

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

MARIE-HÉLÈNE TOUPIN

**NOUVEAU TEST D'ADÉQUATION POUR LES COPULES BASÉ  
SUR LE PROCESSUS DE SPEARMAN**

JANVIER 2008

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.

# REMERCIEMENTS

Je tiens à remercier toutes personnes ayant contribué de près ou de loin à ce mémoire. Tout d'abord, je tiens à exprimer toute ma gratitude envers mes directeurs de recherche, Mhamed Mesfioui et Jean-François Quessy, tous deux professeurs au département de Mathématiques et d'Informatique de l'Université du Québec à Trois-Rivières. En plus de m'avoir proposé ce projet passionnant, ils ont su m'encourager et me soutenir tout au long de ma maîtrise. La confiance qu'ils m'ont témoignée, leur grande disponibilité et leurs judicieux conseils ont assurés la réussite de ce projet. J'ai grandement apprécié travailler avec eux.

Je tiens également à remercier mes parents, Hélène et Réjean, pour leur soutien tout au long de ces années d'études ainsi qu'à mon amoureux, Éric, pour m'avoir encouragé et aidé à garder le sourire.

Merci également à Mme Josée St-Pierre, directrice du Laboratoire de recherche sur la performance des entreprises de l'Université du Québec à Trois-Rivières, pour m'avoir permis de travailler dans son laboratoire ces deux dernières années.

Merci à Christian Genest et à Louis Houde pour avoir lu attentivement mon mémoire et avoir suggéré des modifications qui ont améliorées grandement la version finale.

Une partie de ce travail à été financée par des octrois individuels accordés à MM. Mhamed Mesfioui et Jean-François Quessy par le Conseil de Recherche en Sciences Naturelles et en Génie du Canada et par le Fonds Institutionnel de Recherche de l'Université du Québec à Trois-Rivières. Je remercie également l'Institut des Sciences Mathématiques du Québec pour la bourse d'étude qu'il m'a accordée.

# Table des matières

Liste des tableaux	1
Liste des figures	3
Chapitre 1. Introduction	4
Chapitre 2. Méthodes d'inférence pour les modèles de copules	8
2.1. Définition et exemples de copules . . . . .	8
2.2. Estimation des paramètres d'une copule . . . . .	12
2.2.1. Inversion d'une statistique d'association . . . . .	12
2.2.2. Pseudo-maximum de vraisemblance . . . . .	14
2.3. Description de quelques procédures d'adéquation pour les copules	16
2.3.1. Approche de Malevergne & Sornette . . . . .	17
2.3.2. Approche de Breymann, Dias & Embrechts . . . . .	18
2.3.3. Approche de Fermanian . . . . .	19
2.3.4. Approche de Berg . . . . .	20

	2
2.3.5	Approche de Genest, Quessy & Rémillard . . . . . 21
2.3.6	Approche par la copule empirique . . . . . 22
<b>Chapitre 3.</b>	<b>La fonction de dépendance de Spearman 24</b>
3.1	Rho de Spearman . . . . . 24
3.2	Définition de la fonction de dépendance de Spearman . . . . . 26
3.3	Exemples . . . . . 27
<b>Chapitre 4.</b>	<b>On a new goodness-of-fit process for families of copulas 33</b>
4.1	Introduction . . . . . 34
4.2	Large-sample behavior of Spearman's goodness-of-fit process . . . 36
4.3	Some computations of Spearman functions . . . . . 40
4.3.1	The Gaussian copula . . . . . 41
4.3.2	The Clayton family . . . . . 42
4.3.3	Frank's copula . . . . . 43
4.3.4	The Farlie–Gumbel–Morgenstern distributions . . . . . 43
4.3.5	Durante's class . . . . . 43
4.4	Empirical issues . . . . . 44
4.4.1	Size and power of tests based on the process $\mathcal{S}_n$ . . . . . 44
4.4.2	Comparisons with Kendall and empirical copula processes 49
4.4.3	Testing for a bivariate Gaussian dependence structure . . . 51

	3
4.5 Illustrations . . . . .	53
4.5.1 Modelling of a hydrological data set . . . . .	53
4.5.2 Modelling of a multivariate data set . . . . .	57
4.6 Multivariate extension of $\mathcal{S}_n$ . . . . .	58
<b>Chapitre 5. Conclusion et discussion</b>	<b>60</b>
<b>Annexe A. Proof of proposition 1</b>	<b>62</b>
<b>Annexe B. Proof of lemma 4.1</b>	<b>66</b>
<b>Annexe C. Verification of assumption <math>\mathcal{A}_2</math> for models in Durante's class</b>	<b>67</b>
<b>Annexe D. Some models in Durante's class</b>	<b>68</b>
<b>Annexe E. Quelques ajouts à l'article</b>	<b>69</b>
E.1 Algorithme de simulation pour la famille de Durante . . . . .	69
E.2 Estimateurs déduits d'une inversion du rho de Spearman . . . . .	70
<b>Annexe F. Programmes en C</b>	<b>71</b>
<b>Bibliographie</b>	<b>98</b>

# LISTE DES TABLEAUX

4.1	Percentage of rejection, based on 10 000 replicates, of the null hypothesis of belonging to a given family of copulas for the goodness-of-fit test based on $\mathcal{V}_n$ , $n = 100, 250$ . . . . .	46
4.2	Percentage of rejection, based on 10 000 replicates, of the null hypothesis of belonging to a given family of copulas for the goodness-of-fit test based on $\mathcal{W}_{n,N}$ , $n = 100, 250$ and $N = 500$ . . . . .	48
4.3	Percentage of rejection, based on 10 000 replicates, of the null hypothesis of belonging to a given family of copulas for the Cramér-von Mises goodness-of-fit statistics computed from Spearman, Kendall and empirical copula processes $\mathcal{S}_n$ (Cuadras-Augé), $\mathcal{S}_{n,N}$ (Clayton, Frank, Gauss), $\mathcal{K}_n$ and $\mathcal{C}_{n,N}$ when $n = 250$ . . . . .	51
4.4	Percentage of rejection, based on 10 000 replicates, of the null hypothesis of a Gaussian copula for the Cramér-von Mises goodness-of-fit statistics computed from processes $\mathcal{S}_{n,N}$ and $\mathcal{L}_n$ when $n = 250$	53
4.5	Results of univariate goodness-of-fit tests for $X$ : <i>mean annual flow (<math>m^3/s</math>) at Gouin station</i> and $Y$ : <i>mean annual flow (<math>m^3/s</math>) at Rapide blanc station</i> . . . . .	55
4.6	Results of goodness-of-fit tests for eight copula families for the joint behavior of $X$ : <i>mean annual flow at Gouin station</i> and $Y$ : <i>mean annual flow at Rapide blanc station</i> . . . . .	56

4.7	Results of pairwise goodness-of-fit tests based on $\mathcal{V}_n$ and $\mathcal{W}_{n,N}$ for the SME data set. The upper triangle indicates the value of Spearman's rho for each pair and the lower triangle gives the models among a. Gaussian, b. Clayton, c. Frank, d. Durante A, e. Durante B, f. Cuadras–Augé, g. Fréchet and h. Ordinal sum for which the null hypothesis is accepted . . . . .	58
5.1	Generator, Spearman's dependence function and Spearman's rho for some members of Durante's family. . . . .	68
5.2	Biais (B), biais relatif (BR) et erreur quadratique moyenne (EQM) pour l'estimateur basé sur l'inversion du rho de Spearman . . . . .	70



## LISTE DES FIGURES

4.1	Histograms of $X$ : <i>mean annual flow (<math>m^3/s</math>) at Gouin station</i> and $Y$ : <i>mean annual flow (<math>m^3/s</math>) at Rapide blanc station</i> . . . . .	54
4.2	Mean annual flow at Rapide blanc station versus Gouin station. Left panel: scatter plot of the observations; right panel: scatter plot of the normalized ranks. . . . .	55
4.3	Graphs of $S_n(t)$ versus the parametric versions $S_{\hat{\theta}_n}(t)$ for Gaussian (left panel) and Ordinal sum (right panel) copulas. . . . .	56

# CHAPITRE 1

## INTRODUCTION

Traditionnellement, la modélisation simultanée de variables aléatoires fait appel aux techniques classiques basées sur le modèle gaussien ainsi que sur les extensions multivariées de distributions univariées telles les lois Gamma et Student. Une restriction importante lors de la mise en application de ces distributions est que les comportements marginaux des variables sont entièrement déterminés par le modèle choisi. Par exemple, dans le modèle de Student  $d$ -varié à  $\nu$  degrés de liberté, les marges sont nécessairement Student à  $\nu$  degrés de liberté. L'usage de ces modèles dans nombre d'applications, notamment en finance, a montré que ces lois quelques peu idéalisées sont souvent inaptes à expliquer un phénomène.

Depuis une quinzaine d'années, une nouvelle approche basée sur les copules s'est imposée dans le monde de la modélisation multivariée. La principale utilité de cette théorie est qu'elle permet de définir des lois à plusieurs variables avec les marges désirées. Dans ce cadre, il est facile de construire une distribution dont la structure de dépendance possède les caractéristiques d'une loi de Student, mais dont les marges sont normales, Pareto ou exponentielles, pour ne nommer que quelques-unes des possibilités offertes par l'emploi des copules.

L'origine de cette approche remonte à un désormais célèbre théorème dû à Sklar (1959). Pour citer ce résultat, supposons un couple  $(X, Y)$  de variables aléatoires dont la loi conjointe est définie par  $H(x, y) = P(X \leq x, Y \leq y)$ . On peut évidemment en extraire les lois marginales  $F$  et  $G$  de  $X$  et  $Y$  par  $F(x) = H(x, \infty)$

et  $G(y) = H(\infty, y)$ . Sklar a montré qu'il existe toujours une fonction  $C : [0, 1]^2 \rightarrow [0, 1]$  telle que la représentation

$$H(x, y) = C \{F(x), G(y)\} \quad (1.1)$$

est valide. La fonction  $C$  s'appelle la *copule* de  $(X, Y)$ . Dans le cas le plus courant où les marges  $F$  et  $G$  sont continues, la fonction  $C$  est unique. Il s'ensuit qu'on peut extraire la copule d'une loi bivariée continue par la formule

$$C(u, v) = H \{F^{-1}(u), G^{-1}(v)\}. \quad (1.2)$$

On peut donc en déduire les copules gaussienne et Student, en particulier. La représentation (1.1) met en lumière les trois parties constituantes d'une loi bivariée, à savoir

1. la loi de  $X$ , caractérisée par la fonction de répartition  $F$ ;
2. la loi de  $Y$ , caractérisée par la fonction de répartition  $G$ ;
3. la structure de dépendance  $C$  de  $(X, Y)$ .

Les trois structures précédentes représentent autant d'étapes qu'un statisticien peut accomplir indépendamment les unes des autres pour obtenir le meilleur modèle explicatif possible.

Maintenant, il est clair que toute l'information à propos de la dépendance d'un vecteur  $(X, Y)$  se retrouve dans la copule associée, qui peut être vue comme la *structure de dépendance*. Modéliser efficacement la dépendance dans un vecteur soulève quelques problèmes liés à l'inférence. Ces derniers sont autant d'avenues possibles de recherche. Une des voies très actuelle et prometteuse réside dans le choix d'une copule adéquate pour modéliser la dépendance dans un phénomène donné. Statistiquement parlant, on cherche à construire des tests afin de vérifier l'hypothèse que  $C$  appartient à une certaine famille de copules. Plus formellement, on désire confronter les hypothèses nulle et contre-hypothèses

$$\mathcal{H}_0 : C \in \{C_\theta; \theta \in \mathcal{M}\} \quad \text{vs.} \quad \mathcal{H}_1 : C \notin \{C_\theta; \theta \in \mathcal{M}\}.$$

Ici,  $\{C_\theta; \theta \in \mathcal{M}\}$  représente une famille paramétrique de copules, dont le paramètre devra être estimé sur la base d'un échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  tiré d'une population de loi  $H$ . Quelques auteurs se sont intéressés à la question de l'adéquation à un modèle de copules. Citons notamment Genest & Rivest (1993), Chen & Fan (2005), Fermanian (2005) et Genest, Quessy & Rémillard (2006).

Dans ce projet, le comportement asymptotique d'un processus empirique qui généralisera la très populaire mesure non paramétrique de dépendance de Spearman sera étudié. La convergence en loi de ce processus permettra de construire de nouveaux tests d'adéquation pour des familles de copules. Une particularité de ce processus réside dans le fait qu'il est construit à partir des rangs des observations, ce qui nécessite l'usage de techniques non classiques. Les étapes de réalisation de ce projet de recherche sont énoncées dans la suite.

1. Définir la fonction de dépendance de Spearman et étudier certaines de ses propriétés;
2. Définir le *processus de Spearman* et montrer qu'il converge en loi dans un certain espace de fonctions;
3. Construire de nouvelles statistiques d'adéquation comme fonctionnelles de ce processus;
4. Étudier la puissance des nouveaux tests proposés et les comparer à quelques procédures concurrentes;
5. Utiliser les nouvelles statistiques pour analyser de vrais jeux de données.

La procédure de test proposée dans ce mémoire possède plusieurs avantages :

- elle fournit un outil graphique pour évaluer l'adéquation visuellement;

- plusieurs familles de copules parmi les plus populaires peuvent être testées, incluant les copules gaussienne, Clayton, Frank, Farlie–Gumbel–Morgenstern et les modèles appartenant à la famille de Durante (2007);
- le test proposé est fondé sur une statistique dont la représentation est explicite en fonction des rangs des observations.

Au Chapitre 2, on présente une introduction à la théorie des copules et à l'estimation des paramètres de dépendance et on offre une brève revue de littérature à propos des tests d'adéquation pour les modèles de copules. Au Chapitre 3, on rappelle la définition de la mesure d'association de Spearman et on définit la fonction de dépendance de Spearman. Des exemples de calcul de cette fonction pour plusieurs familles de copules connues sont présentés. Le Chapitre 4 contient l'article soumis à la *revue canadienne de statistique* et intitulé *On a new goodness-of-fit process for families of copulas*. On y retrouve la définition de la version empirique de la fonction de dépendance de Spearman et la vérification de la convergence du processus empirique associé. En outre, les conditions nécessaires à ce résultat de convergence sont démontrées pour plusieurs familles de copules. La puissance d'un test d'adéquation basé sur une statistique de Cramér–von Mises de ce processus est ensuite comparée à celle de quelques procédures concurrentes. Deux exemples d'application sur de vrais jeux de données ainsi qu'une discussion concluent la partie principale de l'article. Les résultats de nature plus technique sont relégués aux Annexes A–D, qui font également partie de l'article qui a été soumis. Une conclusion générale et des idées d'explorations futures sont offertes à la suite de l'article. Finalement, l'Annexe E apporte quelques ajouts à l'article, alors que l'Annexe F présente les programmes informatiques utilisés lors des études de simulation et pour les analyses des jeux de données.

## CHAPITRE 2

# MÉTHODES D'INFÉRENCE POUR LES MODÈLES DE COPULES

### 2.1 Définition et exemples de copules

Le Théorème de Sklar (1959), dont la conséquence est donnée par la formule (1.1), implique qu'on peut étudier toute la dépendance d'un vecteur en tenant compte uniquement de la copule qu'on peut extraire de la loi conjointe d'un vecteur aléatoire. Inversement, le résultat permet de construire des lois avec une copule et des marges choisies. On en arrive donc à la question de la construction d'une copule. En fait, toute fonction  $C : [0, 1]^2 \rightarrow [0, 1]$  qui satisfait aux conditions (i) et (ii) suivantes est une copule.

(i)  $C(u, 0) = C(0, v) = 0$ ,  $C(u, 1) = u$  et  $C(1, v) = v$ ;

(ii) Pour tout  $u_1 \leq u_2$  et pour tout  $v_1 \leq v_2$ ,

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

Ces propriétés sont nécessaires et suffisantes pour assurer que  $C$  soit une fonction de répartition dont les marges sont uniformes sur  $(0, 1)$ .

