIGURE 2.7 – Nuage de points de 1000 paires simulées à partir de la copule de Student à 1 degré de liberté avec un tau de Kendall de 0,6

2.3.6 Copules Khi-deux

Les copules Normale et Student souffrent de certaines limitations en raison du type de structures de dépendance qu'elles peuvent modéliser. De plus, toutes les marges à deux variables du modèle Normal ont une structure échangeable. Par conséquent, ces copules sont incapables de modéliser des structures de dépendance asymétriques, c'est-à-dire non échangeables. La famille des copules Khi-deux est une alternative aux structures de dépendance normales qui ne souffrent pas des limitations susmentionnées. Ces modèles permettent l'asymétrie de queue et le caractère non échangeable, tout en conservant les avantages offerts par les copules Normale et Student.

Pour $\mathbf{Z} = (Z_1, \dots, Z_d) \sim \Phi_\Sigma$, la copule Khi-deux à d dimensions de paramètre de décentralité $a \geq 0$, notée $C_{\Sigma,a}^\chi$, est définie comme la structure de dépendance du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d) = ((Z_1 + a)^2, \dots, (Z_d + a)^2)$. Il a été montré par Quessy et al. (2016) que

$$C_{\Sigma,a}^\chi(\mathbf{u}) = \sum_{\epsilon=(\epsilon_1, \dots, \epsilon_d) \in \{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j \right) \Phi_\Sigma \{ h_a(\epsilon_1 u_1), \dots, h_a(\epsilon_d u_d) \},$$

où pour $G_a(x) = \Phi(\sqrt{x} - a) + \Phi(\sqrt{x} + a) - 1$, $h_a(u) = \text{signe}(u) \sqrt{G_a^{-1}(|u|)} - a$. Une autre expression en fonction de la copule Normale C_Σ^N est

$$C_{\Sigma,a}^\chi(\mathbf{u}) = \sum_{\epsilon=(\epsilon_1, \dots, \epsilon_d) \in \{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j \right) C_\Sigma^N \left(\frac{1 + \epsilon_1 u_1}{2}, \dots, \frac{1 + \epsilon_d u_d}{2} \right).$$

La Figure 2.8 montre un nuage de 1000 paires simulées de la copule $C_{\Sigma,a}^\chi$.

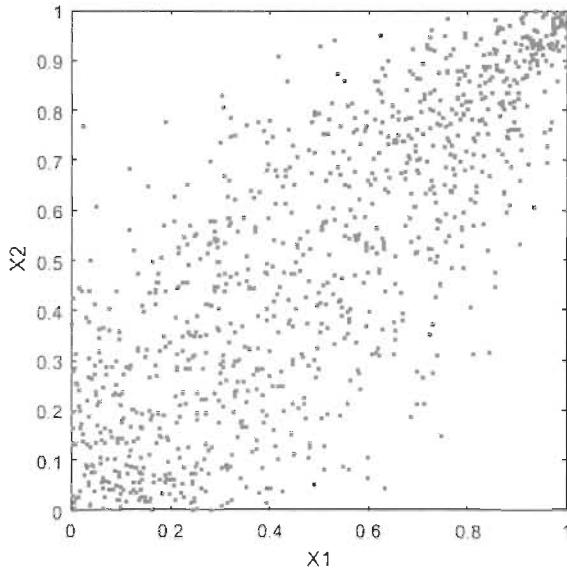


FIGURE 2.8 – Nuage de points de 1000 paires simulées à partir de la copule Khi-deux avec $a = 0$ et un tau de Kendall de 0,6

2.3.7 Copules de Fisher

Les données de précipitations conjointes montrent généralement des asymétries de queue et une dépendance caudale supérieure. Les modèles de copules couramment utilisés en grandes dimensions, tel la Normale, Student et autres copules elliptiques, ne parviennent pas à capturer ces attributs. Afin de combler cette lacune, la nouvelle famille des copules Fisher a été introduite par Favre et al. (2018). Ce modèle de dépendance est asymétrique dans la queue et permet une dépendance caudale supérieure. Pour la construire, soit $\mathbf{X} = (X_1, \dots, X_d)$ de loi de Student d -dimensionnelle de matrice de corrélation Σ et $\nu \in \mathbb{N}$ degrés de liberté. Alors la copule de Fisher correspond à la structure de dépendance du vecteur (X_1^2, \dots, X_d^2) . Il a été montré par Favre et al. (2018) qu'une expression pour la copule de Fisher est

$$C_{\Sigma, \nu}^F(\mathbf{u}) = \sum_{\epsilon=(\epsilon_1, \dots, \epsilon_d) \in \{-1, 1\}^d} \left(\prod_{j=1}^d \epsilon_j \right) C_{\Sigma, \nu}^T \left(\frac{1 + \epsilon_1 u_1}{2}, \dots, \frac{1 + \epsilon_d u_d}{2} \right).$$

Une extension générale de ces modèles à la famille des copules dites *squared* a été proposée par Quessy & Durocher (2019). La Figure 2.9 montre un nuage de 1000 paires simulées de la copule $C_{\Sigma,\nu}^F$ lorsque $\tau_C = 0,6$.

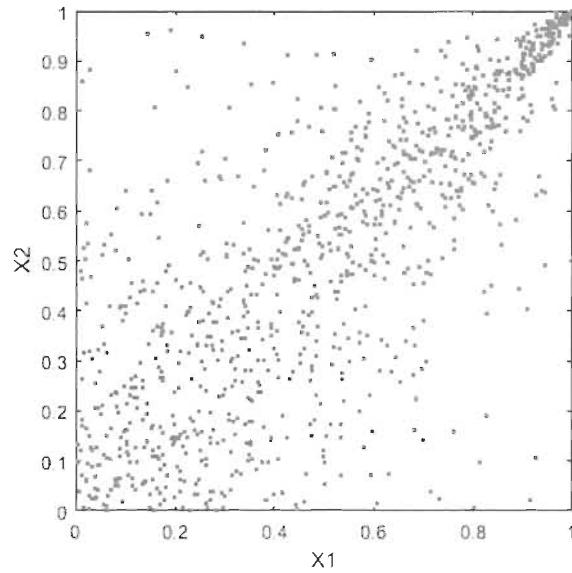


FIGURE 2.9 – Nuage de points de 1000 paires simulées à partir de la copule de Fisher à 1 degré de liberté avec un tau de Kendall de 0,6

Chapitre 3

Extension d -dimensionnelle des tests de Genest et Nešlehová

3.1 Tests de Genest et Nešlehová pour la symétrie bivariée

Les tests de Genest et Nešlehová sont utilisés pour détecter la présence ou non de symétrie radiale dans la structure de dépendance de données bivariées. Ces tests sont très importants dans le processus de modélisation afin d'éclairer le choix vers les bonnes familles de modèles à considérer. Pour un couple de variables aléatoires (X, Y) de marges continues et de copule C , l'hypothèse nulle de symétrie radiale est

$$\mathbb{H}_0 : C(u, v) = C^*(u, v), \text{ pour tout } (u, v) \in (0, 1)^2,$$

où $C^*(u, v) = u + v - 1 + C(1 - u, 1 - v)$ est la copule de survie associée à C .

3.1.1 Copule empirique bivariée et copule de survie empirique

Soit à nouveau un couple (X, Y) de loi H et de marges continues F et G . On rappelle que d'après le Théorème de Sklar (1959), il existe une unique copule $C : [0, 1]^2 \rightarrow [0, 1]$ telle que pour tout $(x, y) \in \mathbb{R}^2$, on $H(x, y) = C\{F(x), G(y)\}$. Estimer la copule C est une tâche compliquée, car une copule n'est pas directement observable à partir d'un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ i.i.d. de loi H . Pour ce faire, soit d'abord l'estimation sans biais de H donnée par la fonction de répartition empirique conjointe

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

Soient également les marges empiriques

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

L'estimateur de C proposé par Deheuvels (1979) est un estimateur de type *plug-in* basé sur $\tilde{C}(u, v) = H\{F^{-1}(u), G^{-1}(v)\}$, à savoir

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

Une autre manière de l'écrire est

$$\tilde{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq F_n^{-1}(u), Y_i \leq G_n^{-1}(v)\}.$$

Ce qui est communément appelé la *copule empirique* est une version asymptotiquement équivalente à \tilde{C}_n , c'est-à-dire

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

Il s'agit de la fonction de répartition conjointe des paires de *pseudo-observations* $(\hat{U}_1, \hat{V}_1), \dots, (\hat{U}_n, \hat{V}_n)$, où $\hat{U}_i = F_n(X_i)$ et $\hat{V}_i = G_n(Y_i)$ pour chaque $i \in \{1, \dots, n\}$.

Enfin, puisque la copule de survie C^* associée à C correspond à la loi conjointe de $(1 - U, 1 - V)$ pour $(U, V) \sim C$, on l'estime avec

$$C_n^*(u, v) = \frac{1}{n} \sum_{i=1}^n I\left(1 - \widehat{U}_i \leq u, 1 - \widehat{V}_i \leq v\right).$$

On remarque que

$$\begin{aligned} C_n^*(u, v) &= \frac{1}{n} \sum_{i=1}^n I\left(\widehat{U}_i \geq 1 - u, \widehat{V}_i \geq 1 - v\right) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ 1 - I\left(\widehat{U}_i < 1 - u\right) - I\left(\widehat{V}_i < 1 - v\right) + I\left(\widehat{U}_i < 1 - u, \widehat{V}_i < 1 - v\right) \right\} \\ &\approx 1 - (1 - u) - (1 - v) + \frac{1}{n} \sum_{i=1}^n I\left(\widehat{U}_i < 1 - u, \widehat{V}_i < 1 - v\right) \\ &= u + v - 1 + C_n(1 - u, 1 - v). \end{aligned} \tag{3.1}$$

3.1.2 Convergence faible de la copule empirique

Soit d'abord la notion de copule régulière définie ci-dessous.

Définition 3.1. Une copule C bivariée est dite régulière si les dérivées partielles $\dot{C}_1(u, v) = \partial C(u, v)/\partial u$ et $\dot{C}_2(u, v) = \partial C(u, v)/\partial v$ existent sur $[0, 1]^2$, et que \dot{C}_1 est continue sur $(0, 1) \times [0, 1]$ et \dot{C}_2 est continue sur $[0, 1] \times (0, 1)$.