Une des copules les plus communes est celle associée à l'indépendance entre deux variables aléatoires. Comme l'indépendance est caractérisée par la factorisation

de la loi conjointe associée  $H$  en le produit de ses marges  $F$  et  $G$ , i.e.  $H(x, y) = F(x)G(y)$ , il est facile de montrer, par une application de la formule (1.2), que la copule associée est  $\Pi(u, v) = uv$ . Cette copule caractérise donc l'indépendance bivariée.

Les bornes de Fréchet permettent de définir un intervalle à l'intérieur duquel toutes les lois bivariées se situent. En effet, on montre que pour toute loi  $H$ ,

$$\max\{F(x) + G(y) - 1, 0\} \leq H(x, y) \leq \min\{F(x), G(y)\}.$$

À noter que les bornes sont elles-mêmes des fonctions de répartition, ce qui veut dire que ces dernières ne peuvent pas être raffinées. Par une application de (1.2), on en déduit que toute copule  $C$  est telle que

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

où

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

sont appelées respectivement les bornes inférieure et supérieure de Fréchet–Hoeffding.

Par une autre application de (1.2), on découvre que la copule normale est

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} h_\rho(s, t) dt ds, \quad (2.1)$$

où

$$h_\rho(s, t) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right\}$$

est la densité d'une loi normale à deux variables de moyennes nulles, de variances unitaires et de corrélation  $\rho$ . Ici,  $\Phi$  est la fonction de répartition d'une variable aléatoire  $\mathcal{N}(0, 1)$ .

Comme autre exemple, on retrouve la famille de modèles bivariés de la forme

$$H_\omega(x, y) = F(x)G(y) + \omega F(x)G(y) \{1 - F(x)\} \{1 - G(y)\},$$

qui a souvent été utilisée pour bâtir des contre-hypothèses dans l'étude de l'efficacité de tests d'indépendance. En effet, pour  $\omega = 0$ , on retrouve l'indépendance et

$\omega \neq 0$  crée un écart à la dépendance et permet donc l'étude de la puissance de tests sous ces contre-hypothèses. La copule extraite de  $H_\omega$  est la copule de Farlie–Gumbel–Morgenstern

$$C_\omega(u, v) = uv - \omega uv(1-u)(1-v), \quad -1 \leq \omega \leq 1. \quad (2.2)$$

Une classe de copules importante, dont ne font pas parti les copules de Farlie–Gumbel–Morgenstern et normale, est la famille des copules archimédiennes. Les modèles qui appartiennent à cette famille sont de la forme

$$C_\phi(u, v) = \phi^{-1} \{ \phi(u) + \phi(v) \}, \quad (2.3)$$

où la fonction  $\phi : [0, 1] \rightarrow \mathbb{R}^+$ , appelée le *générateur*, est convexe, décroissante et telle que  $\phi(1) = 0$ . Quelques propriétés faciles à démontrer pour ces copules sont énumérées dans ce qui suit.

1. La distribution est symétrique, c'est-à-dire  $C_\phi(u, v) = C_\phi(v, u)$  pour tout  $(u, v) \in (0, 1)^2$ ;
2. Si  $r > 0$  est une constante, alors  $\phi^*(t) = r\phi(t)$  engendre la même copule que  $\phi$ , c'est-à-dire  $C_\phi(u, v) = C_{\phi^*}(u, v)$  pour tout  $(u, v) \in (0, 1)^2$ ;
3.  $C_\phi$  est la copule d'indépendance si et seulement si  $\phi(t) \propto -\log t$

Parmi les modèle de copules archimédiennes, celle dont le générateur est

$$\phi_\theta(t) = \frac{t^{-\theta} - 1}{\theta},$$

est la copule de Clayton. La loi jointe associée est

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0. \quad (2.4)$$

Un autre modèle populaire est la copule de Frank, dont le générateur est

$$\phi_\vartheta(t) = -\ln \left( \frac{e^{-\vartheta t} - 1}{e^{-\vartheta} - 1} \right), \quad \vartheta > 0.$$



L'expression de la copule engendrée est donc

$$C_{\vartheta}(u, v) = -\frac{1}{\vartheta} \ln \left\{ 1 + \frac{(e^{-\vartheta u} - 1)(e^{-\vartheta v} - 1)}{e^{-\vartheta} - 1} \right\}. \quad (2.5)$$

Pour les trois modèles précédents, on retrouve la copule d'indépendance à la limite lorsque la valeur du paramètre tend vers 0. Autrement dit,

$$\lim_{\delta \rightarrow 0} C_{\delta}(u, v) = \lim_{\theta \rightarrow 0} C_{\theta}(u, v) = \lim_{\vartheta \rightarrow 0} C_{\vartheta}(u, v) = uv.$$

Une autre famille de lois importante est celle des *valeurs extrêmes*. Ces modèles sont fréquemment utilisés en hydrologie et en finance. Une copule dans cette classe prend la forme générale

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

en terme d'une fonction de dépendance  $A : [0, 1] \rightarrow [0, 1]$  telle que  $A(0) = A(1) = 1$  et  $A(t) \geq \max(t, 1 - t)$ . Dans le cas particulier où  $A(t) \equiv 1$ , on retrouve la copule d'indépendance. Parmi les modèles de dépendance de nature *extrême*, on retrouve la copule de Gumbel de type B, aussi appelée copule logistique, dont la forme est

$$C_r(u, v) = \exp \left[ - \{ (-\log u)^r + (-\log v)^r \}^{1/r} \right], \quad r \geq 1.$$

La fonction de dépendance associée à  $C_r$  est

$$A_r(t) = \{ (1 - t)^r + t^r \}^{1/r},$$

La copule de Galambos, ou copule logistique négative, est

$$C_{\nu}(u, v) = uv \exp \left[ \{ (-\log u)^{-\nu} + (-\log v)^{-\nu} \}^{-1/\nu} \right], \quad \nu \geq 0.$$

La fonction de dépendance associée est

$$A_{\nu}(t) = 1 - \{ t^{-\nu} + (1 - t)^{-\nu} \}^{-1/\nu}.$$

## 2.2 Estimation des paramètres d'une copule

Soit une famille de copules  $\{C_\theta; \theta \in \mathcal{M}\}$ , où  $\mathcal{M} \subseteq \mathbb{R}^p$ . En considérant que le modèle de dépendance du couple aléatoire  $(X, Y)$  provient de cette famille, on désire souvent estimer le paramètre inconnu  $\theta$ . Pour ce faire, on considère  $n$  copies indépendantes  $(X_1, Y_1), \dots, (X_n, Y_n)$  du vecteur  $(X, Y)$ . Cette section présente deux techniques générales d'estimation, à savoir l'inversion d'une statistique d'association et une méthode de pseudo-maximum de vraisemblance.

### 2.2.1 Inversion d'une statistique d'association

Puisqu'une copule contient toute l'information au sujet de la dépendance dans un vecteur aléatoire, une mesure qui tente de saisir l'ampleur de cette association devrait dépendre uniquement de la structure de dépendance. De telles mesures sont données par le tau de Kendall et le rho de Spearman, dont les expressions en terme de la copule  $C$  sont respectivement

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad (2.6)$$

et

$$\rho_S = 12 \int_0^1 \int_0^1 \{C(u, v) - uv\} dudv. \quad (2.7)$$

Sous l'indépendance, on établit facilement que  $\tau = 0$ , alors que pour les copules associées aux bornes de Fréchet–Hoeffding, c'est-à-dire  $W$  et  $M$ , on a respectivement  $\tau = \rho_S = -1$  et  $\tau = \rho_S = 1$ .

Pour les copules archimédiennes, dont la forme générale est donnée à l'équation (2.3), Genest & Mackay (1986) ont mis en lumière la relation simple

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt$$

entre le tau de Kendall et le générateur de la famille. Pour la copule de Clayton, on a  $\phi(t)/\phi'(t) = (t^{\theta+1} - t)/\theta$  et on trouve alors  $\tau = \theta/(\theta + 2)$ . Des calculs un

peu plus complexes permettent de montrer que

$$\tau = 1 - \frac{4}{\vartheta} + \frac{4D_1\{\vartheta\}}{\vartheta}, \quad \text{où } D_1\{\vartheta\} = \frac{1}{\vartheta} \int_0^\vartheta \frac{x}{e^x - 1} dx$$

pour la copule de Frank.

La version empirique du tau de Kendall est donnée par

$$\tau_n = \frac{4Q_n}{n(n-1)} - 1,$$

où  $Q_n$  est le nombre de *paires concordantes* dans un échantillon bivarié de taille  $n$ . On dit que deux paires  $(X_i, Y_i)$  et  $(X_j, Y_j)$ ,  $i \neq j$ , sont concordantes si  $(X_i - X_j)(Y_i - Y_j) > 0$ . La statistique  $\tau_n$  est un cas particulier de U-statistique. Le livre de Lee (1990) donne d'ailleurs une excellente introduction au sujet. En particulier, on peut y déduire que  $\sqrt{n}(\tau_n - \tau)$  est asymptotiquement normale de moyenne nulle et de variance  $\sigma_\tau^2 = \text{var}\{2C(U, V) - U - V\}$ , donc qui dépend seulement de la copule sous-jacente à la loi  $H$  de la population.

Une version empirique du rho de Spearman est quant à elle donnée par

$$\rho_n = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - S_i)^2, \quad (2.8)$$

où  $R_i$  est le rang de l'observation  $X_i$  parmi  $X_1, \dots, X_n$  et  $S_i$  est le rang de l'observation  $Y_i$  parmi  $Y_1, \dots, Y_n$ .

Soit maintenant une famille  $\{C_\theta; \theta \in \mathcal{M}\}$  dont le paramètre  $\theta$  est univarié, i.e.  $\mathcal{M} \subseteq \mathbb{R}$ . Dans ce cas, une façon de l'estimer consiste à employer un analogue de la méthode des moments. Avant de décrire cette technique dans le cas général, supposons que la copule sous-jacente à la paire  $(X, Y)$  appartient à la famille de Farlie–Gumbel–Morgenstern dont les membres sont décrits par l'équation (2.2). Dans ce cas, il existe une relation simple entre  $\tau$  et le paramètre  $\omega$ , à savoir  $\tau = 2\omega/9$ . Ainsi,  $\omega = 9\tau/2$  et une estimation de  $\omega$  est donnée par

$$\tilde{\omega}_n = \frac{9\tau_n}{2}.$$

Ainsi, comme  $\sqrt{n}(\tau_n - \tau)$  converge vers une loi  $\mathcal{N}(0, \sigma_\tau^2)$ , alors

$$\sqrt{n}(\tilde{\omega}_n - \omega) = \frac{9}{2}\sqrt{n}(\tau_n - \tau) \rightsquigarrow \mathcal{N}(0, 81\sigma_\tau^2/4),$$

où ici et dans la suite,  $\rightsquigarrow$  dénotera la convergence en loi. Un résultat similaire s'applique également au rho de Spearman, qui pour ce modèle vaut  $\rho_S = \omega/3$ .

Plus généralement, supposons que pour une certaine famille de copules, on a la relation  $\tau = g(\theta)$ , où  $g$  est une fonction continue dont la dérivée première existe. Une estimation plausible pour  $\theta$  est  $\tilde{\theta}_n = g^{-1}(\tau_n)$ . Par un développement en série de Taylor d'ordre 1 autour de  $\theta$ , on a

$$\sqrt{n}(\tilde{\theta}_n - \theta) = \{g^{-1}(\tau)\}' \sqrt{n}(\tau_n - \tau) + o_p(1) = \frac{1}{g'(\theta)} \sqrt{n}(\tau_n - \tau) + o_p(1).$$

De là, le Théorème de Slutsky assure que

$$\sqrt{n}(\tilde{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, \bar{\sigma}_\tau^2),$$

où  $\bar{\sigma}_\tau^2 = \sigma_\tau^2 / \{g'(\theta)\}^2$ . Une étude de l'efficacité de cette estimation basée sur le tau de Kendall est présentée dans l'article de Genest, Quessy & Rémillard (2006) pour plusieurs modèles de copules.

Une adaptation de cette méthode pour le rho de Spearman est immédiate. En effet, si le rho de Spearman se définit en fonction du paramètre  $\theta$  par la relation  $\rho_S = h(\theta)$ , alors  $\hat{\theta}_n = h^{-1}(\rho_n)$  fournit une estimation convergente. La performance de cet estimateur est étudiée au Chapitre 4.

### 2.2.2 Pseudo-maximum de vraisemblance

Soit un couple  $(X, Y)$  de copule  $C_\theta$  et de marges  $F$  et  $G$ . La fonction de répartition jointe est donc  $H(x, y) = C_\theta\{F(x), G(y)\}$  et la densité

$$h_\theta(x, y) = c_\theta\{F(x), G(y)\}f(x)g(y),$$

où  $f = dF$ ,  $g = dG$  et  $c_\theta(u, v) = \partial^2 C_\theta(u, v) / \partial u \partial v$  sont les densités associées à  $F$ ,  $G$  et  $C_\theta$  respectivement. Dans le cas de marges connues, la méthode classique du maximum de vraisemblance consiste à maximiser en  $\theta$  la fonction de log-vraisemblance

$$m(\theta) = \sum_{i=1}^n \log h_\theta(X_i, Y_i) = \sum_{i=1}^n \log c_\theta\{F(X_i), G(Y_i)\} + K, \quad (2.9)$$

où  $K = \sum_{i=1}^n \log f(X_i)g(Y_i)$  ne dépend pas de  $\theta$ . La solution  $\hat{\theta}_{MV}$  est l'estimateur à maximum de vraisemblance classique de  $\theta$ .

L'intérêt principal de l'utilisation des copules tient au fait qu'il n'est pas nécessaire de connaître les marges pour effectuer de l'inférence sur la fonction de dépendance. Ainsi, comme les marges sont supposées inconnues, on remplace habituellement  $F$  et  $G$  par leurs estimations naturelles

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

c'est-à-dire les marges empiriques. Le choix de diviser par  $n+1$  plutôt que par  $n$  permet d'éviter des problèmes au bord dans l'équation (2.9). Dans le cas de la recherche d'un estimateur issu d'une méthode de la vraisemblance, Genest, Ghoudi & Rivest (1995) ont proposé de maximiser la pseudo log-vraisemblance

$$\ell(\theta) = \sum_{i=1}^n \log c_\theta\{F_n(X_i), G_n(Y_i)\}.$$

À remarquer que  $nF_n(X_i) = R_i$  et  $nG_n(Y_i) = S_i$ , où  $R_i$  et  $S_i$  sont les rangs de  $X_i$  et  $Y_i$  dans leurs échantillons univariés respectifs. La méthode du pseudo maximum de vraisemblance revient donc à maximiser

$$\ell(\theta) = \sum_{i=1}^n \log c_\theta \left( \frac{R_i}{n}, \frac{S_i}{n} \right).$$

L'estimateur  $\hat{\theta}_n$  est en général l'unique racine de l'équation

$$\sum_{i=1}^n \dot{c}_\theta \left( \frac{R_i}{n}, \frac{S_i}{n} \right) / c_\theta \left( \frac{R_i}{n}, \frac{S_i}{n} \right) = 0,$$

où  $\dot{c}_\theta = \partial c_\theta / \partial \theta$ . Il est évident que l'estimateur tiré de cette méthode est basé sur les rangs des observations. Le calcul de cet estimateur exige habituellement un long travail numérique. Cependant, il s'applique aux situations à plusieurs paramètres, ce qui n'est pas le cas pour les méthodes d'inversion.