Quand n tend vers l'infini et en autant que la copule C est régulière, Segers (2012) a montré que le processus de copule empirique $\mathbb{C}_n = \sqrt{n}(C_n - C)$ converge faiblement dans l'espace $\ell^\infty([0, 1]^2)$ des fonctions bornées définies sur $[0, 1]^2$ vers une limite gaussienne de représentation

$$\mathbb{C}(u, v) = \mathbb{B}_C(u, v) - \dot{C}_1(u, v) \mathbb{B}_C(u, 1) - \dot{C}_2(u, v) \mathbb{B}_C(1, v), \tag{3.2}$$

où \mathbb{B}_C est un processus Gaussian centré tel que

$$\text{cov} \{\mathbb{B}_C(u, v), \mathbb{B}_C(u', v')\} = C \{\min(u, u'), \min(v, v')\} - C(u, v) C(u', v').$$

En fait, \mathbb{B}_C correspond à la limite faible du processus empirique

$$\mathbb{B}_n(u, v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\mathbb{I}\{F(X_i) \leq u, G(Y_i) \leq v\} - C(u, v)\}.$$

Pour ce qui est de la copule de survie empirique C_n^* , l'Équation (3.1) permet d'écrire le processus de copule de survie $\mathbb{C}_n^* = \sqrt{n}(C_n^* - C^*)$ sous la forme

$$\mathbb{C}_n^*(u, v) = \sqrt{n} \{(C_n(1-u, 1-v) - C(1-u, 1-v))\} = \mathbb{C}_n(1-u, 1-v).$$

Par conséquent, \mathbb{C}_n^* converge faiblement vers

$$\begin{aligned} \mathbb{C}^*(u, v) &= \mathbb{C}(1-u, 1-v) \\ &= \mathbb{B}_C(1-u, 1-v) - \dot{C}_1(1-u, 1-v) \mathbb{B}_C(1-u, 1) \\ &\quad - \dot{C}_2(1-u, 1-v) \mathbb{B}_C(1, 1-v). \end{aligned} \tag{3.3}$$

3.1.3 Version moltiplicateur du processus de copule

Pour un vecteur $\xi = (\xi_1, \dots, \xi_n)$ de variables aléatoires indépendantes non négatives avec moyenne et variance unitaire, la version moltiplicateur de \mathbb{B}_n est

$$\widehat{\mathbb{B}}_n(u, v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\xi_i}{\bar{\xi}} - 1 \right) \mathbb{I}\{F_n(X_i) \leq u, G_n(Y_i) \leq v\}.$$

La version moltiplicateur de \mathbb{C}_n basée sur la représentation (3.2) est alors

$$\widehat{\mathbb{C}}_n(u, v) = \widehat{\mathbb{B}}_n(u, v) - \widehat{\dot{C}}_1(u, v) \widehat{\mathbb{B}}_n(u, 1) - \widehat{\dot{C}}_2(u, v) \widehat{\mathbb{B}}_n(1, v),$$

où $\widehat{\dot{C}_1}$, $\widehat{\dot{C}_2}$ sont des estimateurs convergents des dérivées partielles \dot{C}_1 , \dot{C}_2 . Généralement, on utilise les estimateurs à différences finies

$$\widehat{\dot{C}_1}(u, v) = \frac{C_n(u + \ell_n, v) - C_n(u - \ell_n, v)}{2\ell_n}$$

et

$$\widehat{\dot{C}_2}(u, v) = \frac{C_n(u, v + \ell_n) - C_n(u, v - \ell_n)}{2\ell_n},$$

où $\ell_n = h/\sqrt{n}$ pour $h \in \mathbb{N}$. Il a été montré par Segers (2012) que $(\mathbb{C}_n, \widehat{\mathbb{C}}_n)$ converge faiblement vers $(\mathbb{C}, \widehat{\mathbb{C}})$, où $\widehat{\mathbb{C}}$ est une copie indépendante de \mathbb{C} . Autrement dit, la version multiplicateur $\widehat{\mathbb{C}}_n$ parvient, asymptotiquement, à reproduire le comportement de \mathbb{C}_n . Enfin, la version multiplicateur de \mathbb{C}_n^* basée sur la représentation (3.3) est

$$\begin{aligned} \widehat{\mathbb{C}}_n^*(u, v) &= \widehat{\mathbb{B}}_n(1-u, 1-v) - \widehat{\dot{C}_1}(1-u, 1-v)\widehat{\mathbb{B}}_n(1-u, 1) \\ &\quad - \widehat{\dot{C}_2}(1-u, 1-v)\widehat{\mathbb{B}}_n(1, 1-v). \end{aligned}$$

3.1.4 Statistiques de test et leurs versions multiplicateur

Pour confronter les hypothèses $\mathbb{H}_0 : C = C^*$ et $\mathbb{H}_1 : C \neq C^*$, Genest & Nešlehová (2014) ont proposé les statistiques de test

$$\begin{aligned} R_n &= n \int_{[0,1]^2} \{C_n(u, v) - C_n^*(u, v)\}^2 du dv, \\ S_n &= n \int_{[0,1]^2} \{C_n(u, v) - C_n^*(u, v)\}^2 dC_n(u, v), \\ T_n &= \sqrt{n} \sup_{(u,v) \in [0,1]^2} |C_n(u, v) - C_n^*(u, v)|. \end{aligned}$$

Des formules explicites pour ces statistiques sont données dans l'Annexe A.1. En rappelant que $\mathbb{C}_n = \sqrt{n}(C_n - C)$ converge en loi vers la limite \mathbb{C} définie à l'Équation (3.2) et que $\mathbb{C}_n^* = \sqrt{n}(C_n^* - C^*)$ converge en loi vers la limite \mathbb{C}^* définie à l'Équation (3.3),

on obtient sous \mathbb{H}_0 que

$$\begin{aligned} R_n &= \int_{[0,1]^2} \{\mathbb{C}_n(u,v) - \mathbb{C}_n^*(u,v)\}^2 du dv, \\ S_n &= \int_{[0,1]^2} \{\mathbb{C}_n(u,v) - \mathbb{C}_n^*(u,v)\}^2 dC_n(u,v), \\ T_n &= \sup_{(u,v) \in [0,1]^2} |\mathbb{C}_n(u,v) - \mathbb{C}_n^*(u,v)|. \end{aligned}$$

Ceci permet alors de déduire que

$$\begin{aligned} R_n \rightsquigarrow \mathbb{R} &= \int_{[0,1]^2} \{\mathbb{C}(u,v) - \mathbb{C}^*(u,v)\}^2 du dv, \\ S_n \rightsquigarrow \mathbb{S} &= \int_{[0,1]^2} \{\mathbb{C}(u,v) - \mathbb{C}^*(u,v)\}^2 dC(u,v), \\ T_n \rightsquigarrow \mathbb{T} &= \sup_{(u,v) \in [0,1]^2} |\mathbb{C}(u,v) - \mathbb{C}^*(u,v)|, \end{aligned}$$

où la notation \rightsquigarrow est synonyme de “convergence en loi”. Les versions multiplicativeurs des statistiques R_n , S_n et T_n sont simplement

$$\begin{aligned} \widehat{R}_n &= \int_{[0,1]^2} \left\{ \widehat{\mathbb{C}}_n(u,v) - \widehat{\mathbb{C}}_n^*(u,v) \right\}^2 du dv, \\ \widehat{S}_n &= \int_{[0,1]^2} \left\{ \widehat{\mathbb{C}}_n(u,v) - \widehat{\mathbb{C}}_n^*(u,v) \right\}^2 dC_n(u,v), \\ \widehat{T}_n &= \sup_{(u,v) \in [0,1]^2} \left| \widehat{\mathbb{C}}_n(u,v) - \widehat{\mathbb{C}}_n^*(u,v) \right|. \end{aligned}$$

3.1.5 Algorithme pour performer les tests et calculer leurs p-valeurs

Pour tester la symétrie radiale d’une copule, il faut effectuer l’algorithme suivant :

- (i) Calculer les statistiques de test R_n , S_n ou T_n ;
- (ii) Définir, pour chaque $h \in \{1, \dots, M\}$,

$$\bar{E}_n^{(h)}(u,v) = \frac{1}{\sqrt{n}} \Xi_n^{(h)} \left\{ P_n(u,v) - \widehat{C}_1(u,v)P_n(u,1) - \widehat{C}_2(u,v)P_n(1,v) \right\},$$

où pour $\xi^{(h)} = (\xi_1^{(h)}, \dots, \xi_n^{(h)})$ un vecteur de variables multiplicateurs,

$$\Xi_n^{(h)} = \left(\frac{\xi_1^{(h)}}{\xi_n^{(h)}} - 1, \dots, \frac{\xi_n^{(h)}}{\xi_n^{(h)}} - 1 \right),$$

alors que $P_n = (P_{1n}, \dots, P_{nn})^\top$, avec

$$P_{in}(u, v) = I(\widehat{U}_i \leq u, \widehat{V}_i \leq v) - I(\widehat{U}_i > 1-u, \widehat{V}_i > 1-v).$$

(iii) Calculer les versions multiplicateurs des statistiques de test via

$$\begin{aligned} \widehat{R}_n^{(h)} &= \int_0^1 \int_0^1 \{ \bar{E}_n^{(h)}(u, v) \}^2 \, du \, dv \approx \frac{1}{N^2} \sum_{k, \ell=1}^N \left\{ \bar{E}_n^{(h)} \left(\frac{k}{N}, \frac{\ell}{N} \right) \right\}^2, \\ \widehat{S}_n^{(h)} &= \int_0^1 \int_0^1 \{ \bar{E}_n^{(h)}(u, v) \}^2 \, dC_n(u, v), \\ \widehat{T}_n^{(h)} &= \sup_{(u, v) \in (0, 1)^2} |\bar{E}_n^{(h)}(u, v)| \approx \max_{(k, \ell) \in \{1, \dots, N\}^2} \left| \bar{E}_n^{(h)} \left(\frac{k}{N}, \frac{\ell}{N} \right) \right|; \end{aligned}$$

(iv) Estimer les p-valeurs des tests avec

$$\begin{aligned} \widehat{PV}_R &= \frac{1}{M} \sum_{h=1}^M I \left(\widehat{R}_n^{(h)} > R_n \right), \\ \widehat{PV}_S &= \frac{1}{M} \sum_{h=1}^M I \left(\widehat{S}_n^{(h)} > S_n \right), \\ \widehat{PV}_T &= \frac{1}{M} \sum_{h=1}^M I \left(\widehat{T}_n^{(h)} > T_n \right). \end{aligned}$$

Une fois cela effectué, la décision concernant l'acceptation ou le rejet de l'hypothèse nulle est basée sur la p-valeur estimée du test. Ainsi, si $\widehat{PV} < \alpha$, alors H_0 est rejetée au seuil de signification α .

3.2 Versions d -dimensionnelles des tests de Genest & Nešlehová (2014)

Dans cette section, les tests de Genest & Nešlehová (2014) permettant de détecter la présence de symétrie radiale d'une copule seront généralisées dans le cadre multidimensionnel. À notre connaissance, une telle extension n'a pas encore été développée dans la littérature. À cette fin, soit un vecteur de variables aléatoires (X_1, \dots, X_d) dans \mathbb{R}^d dont les marges sont continues et l'unique copule est C . L'hypothèse nulle de symétrie radiale dans ce cas est

$$\mathbb{H}_0 : C(\mathbf{u}) = C^*(\mathbf{u}), \text{ pour tout } \mathbf{u} = (u_1, \dots, u_d) \in (0, 1)^d,$$

où C^* est la copule de survie multivariée associée à C . On rappelle que la copule C^* correspond à la loi conjointe de $\mathbf{1} - \mathbf{U} = (1 - U_1, \dots, 1 - U_d)$ pour $(U_1, \dots, U_d) \sim C$.