### 2.3 Description de quelques procédures d'adéquation pour les copules

Dans cette section, une revue de la littérature sur les tests d'adéquation pour les copules sera présentée. Les tests de Malevergne & Sornette (2003), de Breyman, Dias & Embrechts (2003), de Fermanian (2005), de Berg (2007), de Genest, Quessy & Rémillard (2006) ainsi que de Genest, Rémillard & Beaudoin (2008) seront présentés dans un contexte multivarié général. À cette fin, soit un vecteur aléatoire  $\mathbf{X} = (X_1, \dots, X_d)$ ,  $d \geq 2$ , de copule  $C$ , de fonction de répartition multivariée  $H(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d)$  et de lois marginales  $F_i(x) = P(X_i \leq x)$ ,  $1 \leq i \leq d$ . Une extension directe du Théorème de Sklar assure qu'il existe une copule  $C : [0, 1]^d \rightarrow [0, 1]$  telle que

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

À partir d'un échantillon  $(X_{11}, \dots, X_{1d}), \dots, (X_{n1}, \dots, X_{nd})$ , composé de  $n$  copies indépendantes de  $\mathbf{X}$ , nous désirons déterminer si  $C$  appartient ou non à une famille de copule paramétrique donnée  $\{C_\theta; \theta \in \mathcal{M}\}$ . Plus formellement, les hypothèses à confronter sont

$$\mathcal{H}_0 : C \in \{C_\theta; \theta \in \mathcal{M}\} \quad \text{vs.} \quad \mathcal{H}_1 : C \notin \{C_\theta; \theta \in \mathcal{M}\}.$$

Autrement dit, les tests d'adéquation pour les copules permettent de valider une structure de dépendance adéquate pour  $\mathbf{X}$ .

### 2.3.1 Approche de Malevergne & Sornette

Malgré le fait que la normalité multivariée est plus souvent une exception que la norme dans beaucoup de phénomènes naturels et financiers, la structure de dépendance normale demeure un choix populaire lors d'une modélisation multivariée. Dans le cas de la normalité multivariée classique, plusieurs tests existent. Pour un article de synthèse sur le sujet, voir Henze (2002).

Dans un contexte semi-paramétrique où l'on désire établir que la distribution de  $\mathbf{X}$  a une structure de dépendance normale, mais pas nécessairement des marges normales, on utilise généralement des méthodes basées sur les rangs. Une telle procédure est proposée par Malevergne & Sornette (2003). Ces auteurs ont d'abord défini la transformation

$$Z(x_1, \dots, x_d) = (\Phi^{-1}\{F_1(x_1)\}, \dots, \Phi^{-1}\{F_d(x_d)\})^\top.$$

Sous l'hypothèse que la structure de dépendance sous-jacente à la population est la copule normale de matrice de corrélation  $R$  et que les marges sont  $F_1, \dots, F_d$ , on déduit que le vecteur aléatoire  $Z(\mathbf{X})$  suit une loi normale multivariée standard de corrélation  $R$ . Un résultat classique assure alors que  $T = Z(\mathbf{X})^\top R^{-1} Z(\mathbf{X})$  suit une loi khi-deux à  $d$  degrés de liberté.

Dans le cas plus général où la copule de  $\mathbf{X}$  est normale mais dont les lois marginales sont inconnues, Malevergne & Sornette (2003) ont proposé de remplacer chaque  $F_j$  par son homologue empirique

$$F_{nj} = \frac{1}{n+1} \sum_{i=1}^n \mathbf{1}(X_{ij} \leq x), \quad j = 1, \dots, d. \quad (2.10)$$

Ceci revient à considérer la transformation

$$Z_n(x_1, \dots, x_d) = (\Phi^{-1}\{F_{n1}(x_1)\}, \dots, \Phi^{-1}\{F_{nd}(x_d)\})^\top$$

et à former la variable aléatoire  $T_n = Z_n(\mathbf{X})^\top \hat{R}_n^{-1} Z_n(\mathbf{X})$ , où  $\hat{R}_n$  est l'estimateur de  $R$  obtenu par la statistique de van der Waerden. Ils ont ensuite suggéré de

baser un test d'adéquation sur les variables

$$T_{ni} = Z_n(\mathbf{X}_i)^\top R_n^{-1} Z_n(\mathbf{X}_i), \quad i = 1, \dots, n.$$

Selon eux, la distribution de  $T_{n1}, \dots, T_{nn}$  devrait être approximativement une loi khi-carré à  $d$  degrés de liberté, en autant que  $n$  soit suffisamment grand. Ils proposent ensuite de définir des statistiques d'adéquation en terme de distance entre la fonction de répartition de  $T_{n1}, \dots, T_{nn}$  et la fonction de répartition d'une loi khi-carré à  $d$  degrés de liberté.

### 2.3.2 Approche de Breymann, Dias & Embrechts

Définissons pour  $2 \leq i \leq d$  les fonctions de répartitions marginales conditionnelles

$$F_{i|1,\dots,i-1}(x_i|x_1, \dots, x_{i-1}) = P(X_i \leq x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1})$$

La transformation de Rosenblatt des variables  $X_1, \dots, X_d$  est définie par  $Z_1 = F_1(X_1)$  et

$$Z_i = F_{i|1,\dots,i-1}(X_i | X_1, \dots, X_{i-1}), \quad 2 \leq i \leq d. \quad (2.11)$$

L'utilité de cette transformation tient au fait que  $(Z_1, \dots, Z_d)$  est un vecteur de variables aléatoires indépendantes, chacune étant distribuée uniformément sur l'intervalle  $(0, 1)$ .

Breymann, Dias & Embrechts (2003) on exploité cette transformation en remarquant que la loi de

$$Y_G = \sum_{i=1}^d \{\Phi^{-1}(Z_i)\}^2$$

est une loi du chi-deux à  $d$  degrés de liberté. Ceci vient du fait que si  $U \sim \mathcal{U}(0, 1)$  et  $\Phi$  est la fonction de répartition du variable  $\mathcal{N}(0, 1)$ , alors  $\Phi^{-1}(U) \sim \mathcal{N}(0, 1)$ . Sous  $\mathcal{H}_0$ ,  $W_G = F_{\chi_d^2}(Y_G)$  est de loi uniforme sur  $(0, 1)$ , i.e.

$$F_G(w) = P(W_G \leq w) = w \quad \text{et} \quad f_G(w) = dF_G(w) = 1.$$



Pour un échantillon de  $n$  observations du vecteur à  $d$  dimensions  $(Z_1, \dots, Z_d)$ , la version empirique de  $F_G(w)$  est égale à

$$\hat{F}_G(w) = \frac{1}{n+1} \sum_{j=1}^n \mathbf{1}(W_{G_j} \leq w), \quad w = \frac{1}{n+1}, \dots, \frac{n}{n+1}.$$

Breymann, Dias & Embrechts (2003) proposèrent d'utiliser la statistique d'Anderson-Darling appliquée à  $F_G$ , c'est-à-dire

$$\hat{T}^{AD} = \frac{n}{n+1} \sum_{j=1}^n \frac{\left\{ \hat{F}_G\left(\frac{j}{n+1}\right) - \frac{j}{n+1} \right\}^2}{\frac{j}{n+1} \left(1 - \frac{j}{n+1}\right)}.$$

S'inspirant de ces travaux, Chen & Fan (2005) utilisent une statistique basée sur l'estimation par noyau de la densité  $f_G$ , c'est-à-dire qu'ils utilisent la statistique

$$\hat{T}^{L2KDE} = \frac{n}{q} \sum_{i=1}^q \left( \hat{f}_G\left(\frac{i-1/2}{q}\right) - 1 \right)^2,$$

où  $q$  est le nombre de segments utilisés pour l'estimation de  $\hat{f}_G$ .

### 2.3.3 Approche de Fermanian

Le test développé par Fermanian (2005) est basé sur l'estimation de la densité  $c$  associée à une copule  $C$ . Cette estimation utilise la méthode du noyau. Pour effectuer ce test, on doit choisir le noyau ainsi que la pondération, ce qui rend le test difficile à appliquer. On définit d'abord la copule empirique

$$C_n(u) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbf{1}\{F_{n_j}(X_{ij})\}, \quad (2.12)$$

où  $F_{n1}, \dots, F_{nd}$  sont les fonctions de répartition marginales empiriques données par l'équation (2.10). Cette fonction a été introduite par Deheuvels (1979) et étudiée ultérieurement par Gänssler & Stute (1987) et Tsukahara (2005), entre autres. Par définition, l'estimateur par noyau de la densité  $c_\theta$  au point  $u$  est

$$c_n(u) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{u - Y_{ni}}{h}\right),$$

où  $Y_{ni} = (F_{n1}(X_{i1}), \dots, F_{nd}(X_{id}))$ ,  $1 \leq i \leq n$ ,  $K$  est un noyau à  $d$  dimensions étant le produit de  $d$  noyaux uniformes univariés  $K_r$ ,  $r = 1, \dots, d$  et  $h = h(n)$  est la largeur de bande utilisée. Soit ensuite

$$J_n = \int \{c_n(u) - K_h * c_{\hat{\theta}_n}(u)\}^2 \omega(u) du,$$

où  $\omega : [0, 1]^d \rightarrow \mathbb{R}^+$  est une fonction de pondération,  $\hat{\theta}_n$  est un estimateur de  $\theta$  et  $K_h(\cdot) = K(\cdot/h)/h^d$ . Fermanian (2005) montre que

$$nh^{d/2} \left( J_n - \frac{1}{nh^d} \int K^2(\mathbf{t})(c\omega)(u - ht) dt du + \frac{1}{nh} \int c^2 \omega \sum_{r=1}^d \int K_r^2 \right)$$

converge en loi vers une  $\mathcal{N}(0, 2\sigma^2)$ , où

$$\sigma^2 = \int c^2 \omega \int \left\{ \int K(u)K(u+v) du \right\}^2 dv.$$

Ainsi, une statistique de test possible est

$$\mathcal{T}_n = \frac{n^2 h^d \left\{ J_n - (nh^d)^{-1} \int K^2(t)(c_{\hat{\theta}_n} \omega)(t - ht) dt du + (nh)^{-1} \int c_{\hat{\theta}_n}^2 \omega \sum_{r=1}^d \int K_r^2 \right\}^2}{2 \int c_{\hat{\theta}_n}^2 \omega \int \left\{ \int K(u)K(u+v) du \right\}^2 dv}$$

### 2.3.4 Approche de Berg

L'approche proposée par Berg (2007) s'inspire des travaux de Breymann, Dias & Embrechts (2003). Soit donc le vecteur  $(Z_1, \dots, Z_d)$  obtenu suite à la transformation de Rosenblatt donnée à l'équation (2.11). Soit maintenant la transformation  $Z_i \rightarrow Z_i^*$ ,  $1 \leq i \leq d$  définie par

$$Z_i^* = \left\{ 1 - \left( \frac{1 - \tilde{Z}_i}{1 - r_{i-1}} \right)^{d+1-i} \right\}, \quad 1 \leq i \leq d,$$

où  $(\tilde{Z}_1, \dots, \tilde{Z}_d)$  est obtenu du vecteur  $(Z_1, \dots, Z_d)$  en triant les composantes en ordre croissant et  $r_i$  est le rang de  $Z_i$  parmi  $Z_1, \dots, Z_d$ . Le test développé s'appuie sur la variable

$$\xi = \sum_{i=1}^d \gamma_\alpha(Z_i) \{ \Phi^{-1}(Z_i^*) \}^2,$$

où  $\gamma_\alpha$  est une fonction utilisée pour pondérer  $\{\Phi^{-1}(Z_i^*)\}^2$ . Si  $F_\xi$  est la fonction de répartition de  $\xi$ , alors la transformation intégrale de probabilité implique que  $U = F_\xi(\xi)$  suit une loi uniforme sur  $(0, 1)$ , en autant que l'hypothèse nulle  $\mathcal{H}_0$  soit vraie. On remarque cependant que comme  $\xi$  est une combinaison linéaire de variables normales au carré, la loi  $F_\xi$  ne se trouve pas explicitement en général.

Berg (2007) proposent d'estimer la variable  $U = F_\xi(\xi)$  sur la base d'un échantillon. Sous l'hypothèse nulle,  $F_\xi(w) = w$  et la fonction de densité qui y est associée est  $f_\xi(w) = 1$ . Une version empirique de  $F_\xi(w)$  est

$$\hat{F}_\xi(w) = \frac{1}{n+1} \sum_{j=1}^n \mathbf{1}(W_{\xi_j} \leq w), \quad w = \frac{1}{n+1}, \dots, \frac{n}{n+1}.$$

### 2.3.5 Approche de Genest, Quessy & Rémillard

Le processus de Kendall a d'abord été introduit par Genest & Rivest (1993) pour vérifier visuellement si un modèle de copules archimédiennes s'ajuste bien à un échantillon de données bivariées et son comportement général a été étudié en détail par Barbe, Genest, Ghoudi & Rémillard (1996). Le processus de Kendall est donné pour  $t \in [0, 1]$  par

$$\mathcal{K}_n(t) = \sqrt{n} \{K_n(t) - K(t)\},$$

où  $K$  est la transformation intégrale de probabilité multivariée définie par

$$K(t) = P\{C(U_1, \dots, U_p) \leq t\},$$

où la loi de  $(U_1, \dots, U_d)$  est  $C$ . L'estimation  $K_n$  de  $K$  est

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(V_{i,n} \leq t),$$

où pour chaque  $1 \leq i \leq n$ ,  $V_{i,n} = H_n(X_{i1}, \dots, X_{id})$  et

$$H_n(x_1, \dots, x_d) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}(X_{j1} \leq x_1, \dots, X_{jd} \leq x_d).$$

Wang & Wells (2000) furent les premiers à utiliser le processus de Kendall afin de tester l'adéquation à une famille de copules. Ils ont établi une procédure de sélection parmi un ensemble de copules archimédiennes dans le cas de données bivariées censurées. Leur idée fut de mesurer la distance entre l'estimateur non paramétrique  $K_n$  et la fonction  $K_\theta$  calculée sous l'hypothèse nulle  $\mathcal{H}_0$ . Ici, le paramètre inconnu  $\theta$  est estimé par  $\hat{\theta}_n$ . Leur statistique de test est toutefois basée sur une version tronquée de  $\mathbb{K}_n$ , à savoir

$$S_{\xi n} = n \int_{\xi}^1 \{K_n(t) - K_{\hat{\theta}_n}(t)\}^2 dt,$$

où  $\xi > 0$ . Par la suite, Genest, Quessy & Rémillard (2006) ont suggéré deux nouveaux tests d'adéquation basés sur une version non tronquée du processus  $\tilde{\mathbb{K}}_n = \sqrt{n}(K_n - K_{\hat{\theta}_n})$ , soit

$$S_n = \int_0^1 \{\tilde{\mathbb{K}}_n(t)\}^2 dK_{\hat{\theta}_n}(t) \quad \text{et} \quad T_n = \sup_{0 \leq t \leq 1} |\tilde{\mathbb{K}}_n(t)|.$$

La statistique  $S_n$  est basée sur la distance de Cramér–von Mises, tandis que  $T_n$  est basée sur la distance de Kolmogorov–Smirnov. Contrairement au test de Wang & Wells (2000), ce test s'applique à une grande variété de copules, pas nécessairement archimédiennes, il n'a pas à s'embêter d'un paramètre  $\xi$  dont la sélection est arbitraire, et s'applique à des familles de dimension  $d \geq 2$ . De plus, ces auteurs montrent comment obtenir des valeurs critiques asymptotiquement convergentes par l'usage d'une adaptation adéquate de la méthode du *bootstrap paramétrique*.

### 2.3.6 Approche par la copule empirique

Plutôt que d'utiliser une projection d'une copule multivariée dans un espace de fonctions univariées, comme pour les tests d'adéquation présentés précédemment, on peut travailler directement avec une estimation de la copule inconnue  $C$ , à savoir la copule empirique (2.12). On sait que  $\sqrt{n}(C_n - C)$  converge faiblement vers un processus gaussien continu lorsque les dérivés partielles existent et sont continues.

Deheuvels (1979) fut le premier à étudier ce processus dans le cas spécial de l'indépendance et à démontrer sa convergence dans ce contexte particulier. Ce résultat fut étendu au cas général par Gänssler & Stute (1987), et fut récemment reconsidéré par Fermanian, Radulović & Wegkamp (2004) et Tsukahara (2005).

L'idée de baser un test d'adéquation sur  $C_n$  se retrouve dans Quessy (2005) et a été étudiée sous divers scénarios de dépendance par Genest, Rémillard & Beaudoin (2008). Ces derniers suggèrent d'utiliser plutôt la version

$$C_{n,N}(u, v) = \sqrt{n} \{C_n(u, v) - C_N(u, v)\},$$

où  $C_N$  est une approximation par bootstrap de  $C_{\hat{\theta}_n}$  basé sur un échantillon aléatoire artificiel provenant de  $C_{\hat{\theta}_n}$ .

## CHAPITRE 3

# LA FONCTION DE DÉPENDANCE DE SPEARMAN

Le présent chapitre a pour but de définir la nouvelle fonction de dépendance de Spearman. La section 3.1 se veut un rappel sur le rho de Spearman et son lien avec les copules. À la section 3.2, la fonction de dépendance de Spearman est introduite et des exemples de calculs de celle-ci pour différentes familles de copules y sont présentés.

### 3.1 Rho de Spearman

Comme il a été vu au Chapitre 2, le rho de Spearman est une mesure non-paramétrique qui permet d'évaluer la force de la dépendance entre deux variables aléatoires. Plutôt que de calculer la corrélation en  $X$  et  $Y$ , ce coefficient mesure la corrélation entre les variables uniformisées  $U = F(X)$  et  $V = G(Y)$ . À noter que ces deux variables ont pour espérance  $1/2$ . Ainsi, la définition du rho de Spearman est

$$\rho_S = \text{cor} \{F(X), G(Y)\} = \frac{E(UV) - E(U)E(V)}{\{\text{var}(U)\}^{1/2} \{\text{var}(V)\}^{1/2}}.$$

Comme les variables  $F(X)$  et  $G(Y)$  sont uniformes sur  $(0, 1)$ , on calcule aisément  $\text{var}(U) = \text{var}(V) = 1/12$ . Ainsi,  $\rho_S = 12 \{E(UV) - E(U)E(V)\}$ .

La prochaine définition, qui permettra une représentation de  $\rho_S$  en terme de la copule associée à un couple de variables aléatoires, s'appelle *l'identité de Hoeffding*, due à Hoeffding (1948).

**Définition 3.1.** Soit  $(X, Y)$  de loi  $H$  et  $(X^*, Y^*)$  de loi  $H^*$ . Alors

$$E(XY) - E(X^*Y^*) = \int_0^1 \int_0^1 \{H(x, y) - H^*(x, y)\} dx dy.$$

Par définition, le couple  $(U, V)$  est distribué selon la copule  $C$ . Soit maintenant  $(U^*, V^*)$  de loi  $\Pi(u, v) = uv$ . Par l'identité 3.1, on peut donc écrire

$$\rho_S = 12 \{E(UV) - E(U^*V^*)\} = 12 \int_0^1 \int_0^1 \{C(u, v) - uv\} dudv. \quad (3.1)$$

**Remarque 3.1.** Supposons une certaine famille paramétrique  $\{C_\theta; \theta \in \Theta\}$  de copules dont les membres sont ordonnés par  $\theta$ , c'est-à-dire que pour tout  $(u, v) \in (0, 1)^2$ , on a  $C_\theta(u, v) \leq C_{\theta'}(u, v)$  dès lors que  $\theta < \theta'$ . On dit alors que  $\theta$  est un paramètre de dépendance. Dans ce cas, on montre que  $\rho_S$  est monotone croissante en fonction de  $\theta$ . En effet,

$$\begin{aligned} \rho_S(\theta) &= 12 \int_0^1 \int_0^1 \{C_\theta(u, v) - uv\} dudv \\ &\leq 12 \int_0^1 \int_0^1 \{C_{\theta'}(u, v) - uv\} dudv \\ &= \rho_S(\theta'). \end{aligned}$$

**Exemple 3.1.** Pour la copule de Farlie-Gumbel-Morgenstern, dont l'expression est donnée à l'équation (2.2), on a déjà mentionné au Chapitre 2 que  $\rho_S(\omega) = \omega/3$ . En effet, un calcul direct en utilisant (3.1) permet d'obtenir

$$\begin{aligned} \rho_S(\omega) &= 12 \int_0^1 \int_0^1 \{C_\omega(u, v) - uv\} dudv \\ &= 12 \int_0^1 \int_0^1 \omega uv(1-u)(1-v) dudv \\ &= 12\omega \left\{ \int_0^1 u(1-u) du \right\}^2 \\ &= \frac{\omega}{3}. \end{aligned}$$

**Exemple 3.2.** Si  $(X, Y)$  est un couple de loi normale bivariée de coefficient de corrélation  $\rho$ , alors

$$\rho_S(\rho) = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right).$$

**Exemple 3.3.** Pour la copule de Frank exprimée à l'équation (2.5), Frees & Valdez (1998) ont présenté les formules

$$\rho_S(\vartheta) = 1 + \frac{12}{\vartheta} [D_2\{\vartheta\} - D_1\{\vartheta\}], \quad \text{où } D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt.$$

Une façon compacte d'écrire cette expression est

$$\rho_S(\vartheta) = 1 + \frac{12}{\vartheta^3} \int_0^\vartheta \frac{t\{2t - \vartheta\}}{e^t - 1} dt.$$

L'intégrale dans l'expression précédente ne s'évaluant pas de façon explicite, une procédure numérique est nécessaire pour calculer le rho de Spearman.

**Exemple 3.4.** Pour la copule de Clayton, donnée à l'équation (2.4), aucune expression simple n'est envisageable pour  $\rho_S$ . Dans les applications ultérieures, une procédure d'intégration numérique sur le carré  $(0, 1)^2$  sera employée. En effet,  $\rho_S(\theta) = 12I(\theta) - 3$ , où

$$I(\theta) = \int_0^1 \int_0^1 (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} dudv.$$

## 3.2 Définition de la fonction de dépendance de Spearman

La version théorique de  $\rho_S$  est basée sur la moyenne de la variable aléatoire  $\xi = UV$ , où  $U = F(X)$  et  $V = G(Y)$ , à savoir  $E(\xi)$ . Plutôt que de baser une notion de dépendance sur la moyenne de  $\xi$ , on pourrait de façon plus générale considérer la distribution de cette variable aléatoire. En d'autres termes, on pourrait mesurer la dépendance par la fonction

$$S(t) = P(UV \leq t).$$



Cette manière de voir la dépendance généralise donc le rho de Spearman puisque

$$E(UV) = \int_0^1 t dS(t) = \int_0^1 \{1 - S(t)\} dt.$$

Comme  $\rho_S = 12E(UV) - 3$ , alors

$$\rho_S = 9 - 12 \int_0^1 S(t) dt.$$

Ainsi, si  $C$  est la copule de  $(X, Y)$ ,

$$S(t) = t + \int_t^1 C_{01} \left( \frac{t}{v}, v \right) dv, \quad (3.2)$$

où  $C_{01} = \partial C / \partial v$  est la loi conditionnelle de  $U$  étant donné  $V = v$ . En effet,

$$\begin{aligned} P(UV \leq t) &= \int_0^1 P(UV \leq t | V = v) dv \\ &= \int_0^1 P \left( U \leq \frac{t}{v} \middle| V = v \right) dv \\ &= \int_0^1 C_{01} \left( \frac{t}{v}, v \right) dv \\ &= t + \int_t^1 C_{01} \left( \frac{t}{v}, v \right) dv. \end{aligned}$$

### 3.3 Exemples

Dans cette section, des exemples de calculs de  $S(t)$  sont présentés pour un grand nombre de modèles de copules.

**Exemple 3.5.** Pour la copule d'indépendance  $\Pi(u, v) = uv$ , on a facilement  $\Pi_{01}(u, v) = u$ . Il s'ensuit que

$$S_{\Pi}(t) = t - \int_t^1 \frac{t}{v} dv = t - t \log t.$$

À noter que dans le cas particulier de l'indépendance, cette fonction est la même que la fonction de dépendance de Kendall considérée pour la première fois par Genest & Rivest (1993) et étudié ultérieurement par Barbe et coll. (1996). Ces derniers ont en effet défini  $K(t) = P\{C(U, V) \leq t\}$ , où  $(U, V) \sim C$ , si bien que sous l'indépendance, on retrouve  $K_{\Pi}(t) = S_{\Pi}(t)$ .

**Exemple 3.6.** Pour la famille de copules de Farlie–Gumbel–Morgenstern, dont les membres sont représentés à l'équation (2.2), on a  $C_{01}(u, v) = u + \omega u(1-u)(1-2v)$ .

Ainsi,

$$\begin{aligned}
 S_\omega(t) &= t + \int_t^1 \left\{ \frac{t}{v} + \omega \frac{t}{v} \left( 1 - \frac{t}{v} \right) (1-2v) \right\} dv \\
 &= t + t \int_t^1 \frac{dv}{v} + \omega \left\{ t \int_t^1 \frac{dv}{v} - 2t \int_t^1 dv - t^2 \int_t^1 \frac{dv}{v} + 2t^2 \int_t^1 dv \right\} \\
 &= t - t \log t + \omega \left\{ -t \log t - 2t(1-t) + t^2 \log t + 2t^2(1-t) \right\} \\
 &= t - 2\omega t(1-t)^2 + t \log(1/t) \{1 + \omega(1-t)\}.
 \end{aligned}$$

Sous l'indépendance, c'est-à-dire lorsque  $\omega = 0$ , on retrouve  $S_0(t) = t - t \log t$ .

**Exemple 3.7.** Pour la copule normale à deux variables, on déduit de l'équation (2.1)

que  $C_{01}(u, v)$

$$\begin{aligned}
 &= \int_{-\infty}^{\Phi^{-1}(u)} \varphi_\rho \{s, \Phi^{-1}(v)\} \{\Phi^{-1}(v)\}' ds \\
 &= \frac{1}{\varphi \{\Phi^{-1}(v)\}} \int_{-\infty}^{\Phi^{-1}(u)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[ -\frac{1}{2} \left\{ \frac{s^2 - 2\rho s\Phi^{-1}(v) + (\Phi^{-1}(v))^2}{\sqrt{1-\rho^2}} \right\} \right] ds \\
 &= \frac{e^{-(\Phi^{-1}(v))^2/2}}{\varphi \{\Phi^{-1}(v)\} \sqrt{2\pi}} \int_{-\infty}^{\Phi^{-1}(u)} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2} \left( \frac{s - \rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}} \right)^2 \right\} ds \\
 &= \int_{-\infty}^{\Phi^{-1}(u)} \frac{1}{\sqrt{1-\rho^2}} \varphi \left\{ \frac{s - \rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}} \right\} ds,
 \end{aligned}$$

où on a utilisé le fait que

$$\varphi \{\Phi^{-1}(v)\} = \frac{e^{-(\Phi^{-1}(v))^2/2}}{\sqrt{2\pi}}.$$

En effectuant le changement de variable

$$r = \frac{s - \rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}},$$

on obtient

$$C_{01}(u, v) = \Phi \left\{ \frac{\Phi^{-1}(u) - \rho\Phi^{-1}(v)}{\sqrt{1-\rho^2}} \right\}.$$

Ainsi l'expression de la fonction de dépendance de Spearman pour la copule gaussienne est

$$S_\rho(t) = t + \int_t^1 \Phi \left\{ \frac{\Phi^{-1}(t/v) - \rho \Phi^{-1}(v)}{\sqrt{1-\rho^2}} \right\} dv.$$

Il est facile de constater que  $S_0(t) = t - t \log t$ , tel qu'attendu.

**Exemple 3.8.** Pour la copule de Clayton,

$$C_{01} \left( \frac{t}{v}, v \right) = v^{-\theta-1} \left\{ \left( \frac{t}{v} \right)^{-\theta} + v^{-\theta} - 1 \right\}^{-1/\theta-1}.$$

La fonction de dépendance de Spearman pour cette copule est donc

$$S_\theta(t) = t - \int_t^1 v^{-\theta-1} \left\{ \left( \frac{t}{v} \right)^{-\theta} + v^{-\theta} - 1 \right\}^{-1/\theta-1} dv.$$

**Exemple 3.9.** Pour la copule de Frank décrite à l'équation (2.5), on trouve

$$C_{01} \left( \frac{t}{v}, v \right) = -\frac{1}{\vartheta} \frac{f_{\vartheta,01}(t/v, v)}{f_\vartheta(t/v, v)},$$

où

$$f_\vartheta(u, v) = 1 + \frac{(e^{-\vartheta u} - 1)(e^{-\vartheta v} - 1)}{e^{-\vartheta} - 1} \quad \text{et} \quad f_{\vartheta,01}(u, v) = \frac{\partial}{\partial v} f_\vartheta(u, v).$$

On obtient

$$S_\vartheta(t) = t - \frac{1}{\vartheta} \int_t^1 \frac{f_{\vartheta,01}(t/v, v)}{f_\vartheta(t/v, v)} dv.$$

**Exemple 3.10.** Soit  $(U, V) \sim M$ , où  $M$  est la copule de Fréchet-Hoeffding supérieure. Alors  $U = V$  avec probabilité 1. Ainsi,

$$S_M(t) = P(U^2 \leq t) = P(U \leq \sqrt{t}) = \sqrt{t}.$$

Pour la copule de Fréchet-Hoeffding inférieure  $W$ , l'événement  $V = 1 - U$  survient avec probabilité 1. Ainsi

$$P(UV \leq t) = P\{U(1-U) \leq t\} = P\left\{ \left( U - \frac{1}{2} \right)^2 \geq \frac{1}{4} - t \right\}.$$

Quand  $t \geq 1/4$ , la probabilité précédente vaut 1, alors que lorsque  $t < 1/4$ ,

$$\begin{aligned} \mathbb{P} \left\{ \left( U - \frac{1}{2} \right)^2 \geq \frac{1}{4} - t \right\} &= 1 - \mathbb{P} \left\{ \left( U - \frac{1}{2} \right)^2 \leq \frac{1}{4} - t \right\} \\ &= 1 - \mathbb{P} \left\{ \frac{1}{2} - \sqrt{\frac{1}{4} - t} \leq U \leq \frac{1}{2} + \sqrt{\frac{1}{4} - t} \right\} \\ &= 1 - \sqrt{1 - 4t}. \end{aligned}$$

Donc,

$$S_W(t) = \begin{cases} 1 - \sqrt{1 - 4t}, & t < 1/4 \\ 1, & t \geq 1/4. \end{cases}$$

Supposons que les membres d'une famille de copules s'écrivent par la combinaison linéaire  $C_p(u, v) = (1-p)C(u, v) + pD(u, v)$ , où  $p \in [0, 1]$  et  $C, D$  sont des copules. Alors la dérivée partielle de  $C_p$  par rapport à  $v$  est

$$C_{p,01}(u, v) = (1-p)C_{01}(u, v) + pD_{01}(u, v),$$

ce qui amène

$$\begin{aligned} S_p(t) &= t + (1-p) \int_t^1 C_{01} \left( \frac{t}{v}, v \right) dv + p \int_t^1 D_{01} \left( \frac{t}{v}, v \right) dv \\ &= t + (1-p) \{S_C(t) - t\} + p \{S_D(t) - t\} \\ &= (1-p)S_C(t) + pS_D(t), \end{aligned} \tag{3.3}$$

où  $S_C$  et  $S_D$  sont respectivement les fonctions de dépendance de Spearman pour les copules  $C$  et  $D$ . Les deux exemples suivants constituent des illustrations de cette formule.

**Exemple 3.11.** Pour la combinaison linéaire de la copule d'indépendance et de la borne supérieure de Fréchet-Hoeffding, i.e.

$$C_\delta(u, v) = (1-\delta)\Pi(u, v) + \delta M(u, v),$$

on déduit de l'équation (3.3) que

$$S_\delta(t) = (1-\delta)L_\Pi(t) + \delta L_M(t).$$

Les copules  $W$  et  $M$  sont entièrement singulières. Il existe aussi des copules qui comportent une partie absolument continue et une partie singulière. En fait, toute copule  $C$  s'écrit sous la forme  $C(u, v) = A_c(u, v) + A_s(u, v)$ , où

$$A_c(u, v) = \int_0^u \int_0^v \frac{\partial^2}{\partial s \partial t} C(s, t) ds dt$$

est la partie absolument continue et  $A_s(u, v) = C(u, v) - A_c(u, v)$  est la partie singulière. La plupart des copules utilisées dans la modélisation de phénomènes sont absolument continues, i.e.  $A_s \equiv 0$ .