3.2.1 Copule empirique multivariée et sa version survie

Considérons à nouveau un vecteur $\mathbf{X} = (X_1, \dots, X_d)$ de loi H et de marges continues F_1, \dots, F_d dont on souhaite estimer l'unique copule C . On rappelle que du Théorème de Sklar (1959),

$$H(\mathbf{x}) = C \{F_1(x_1), \dots, F_d(x_d)\}.$$

Étant donné qu'il serait difficile d'estimer directement la copule C , soit d'abord l'estimation sans biais de H donnée par la fonction de répartition empirique conjointe

$$H_n(x_1, \dots, x_d) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(X_{i1} \leq x_1, \dots, X_{id} \leq x_d).$$

Soit également les marges empiriques

$$F_{n\ell}(x_\ell) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(X_{i\ell} \leq x_\ell), \quad \ell \in \{1, \dots, d\}.$$

À l'instar du cas bivarié, soit l'estimateur

$$\begin{aligned} \tilde{C}_n(\mathbf{u}) &= H_n \left\{ F_{n1}^{-1}(u_1), \dots, F_{nd}^{-1}(u_d) \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{I} \left\{ X_{i1} \leq F_{n1}^{-1}(u_1), \dots, X_{id} \leq F_{nd}^{-1}(u_d) \right\}. \end{aligned}$$

Puisque lorsque la taille d'échantillon n est grande, l'inégalité $X_{il} \leq F_{nl}^{-1}(u_\ell)$ est à peu près équivalente à $F_{n\ell}(X_{il}) \leq u_\ell$, l'estimateur \tilde{C}_n est asymptotiquement équivalent à

$$\begin{aligned} C_n(\mathbf{u}) &= \frac{1}{n} \sum_{i=1}^n \mathbf{I} \left\{ F_{n1}(X_{i1}) \leq u_1, \dots, F_{nd}(X_{id}) \leq u_d \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{I} \left(\widehat{U}_{i1} \leq u_1, \dots, \widehat{U}_{id} \leq u_d \right). \end{aligned}$$

La fonction empirique C_n est appelée couramment la copule empirique d -dimensionnelle. Elle est calculée à partir des vecteurs de pseudo-observations $\widehat{\mathbf{U}}_1, \dots, \widehat{\mathbf{U}}_n$, où pour chaque $j \in \{1, \dots, n\}$,

$$\widehat{\mathbf{U}}_j = (F_{n1}(X_{i1}), \dots, F_{nd}(X_{id})).$$

Parallèlement, puisque la copule de survie multivariée C^* associée à C correspond à la loi conjointe de $\mathbf{1} - \mathbf{U} = (1 - U_1, \dots, 1 - U_d)$ pour $(U_1, \dots, U_d) \sim C$, on l'estime avec la copule empirique d -dimensionnelle de survie, à savoir

$$\begin{aligned} C_n^*(\mathbf{u}) &= \frac{1}{n} \sum_{i=1}^n \mathbf{I} \left(1 - \widehat{U}_{i1} \leq u_1, \dots, 1 - \widehat{U}_{id} \leq u_d \right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{I} \left(\widehat{U}_{i1} \geq 1 - u_1, \dots, \widehat{U}_{id} \geq 1 - u_d \right). \end{aligned}$$

3.2.2 Convergence faible de la copule empirique multivariée

Tel que dans le cas bivarié, la notion de copule régulière est d'abord définie ; il s'agit d'une simple extension d -dimensionnelle de la Définition 3.1.

Définition 3.2. Une copule d -dimensionnelle $C : [0, 1]^d \rightarrow [0, 1]$ est régulière si pour chaque $\ell \in \{1, \dots, d\}$, la dérivée partielle $\dot{C}_\ell(\mathbf{u}) = \partial C(\mathbf{u}) / \partial \ell$ existe sur $[0, 1]^d$ et \dot{C}_ℓ est continue sur $\{\mathbf{u} \in [0, 1]^d : 0 < u_\ell < 1\}$.

Sous l'hypothèse nulle, les statistiques S_n , R_n et T_n sont calculées en utilisant le comportement du processus de copule empirique $\mathbb{C}_n = \sqrt{n}(C_n - C)$. Quand n tend vers l'infini et en autant que la copule C est régulière, Segers (2012) a montré que le processus de copule empirique \mathbb{C}_n converge faiblement dans l'espace $\ell^\infty([0, 1]^d)$ des fonctions bornées définies sur $[0, 1]^d$ vers une limite gaussienne de représentation

$$\mathbb{C}(\mathbf{u}) = \mathbb{B}_C(\mathbf{u}) - \sum_{\ell=1}^d \dot{C}_\ell(\mathbf{u}) \mathbb{B}_C(\mathbf{u}_\ell^*), \quad (3.4)$$

où \mathbf{u}_ℓ^* est le vecteur $\mathbf{u} = (u_1, \dots, u_d)$ dont toutes les composantes sont égales à 1, sauf u_ℓ , et \mathbb{B}_C est un processus gaussien centré tel que

$$\text{cov}\{\mathbb{B}_C(\mathbf{u}), \mathbb{B}_C(\mathbf{u}')\} = C\{\min(\mathbf{u}, \mathbf{u}')\} - C(\mathbf{u})C(\mathbf{u}').$$

Dans l'expression précédente de $\mathbb{C}(\mathbf{u})$, \mathbb{B}_C est la limite de

$$\mathbb{B}_n(u, v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\prod_{\ell=1}^d \mathbb{I}\{F_\ell(X_{il}) \leq u_\ell\} - C(u, v) \right].$$

Enfin, comme cas particulier d'un résultat de Quesy (2016), on déduit que le processus de copule de survie multivariée $\mathbb{C}_n^* = \sqrt{n}(C_n^* - C^*)$ converge faiblement vers

$$\mathbb{C}^*(\mathbf{u}) = \mathbb{B}_C(\mathbf{1} - \mathbf{u}) - \sum_{\ell=1}^d \dot{C}_\ell(\mathbf{1} - \mathbf{u}) \mathbb{B}_C(\mathbf{1} - \mathbf{u}_\ell^*), \quad (3.5)$$

où \mathbf{u}_ℓ^* est le vecteur $\mathbf{u} = (u_1, \dots, u_d)$ dont toutes les composantes sont 0, sauf u_ℓ .

3.2.3 Versions multiplicateurs de \mathbb{C}_n et \mathbb{C}_n^* quand $d \geq 2$

Pour un vecteur $\xi = (\xi_1, \dots, \xi_n)$ de variables aléatoires indépendantes non négatives avec moyenne et variance unitaire, la version multiplicateur de \mathbb{B}_n s'écrit

$$\widehat{\mathbb{B}}_n(\mathbf{u}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\xi_i}{\bar{\xi}^n} - 1 \right) \prod_{\ell=1}^d \mathbf{I}\{F_{n\ell}(X_{i\ell}) \leq u_\ell\}.$$

La version multiplicateur de \mathbb{C}_n basée sur la représentation (3.4) est alors

$$\widehat{\mathbb{C}}_n(\mathbf{u}) = \widehat{\mathbb{B}}_n(\mathbf{u}) - \sum_{\ell=1}^d \widehat{C}_\ell(\mathbf{u}) \widehat{\mathbb{B}}_n(\mathbf{u}_\ell^*),$$

où $\widehat{C}_1, \dots, \widehat{C}_d$ sont les estimateurs des dérivées partielles $\dot{C}_1, \dots, \dot{C}_d$ définis pour chaque $\ell \in \{1, \dots, d\}$ par

$$\widehat{C}_\ell(\mathbf{u}) = \frac{C_n(\mathbf{u} + \ell_n e_\ell) - C_n(\mathbf{u} - \ell_n e_\ell)}{2\ell_n},$$

où $\ell_n = h/\sqrt{n}$ pour $h \in \mathbb{N}$ et e_1, \dots, e_d sont les vecteurs de base unitaires; par exemple, $e_1 = (1, 0, \dots, 0)$ et $e_2 = (0, 1, 0, \dots, 0)$. Il a été démontré par Segers (2012) que $(\mathbb{C}_n, \widehat{\mathbb{C}}_n)$ converge faiblement vers $(\mathbb{C}, \widehat{\mathbb{C}})$, où $\widehat{\mathbb{C}}$ est une copie indépendante de \mathbb{C} . Autrement dit, la version multiplicateur $\widehat{\mathbb{C}}_n$ parvient, asymptotiquement, à reproduire le comportement de \mathbb{C}_n . Enfin, la version multiplicateur de \mathbb{C}_n^* basée sur (3.5) est

$$\widehat{\mathbb{C}}_n^*(\mathbf{u}) = \widehat{\mathbb{B}}_n(\mathbf{1} - \mathbf{u}) - \sum_{\ell=1}^d \widehat{C}_\ell(\mathbf{1} - \mathbf{u}) \widehat{\mathbb{B}}_n(\mathbf{1} - \mathbf{u}_\ell^*).$$

3.2.4 Statistiques de test et leurs versions multiplicateur

Afin de confronter les hypothèses de symétrie radiale $\mathbb{H}_0 : C = C^*$ et $\mathbb{H}_1 : C \neq C^*$, on propose les statistiques de test

$$\begin{aligned} R_n &= n \int_{[0,1]^d} \{C_n(\mathbf{u}) - C_n^*(\mathbf{u})\}^2 d\mathbf{u}, \\ S_n &= n \int_{[0,1]^d} \{C_n(\mathbf{u}) - C_n^*(\mathbf{u})\}^2 dC_n(\mathbf{u}), \\ T_n &= \sqrt{n} \sup_{\mathbf{u} \in [0,1]^d} |C_n(\mathbf{u}) - C_n^*(\mathbf{u})|. \end{aligned}$$

Tel que démontré dans l'Annexe A.2,

$$\begin{aligned} R_n &= \frac{1}{n} \sum_{i,j=1}^n \left[\prod_{\ell=1}^d \min \left(1 - \widehat{U}_{i\ell}, 1 - \widehat{U}_{j\ell} \right) \right. \\ &\quad \left. - 2 \prod_{\ell=1}^d \min \left(1 - \widehat{U}_{i\ell}, \widehat{U}_{j\ell} \right) + \prod_{\ell=1}^d \min \left(1 - \widehat{U}_{i\ell}, 1 - \widehat{U}_{j\ell} \right) \right]. \end{aligned}$$