**Exemple 3.12.** Durante (2007) a décrit une famille générale de copules ayant une partie singulière non nulle en général. Ses membres s'écrivent

$$C_f(u, v) = (u \wedge v) f(u \vee v),$$

où la fonction génératrice  $f : [0, 1] \rightarrow [0, 1]$  est croissante,  $f(1) = 1$  et  $f(t)/t$  est décroissante sur  $(0, 1]$ . Ces caractéristiques assurent que  $C_f$  est une copule. Comme cas particuliers, on retrouve l'indépendance  $\Pi$  quand  $f(v) = v$  et la borne supérieure de Fréchet  $M$  lorsque  $f \equiv 1$ . Également, la classe des copules de Fréchet s'obtient pour  $f_\delta(v) = (1 - \delta)v + \delta$ . Comme

$$C_{01}(u, v) = \begin{cases} u f'(v), & u \leq v \\ f(u), & u > v, \end{cases}$$

on déduit que la fonction de dépendance de Spearman associée à  $C_f$  est

$$S_f(t) = 2t - \sqrt{t} f(\sqrt{t}) + 2t \int_{\sqrt{t}}^1 \frac{f(v)}{v^2} dv. \quad (3.4)$$

En effet, notons d'abord que

$$C_{01}\left(\frac{t}{v}, v\right) = \begin{cases} f\left(\frac{t}{v}\right), & t \leq v \leq \sqrt{t} \\ \frac{t f'(v)}{v}, & \sqrt{t} < v \leq 1. \end{cases}$$

De la formule (4.4), on déduit

$$S_f(t) = t + \int_t^{\sqrt{t}} f\left(\frac{t}{v}\right) dv + t \int_{\sqrt{t}}^1 \frac{f'(v)}{v} dv.$$

D'abord, en effectuant le changement de variable  $u = t/v$  dans la première intégrale, on obtient

$$\int_t^{\sqrt{t}} f\left(\frac{t}{v}\right) dv = -t \int_{\sqrt{t}}^t \frac{f(u)}{u^2} du = t \int_{\sqrt{t}}^1 \frac{f(u)}{u^2} du.$$

En intégrant par partie la deuxième intégrale, on trouve

$$\int_{\sqrt{t}}^1 \frac{f'(v)}{v} dv = \frac{f(v)}{v} \Big|_{\sqrt{t}}^1 + \int_{\sqrt{t}}^1 \frac{f(v)}{v^2} dv = 1 - \frac{f(\sqrt{t})}{\sqrt{t}} + \int_{\sqrt{t}}^1 \frac{f(v)}{v^2} dv.$$

En rassemblant le tout, on découvre la fonction  $S_f$  annoncée précédemment.

Quelques calculs permettent de montrer que

$$\int_0^1 S_f(t) dt = 1 - \int_0^1 v^2 f(v) dv.$$

Ainsi, le rho de Spearman est donné par la formule

$$\rho_f = 12 \int_0^1 v^2 f(v) dv - 3,$$

tel qu'établi par Durante (2007).

**Exemple 3.13.** La copule de la famille à deux paramètres de Marshall–Olkin est définie par

$$C_{\alpha,\beta}(u,v) = \begin{cases} u^{1-\alpha}v, & u^\alpha > v^\beta \\ uv^{1-\beta}, & u^\alpha \leq v^\beta. \end{cases}$$

On calcule

$$S_{\alpha,\beta}(t) = t - \left( \frac{1}{\alpha} + \frac{1}{\beta} - 1 \right) \{1 - t^{-\alpha\beta/(\alpha+\beta)}\} t.$$

## CHAPITRE 4

### ON A NEW GOODNESS-OF-FIT PROCESS FOR FAMILIES OF COPULAS

*Mots-clés:* Copule; statistique de Cramér–von Mises; processus empirique; test d'adéquation; pseudo-observation; fonction de dépendance de Spearman; coefficient de corrélation de rangs de Spearman.

**Résumé :** Une procédure d'adéquation est proposée pour des familles paramétriques de copules bivariées. Les nouvelles statistiques de test sont des fonctionnelles d'un processus empirique fondé sur les versions théorique et empirique de la fonction de dépendance de Spearman. Les conditions sous lesquelles ce processus empirique est convergent sont satisfaites pour de nombreuses familles dont celles de Gauss, Clayton, Frank, Farlie–Gumbel–Morgenstern ainsi que les modèles à section singulière décrits par Durante (2007). Grâce à une méthode de ré-échantillonnage paramétrique qui permet une estimation convergente des seuils observés, il est montré empiriquement que les tests basés sur la distance de Cramér–von Mises gardent leur seuil sous l'hypothèse nulle. Des simulations attestant de la puissance des nouveaux tests, des comparaisons avec des procédures concurrentes et des analyses de vraies jeux de données sont présentées.

## 4.1 Introduction

Many inferential problems involving two random variables with continuous marginal distributions are fruitfully resolved considering the dependence structure and the marginal behaviors separately. This idea lies in the representation theorem of Sklar (1959) that enables to write a bivariate distribution function  $H$  as a function of its marginals  $F$  and  $G$  via

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

Here,  $C : [0, 1]^2 \rightarrow [0, 1]$  is called the copula of  $H$ . This formula suggests to develop statistical techniques in order to select a dependence function, i.e. a copula, that fits correctly to a set of bivariate observations. In other words, one wants to choose between the null and alternative hypotheses of belonging or not to a given parametric family, namely

$$\mathcal{H}_0 : C \in \{C_\theta; \theta \in \mathcal{M}\} \quad \text{vs.} \quad \mathcal{H}_1 : C \notin \{C_\theta; \theta \in \mathcal{M}\}.$$

Many goodness-of-fit procedures to confront  $\mathcal{H}_0$  and  $\mathcal{H}_1$  have been developed recently, e.g. Breymann, Dias & Embrechts (2003), Chen & Fan (2005), Fermaian (2005), Genest, Quessy & Rémillard (2006) and Scaillet (2006). These methods consider the univariate marginal distributions as infinite-dimensional nuisance parameters, and typically the observations are replaced by the maximally invariant statistics in that context, i.e. the ranks.

The fact that the ranks are used instead of the observations themselves can have a dramatic impact on the limiting distribution of a test statistic. For example, a goodness-of-fit statistic proposed by Breymann, Dias & Embrechts (2003), which is distributed as a chi-squared variable in the case of known marginals, seems to get away from this assumed distribution when using its rank-based version. As noted by the authors themselves, this had a significant impact on the observed size of the test. Further numerical illustrations of this phenomenon are provided in Dobrić & Schmid (2007).



In order to enrich the class of possible goodness-of-fit procedures that enable to choose an appropriate copula family that fits the data at hand, a new rank-based empirical process is proposed in this paper. The latter is developed around a sample version of Spearman's dependence function

$$S(t) = P \{F(X)G(Y) \leq t\}, \quad (4.1)$$

where  $(X, Y) \sim H$ . Its name derived from the fact that  $S$  is the distribution function of the random variable  $\xi = F(X)G(Y)$  whose mean is an affine transformation of Spearman's rho. The testing procedure developed in this work has many advantages :

- Many of the most popular copula models are testable, including the Gaussian copula and the members of the newly proposed family of Durante (2007);
- The test statistics have explicit representations in terms of the ranks of the observations;
- It is possible to compute valid p-values for the test statistics.

Moreover, the asymptotic tools used in this paper are typical of the kind of arguments that are necessary to characterize the large sample behavior of goodness-of-fit statistics in semi-parametric models. In this regard, the works of Ghoudi & Rémillard (1998, 2004) about the weak consistency of empirical processes based on pseudo-observations are well adapted.

The paper is organized as follows. In Section 4.2, an empirical process based on a functional estimation of  $S$  is defined and the weak consistency of a suitably normalized version is established. In Section 4.3, the assumptions leading to this asymptotic result are seen to be verified for a large number of popular copula models. In Section 4.4, the power of a Cramér-von Mises statistic is compared through simulations to that of competing procedures including the empirical copula goodness-of-fit and Kendall processes as well as the method of Malevergne &

Sornette (2003). In Section 4.5, the newly proposed methodology is illustrated on two real data sets, whereas a multivariate extension of  $\mathcal{S}_n$  is briefly described in Section 4.6.

## 4.2 Large-sample behavior of Spearman's goodness-of-fit process

Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be a random sample from a population with continuous marginal distributions  $F$  and  $G$ . If  $F$  and  $G$  are known, a natural estimation of Spearman's dependence function defined in (4.1) is given by the empirical distribution  $S_n^*$  of the i.i.d. observations  $\xi_1, \dots, \xi_n$ , where  $\xi_i = F(X_i)G(Y_i) \sim S$ . Standard arguments in the theory of empirical processes would then enable to conclude that  $\mathcal{D}_n = \sqrt{n}(S_n^* - S)$  converges to an  $S$ -Brownian bridge  $\mathcal{D}$ , i.e. a continuous stochastic process on  $[0, 1]$  such that  $\mathcal{D}(0) = \mathcal{D}(1) = 0$  and whose covariance function is  $\text{cov}\{\mathcal{D}(x), \mathcal{D}(y)\} = S(x \wedge y) - S(x)S(y)$ . Here and in the sequel,  $x \wedge y = \min(x, y)$ .

Since in the context of copula inference, one usually wants to keep the marginal distributions unknown, one has to rely on estimations of the non-observable variables  $\xi_i$ ,  $1 \leq i \leq n$ . Such estimations are provided by replacing  $F$  and  $G$  by their natural empirical versions

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

which yields the sample of dependent pseudo-observations  $\xi_{1,n}, \dots, \xi_{n,n}$ , where  $\xi_{i,n} = F_n(X_i)G_n(Y_i)$ . Under assumptions that are verified in many models, it will be seen that

$$S_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\xi_{i,n} \leq t)$$

is a good estimator of Spearman's dependence function  $S_\theta$  associated to a member  $C_\theta$  of a given parametric copula family. Specifically, it will be proven that the known parameter process  $\mathcal{S}_{n,\theta} = \sqrt{n}(S_n - S_\theta)$  converges weakly to a zero-mean Gaussian process under  $\mathcal{H}_0$ . This suggests to define the goodness-of-fit version of  $\mathcal{S}_{n,\theta}$  as  $\mathcal{S}_n = \sqrt{n}(S_n - S_{\hat{\theta}_n})$ , where  $\hat{\theta}_n$  is an estimator of  $\theta$ .

The weak consistency of  $\mathcal{S}_n$  will be obtained in two steps. Firstly, the asymptotic behavior of  $\mathcal{S}_{n,\theta}$  will be derived. This will be accomplished using a particular case of the general result of Ghoudi & Rémillard (1998) about empirical processes built from pseudo-observations. This is the subject of Proposition 1, where Assumptions  $\mathcal{A}_1$  and  $\mathcal{A}_2$  described next are needed. The proof is deferred to Appendix A.

- $\mathcal{A}_1$ . For all  $\theta \in \mathcal{M}$ , the density  $s_\theta(t) = \partial S_\theta(t)/\partial t$  is bounded on any compact subset of  $(0, 1]$  and for all continuous  $h : [0, 1]^2 \rightarrow [0, 1]$ , the mapping

$$t \mapsto \mu_\theta(t, h) = s_\theta(t) \text{E} \{h(U, V) | \xi = t\} \quad (4.2)$$

is continuous, where  $(U, V) \sim C_\theta(u, v)$  and  $\xi = UV$ .

- $\mathcal{A}_2$ . For all  $\theta \in \mathcal{M}$ ,  $s_\theta(t)$  is  $o\{t^{-1/2} \log^{-1}(1/t)\}$  as  $t \rightarrow 0$ .

PROPOSITION 1. If  $\mathcal{A}_1$  and  $\mathcal{A}_2$  hold, the empirical process  $\mathcal{S}_{n,\theta} = \sqrt{n}(S_n - S_\theta)$  converges in the Skorohod space of càdlàg functions to a Gaussian process having representation  $\mathcal{S}_\theta(t) = \mathcal{D}_\theta(t) - \mu_\theta(t, \mathbb{H}^*)$  on  $T = [0, 1]$ , and vanishing outside  $T$ . Here,  $\mathcal{D}_\theta$  is the limit of the process

$$\mathcal{D}_{n,\theta}(t) = n^{-1/2} \sum_{i=1}^n \{\mathbf{1}(\xi_i \leq t) - S_\theta(t)\}$$

and  $\mathbb{H}^*(u, v) = v\beta_1^*(u) + u\beta_2^*(v)$ , where  $\beta_1^*$ ,  $\beta_2^*$  are Brownian bridges such that  $\text{E}\{\beta_1^*(u)\beta_2^*(v)\} = C(u, v) - uv$ .

REMARK 1. The limit  $\mathcal{S}_\theta$  appearing in Proposition 1 does not depend on the marginal distributions. To see this, note that  $\mathbb{H}^*$  is free of  $F$  and  $G$  and the

process  $\mathcal{D}_{n,\theta}$  is based on the independent random variables  $\xi_i = F(X_i)G(Y_i) = U_i V_i$ ,  $i = 1, \dots, n$ , where  $(U_i, V_i) \sim C$ .

The study of  $\mathcal{S}_{n,\theta}$ , in particular the conclusion of Proposition 1, is of an independent interest since, for example, this process could be used to define statistics for a simple hypothesis such as bivariate independence. In our context where  $\theta$  is unknown, however, note that

$$\mathcal{S}_n(t) = \mathcal{S}_{n,\theta}(t) - \sqrt{n} \{S_{\hat{\theta}_n}(t) - S_\theta(t)\}.$$

The convergence of the first summand was the subject of Proposition 1. To ensure that the second term is consistent as well, and jointly consistent with the first term, two more assumptions are needed. To this end, let  $\dot{S}_\theta = \partial S_\theta / \partial \theta$  and  $\Theta_n = \sqrt{n}(\hat{\theta}_n - \theta)$ .

- $\mathcal{A}_3$ . For all  $\theta \in \mathcal{M}$ ,  $\dot{S}_\theta = \partial S_\theta / \partial \theta$  exists and is continuous on  $[0, 1]$ . Moreover, as  $\epsilon \rightarrow 0$ ,

$$\sup_{|\theta^* - \theta| < \epsilon} \sup_{t \in [0, 1]} |\dot{S}_{\theta^*}(t) - \dot{S}_\theta(t)| \rightarrow 0.$$

- $\mathcal{A}_4$ . The vector  $(\mathcal{S}_{n,\theta}, \Theta_n)$  converges in law to  $(S_\theta, \Theta)$ .

The main theoretical result of the paper is now stated.

**PROPOSITION 2.** Under Assumptions  $\mathcal{A}_1$ ,  $\mathcal{A}_2$ ,  $\mathcal{A}_3$  and  $\mathcal{A}_4$ , the empirical process  $\mathcal{S}_n = \sqrt{n}(S_n - S_{\hat{\theta}_n})$  converges in the Skorohod space of càdlàg functions to a Gaussian process whose representation is  $\mathcal{S}(t) = S_\theta(t) - \Theta \dot{S}_\theta(t)$  on  $T = [0, 1]$  and vanishing outside  $T$ .

**Proof.** Let  $\mathcal{S}_n = \mathcal{S}_{n,\theta} - \mathcal{E}_n$ , where  $\mathcal{E}_n(t) = \sqrt{n}\{S_{\hat{\theta}_n}(t) - S_\theta(t)\}$  for  $t \in [0, 1]$ . First,  $\mathcal{A}_1$  and  $\mathcal{A}_2$  ensure that  $\mathcal{S}_{n,\theta}$  converges to  $S_\theta$ , from Proposition 1. It remains to

show that

$$\sup_{t \in [0,1]} |\mathcal{E}_n(t) - \Theta \dot{S}_\theta(t)|$$

converges in probability to 0, which follows by arguments similar to that given in Appendix A of Genest, Quessy & Rémillard (2006). More details can be found in Toupin (2008).