Également, on a

$$\begin{aligned} S_n &= \frac{1}{n^2} \sum_{i,j,k=1}^n \left[\prod_{\ell=1}^d I \left\{ \widehat{U}_{k\ell} \geq \max \left(\widehat{U}_{i\ell}, \widehat{U}_{j\ell} \right) \right\} \right. \\ &\quad \left. - 2 \prod_{\ell=1}^d I \left\{ \widehat{U}_{k\ell} \geq \max \left(\widehat{U}_{i\ell}, 1 - \widehat{U}_{j\ell} \right) \right\} \right. \\ &\quad \left. + \prod_{\ell=1}^d I \left\{ \widehat{U}_{k\ell} \geq \max \left(1 - \widehat{U}_{i\ell}, 1 - \widehat{U}_{j\ell} \right) \right\} \right]. \end{aligned}$$

Enfin, il est montré que

$$T_n = \sqrt{n} \max_{i_1, \dots, i_d \in \{1, \dots, n\}} \left| C_n \left(\frac{i_1}{n+1}, \dots, \frac{i_d}{n+1} \right) - C_n^* \left(\frac{i_1}{n+1}, \dots, \frac{i_d}{n+1} \right) \right|.$$

Étant donné que $\mathbb{C}_n = \sqrt{n}(C_n - C)$ converge en loi vers la limite \mathbb{C} définie à l'Équation (3.4) et que $\mathbb{C}_n^* = \sqrt{n}(C_n^* - C^*)$ converge en loi vers la limite \mathbb{C}^* définie à

l'Équation (3.5), on obtient sous \mathbb{H}_0 que

$$\begin{aligned} R_n &= \int_{[0,1]^d} \{\mathbb{C}_n(\mathbf{u}) - \mathbb{C}_n^*(\mathbf{u})\}^2 d\mathbf{u}, \\ S_n &= \int_{[0,1]^d} \{\mathbb{C}_n(\mathbf{u}) - \mathbb{C}_n^*(\mathbf{u})\}^2 dC_n(\mathbf{u}), \\ T_n &= \sup_{\mathbf{u} \in [0,1]^d} |\mathbb{C}_n(\mathbf{u}) - \mathbb{C}_n^*(\mathbf{u})|. \end{aligned}$$

On peut alors déduire que sous l'hypothèse nulle \mathbb{H}_0 ,

$$\begin{aligned} R_n \rightsquigarrow \mathbb{R} &= \int_{[0,1]^d} \{\mathbb{C}(\mathbf{u}) - \mathbb{C}^*(\mathbf{u})\}^2 d\mathbf{u}, \\ S_n \rightsquigarrow \mathbb{S} &= \int_{[0,1]^d} \{\mathbb{C}(\mathbf{u}) - \mathbb{C}^*(\mathbf{u})\}^2 dC(\mathbf{u}), \\ T_n \rightsquigarrow \mathbb{T} &= \sup_{\mathbf{u} \in [0,1]^d} |\mathbb{C}(\mathbf{u}) - \mathbb{C}^*(\mathbf{u})|. \end{aligned}$$

Les versions muplicateurs des statistiques R_n , S_n et T_n sont ainsi

$$\begin{aligned} \widehat{R}_n &= \int_{[0,1]^d} \left\{ \widehat{\mathbb{C}}_n(\mathbf{u}) - \widehat{\mathbb{C}}_n^*(\mathbf{u}) \right\}^2 d\mathbf{u}, \\ \widehat{S}_n &= \int_{[0,1]^d} \left\{ \widehat{\mathbb{C}}_n(\mathbf{u}) - \widehat{\mathbb{C}}_n^*(\mathbf{u}) \right\}^2 dC_n(\mathbf{u}), \\ \widehat{T}_n &= \sup_{\mathbf{u} \in [0,1]^d} \left| \widehat{\mathbb{C}}_n(\mathbf{u}) - \widehat{\mathbb{C}}_n^*(\mathbf{u}) \right|. \end{aligned}$$

3.2.5 Algorithme pour performer les tests et calculer leurs p-valeurs

Pour tester la symétrie radiale d'une copule, il faut :

- (i) calculer les statistiques de test R_n , S_n ou T_n ;
- (ii) pour $P_n = (P_{1n}, \dots, P_{nn})^\top$, où

$P_{in}(\mathbf{u}) = \text{I}(\widehat{U}_{1i} \leq u_1, \dots, \widehat{U}_{di} \leq u_d) - \text{I}(\widehat{U}_{1i} > 1 - u_1, \dots, \widehat{U}_{di} > 1 - u_d)$, et

$$\Xi_n^{(h)} = \left(\frac{\xi_1^{(h)}}{\xi_n^{(h)}} - 1, \dots, \frac{\xi_n^{(h)}}{\xi_n^{(h)}} - 1 \right),$$

où $\xi^{(h)} = (\xi_1^{(h)}, \dots, \xi_n^{(h)})$ est un vecteur de variables multiplicateurs, définir

$$\bar{E}_n^{(h)}(\mathbf{u}) = \frac{1}{\sqrt{n}} \Xi_n^{(h)} \left\{ P_n(\mathbf{u}) - \dot{C}_{1n}(\mathbf{u}) P_n(u_1, 1) - \dots - \dot{C}_{dn}(\mathbf{u}) P_n(1, u_d) \right\}$$

(iii) calculer, pour chaque $h \in \{1, \dots, M\}$, les versions multiplicateurs des statistiques de test via

$$\begin{aligned} \widehat{R}_n^{(h)} &= \int_0^1 \dots \int_0^1 \left\{ \bar{E}_n^{(h)}(\mathbf{u}) \right\}^2 du_1 \dots du_d \approx \frac{1}{N^2} \sum_{k,\ell=1}^N \left\{ \bar{E}_n^{(h)} \left(\frac{k}{N}, \frac{\ell}{N} \right) \right\}^2, \\ \widehat{S}_n^{(h)} &= \int_0^1 \dots \int_0^1 \left\{ \bar{E}_n^{(h)}(\mathbf{u}) \right\}^2 dC_n(u_1, \dots, u_d), \\ \widehat{T}_n^{(h)} &= \sup_{\mathbf{u} \in (0,1)^d} |\bar{E}_n^{(h)}(\mathbf{u})| \approx \max_{(k,\ell) \in \{1, \dots, N\}^2} \left| \bar{E}_n^{(h)} \left(\frac{k}{N}, \frac{\ell}{N} \right) \right|; \end{aligned}$$

(iv) Estimer les p-valeurs des tests avec

$$\begin{aligned} \widehat{\text{PV}}_R &= \frac{1}{M} \sum_{h=1}^M \text{I} \left(\widehat{R}_n^{(h)} > R_n \right), \\ \widehat{\text{PV}}_S &= \frac{1}{M} \sum_{h=1}^M \text{I} \left(\widehat{S}_n^{(h)} > S_n \right), \\ \widehat{\text{PV}}_T &= \frac{1}{M} \sum_{h=1}^M \text{I} \left(\widehat{T}_n^{(h)} > T_n \right). \end{aligned}$$

À l'instar du cas bidimensionnel, la décision concernant l'acceptation ou le rejet de l'hypothèse nulle est basée sur la p-valeur estimée du test, c'est-à-dire que \mathbb{H}_0 est rejetée au seuil de signification α si $\widehat{\text{PV}} < \alpha$.

Chapitre 4

Étude de puissance des tests de symétrie radiale

Cette section est consacrée à l'étude des propriétés échantillonnelles des statistiques de test proposés au Chapitre 3. On s'intéressera, d'une part, à la capacité de ces tests à maintenir leur seuil nominal sous l'hypothèse nulle de symétrie radiale. D'autre part, on étudiera à quel point ces procédures de test sont efficaces pour rejeter l'hypothèse de symétrie radiale dans des situations où cette hypothèse n'est pas satisfaite. On en profitera pour mettre cette performance en relief avec celle de tests récemment proposés par Bahraoui & Quessy (2017) ; ces tests sont d'abord décrits à la section suivante. Le reste du chapitre présente l'étude de simulation et la commente, et fournit une petite illustration sur de vraies données multidimensionnelles.

4.1 Procédure de test de Bahraoui & Quessy (2017)

Une nouvelle classe de statistiques pour tester formellement la présence ou non de symétrie radiale de la copule d'une loi multidimensionnelle a été introduite par Bahraoui & Quessy (2017). Pour la décrire, considérons des vecteurs observés $\mathbf{X}_1, \dots, \mathbf{X}_n$, où $\mathbf{X}_j = (X_{j1}, \dots, X_{jd})$, qui sont des copies indépendantes d'un vecteur aléatoire \mathbf{X} de loi d -dimensionnelle F . On suppose que les distributions marginales F_1, \dots, F_d sont continues, de telle sorte qu'il existe une unique copule C telle que $F(\mathbf{x}) = C\{F_1(x_1), \dots, F_d(x_d)\}$.

On a vu que l'hypothèse nulle $\mathbb{H}_0 : C = C^*$ de symétrie radiale s'écrit, de manière équivalente, $\mathbb{H}_0 : \mathbf{U} \stackrel{\mathcal{D}}{=} \mathbf{1} - \mathbf{U}$, où $\mathbf{U} \sim C$. En recentrant l'égalité en loi dans cette expression, on peut également écrire, avec $\mathbf{W} = \mathbf{U} - \mathbf{1}/2 = (U_1 - 1/2, \dots, U_d - 1/2)$,

$$\mathbb{H}_0 : \mathbf{W} \stackrel{\mathcal{D}}{=} -\mathbf{W}.$$

Les versions échantillonnelles du vecteur aléatoire \mathbf{W} sont données par les vecteurs de pseudo-observations $\widehat{\mathbf{W}}_1, \dots, \widehat{\mathbf{W}}_n$, où pour chaque $j \in \{1, \dots, n\}$,

$$\widehat{\mathbf{W}}_j = \widehat{\mathbf{U}}_j - \frac{\mathbf{1}}{2} = \left(\widehat{U}_{j1} - \frac{1}{2}, \dots, \widehat{U}_{jd} - \frac{1}{2} \right), \quad \text{avec } \widehat{U}_{j\ell} = F_{n\ell}(X_{j\ell}).$$

On rappelle que F_{n1}, \dots, F_{nd} sont les marges empiriques, *i.e.*

$$F_{n\ell}(x) = \frac{1}{n+1} \sum_{k=1}^n \mathbf{I}(X_{k\ell} \leq x), \quad \ell \in \{1, \dots, d\}.$$

Pour tester si la copule C d'une population d -dimensionnelle est symétrique ou non, les statistiques proposées par Bahraoui & Quessy (2017) sont de la forme

$$R_{n,\omega} = n \int_{\mathbb{R}^d} \{L_n(\mathbf{t})\}^2 \omega(\mathbf{t}) d\mathbf{t}, \tag{4.1}$$

où pour $\widehat{\mathbf{W}}_j^\top$ qui est la transposée de $\widehat{\mathbf{W}}_j$ et $\mathbf{t} = (t_1, \dots, t_d)$,

$$L_n(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^n \sin \left(\mathbf{t} \widehat{\mathbf{W}}_j^\top \right).$$

Dans l'expression de la statistique à l'Équation (4.1), ω est une fonction de poids qui est non-négative et intégrable. Ceci assure l'existence de la statistique $R_{n,\omega}$. Il a été montré par Bahraoui & Quessy (2017) qu'une expression explicite pour cette statistique de test est

$$R_{n,\omega} = \frac{1}{n} \sum_{j,j'=1}^n B_\omega \left(\widehat{\mathbf{W}}_j, \widehat{\mathbf{W}}_{j'} \right),$$

où pour $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$,

$$B_\omega(\mathbf{a}, \mathbf{b}) = \int_{\mathbb{R}^d} \sin(\mathbf{ta}^\top) \sin(\mathbf{tb}^\top) \omega(\mathbf{t}) d\mathbf{t}.$$

La statistique de test $R_{n,\omega}$ peut donc être facilement calculée pour toute taille échantillonnale n et toute dimension d . La décision concernant le rejet ou non de \mathbb{H}_0 est basée sur des versions multiplicateurs $\widehat{R}_{n,\omega}^{(1)}, \dots, \widehat{R}_{n,\omega}^{(M)}$ de la statistique de test $R_{n,\omega}$.

Plusieurs fonctions de poids ω peuvent être utilisées. Un cas particulier survient lorsque ω est un produit de densités de la forme

$$\omega(\mathbf{t}) = g \left(\frac{t_1}{\sigma} \right) \times \cdots \times g \left(\frac{t_d}{\sigma} \right),$$

où g est une densité de probabilité sur \mathbb{R} et $\sigma > 0$ est un paramètre d'échelle. Dans ce cas, en posant $\zeta(a) = \int_{\mathbb{R}} \cos(xa) g(x) dx$, on montre que

$$B_\omega(\mathbf{a}, \mathbf{b}) = \prod_{j=1}^d \zeta \{ \sigma(a_\ell - b_\ell) \} - \prod_{j=1}^d \zeta \{ \sigma(a_\ell + b_\ell) \}.$$

En prenant g comme étant les densités Normale, Double-Exponentielle et Double-

Gamma, on a respectivement

$$\zeta(a) \propto e^{-a^2/2}, \quad \zeta(a) \propto \frac{1}{a^2 + 4} \quad \text{et} \quad \zeta(a) \propto \frac{4 - a^2}{(a^2 + 4)^2}.$$

La statistique $R_{n,\omega}$ est notée respectivement par R_n^N , R_n^{DE} et R_n^{DG} pour ces trois possibilités. Enfin, en considérant $\omega(\mathbf{t}) = \varphi_R(\mathbf{t}/\sigma)$, où φ_R est la densité Normal de vecteur de moyennes nulles et de matrice de corrélation R , la statistique est notée R_n^{BN} et on montre que

$$B_\omega(\mathbf{a}, \mathbf{b}) = \varphi_R\{\sigma(\mathbf{a} - \mathbf{b})\} - \varphi_R\{\sigma(\mathbf{a} + \mathbf{b})\}.$$

La performance des statistiques R_n^N , R_n^{DE} , R_n^{DG} et R_n^{BN} , de même que celle de S_n , est étudiée dans la suite. Les cas $d = 2$ et $d > 2$ sont traités séparément. Pour les statistiques R_n^N , R_n^{DE} et R_n^{DG} , on posera $\sigma = 1$, tel que le recommandent Bahraoui & Quessy (2017) ; on prendra $\sigma = 5$ pour calculer R_n^{BN} .

4.2 Étude de simulations dans le cas $d = 2$

4.2.1 Plan de simulation

Genest & Nešlehová (2014) ont réalisé une étude de simulation afin d'étudier la précision de la méthode du multiplicateur et la puissance des tests de symétrie radiale basés sur les statistiques R_n , S_n et T_n . Leurs résultats montrent que la puissance de tous les tests augmente avec la taille de l'échantillon. Ils indiquent également que le test basé sur la statistique S_n est généralement le meilleur du point de vue de la puissance. De leur côté, Bahraoui & Quessy (2017) ont étudié la puissance des tests basés sur les statistiques R_n^N , R_n^{DE} , R_n^{DG} , R_n^{BN} et S_n . Selon leurs résultats de simulations, le test basé sur R_n^{BN} est systématiquement moins puissant que ceux basés sur R_n^N , R_n^{DE}

et R_n^{DG} ; la performance de ces derniers est similaire à celle de S_n . Sous les alternatives Gumbel, Khi-deux et Student asymétrique, les performances de R_n^{N} , R_n^{DE} et R_n^{DG} sont nettement meilleures que celles de S_n , tandis que sous les alternatives Clayton, les puissances de R_n^{N} , R_n^{DE} et R_n^{DG} sont très semblables à celle de S_n .

La statistique S_n étant meilleure que T_n et R_n , elle sera alors utilisée dans ce présent mémoire pour étudier la puissance des tests de symétrie radiale sous des contre-hypothèses de copules de Fisher. Outre S_n , on a aussi inclus dans l'étude les statistiques R_n^{N} , R_n^{DE} , R_n^{DG} et R_n^{BN} proposées par Bahraoui & Quessy (2017). La performance de ces statistiques a été étudiée pour les tailles d'échantillon $n \in \{125, 250\}$. À noter que le paramètre ℓ_n nécessaire à l'estimation des dérivées partielles a été posé à $\ell = 3$ quand $n = 125$ et à $\ell_n = 2$ lorsque $n = 250$.

Les modèles considérés dans l'étude appartiennent à la famille des copules de Fisher à $\nu \in \{1, 4, 7\}$ degrés de liberté dont les niveaux de dépendance, tels que mesurés par le tau de Kendall, sont $\tau_C \in \{1/4, 1/2, 3/4\}$. Des nuages de points générés à partir de données simulées de ces neuf modèles sont présentés à la Figure 4.1

4.2.2 Résultats et commentaires

La famille des copules de Fisher construite par Favre et al. (2018) possède la propriété d'asymétrie radiale. A cet effet, le pourcentage de rejet estimés à partir de 1000 échantillons de la copule permettra d'avoir une idée claire de la puissance des tests établis par Bahraoui & Quessy (2017) et Genest & Nešlehová (2014). Les résultats des différentes simulations sont présentés aux Tableau 4.1, Tableau 4.2, et Tableau 4.3 respectivement pour les degrés de liberté $\nu = 1$, $\nu = 4$ et $\nu = 7$.

Sous les copules de Fisher à 1 degré de liberté, la puissance des tests augmente en fonction de la taille de l'échantillon, ce qui était prévisible. Mis à part la statistique

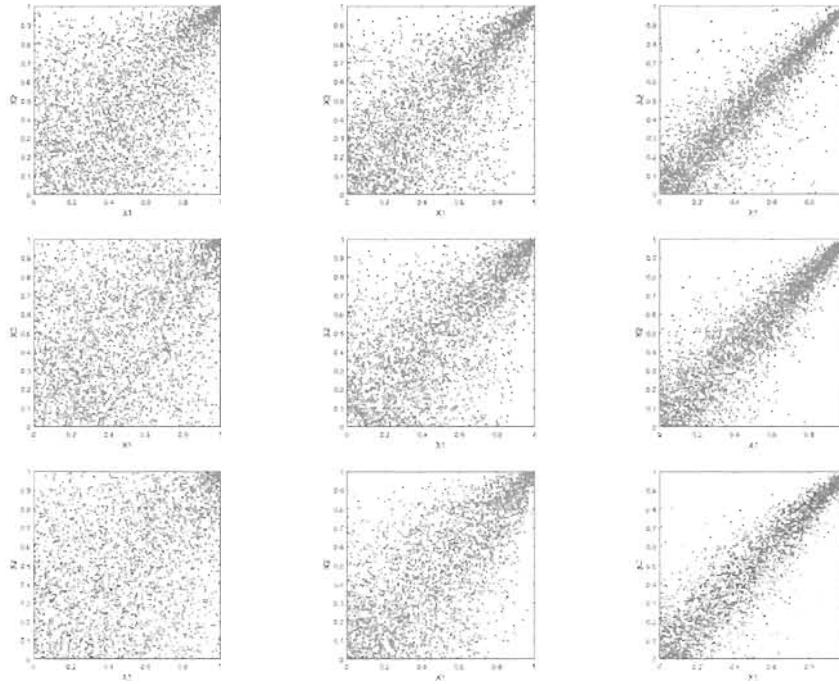


FIGURE 4.1 – Nuage de points 3000 paires simulées des copules de Fisher à 1, 4 et 7 degrés de liberté (respectivement de haut en bas) pour trois niveaux de dépendance (0,25 à gauche, 0,5 au milieu et 0,75 à droite)

R_n^{BN} , qui s'avère la moins puissante peu importe la taille de l'échantillon et le tau de Kendall, les tests basés sur les statistiques R_n^{N} , R_n^{DE} et R_n^{DG} sont généralement meilleurs que celui basé sur S_n . Globalement, la statistique R_n^{DE} est la plus puissante, suivie par R_n^{DE} , et ensuite par R_n^{DG} . Cependant, pour une taille $n = 250$ et un tau de Kendall de 0,75, la statistique S_n est meilleure que les tests basés sur la fonction caractéristique de copule. Pour $\mu = 4$ et $\nu = 7$ degrés de liberté, on constate des résultats très similaires au cas $\mu = 1$. La statistique R_n^{DE} est ainsi la plus puissante, tandis que R_n^{BN} est la moins performante.

Donc, globalement, les tests de Bahraoui & Quessy (2017) sont meilleurs que ceux de Genest & Nešlehová (2014) sous la nouvelle famille des copules de Fisher dans le cas à $d = 2$ dimensions. Ces résultats confirment donc la supériorité des tests de symétrie radiale basés sur la fonction caractéristique de copule sur ceux basés sur la copule empirique sous une panoplie d'alternatives.