In a semi-parametric context, a popular strategy to estimate  $\theta$  when the latter is real-valued consists in inverting a rank statistic. To illustrate in our context, suppose there exists a one-to-one relationship between  $\theta$  and the population version of Spearman's rho for a given copula family  $\{C_\theta, \theta \in \mathcal{M}\}$ . This dependence measure can be expressed as  $\rho(\theta) = 12\mathbb{E}(\xi) - 3$ , where  $\xi = F(X)G(Y)$  and the underlying copula of  $(X, Y)$  is  $C_\theta$ . In term of the associated Spearman's dependence function, one can write

$$\rho(\theta) = 12 \int_0^1 \{1 - S_\theta(t)\} dt - 3. \quad (4.3)$$

Since by assumption,  $\rho(\cdot)$  is a bijective application from  $\mathcal{M}$  to  $[-1, 1]$ , a method-of-moment estimator is given by  $\hat{\theta}_n = \rho^{-1}(\rho_n)$ , where

$$\rho_n = \frac{12n^2}{n^2 - 1} \bar{\xi}_n - 3 \left( \frac{n+1}{n-1} \right)$$

is the sample rank correlation coefficient and  $\bar{\xi}_n$  is the mean of the pseudo-observations  $\xi_{1,n}, \dots, \xi_{n,n}$ . Next, since

$$\rho_n = 12\bar{\xi}_n - 3 + \left( \frac{12\bar{\xi}_n}{n^2 - 1} - \frac{6}{n - 1} \right) = 12 \int_0^1 \{1 - S_n(\omega)\} d\omega - 3 + o_{\mathbb{P}}(1/\sqrt{n}),$$

one obtains easily

$$\sqrt{n} \{\rho_n - \rho(\theta)\} = -12 \int_0^1 \mathcal{S}_{n,\theta}(\omega) d\omega + o_{\mathbb{P}}(1).$$

Furthermore, a Taylor expansion of order one enables to write

$$\Theta_n = \sqrt{n}(\hat{\theta}_n - \theta) = \{\rho'(\theta)\}^{-1} \sqrt{n} \{\rho_n - \rho(\theta)\} + o_{\mathbb{P}}(1),$$

so that provided  $\mathcal{A}_1$  and  $\mathcal{A}_2$  hold, Assumption  $\mathcal{A}_4$  is satisfied with

$$\Theta = -12\{\rho'(\theta)\}^{-1} \int_0^1 \mathcal{S}_\theta(\omega) d\omega.$$

Thus, in view of Proposition 2, the limiting representation of the process in this special case is

$$\mathcal{S}(t) = \mathcal{S}_\theta(t) + 12\{\rho'(\theta)\}^{-1} \dot{\mathcal{S}}_\theta(t) \int_0^1 \mathcal{S}_\theta(\omega) d\omega.$$

### 4.3 Some computations of Spearman functions

Let  $C$  be the copula of a random couple  $(X, Y)$ . Then one can show that

$$S(t) = t + \int_t^1 C_{01} \left( \frac{t}{\omega}, \omega \right) d\omega, \quad (4.4)$$

where  $C_{01}$  is the conditional distribution of  $F(X)$  given  $G(Y) = v$ , i.e. the partial derivative of  $C$  with respect to the second component. For absolutely continuous copulas, straightforward computations show that

$$s(t) = \int_t^1 \frac{1}{\omega} c \left( \frac{t}{\omega}, \omega \right) d\omega, \quad (4.5)$$

where  $c$  is the density of  $C$ . The simplest of cases occurs when  $X$  and  $Y$  are independent, i.e. when the associated copula is  $\Pi(u, v) = uv$ . Using equation (4.4) with  $C_{01}(u, v) = u$ , it is then straightforward to show that the associated Spearman's dependence function is  $S_\Pi(t) = t - t \log t$ .

In this section, Assumptions  $\mathcal{A}_1$  and  $\mathcal{A}_2$  are shown to hold for the parametric copulas in the families of Gauss, Clayton, Frank and Farlie–Gumbel–Morgenstern. For the latter models, the associated densities  $c$  are bounded on any compact subset of  $(0, 1]^2$ , so that in view of equation (4.5),  $s(t)$  is bounded on any compact subset of  $T = (0, 1]$ . Next, one deduces from equation (4.2) that for absolutely continuous copulas,

$$\mu(t, h) = \int_t^1 h \left( y, \frac{t}{y} \right) c \left( y, \frac{t}{y} \right) \frac{1}{y} dy.$$

This mapping is continuous since for any sequence  $t_n \in T$  that converges to  $t \in T$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu(t_n, h) &= \lim_{n \rightarrow \infty} \int_{t_n}^1 h\left(y, \frac{t_n}{y}\right) c\left(y, \frac{t_n}{y}\right) \frac{1}{y} dy \\ &= \int_0^1 \lim_{n \rightarrow \infty} \mathbf{1}(y \geq t_n) h\left(y, \frac{t_n}{y}\right) c\left(y, \frac{t_n}{y}\right) \frac{1}{y} dy \end{aligned}$$

by Lebesgue's dominated convergence Theorem and the fact that the integrand is bounded. Thus,  $\lim_{n \rightarrow \infty} \mu(t_n, h) = \mu(t, h)$  and Assumption  $\mathcal{A}_1$  is satisfied for the four listed models. The remaining of the section is devoted to the checking of Assumption  $\mathcal{A}_2$  for these models and the members of a family of copulas with singular components introduced by Durante (2007).

### 4.3.1 The Gaussian copula

Let  $\varphi_\rho$  be the probability density of a bivariate normal distribution with zero means, unit variances and correlation coefficient  $\rho$ . The density of the Gaussian copula is given by  $c_\rho(u, v) = \varphi_\rho(x, y) / \varphi(x)\varphi(y)$ , where  $\varphi$  is the density of a  $\mathcal{N}(0, 1)$  random variable and  $\Phi^{-1}$  is the percentile function of a  $\mathcal{N}(0, 1)$  variable. If  $\rho = 0$ , one has  $s_\rho(t) = -\log t$ , which clearly satisfies  $\mathcal{A}_2$ . For  $0 < \rho < 1$ , equation (4.5) and a simple change of variable yields

$$s_\rho(t) = 2 \int_t^{\sqrt{t}} \frac{1}{\omega} \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{\rho^2 \Lambda_\rho(t, \omega)}{2(1-\rho^2)}\right\} d\omega,$$

where  $\Lambda_\rho(t, \omega) = \{\Phi^{-1}(\omega)\}^2 + \{\Phi^{-1}(t/\omega)\}^2 - 2\rho^{-1}\Phi^{-1}(\omega)\Phi^{-1}(t/\omega)$ . The next lemma, whose proof is deferred to Appendix B, will be instrumental to show that  $s_\rho$  satisfies Assumption  $\mathcal{A}_2$ .

**LEMMA 4.1** *For  $0 \leq t < \{\Phi(-1)\}^2$ , the minimum value of the mapping  $\omega \mapsto \Lambda_\rho(t, \omega)$ , for  $\omega \in [t, \sqrt{t}]$ , is attained when  $\omega = \sqrt{t}$ , so that the minimum value of  $\Lambda_\rho$  on this interval is*

$$\Lambda_\rho(t, \sqrt{t}) = \frac{-2(1-\rho)}{\rho} \left\{ \Phi^{-1}(\sqrt{t}) \right\}^2.$$

Lemma 4.1 implies that for all  $0 \leq t < \{\Phi(-1)\}^2$ ,

$$\begin{aligned} s_\rho(t) &\leq \frac{2}{\sqrt{1-\rho^2}} \exp\left\{-\frac{\rho^2 \Lambda_\rho(t, \sqrt{t})}{2(1-\rho^2)}\right\} \int_t^{\sqrt{t}} \frac{d\omega}{\omega} \\ &= M_\rho \left[ \varphi \left\{ \Phi^{-1}(\sqrt{t}) \right\} \right]^{-2\rho/(\rho+1)} (-\log t), \end{aligned}$$

where  $M_\rho = (1-\rho^2)^{-1/2} (2\pi)^{-\rho/(\rho+1)} < \infty$ . Since for all  $x < -1$ ,  $\varphi(x) \leq \Phi(x)$ , it follows for all  $t < \{\Phi(-1)\}^2$  that  $\varphi\{\Phi^{-1}(\sqrt{t})\} \leq \sqrt{t}$ . One concludes that  $s_\rho(t) \leq M_\rho t^{-\rho/(\rho+1)}(-\log t)$  for all  $t < \{\Phi(-1)\}^2$ . Assumption  $\mathcal{A}_2$  is hence satisfied for all  $0 \leq \rho < 1$ .

### 4.3.2 The Clayton family

The copulas in this class are of the form  $C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$ , where  $\theta \geq -1$ . In this article, however, only parameter values corresponding to positive dependence, i.e.  $\theta \geq 0$ , will be considered. Differentiating with respect to both  $u$  and  $v$  and by straightforward algebra, the density can be written

$$c_\theta(u, v) = (\theta + 1) \left[ \{C_\theta(u, v)\}^{a(\theta)} \right]^{\theta+1} (uv)^{-(\theta+1)},$$

where  $a(\theta) = (2\theta+1)/(\theta+1)$  is a positive increasing function such that  $\lim_{\theta \rightarrow \infty} a(\theta) = 2$  and  $a(0) = 1$ . Since the mapping  $\theta \mapsto C_\theta(u, v)$  is increasing for all  $u, v \in [0, 1]^2$  and the mapping  $\theta \mapsto x^{a(\theta)}$  is decreasing for  $0 \leq x \leq 1$ , one deduces that  $\theta \mapsto \{C_\theta(u, v)\}^{a(\theta)}$  is decreasing. Consequently,  $\{C_\theta(u, v)\}^{a(\theta)} \leq \{C_0(u, v)\}^{a(0)} = uv$ , so that  $c_\theta(u, v) \leq (\theta + 1)$  for all  $(u, v) \in [0, 1]^2$ . An application of formula (4.5) then enables to write

$$s_\theta(t) \leq (\theta + 1) \int_t^1 \frac{1}{\omega} d\omega = (\theta + 1) \log(1/t),$$

which establishes that  $s_\theta(t) = o\{t^{-1/2} \log^{-1}(1/t)\}$  when  $t \rightarrow 0$  for all  $\theta < \infty$ .



### 4.3.3 Frank's copula

For  $\vartheta \neq 0$ , the density associated to Frank's copula is

$$c_{\vartheta}(u, v) = \frac{\vartheta e^{-\vartheta u} e^{-\vartheta v} (e^{-\vartheta} - 1)}{\{1 - e^{-\vartheta} - (1 - e^{-\vartheta u})(1 - e^{-\vartheta v})\}^2}.$$

Note that when  $\vartheta \rightarrow 0$ , one recovers the density associated to independence. Consequently,  $c_{\vartheta}$  is continuous on  $[0, 1]^2$  for all  $\vartheta \geq 0$  and  $c_{\vartheta}(u, v) \leq M$  for any  $(u, v) \in [0, 1]^2$ . As a consequence,

$$s_{\vartheta}(t) \leq M \int_t^1 \frac{1}{v} dv = M(-\log t) = o\{t^{-1/2} \log^{-1}(1/t)\} \quad \text{as } t \rightarrow 0.$$

### 4.3.4 The Farlie–Gumbel–Morgenstern distributions

The members of this family are of the form  $C_{\omega}(u, v) = uv + \omega uv(1-u)(1-v)$ , where  $|\omega| \leq 1$ . These copulas can only be used to model small departures from independence, as can be seen from the value of Spearman's rho  $|\rho_S(\omega)| = |\omega|/3 \leq 1/3$ . Assumption  $\mathcal{A}_2$  holds since it is easy to see that  $s_{\omega}(t) = (1 - \omega + 2\omega t) \log(1/t) + 3\omega(1-t)(1-2t)$  is  $o\{t^{-1/2} \log^{-1}(1/t)\}$  as  $t \rightarrow 0$ .

### 4.3.5 Durante's class

Durante (2007) described a family of copulas with a non vanishing singular part in general. The members are written as  $C_{\psi}(u, v) = \min(u, v)\psi\{\max(u, v)\}$  in term of a non-decreasing generator function  $\psi : [0, 1] \rightarrow [0, 1]$  such that  $\psi(1) = 1$  and  $\psi(t)/t$  is non-increasing on  $(0, 1]$ . The class includes, as special cases, the independence copula when  $\psi(t) = t$  and the Fréchet–Hoeffding upper bound when  $\psi \equiv 1$ . See Table 5.1 in Appendix D for a description of some models.

Since  $C_{01}(u, v) = u\psi'(v)$  when  $u \leq v$  and  $C_{01}(u, v) = \psi(u)$  otherwise, one deduces from (4.4) that the associated Spearman's dependence function is

$$S_\psi(t) = 2t - t^{1/2} \psi(t^{1/2}) + 2t \int_{t^{1/2}}^1 \frac{\psi(\omega)}{\omega^2} d\omega. \quad (4.6)$$

Some computations enable to show that  $\int_0^1 \{1 - S_\psi(\omega)\} d\omega = \int_0^1 \omega^2 \psi(\omega) d\omega$  and an application of equation (4.3) yields the value of Spearman's rho as

$$\rho_\psi = 12 \int_0^1 \omega^2 \psi(\omega) d\omega - 3,$$

already reported by Durante (2007). A proof that the models in this class satisfy Assumption  $\mathcal{A}_2$  is given in Appendix C.

## 4.4 Empirical issues

### 4.4.1 Size and power of tests based on the process $\mathcal{S}_n$

Proposition 2 implies that all continuous functionals of the process  $\mathcal{S}_n$  will be consistent. This is a consequence of the continuous mapping Theorem. In particular, the Cramér–von Mises distance

$$\mathcal{V}_n = \int_0^1 \mathcal{S}_n^2(\omega) dS_{\hat{\theta}_n}(\omega) \quad (4.7)$$

will provide a consistent goodness-of-fit test statistic for families of copulas for which Assumptions  $\mathcal{A}_1$ – $\mathcal{A}_4$  hold. Another choice for a goodness-of-fit measure is the Kolmogorov–Smirnov distance, but evidence provided in the works of Genest, Quessy & Rémillard (2002, 2006) suggests that the latter is generally not as powerful as the Cramér–von Mises distance. Hence, for simplicity, only  $\mathcal{V}_n$  will be considered in the following. Note that in all subsequent applications, only families of copulas having one parameter will be considered, and the latter will be estimated by the inversion of Spearman's rho. Simulations to be found in Toupin

(2008) show that this method is good in terms of mean squared error, even for moderately small sample sizes.

An interesting feature of  $\mathcal{V}_n$  is the fact that this goodness-of-fit statistic admits an explicit representation as a finite sum involving the ranks of the observations. Specifically, developing the square of  $\mathcal{S}_n$  in equation (4.7) and using standard integration taking into account the jumps in this empirical distribution function, one finds

$$\mathcal{V}_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \{1 - S_{\hat{\theta}_n}(\xi_{i,n} \vee \xi_{j,n})\} - \sum_{i=1}^n \{1 - S_{\hat{\theta}_n}^2(\xi_{i,n})\} + \frac{n}{3},$$

where  $\xi_{i,n} = n^{-2}R_{i,n}S_{i,n}$ ,  $R_{i,n}$  is the rank of  $X_i$  among  $X_1, \dots, X_n$  and  $S_{i,n}$  is the rank of  $Y_i$  among  $Y_1, \dots, Y_n$ . This formula is easily computed when  $S_\theta$  admits an explicit representation. This is the case for the families of Durante B, Cuadras-Augé, Fréchet and the Ordinal sum defined in Table 5.1. In Table 4.1, the power of the test based on  $\mathcal{V}_n$  is evaluated with the help of 10 000 replicates. Since the distribution of  $\mathcal{V}_n$  under  $\mathcal{H}_0$  depends heavily on the unknown parameter value, a parametric bootstrap procedure is proposed. Once  $\mathcal{V}_n$  and  $\hat{\theta}_n$  are computed from the original sample, the idea is to generate  $M$  samples from  $C_{\hat{\theta}_n}$  and compute the goodness-of-fit statistics  $\mathcal{V}_{n,i}$ ,  $1 \leq i \leq M$  each time. By suitable adaptations of Genest & Rémillard (2008), it should be possible to prove that

$$\hat{p}_{M,n} = \frac{1}{M} \sum_{i=1}^M \mathbf{1}(\mathcal{V}_{n,i} > \mathcal{V}_n)$$

yields a valid p-value as  $n, M \rightarrow \infty$ . A formal justification is however beyond the scope of this article, though empirical evidence suggests that the result holds true. This method was used in the power study presented in Table 4.1, where the ability of  $\mathcal{V}_n$  to reject the null hypothesis under several copula alternatives is investigated.