TABLE 4.1 – Pourcentages de rejet estimés à partir de 1 000 échantillons de la copule de Fisher à $\nu = 1$ degré de liberté (partie du haut : $n = 125$; partie du bas : $n = 250$)

Statistique de test	$\tau_C = 1/4$	$\tau_C = 1/2$	$\tau_C = 3/4$
R_n^N	85,3	86,6	69,4
R_n^{DE}	86,8	86,0	69,0
R_n^{DG}	84,7	85,4	66,0
R_n^{BN}	60,2	65,3	48,5
S_n	76,9	82,1	67,4
<hr/>			
R_n^N	99,3	99,1	94,6
R_n^{DE}	99,6	99,2	96,3
R_n^{DG}	99,5	99,3	94,9
R_n^{BN}	91,6	96,5	88,0
S_n	98,9	99,1	99,0

4.3 Étude de simulations dans le cas $d > 2$

4.3.1 Plan de simulation

Les modèles considérés dans l'étude appartiennent à la famille des copules d -dimensionnelles Normales, de même qu'aux familles de Student et de Fisher à $\nu \in \{1, 4, 7\}$ degrés de liberté. Les tailles d'échantillons considérées sont $n \in \{125, 250\}$, alors que les dimensions étudiées sont $d \in \{3, 5\}$. On se concentrera ici sur les statistiques de test S_n et R_n^N . Pour mettre en œuvre le tests basé sur S_n , le paramètre h tel que $\ell_n = h/\sqrt{n}$ utilisé dans l'estimation des dérivées partielles de la copule a été fixé à

TABLE 4.2 – Pourcentages de rejet estimés à partir de 1 000 échantillons de la copule de Fisher à $\nu = 4$ degrés de liberté (partie du haut : $n = 125$; partie du bas : $n = 250$)

Statistique de test	$\tau_C = 1/4$	$\tau_C = 1/2$	$\tau_C = 3/4$
R_n^N	53,1	80,5	74,0
R_n^{DE}	58,3	81,2	72,6
R_n^{DG}	53,0	79,6	73,3
R_n^{BN}	22,0	57,0	61,0
S_n	39,0	71,7	56,9
<hr/>			
R_n^N	85,8	98,9	98,5
R_n^{DE}	84,2	99,3	98,5
R_n^{DG}	86,2	98,3	97,9
R_n^{BN}	55,2	92,1	97,1
S_n	75,2	98,3	97,5

$h = 3$ lorsque $n = 125$ et à $h = 2$ quand $n = 250$.

Les modèles de copules Normale et Student représentent des situations sous l'hypothèse nulle de symétrie radiale ; on s'attend ainsi, ou du moins on le souhaite, que les pourcentages de rejet soient autour de la valeur seuil du test, c'est-à-dire $100 \times \alpha = 5\%$. À l'opposé, on s'attend à ce que les tests affichent de forts pourcentages de rejet de l'hypothèse nulle lorsque les données sont générées de la copule de Fisher, puisqu'il s'agit d'une situation d'asymétrie radiale.

TABLE 4.3 – Pourcentages de rejet estimés à partir de 1 000 échantillons de la copule de Fisher à $\nu = 7$ degrés de liberté (partie du haut : $n = 125$; partie du bas : $n = 250$)

Statistique de test	$\tau_C = 1/4$	$\tau_C = 1/2$	$\tau_C = 3/4$
R_n^N	54,7	81,1	78,6
R_n^{DE}	55,3	83,7	72,0
R_n^{DG}	52,7	82,6	74,0
R_n^{BN}	22,6	55,6	66,3
S_n	39,7	71,7	55,3
<hr/>			
R_n^N	85,1	98,6	98,4
R_n^{DE}	85,9	98,5	98,0
R_n^{DG}	84,8	98,4	97,1
R_n^{BN}	54,2	92,1	98,0
S_n	74,5	97,7	98,4

4.3.2 Résultats et commentaires

Les résultats des différentes simulations sont présentés aux Tableau 4.4, Tableau 4.5 et Tableau 4.6 respectivement pour $\nu = 1$, $\nu = 4$ et $\nu = 7$ degrés de liberté. Globalement, ces trois tableaux montrent que les deux tests tiennent raisonnablement bien leur seuil nominal de 5%, surtout lorsque $n = 250$.

On peut affirmer que la puissance des tests, sous des alternatives de Fisher, augmente en fonction de la taille de l'échantillon, tel qu'attendu. Globalement, le test basé sur la statistique S_n est généralement plus puissant que celui basé sur R_n^N , et ce pour les trois degrés de liberté, à savoir $\nu \in \{1, 4, 7\}$. Cette constatation est intéressante, dans

la mesure où les résultats obtenus dans le cas $d = 2$ ne se transposent pas au cas multidimensionnel.

TABLE 4.4 – Pourcentages de rejet estimés à partir de 1 000 échantillons des copules Normale, Fisher et Student à $\nu = 1$ degré de liberté et à $d \in \{3, 5\}$ (partie du haut : $n = 125$; partie du bas : $n = 250$)

Copule	Statistique de test	$d = 3$	$d = 5$
Fisher	R_n^N	88,1	92,0
	S_n	97,3	99,7
Normale	R_n^N	3,9	1,3
	S_n	4,7	2,9
Student	R_n^N	2,8	1,4
	S_n	3,3	1,9
Fisher	R_n^N	100	100
	S_n	100	100
Normale	R_n^N	4,4	3,6
	S_n	4,3	4,2
Student	R_n^N	4,9	2,6
	S_n	5,9	3,4

4.4 Analyse de la symétrie radiale de données multidimensionnelles

L'étude qui suit porte sur les éléments nutritifs et les habitudes alimentaires de femmes âgées de 25 à 50 ans. Les données proviennent du département de l'agriculture des États-Unis. L'échantillon de taille $n = 737$ comporte cinq variables, à savoir les apports quotidiens en calcium (en mg), fer (en mg), protéines (en g), vitamine A (en μg)

TABLE 4.5 – Pourcentages de rejet estimés à partir de 1 000 échantillons des copules Normale, Fisher et Student à $\nu = 4$ degrés de liberté et à $d \in \{3, 5\}$ (partie du haut : $n = 125$; partie du bas : $n = 250$)

Copule	Statistique de test	$d = 3$	$d = 5$
Fisher	R_n^N	42,1	50,5
	S_n	55,0	78,6
Normale	R_n^N	3,3	1,5
	S_n	4,8	2,6
Student	R_n^N	3,4	0,9
	S_n	3,1	3,3
<hr/>			
Fisher	R_n^N	89,6	98,1
	S_n	90,8	99,4
Normale	R_n^N	4,9	3,1
	S_n	3,1	3,6
Student	R_n^N	3,7	3,3
	S_n	5,0	4,0

et vitamine C (en mg). Les nuages de points ci-après proviennent des données brutes et rangs standardisés de la série. En visualisant les paires de variables deux à deux, les données semblent être asymétriques.

Pour confirmer ou informer cette hypothèse, le test basé sur la nouvelle statistique de test S_n pour données multidimensionnelles a été mis en œuvre. On a obtenu une p -valeur inférieure à 0,1%, ce qui rejette l'hypothèse nulle de symétrie radiale. Ainsi, la population à partir de laquelle les données proviennent a vraisemblablement une structure de dépendance asymétrique en regard des queues de sa distribution. Afin de trouver un modèle paramétrique adéquat, on pourrait alors envisager des familles de copules comme la Khi-deux, la Fisher, ou plus généralement une copule *squared* telle que décrite par Quessy & Durocher (2019). Une autre possibilité consiste à considérer

TABLE 4.6 – Pourcentages de rejet estimés à partir de 1 000 échantillons des copules Normale, Fisher et Student à $\nu = 7$ degrés de liberté et à $d \in \{3, 5\}$ (partie du haut : $n = 125$; partie du bas : $n = 250$)

Copule	Statistique de test	$d = 3$	$d = 5$
Fisher	R_n^N	32	38,9
	S_n	39,1	56
Normale	R_n^N	3,6	1,7
	S_n	3,4	3,0
Student	R_n^N	2,9	1,0
	S_n	4,5	2,3
<hr/>			
Fisher	R_n^N	76,7	96,3
	S_n	77,1	97,3
Normale	R_n^N	4,2	4,6
	S_n	4,4	2,9
Student	R_n^N	2,9	3,1
	S_n	4,8	3,5

la famille des copules à valeurs extrêmes.

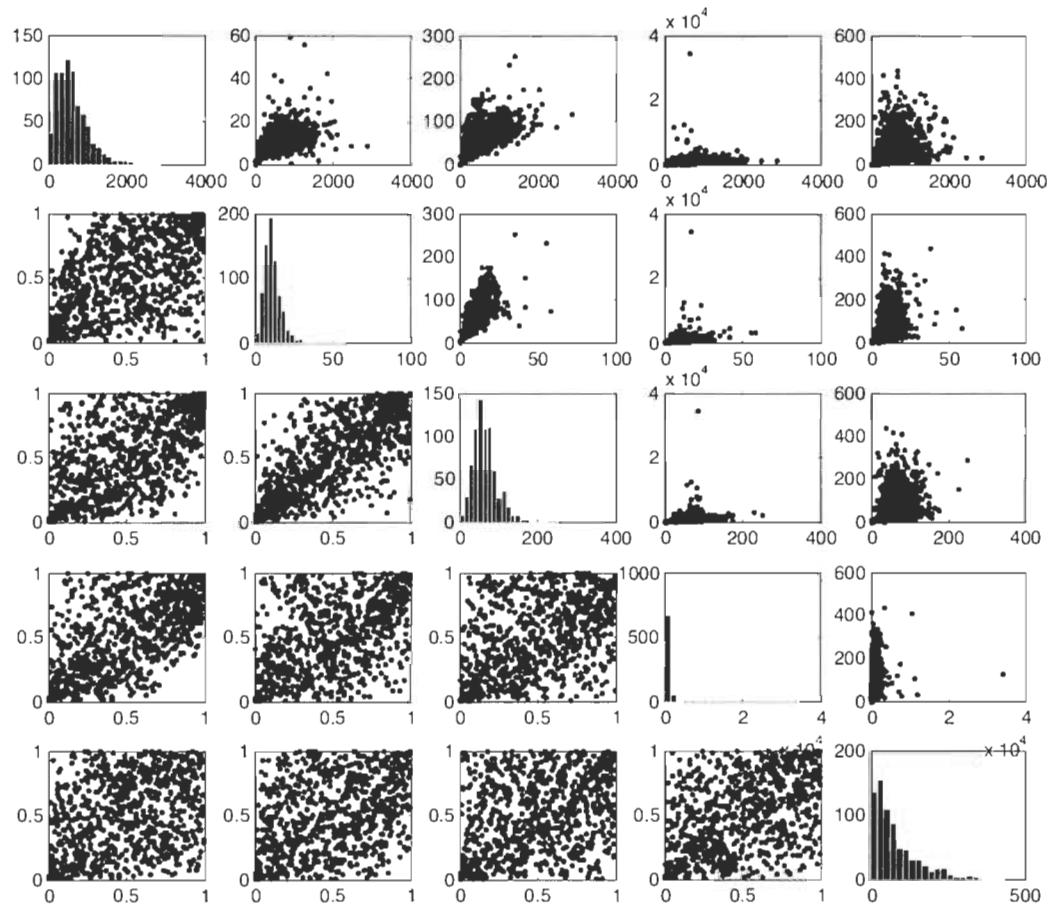


FIGURE 4.2 – Graphiques pour les données sur les habitudes alimentaires de $n = 737$ femmes aux États-Unis ; sur la diagonale, les histogrammes ; dans le triangle supérieur, les nuages de points des données brutes ; sur le triangle inférieur, les nuages de points des rangs standardisés