Table 4.1: Percentage of rejection, based on 10 000 replicates, of the null hypothesis of belonging to a given family of copulas for the goodness-of-fit test based on  $\mathcal{V}_n$ ,  $n = 100, 250$

Families under $\mathcal{H}_1$		Families under $\mathcal{H}_0$								
		$\rho$	Fréchet		Cuadras-Augé		Durante B		Ordinal sum	
			100	250	100	250	100	250	100	250
Fréchet	0.1	5.0	4.7	6.8	7.3	8.9	10.0	17.0	24.4	
	0.3	5.7	5.2	10.3	13.2	18.7	30.6	61.5	93.2	
	0.5	5.4	5.2	10.7	17.0	25.5	51.1	94.5	100.0	
	0.7	5.8	5.5	9.4	15.5	25.1	57.2	99.8	100.0	
	0.9	7.5	6.4	8.1	9.5	10.7	31.7	99.7	100.0	
Cuadras-Augé	0.1	4.2	4.4	5.3	5.1	5.7	6.6	12.9	18.8	
	0.3	4.2	6.0	5.0	5.0	9.1	11.3	50.6	86.3	
	0.5	6.2	11.0	5.6	5.8	10.8	16.7	88.5	99.9	
	0.7	7.4	12.8	5.5	5.4	11.4	20.2	99.2	100.0	
	0.9	10.4	11.9	7.7	6.3	6.6	13.3	99.7	100.0	
Durante B	0.1	3.7	4.5	4.5	4.4	4.9	4.6	11.0	14.9	
	0.3	6.1	15.0	4.5	6.1	5.0	5.2	38.9	76.7	
	0.5	13.0	37.1	5.5	10.7	5.0	5.2	78.4	99.7	
	0.7	19.9	53.6	8.1	16.3	5.3	5.6	96.0	100.0	
	0.9	24.3	48.6	14.2	21.5	5.9	6.6	98.4	100.0	
Ordinal sum	0.1	4.6	11.6	5.3	10.1	5.2	9.3	6.1	5.4	
	0.3	48.5	97.1	39.5	93.5	31.3	87.5	5.4	5.4	
	0.5	97.6	100.0	93.7	100.0	85.9	100.0	6.1	5.9	
	0.7	100.0	100.0	100.0	100.0	99.8	100.0	5.5	5.3	
	0.9	100.0	100.0	100.0	100.0	100.0	100.0	6.0	6.0	
Clayton	0.1	9.1	12.1	12.9	19.5	15.5	26.1	26.1	41.8	
	0.3	31.9	58.0	49.8	83.8	65.4	95.2	90.6	99.8	
	0.5	64.2	94.2	83.0	99.5	94.2	100.0	100.0	100.0	
	0.7	90.4	99.8	96.7	100.0	99.6	100.0	100.0	100.0	
	0.9	99.8	100.0	99.9	100.0	99.9	100.0	100.0	100.0	
Frank	0.1	4.6	5.4	6.0	7.1	7.4	10.1	17.3	26.5	
	0.3	9.2	15.7	12.4	20.7	20.5	38.8	74.0	98.4	
	0.5	21.7	42.6	23.7	44.6	36.3	66.0	98.2	100.0	
	0.7	49.2	79.7	45.5	74.6	50.7	83.4	99.9	100.0	
	0.9	93.2	99.7	87.9	98.5	80.7	97.1	99.9	100.0	
Gauss	0.1	5.3	4.8	7.0	7.1	8.0	11.2	17.6	27.5	
	0.3	7.5	8.8	13.1	19.1	21.5	38.8	70.4	97.1	
	0.5	13.9	17.5	21.1	33.3	38.9	68.2	97.7	100.0	
	0.7	30.7	37.4	36.2	51.2	53.8	84.5	99.9	100.0	
	0.9	82.9	90.5	78.7	90.1	81.0	97.0	100.0	100.0	
Durante A	0.1	3.9	6.0	4.3	4.8	4.4	4.6	8.4	10.0	
	0.3	13.1	46.0	7.9	23.6	6.0	11.8	19.6	40.4	
	0.5	43.4	93.7	23.9	70.2	11.7	36.1	36.5	79.1	
	0.7	81.1	99.9	55.1	97.7	27.8	75.2	54.9	96.1	
	0.9	95.7	100.0	83.8	100.0	53.2	97.1	59.8	98.3	

For the members of the families of Gauss, Clayton, Frank and Durante A, Spearman's dependence function  $S_\theta$  has an expression that involves an integral to be computed by numerical routines. A way to avoid this time-consuming method is to rely on a bootstrap version of  $S_{\hat{\theta}_n}$ , which consists of generating an artificial sample  $(X_1^*, Y_1^*), \dots, (X_N^*, Y_N^*)$ , with  $N$  sufficiently large, from the bivariate

distribution  $C_{\hat{\theta}_n} \in \{C_\theta; \theta \in \mathcal{M}\}$ . Then,  $S_{\hat{\theta}_n}$  is approximated by

$$S_N(t) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\xi_{i,N}^* \leq t),$$

where  $\xi_{i,N}^* = N^{-2} R_{i,N}^* S_{i,N}^*$  and  $(R_{i,N}^*, S_{i,N}^*)$  are the pairs of ranks deduced from the artificial sample. Under conditions to be found in Genest & Rémillard (2008), the asymptotic distribution when  $n, N \rightarrow \infty$  of the Cramér–von Mises statistic computed from a bootstrap version  $\mathcal{S}_{n,N} = \sqrt{n}(S_n - S_N)$  of  $\mathcal{S}_n$ , namely

$$\begin{aligned} \mathcal{W}_{n,N} &= n \int_0^1 \{S_n(\omega) - S_N(\omega)\}^2 d\omega \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (1 - \xi_{i,n} \vee \xi_{j,n}) - \frac{2}{N} \sum_{i=1}^n \sum_{j=1}^N (1 - \xi_{i,n} \vee \xi_{j,N}^*) \\ &\quad + \frac{n}{N^2} \sum_{i=1}^N \sum_{j=1}^N (1 - \xi_{i,N}^* \vee \xi_{j,N}^*), \end{aligned}$$

is the same as that of

$$\mathcal{W}_n = n \int_0^1 \{S_n(\omega) - S_{\hat{\theta}_n}(\omega)\}^2 d\omega,$$

where  $x \vee y = \max(x, y)$ . Thus,  $\mathcal{W}_{n,N}$  provides an asymptotically equivalent and more computationally tractable version of  $\mathcal{W}_n$ . Note that the latter differs from  $\mathcal{V}_n$  in that the weight function  $dS_{\hat{\theta}_n}(\omega)$  is replaced by  $d\omega$ . In Table 4.2, the power of  $\mathcal{W}_{n,N}$  is evaluated under many copula alternatives.

Table 4.2: Percentage of rejection, based on 10 000 replicates, of the null hypothesis of belonging to a given family of copulas for the goodness-of-fit test based on  $\mathcal{W}_{n,N}$ ,  $n = 100, 250$  and  $N = 500$

Families under $\mathcal{H}_1$		Families under $\mathcal{H}_0$							
		Clayton		Frank		Gauss		Durante A	
		$\rho$	100	250	100	250	100	250	100
Clayton	0.1	4.8	4.7	6.9	8.0	7.1	7.3	17.9	23.8
	0.3	5.4	4.7	19.1	29.5	16.3	25.6	75.0	95.5
	0.5	5.2	4.8	30.3	55.2	28.6	50.5	98.9	100.0
	0.7	4.9	4.8	33.5	66.2	39.1	69.6	100.0	100.0
	0.9	5.0	5.3	21.1	44.2	34.8	69.3	100.0	100.0
Frank	0.1	5.6	7.1	4.5	4.5	5.1	5.2	11.2	14.8
	0.3	13.0	25.3	5.0	4.9	6.4	6.9	46.5	73.0
	0.5	22.9	49.7	5.0	5.1	8.0	9.9	81.2	98.2
	0.7	29.6	58.9	5.1	5.5	10.1	16.7	95.1	99.9
	0.9	12.6	31.4	4.0	5.3	8.7	17.6	98.7	100.0
Gauss	0.1	4.9	6.5	4.7	5.0	4.9	4.9	11.8	13.6
	0.3	10.9	21.3	5.6	6.5	5.2	5.2	42.1	65.3
	0.5	20.9	44.8	6.8	8.6	5.1	5.1	76.4	95.9
	0.7	28.8	62.3	7.8	14.4	5.1	4.8	93.6	99.8
	0.9	19.2	54.9	7.9	16.8	4.0	4.8	99.3	100.0
Durante A	0.1	9.5	19.2	6.5	12.0	6.5	11.2	5.1	4.6
	0.3	59.7	94.0	30.9	68.0	26.0	58.1	5.4	5.6
	0.5	94.6	100.0	58.4	95.7	46.8	89.9	5.7	5.2
	0.7	99.3	100.0	62.6	98.7	53.3	96.1	5.1	5.4
	0.9	63.9	100.0	8.1	62.3	11.9	69.6	4.4	5.1
Fréchet	0.1	5.2	7.0	4.9	5.3	5.2	5.3	10.0	10.8
	0.3	12.3	28.6	6.3	9.9	4.3	5.5	30.4	47.9
	0.5	22.8	59.1	5.6	15.2	2.3	3.9	57.4	85.8
	0.7	27.6	75.9	3.4	16.2	0.6	1.5	74.5	97.9
	0.9	5.8	46.6	0.3	3.8	0.0	0.1	76.6	98.6
Cuadras-Augé	0.1	6.6	9.8	4.8	5.8	4.7	5.4	8.2	8.4
	0.3	21.7	49.1	7.3	13.7	5.1	9.5	20.0	29.2
	0.5	40.4	83.6	7.1	20.4	3.9	9.7	39.1	64.1
	0.7	44.9	91.9	3.8	19.5	4.5	5.1	58.3	88.9
	0.9	10.1	58.9	0.3	4.0	0.0	0.2	63.3	94.6
Durante B	0.1	7.3	12.4	4.6	7.3	5.3	6.4	7.3	6.5
	0.3	33.2	69.1	10.4	24.1	8.6	18.4	12.9	16.7
	0.5	62.5	96.4	12.9	37.4	9.5	25.9	23.0	36.6
	0.7	68.2	98.8	7.4	33.4	4.4	16.9	35.3	62.8
	0.9	16.0	73.1	0.2	4.3	0.1	1.0	42.8	76.8
Ordinal sum	0.1	19.3	42.6	15.5	36.1	14.8	33.4	5.6	9.5
	0.3	93.6	100.0	89.2	99.9	84.4	99.8	24.2	58.8
	0.5	100.0	100.0	99.9	100.0	99.6	100.0	53.0	92.1
	0.7	100.0	100.0	99.7	100.0	99.9	100.0	71.2	98.5
	0.9	91.2	100.0	68.1	100.0	79.3	100.0	66.9	98.5

Looking at Table 4.1 and Table 4.2, one sees that the statistics  $\mathcal{V}_n$  and  $\mathcal{W}_{n,N}$  keep their 5% level in most cases, taking into account the  $\pm 1\%$  uncertainty in estimating proportions with 10 000 replicates. As expected, the powers increase with the sample size. From Table 4.1, one sees that the bivariate distributions of Fréchet, Cuadras-Augé and Durante B are similar models, at least in terms of their Spearman dependence function, as can be seen from the low power that  $\mathcal{V}_n$  exhibits. The Ordinal sum, however, seems to be very different from all the

models considered under  $\mathcal{H}_0$ . Conversely, all models are highly rejected under the null hypothesis of belonging to the family of Ordinal sum copulas. A curious phenomenon occurs in Table 4.2 for Frank and Gaussian null hypotheses under Fréchet, Cuadras–Augé and Durante B alternatives, especially when these singular models are simulated under moderate and high levels of dependence. Indeed, some of the estimated powers are less than the 5% nominal level of the tests. This could probably be explained by the particular form of the underlying distribution of  $\mathcal{W}_{n,N}$  under these alternatives. As the sample size increases, however, this problem is less noteworthy.

#### 4.4.2 Comparisons with Kendall and empirical copula processes

Wang & Wells (2000) and Genest, Quessy & Rémillard (2006) based a goodness-of-fit test for copulas on Kendall's dependence function  $K(t) = P\{H(X, Y) \leq t\}$ , where  $(X, Y) \sim H$ , whose name derived from the relationship between Kendall's measure of association and the mean of  $K$ . Note that  $K$  depends only on the underlying copula  $C$  of  $H$  since one can show easily that  $K(t) = P\{C(U, V) \leq t\}$ , with  $U = F(X)$  and  $V = G(Y)$  having joint distribution  $C$ . These authors used the empirical process  $\mathcal{K}_n = \sqrt{n}(K_n - K_{\hat{\theta}_n})$  defined on  $[0, 1]$  as a base to define goodness-of-fit statistics for families of copulas. Here,  $K_n$  is the empirical distribution function of the pseudo-observations

$$\kappa_{i,n} = \frac{1}{n-1} \sum_{j=1}^n \mathbf{1}(X_j < X_i, Y_j < Y_i), \quad 1 \leq i \leq n.$$

Under conditions that are satisfied for a large number of commonly used copula models, Genest, Quessy & Rémillard (2006) show that  $\mathcal{K}_n$  converges weakly to a limiting centered Gaussian process.

Rather than relying on a projection of a bivariate copula into a space of univariate functions, as is the case when using Spearman or Kendall functions  $S(\cdot)$  and

$K(\cdot)$ , one can work directly with a consistent estimation of an unknown copula  $C$ . Specifically, the empirical copula

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{1} \left( \frac{R_{i,n}}{n} \leq u, \frac{S_{i,n}}{n} \leq v \right)$$

is such that  $\sqrt{n}(C_n - C)$  converges weakly to a continuous Gaussian process provided the partial derivatives  $C_{10}$  and  $C_{01}$  exist and are continuous. Deheuvels (1979) was the first to study this process in the special case of independence and obtained its weak consistency in that context. This result was extended in the general case by Gaenssler & Stute (1987) and have recently been reconsidered by Fermanian, Radulović & Wegkamp (2004) and Tsukahara (2005). In a goodness-of-fit setting, Genest, Rémillard & Beaudoin (2008) suggested to use  $C_{n,N} = \sqrt{n}(C_n - C_N)$ , where  $C_N$  is a bootstrap estimation of  $C_{\hat{\theta}_n}$  based on an artificial random sample from  $C_{\hat{\theta}_n}$ .

In Table 4.3, the results of a power study comparing the Cramér–von Mises statistics derived from  $\mathcal{S}_n$ ,  $\mathcal{S}_{n,N}$ ,  $\mathcal{K}_n$  and  $\mathcal{C}_{n,N}$  are offered. To apply the test based on  $\mathcal{K}_n$  for a given member of Durante’s class with generator function  $\psi$ , note that

$$K_\psi(t) = t - t \log t + t \log \left\{ \frac{\Psi^{-1}(t)}{\psi \circ \Psi^{-1}(t)} \right\},$$

where  $\Psi(x) = x\psi(x)$ . In particular,  $K_\psi(t) = t - 2(1 - \gamma)(2 - \gamma)^{-1}t \log t$  for the Cuadras–Augé family. However, the null hypothesis of belonging to this singular copula cannot be tested using  $\mathcal{C}_{n,N}$  since the first-order partial derivatives are not continuous. For the test based on  $\mathcal{K}_n$ , it is not possible to use it for the Gaussian copula, at least in the form presented herein, since no explicit expression is known for the associated Kendall’s dependence function. A bootstrap version  $\mathcal{K}_{n,N}$  of  $\mathcal{K}_n$  has however been proposed by Genest & Rémillard (2008) but will not be studied here.