Conclusion

Cette recherche avait pour but d'étudier la puissance empirique des tests de symétrie radiale proposés par Genest & Nešlehová (2014) et par Bahraoui & Quessy (2017) sous des contre-hypothèses issues de la famille de Fisher. Les tests suggérés par Genest & Nešlehová (2014) n'ont été élaborés que dans le cas bivarié, alors que ceux de Bahraoui & Quessy (2017) sont applicables à la fois dans les cas bivarié et multidimensionnel. En premier lieu, ce mémoire a développé des versions d -dimensionnelles des statistiques de test de Genest & Nešlehová (2014), à savoir

$$\begin{aligned} R_n &= n \int_0^1 \cdots \int_0^1 \{C_n(\mathbf{u}) - C_n^*(\mathbf{u})\}^2 d\mathbf{u}_1 \cdots d\mathbf{u}_d, \\ S_n &= n \int_0^1 \cdots \int_0^1 \{C_n(\mathbf{u}) - C_n^*(\mathbf{u})\}^2 dC_n(\mathbf{u}), \\ T_n &= \sqrt{n} \sup_{\mathbf{u} \in [0,1]^d} |C_n(\mathbf{u}) - C_n^*(\mathbf{u})|. \end{aligned}$$

La méthode du multiplicateur, nécessaire pour estimer la p-valeur des tests basés sur ces statistiques, a également été mise au point. Une étude de simulation Monte-Carlo a ensuite permis de comprendre le comportement de ces statistiques lorsque les données proviennent d'une structure de dépendance asymétrique de type *Fisher*. La copule de Fisher a fait son apparition récemment dans la littérature, avec les travaux de Favre et al. (2018). Les résultats obtenus montrent que dans le cas bivarié, la supériorité des tests de Bahraoui & Quessy (2017) sur ceux de Genest & Nešlehová (2014) est confirmée. Cependant, et contrairement à ce qui aurait pu être anticipé, c'est l'extension multidimensionnelle de la statistique S_n de Genest & Nešlehová (2014) développée dans ce travail qui est la plus performante.

Bibliographie

- BAHRAOUI, T. & QUÉSSY, J.-F. (2017). Tests of radial symmetry for multivariate copulas based on the copula characteristic function. *Electron. J. Stat.* **11**, 2066–2096.
- DEHEUVELS, P. (1979). La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance. *Académie Royale de Belgique, Bulletin de la Classe des Sciences, 5e série* **65**, 274–292.
- FAVRE, A.-C., QUÉSSY, J.-F. & TOUPIN, M.-H. (2018). A new family of copulas to model upper tail dependence and radial asymmetry : properties and application to high dimensional rainfall data. *Environmetrics* .
- GENEST, C. & MACKAY, J. (1986). The joy of copulas : bivariate distributions with uniform marginals. *Amer. Statist.* **40**, 280–283.
- GENEST, C., NESLEHOVÁ, J. & ZIEGEL, J. (2011). Inference in multivariate archimedean copula models. *TEST* **20**, 223–256.
- GENEST, C. & NEŠLEHOVÁ, J. G. (2014). On tests of radial symmetry for bivariate copulas. *Statist. Papers* **55**, 1107–1119.
- JOE, H. (2015). *Dependence modeling with copulas*, vol. 134 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL.
- KLAASSEN, C. A. J. & WELLNER, J. A. (1997). Efficient estimation in the bivariate normal copula model : normal margins are least favourable. *Bernoulli* **3**, 55–77.

- MAI, J.-F. & SCHERER, M. (2012). *Simulating copulas*, vol. 4 of *Series in Quantitative Finance*. Imperial College Press, London. Stochastic models, sampling algorithms, and applications.
- MCNEIL, A. J. & NEŠLEHOVÁ, J. (2009). Multivariate Archimedean copulas, d -monotone functions and l_1 -norm symmetric distributions. *Ann. Statist.* **37**, 3059–3097.
- MEYER, C. (2013). The bivariate normal copula. *Comm. Statist. Theory Methods* **42**, 2402–2422.
- MOLENBERGHS, G. & LESAFFRE, E. (1994). Marginal modeling of correlated ordinal data using a multivariate plackett distribution. *Journal of the American Statistical Association* **89**, 633–644.
- NELSEN, R. B. (1997). Dependence and order in families of Archimedean copulas. *J. Multivariate Anal.* **60**, 111–122.
- NELSEN, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. New York : Springer, 2nd ed.
- PLACKETT, R. L. (1965). A class of bivariate distributions. *J. Amer. Statist. Assoc.* **60**, 516–522.
- QUESSY, J.-F. (2016). A general framework for testing homogeneity hypotheses about copulas. *Electron. J. Stat.* **10**, 1064–1097.
- QUESSY, J.-F. & DUROCHER, M. (2019). The class of copulas arising from squared distributions : Properties and inference. *Econom. Stat.* **12**, 148–166.
- QUESSY, J.-F., RIVEST, L.-P. & TOUPIN, M.-H. (2016). On the family of multivariate chi-square copulas. *J. Multivariate Anal.* **152**, 40–60.
- SEGERS, J. (2012). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli* **18**, 764–782.

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8, 229–231.

Annexe A

Formules explicites pour les statistiques de test

A.1 Cas bidimensionnel

A.1.1 Formule explicite pour R_n

On sait que

$$R_n = n \int_0^1 \int_0^1 \{C_n(u, v) - C_n^*(u, v)\}^2 \, du \, dv,$$

où

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n I(\widehat{U}_i \leq u, \widehat{V}_i \leq v) \quad \text{et} \quad C_n^*(u, v) = \frac{1}{n} \sum_{i=1}^n I(1 - \widehat{U}_i \leq u, 1 - \widehat{V}_i \leq v).$$

On a donc

$$\begin{aligned} R_n &= \frac{1}{n} \int_0^1 \int_0^1 \left\{ \sum_{i=1}^n \left(I(\widehat{U}_i \leq u, \widehat{V}_i \leq v) - I(1 - \widehat{U}_i \leq u, 1 - \widehat{V}_i \leq v) \right) \right\}^2 \, du \, dv \\ &= \frac{1}{n} \int_0^1 \int_0^1 \{A_1(u, v) - 2A_2(u, v) + A_3(u, v)\} \, du \, dv. \end{aligned} \tag{A.1}$$

Dans l'expression (A.1),

$$\begin{aligned}
A_1(u, v) &= \left\{ \sum_{i=1}^n I(\widehat{U}_i \leq u, \widehat{V}_i \leq v) \right\}^2 \\
&= \sum_{i,j=1}^n I(\widehat{U}_i \leq u, \widehat{V}_i \leq v) I(\widehat{U}_j \leq u, \widehat{V}_j \leq v) \\
&= \sum_{i,j=1}^n I(u \geq \widehat{U}_i, v \geq \widehat{V}_i) I(u \geq \widehat{U}_j, v \geq \widehat{V}_j) \\
&= \sum_{i,j=1}^n I\{u \geq \max(\widehat{U}_i, \widehat{U}_j), v \geq \max(\widehat{V}_i, \widehat{V}_j)\}.
\end{aligned}$$

Par des calculs très similaires, on arrive à établir que

$$\begin{aligned}
A_2(u, v) &= \sum_{i,j=1}^n I\{u \geq \max(\widehat{U}_i, 1 - \widehat{U}_j), v \geq \max(\widehat{V}_i, 1 - \widehat{V}_j)\}, \\
A_3(u, v) &= \sum_{i,j=1}^n I\{u \geq \max(1 - \widehat{U}_i, 1 - \widehat{U}_j), v \geq \max(1 - \widehat{V}_i, 1 - \widehat{V}_j)\}.
\end{aligned}$$

Par conséquent, en injectant dans l'expression (A.1) les formules développées pour A_1 , A_2 et A_3 , on obtient

$$\begin{aligned}
R_n &= \frac{1}{n} \sum_{i,j=1}^n \int_0^1 \int_0^1 \left[I\{u \geq \max(\widehat{U}_i, \widehat{U}_j), v \geq \max(\widehat{V}_i, \widehat{V}_j)\} \right. \\
&\quad - 2 I\{u \geq \max(\widehat{U}_i, 1 - \widehat{U}_j), v \geq \max(\widehat{V}_i, 1 - \widehat{V}_j)\} \\
&\quad \left. + I\{u \geq \max(1 - \widehat{U}_i, 1 - \widehat{U}_j), v \geq \max(1 - \widehat{V}_i, 1 - \widehat{V}_j)\} \right] du dv \\
&= \frac{1}{n} \sum_{i,j=1}^n \left[\left\{ 1 - \max(\widehat{U}_i, \widehat{U}_j) \right\} \left\{ 1 - \max(\widehat{V}_i, \widehat{V}_j) \right\} \right. \\
&\quad - 2 \left\{ 1 - \max(\widehat{U}_i, 1 - \widehat{U}_j) \right\} \left\{ 1 - \max(\widehat{V}_i, 1 - \widehat{V}_j) \right\} \\
&\quad \left. + \left\{ 1 - \max(1 - \widehat{U}_i, 1 - \widehat{U}_j) \right\} \left\{ 1 - \max(1 - \widehat{V}_i, 1 - \widehat{V}_j) \right\} \right] \\
&= \frac{1}{n} \sum_{i,j=1}^n \left[\min(1 - \widehat{U}_i, 1 - \widehat{U}_j) \min(1 - \widehat{V}_i, 1 - \widehat{V}_j) \right. \\
&\quad \left. - 2 \min(1 - \widehat{U}_i, \widehat{U}_j) \min(1 - \widehat{V}_i, \widehat{V}_j) + \min(\widehat{U}_i, \widehat{U}_j) \min(\widehat{V}_i, \widehat{V}_j) \right],
\end{aligned}$$

où la dernière égalité provient du simple fait que $1 - \max(a, b) = \min(1 - a, 1 - b)$.

A.1.2 Formule explicite pour S_n

En partant de l'équation (A.1), on peut écrire

$$S_n = \frac{1}{n} \int_0^1 \int_0^1 \{A_1(u, v) - 2A_2(u, v) + A_3(u, v)\} dC_n(u, v).$$

Puisque dC_n accorde un poids de $1/n$ à chaque paire $(\widehat{U}_k, \widehat{V}_k)$ de pseudo-observations,

$$\begin{aligned} S_n &= \frac{1}{n^2} \sum_{k=1}^n \left\{ A_1(\widehat{U}_k, \widehat{V}_k) - 2A_2(\widehat{U}_k, \widehat{V}_k) + A_3(\widehat{U}_k, \widehat{V}_k) \right\} \\ &= \frac{1}{n^2} \sum_{i,j,k=1}^n \left[I\left\{\widehat{U}_k \geq \max(\widehat{U}_i, \widehat{U}_j), \widehat{V}_k \geq \max(\widehat{V}_i, \widehat{V}_j)\right\} \right. \\ &\quad - 2I\left\{\widehat{U}_k \geq \max(\widehat{U}_i, 1 - \widehat{U}_j), \widehat{V}_k \geq \max(\widehat{V}_i, 1 - \widehat{V}_j)\right\} \\ &\quad \left. + I\left\{\widehat{U}_k \geq \max(1 - \widehat{U}_i, 1 - \widehat{U}_j), 1 - \widehat{V}_k \geq \max(\widehat{V}_i, 1 - \widehat{V}_j)\right\} \right]. \end{aligned}$$

A.1.3 Formule explicite pour T_n

Puisque C_n et C_n^* sont des fonctions avec des sauts aux valeurs des paires de pseudo-observations $(\widehat{U}_i, \widehat{V}_i)$, $i \in \{1, \dots, n\}$, et que $\widehat{U}_i, \widehat{V}_i \in \{1/(n+1), 2/(n+1), \dots, n/(n+1)\}$, on peut écrire

$$\begin{aligned} T_n &= \sqrt{n} \sup_{(u,v) \in [0,1]^2} |C_n(u, v) - C_n^*(u, v)| \\ &= \sqrt{n} \max_{i,j \in \{1, \dots, n\}} \left| C_n\left(\frac{i}{n+1}, \frac{j}{n+1}\right) - C_n^*\left(\frac{i}{n+1}, \frac{j}{n+1}\right) \right|. \end{aligned}$$

A.2 Cas multidimensionnel

Les calculs sont très semblables à ceux du cas bivarié. Il s'agit surtout de noter que

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{\ell=1}^d I\left(\widehat{U}_{i\ell} \leq u_\ell\right) \quad \text{et} \quad C_n^*(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{\ell=1}^d I\left(1 - \widehat{U}_{i\ell} \leq u_\ell\right).$$

A.2.1 Formule explicite pour R_n

En adoptant la même démarche que pour R_n dans le cas $d = 2$, on tire

$$\begin{aligned} R_n &= \frac{1}{n} \sum_{i,j=1}^n \left[\prod_{\ell=1}^d \min\left(1 - \widehat{U}_{i\ell}, 1 - \widehat{U}_{j\ell}\right) \right. \\ &\quad \left. - 2 \prod_{\ell=1}^d \min\left(1 - \widehat{U}_{i\ell}, \widehat{U}_{j\ell}\right) + \prod_{\ell=1}^d \min\left(1 - \widehat{U}_{i\ell}, 1 - \widehat{U}_{j\ell}\right) \right]. \end{aligned}$$

A.2.2 Formule explicite pour S_n

En adoptant la même démarche que pour S_n dans le cas $d = 2$, on tire

$$\begin{aligned} S_n &= \frac{1}{n^2} \sum_{i,j,k=1}^n \left[\prod_{\ell=1}^d I\left\{\widehat{U}_{k\ell} \geq \max\left(\widehat{U}_{i\ell}, \widehat{U}_{j\ell}\right)\right\} \right. \\ &\quad \left. - 2 \prod_{\ell=1}^d I\left\{\widehat{U}_{k\ell} \geq \max\left(\widehat{U}_{i\ell}, 1 - \widehat{U}_{j\ell}\right)\right\} \right. \\ &\quad \left. + \prod_{\ell=1}^d I\left\{\widehat{U}_{k\ell} \geq \max\left(1 - \widehat{U}_{i\ell}, 1 - \widehat{U}_{j\ell}\right)\right\} \right]. \end{aligned}$$

A.2.3 Formule explicite pour T_n

Dans le cas d -dimensionnel, C_n et C_n^* sont des fonctions avec des sauts aux valeurs des pseudo-observations $\widehat{\mathbf{U}}_1, \dots, \widehat{\mathbf{U}}_n$, où $\widehat{U}_{i\ell} \in \{1/(n+1), 2/(n+1), \dots, n/(n+1)\}$. On peut écrire, similairement au cas bidimensionnel,

$$\begin{aligned} T_n &= \sqrt{n} \sup_{\mathbf{u} \in [0,1]^d} |C_n(\mathbf{u}) - C_n^*(\mathbf{u})| \\ &= \sqrt{n} \max_{i_1, \dots, i_d \in \{1, \dots, n\}} \left| C_n \left(\frac{i_1}{n+1}, \dots, \frac{i_d}{n+1} \right) - C_n^* \left(\frac{i_1}{n+1}, \dots, \frac{i_d}{n+1} \right) \right|. \end{aligned}$$

Annexe B

Programmes Matlab utilisés pour les simulations

B.1 Test de symétrie radiale basé sur la fonction empirique

```
function [Sn,PV,Test] = RadialSym_MultEmpCopula(X,M,bn)

% Author
% Babacar Ndoye, UQTR.
% Description
% Performs the test of multivariate radial symmetry based on the statistic
% Sn
% Input
% X: n x d data matrix
% ell : integer acting as a smoothing parameter for the estimation
%       of the partial derivatives
% M: number of multiplier bootstrap samples
```

```
% Output
%     Sn : value of the test statistic
%     PV : P-value of the test
%     Test: =1 if H0 is rejected, =0 otherwise
%
% Necessary procedure
%     VectorOfRanks
%     MultDerivatives_Point *****

[n,d] = size(X);
U = VectorOfRanks(X) / (n+1);

% Test statistic
A = zeros(n,n);
for j=1:n
    for k=1:n
        t1 = prod(double(U(j,:)<=U(k,:)));
        t2 = prod(double(1-U(j,:)<=U(k,:)));
        A(j,k) = t1 - t2;
    end
end
Sn = sum(sum(A*A.')) / n^2;

% Computation of the multiplier bootstrap replicates
Cdot = zeros(n,d);
for j=1:d
    for k=1:n
        Cdot(k,j) = MultDerivatives_Point(U,bn,j,U(k,:));
    end
end
```

```

A = zeros(n,n);
for j=1:n
  for k=1:n
    t1 = prod(double(U(j,:)<=U(k,:)));
    t2 = prod(double(1-U(j,:)<=U(k,:)));
    t = Cdots(k,:)*( (U(j,:)<=U(k,:)) - (1-U(j,:)<=U(k,:)) ) .';
    A(j,k) = t1 - t2 - t;
  end
end

Mult = zeros(M,1);
B = A*A.';

parfor b=1:M
  Delta = exprnd(1,1,n); xi = (Delta/mean(Delta)) - 1;
  Mult(b) = xi*B*xi.' / n^2;
end

PV = sum(Mult > Sn) / M;
Test = (PV<.05);

```

B.2 Estimation de la puissance des tests

```

function P = RadialSym_BivCopulaCf_Power(n,C,tauC,betaC,G,sigma,rho,M,IT)

K = length(sigma);

P = zeros(1,K);
for k=1:K

```

```

L = sigma(k);

Test = zeros(1,IT);

parfor i=1:IT

X = BivCopula_Simulation(n,C,tauC,betaC);

[~,~,Test(i)] = RadialSym_BivCopulaCf(X,G,L,rho,M);

end

P(k) = sum(Test) / IT;

end


function P = RadialSym_BivEmpCopula_Power(n,C,tauC,betaC,M,bn,IT)

Test = zeros(1,IT);

parfor i=1:IT

X = BivCopula_Simulation(n,C,tauC,betaC);

[~,~,Test(i)] = RadialSym_BivEmpCopula(X,M,bn);

end

P = sum(Test) / IT;




function P = RadialSym_MultCopulaCf_Power(n,C,R,betaC,sigma,M,IT)

K = length(sigma);

P = zeros(1,K);

for k=1:K

L = sigma(k);

Test = zeros(1,IT);

parfor i=1:IT

X = PairwiseCopula_Simulation(n,C,R,betaC)

[~,~,Test(i)] = RadialSym_MultCopulaCf(X,L,M);

end

```

```

P(k) = sum(Test) / IT;
end

function P = RadialSym_MultEmpCopula_Power(n,C,Sigma,betaC,M,bn,IT)
% Sigma: d x d parameter matrix
Test = zeros(1,IT);
parfor i=1:IT
X = PairwiseCopula_Simulation(n,C,Sigma,betaC);
[~,~,Test(i)] = RadialSym_MultEmpCopula(X,M,bn);
end
P = sum(Test) / IT;

function Table = Babacar_TablesBiv(SampleSize,C,Kendall,betaC,M,bn,IT)

J = length(SampleSize);
K = length(Kendall);

Table = zeros(5*J,K);
for j=1:J
n = SampleSize(j);
if (n==125)
bn = 3;
elseif (n==250)
bn = 2;
end

Dim = 5*j-4;
for k=1:K
tauC = Kendall(k);

```

```

P1 = RadialSym_BivCopulaCf_Power(n,C,tauC,betaC,'N',1,0,M,IT);
P2 = RadialSym_BivCopulaCf_Power(n,C,tauC,betaC,'DE',1,0,M,IT);
P3 = RadialSym_BivCopulaCf_Power(n,C,tauC,betaC,'DG',1,0,M,IT);
P4 = RadialSym_BivCopulaCf_Power(n,C,tauC,betaC,'bivN',5,.5,M,IT);
P5 = RadialSym_BivEmpCopula_Power(n,C,tauC,betaC,M,bn,IT);
Table(Dim:Dim+4,k) = [P1 P2 P3 P4 P5].';
end
end

***

function Table = Babacar_TablesMultiv(SampleSize,R,dim,betaC,M,bn,IT)
% dim : dimension [3 5] dans notre cas
% SampleSize : taille de l'echantillon [125 250] dans ce cas
% R : matrice equi-correlation

J = length(SampleSize);
K = length(dim);

Table = zeros(6*J,K);
for j=1:J
n = SampleSize(j);
if (n==125)
bn = 3;
elseif (n==250)
bn = 2;
end

Dim = 6*j-5;

```

```
for k=1:K
m=dim(k);
if (m==3)
R = [1 0.5 0.5 ; 0.5 1 0.5 ; 0.5 0.5 1];
elseif (m==5)
R = [1 0.5 0.5 0.5 0.5 ; 0.5 1 0.5 0.5 0.5; 0.5 0.5 1 0.5 0.5;
0.5 0.5 0.5 1 0.5; 0.5 0.5 0.5 0.5 1];
end

% Fisher
P1 = RadialSym_MultCopulaCf_Power(n,4,R,betaC,bn,M,IT) ;
P2 = RadialSym_MultEmpCopula_Power(n,4,R,betaC,M,bn,IT);
% Normale
P3 = RadialSym_MultCopulaCf_Power(n,1,R,betaC,bn,M,IT) ;
P4 = RadialSym_MultEmpCopula_Power(n,1,R,betaC,M,bn,IT);
% Student
P5 = RadialSym_MultCopulaCf_Power(n,3,R,betaC,bn,M,IT) ;
P6 = RadialSym_MultEmpCopula_Power(n,3,R,betaC,M,bn,IT);
Table(Dim:Dim+5,k) = [P1 P2 P3 P4 P5 P6].';

end
end
```