Table 4.3: Percentage of rejection, based on 10 000 replicates, of the null hypothesis of belonging to a given family of copulas for the Cramér–von Mises goodness-of-fit statistics computed from Spearman, Kendall and empirical copula processes  $\mathcal{S}_n$  (Cuadras–Augé),  $\mathcal{S}_{n,N}$  (Clayton, Frank, Gauss),  $\mathcal{K}_n$  and  $\mathcal{C}_{n,N}$  when  $n = 250$

Families under $\mathcal{H}_1$		$\rho$		Families under $\mathcal{H}_0$							
				Clayton			Frank			Cuadras–Augé	
		$\mathcal{S}_{n,N}$	$\mathcal{K}_n$	$\mathcal{C}_{n,N}$	$\mathcal{S}_{n,N}$	$\mathcal{K}_n$	$\mathcal{C}_{n,N}$	$\mathcal{S}_n$	$\mathcal{K}_n$	$\mathcal{S}_{n,N}$	$\mathcal{C}_{n,N}$
Clayton	0.3	4.7	5.1	7.9	29.5	86.4	23.8	83.8	93.1	25.6	22.1
	0.5	4.8	4.8	7.8	55.2	99.3	53.0	99.5	99.9	50.5	49.0
	0.7	4.8	4.7	7.5	66.2	100.0	77.3	100.0	100.0	69.6	74.1
Frank	0.3	25.3	51.4	26.1	4.9	5.7	8.2	20.7	14.0	6.9	9.4
	0.5	49.7	91.7	53.3	5.1	5.2	7.6	44.6	28.8	9.9	11.8
	0.7	58.9	99.4	76.4	5.5	4.5	7.8	74.6	98.4	16.7	18.9
Cuadras–Augé	0.3	49.1	66.8	55.6	13.7	10.4	19.9	5.0	5.1	9.5	22.7
	0.5	83.6	98.2	94.3	20.4	49.0	48.8	5.8	5.0	9.7	57.1
	0.7	91.9	100.0	99.9	19.5	100.0	88.1	5.4	4.4	5.1	92.2
Gauss	0.3	21.3	34.3	23.5	6.5	12.2	8.2	19.1	19.0	5.2	7.8
	0.5	44.8	77.4	48.2	8.6	21.0	9.5	33.3	36.1	5.1	7.6
	0.7	62.3	97.0	75.5	14.4	35.9	16.0	51.2	99.6	4.8	7.6

A quick look at Table 4.3 leads to the conclusion that the test based on  $\mathcal{K}_n$  is the best in all situations, except maybe when testing the null hypothesis of a Cuadras–Augé copula under Frank alternatives. The tests based on  $\mathcal{S}_n$  and  $\mathcal{C}_{n,N}$  often show equivalent power, but the latter seems to have problem keeping its nominal level under  $\mathcal{H}_0$  in all three situations where it is considered.

#### 4.4.3 Testing for a bivariate Gaussian dependence structure

Despite much evidence suggesting that multivariate normality is rather an exception than the norm in many natural and financial phenomena, the Gaussian dependence structure is still a popular choice for modelers. In the case of classical multivariate normality, many invariant test statistics exist. For a review article on the subject, see Henze (2002).

In a more flexible, semi-parametric context where one wants to assess that the

distribution of a pair  $(X, Y)$  has a Gaussian dependence structure but not necessarily normal marginals, one has to rely in general on rank-based methods. One such procedure was introduced by Malevergne & Sornette (2003). These authors first defined  $m_R(x, y) = (x, y)R^{-1}(x, y)'$  and noted that  $m_R(X, Y)$  follows a chi-square distribution with two degrees of freedom whenever  $(X, Y)$  follows a standard bivariate normal law with positive definite correlation matrix  $R$ . More generally, when the copula of  $(X, Y)$  is Gaussian with arbitrary marginals  $F$  and  $G$ ,  $m_R(\tilde{X}, \tilde{Y})$  follows a chi-square law with two degrees of freedom, where  $\tilde{X} = \Phi^{-1}\{F(X)\}$  and  $\tilde{Y} = \Phi^{-1}\{G(Y)\}$ . Hence, provided  $F$ ,  $G$  and  $R$  are known,  $\zeta_i = m_R(\tilde{X}_i, \tilde{Y}_i)$ ,  $1 \leq i \leq n$  are i.i.d.  $\chi_2^2$ . The idea of Malevergne & Sornette (2003) was to replace  $F$  and  $G$  by their empirical counterparts  $F_n$  and  $G_n$ , thus implicitly defining the pseudo-observations  $\zeta_{i,n} = m_{\hat{R}_n}(\tilde{X}_{i,n}, \tilde{Y}_{i,n})$ ,  $1 \leq i \leq n$ , where

$$\tilde{X}_{i,n} = \Phi^{-1}\{F_n(X_i)\}, \quad \tilde{Y}_{i,n} = \Phi^{-1}\{G_n(Y_i)\}$$

and  $\hat{R}_n$  is an estimator of  $R$ . They then defined  $L_n$  as the empirical distribution of  $\zeta_{1,n}, \dots, \zeta_{n,n}$  and  $L(x) = P(\chi_2^2 \leq x)$ . They argued that despite the use of the ranks instead of the observations themselves,  $L_n$  should accurately estimate  $L$ . They then proposed to define goodness-of-fit statistics as functional distances between  $L_n$  and  $L$ . However, showing the weak consistency of the process  $\mathcal{L}_n = \sqrt{n}(L_n - L)$  seems to be a very difficult problem that is still unsolved.

Table 4.4 reports the powers of the Cramér–von Mises statistics of  $\mathcal{S}_{n,N}$  and  $\mathcal{L}_n$  under observations simulated from the Student copula with  $\nu = 3$  degrees of freedom, as well as under Clayton, Ordinal sum, Durante A and Durante B alternatives. Note that the Student copula, which is derived from the bivariate Student distribution, is parameterized in terms of the usual correlation coefficient since no closed form expression exists for the relationship between the latter and Spearman's rho.

Table 4.4: Percentage of rejection, based on 10 000 replicates, of the null hypothesis of a Gaussian copula for the Cramér-von Mises goodness-of-fit statistics computed from processes  $\mathcal{S}_{n,N}$  and  $\mathcal{L}_n$  when  $n = 250$

under $\mathcal{H}_1$	$\rho$	$\mathcal{S}_{n,N}$	$\mathcal{L}_n$	under $\mathcal{H}_1$	$\rho$	$\mathcal{S}_{n,N}$	$\mathcal{L}_n$
Clayton	0.25	25.6	10.2	Ordinal sum	0.25	99.8	33.8
	0.50	50.5	20.5		0.50	100.0	46.0
	0.75	69.6	63.5		0.75	100.0	98.9
Student $\nu = 3$	*0.25	9.7	90.3	Durante A	0.25	58.1	66.0
	*0.50	6.9	89.5		0.50	89.9	97.0
	*0.75	4.5	87.1		0.75	96.1	100.0
Student $\nu = \infty$ (Gauss)	0.25	5.2	5.4	Durante B	0.25	18.4	85.9
	0.50	5.1	4.7		0.50	25.9	100.0
	0.75	4.8	4.6		0.75	16.9	100.0

\* : classical correlation coefficient

From Table 4.4, one can see that these two competitors exhibit some kind of complementarity. In other words, the power of the test based on  $\mathcal{S}_{n,N}$  is often high while that using  $\mathcal{L}_n$  is poor, and vice versa. In future investigations, it could be interesting to study goodness-of-fit procedures that combine several test statistics. One such possibility, for example, would be to take the maximum among a set of goodness-of-fit statistics or to consider linear combinations of the latter.

## 4.5 Illustrations

### 4.5.1 Modelling of a hydrological data set

This section is devoted to the complete modelling of a data set consisting of thirty-eight pairs of mean annual flows (in  $\text{m}^3 / \text{sec.}$ ) taken between 1951 and 1988 at two points on St-Maurice's river, located in the south region of the Province of Quebec, in Canada : at the *Gouin station* and at the *Rapide-blanc station*. Specifically, goodness-of-fit procedures will be employed in order to find a bivariate model with satisfactory marginals and copula for  $X$ : *annual flow at Gouin station* and  $Y$ : *annual flow at Rapide-blanc station*.

Sklar's theorem puts in highlight the fact that one can fit univariate distributions

to  $X$  and  $Y$  and independently try to find an appropriate copula for the dependence structure of  $(X, Y)$ . The theory of goodness-of-fit tests in the univariate case is well developed. See, for instance, Durbin (1973) and Stephens (1974). To test for  $\mathcal{H}_0^{(X)} : F \in \{F_\alpha; \alpha \in A\}$  and  $\mathcal{H}_0^{(Y)} : G \in \{G_\beta; \beta \in B\}$ , Cramér–von Mises functionals of the empirical processes with unknown parameters  $\mathcal{F}_n = \sqrt{n}(F_n - F_{\hat{\alpha}_n})$  and  $\mathcal{G}_n = \sqrt{n}(G_n - G_{\hat{\beta}_n})$  will be used as test statistics. Here,  $\hat{\alpha}_n$  and  $\hat{\beta}_n$  are estimators of  $\alpha$  and  $\beta$  respectively, and subject to some regularity conditions. As a first step, four univariate densities have been proposed to fit the histograms of  $X$  and  $Y$  in Figure 4.1, namely the normal, log-normal, logistic and Pareto distributions. The value of the goodness-of-fit statistics are reported in Table 4.5 together with an associated p-value estimated with the help of 10 000 parametric bootstrap samples. Many models in the list are accepted at the 5 % level, and the one having the highest p-value, namely the log-normal distribution, is selected in both cases.

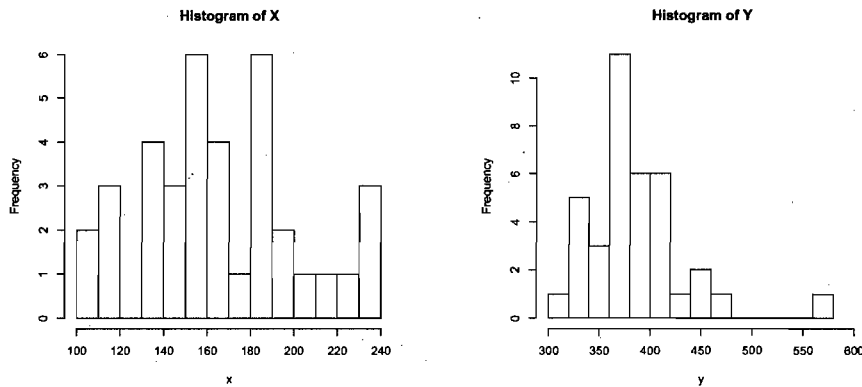


Figure 4.1: Histograms of  $X$ : mean annual flow ( $m^3/s$ ) at Gouvin station and  $Y$ : mean annual flow ( $m^3/s$ ) at Rapide blanc station

In Figure 4.2, the scatter plot of the original observations, as well as that of the standardized ranks, i.e. the pairs  $(F_n(X_i), G_n(Y_i))$ , are presented. While it is not entirely clear which strength of dependence is present in the data looking at the left panel alone, one can suspect that the dependence between  $X$  and  $Y$  is

Table 4.5: Results of univariate goodness-of-fit tests for  $X$ : mean annual flow ( $m^3/s$ ) at Gouin station and  $Y$ : mean annual flow ( $m^3/s$ ) at Rapide blanc station

Model	$X$		$Y$	
	$T_n$	p-value	$T_n$	p-value
Normal	0.0363	0.7438	0.1533	0.0210
Log-normal	0.0267	0.9036	0.0836	0.1853
Logistique	0.0519	0.5951	0.1210	0.1097
Pareto	1.5081	0.0004	1.2444	0.0002

rather high, as indicated by the value of Spearman's rho, namely  $\rho_n = 0.62$ . The latter is nothing but the slope of the linear regression using the standardized ranks plotted on the right panel of Figure 4.2 and deduced from the pairs plotted on the left panel. In order to choose a suitable copula for  $(X, Y)$ , the goodness-of-fit tests based on  $\mathcal{V}_n$  and  $\mathcal{W}_{n,N}$  have been considered for eight families of dependence models. The results are reported in Table 4.6. All models are accepted at the 5% level, except the Ordinal sum copula.

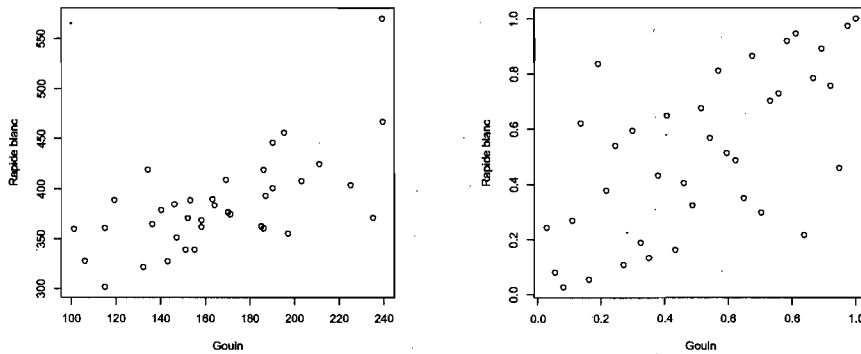


Figure 4.2: Mean annual flow at Rapide blanc station versus Gouin station. Left panel: scatter plot of the observations; right panel: scatter plot of the normalized ranks.

An interesting feature of the empirical and theoretical counterparts of Spearman's dependence function is the possibility to define a graphical tool for goodness-of-fit purposes. The idea is to graph both the nonparametric and parametric versions

Table 4.6: Results of goodness-of-fit tests for eight copula families for the joint behavior of  $X$ : mean annual flow at *Gouin* station and  $Y$ : mean annual flow at *Rapide blanc* station

Model under $\mathcal{H}_0$	$\hat{\theta}_n$	$\mathcal{V}_n$ or $\mathcal{W}_{n,N}$	Estimated p-value
Gauss	0.6346	0.0358	0.6443
Clayton	1.6032	0.0658	0.1331
Frank	4.6692	0.0382	0.5838
Durante A	2.0559	0.0502	0.0910
Durante B	2.9177	0.0546	0.1454
Cuadras–Augé	0.6820	0.0450	0.2509
Fréchet	0.6166	0.0388	0.3263
Ordinal sum	1.3766	0.1255	0.0067

$S_n$  and  $S_{\hat{\theta}_n}$ , which should be close in all points of  $(0, 1)$  under  $\mathcal{H}_0$ . For the mean annual flows at *Gouin* and *Rapide-blanc* stations, the normal copula was accepted as an appropriate model, while the Ordinal sum was strongly rejected. This could have been detected looking at the left and right panels of Figure 4.3.

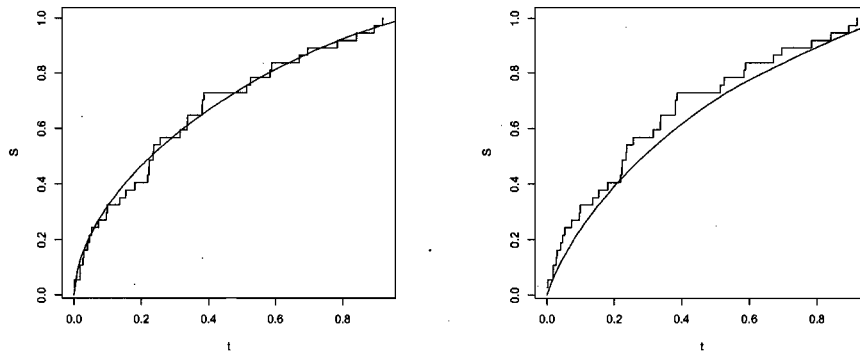


Figure 4.3: Graphs of  $S_n(t)$  versus the parametric versions  $S_{\hat{\theta}_n}(t)$  for Gaussian (left panel) and Ordinal sum (right panel) copulas.

### 4.5.2 Modelling of a multivariate data set

Few methods to model efficiently high dimensional data are available in the literature about copulas. In fact, many of the proposed multivariate extensions of some of the most popular bivariate copula models have very few parameters. For example, a possible generalization of the members of the two-dimensional Archimedean family is

$$C_{\phi_{\theta}}(u_1, \dots, u_d) = \phi_{\theta}^{-1} \left\{ \sum_{i=1}^d \phi_{\theta}(u_i) \right\}.$$

In this latter model, all degrees of dependence, i.e. two-dimensional, three-dimensional, and so on, are modelled by the same parameter. This is very restricting since most situations in practice exhibit more complex dependence patterns. In order to alleviate this problem, some attempts have been made to extend bivariate Archimedean copulas to more complex multivariate structures, e.g. Whelan (2004) and Savu & Trede (2006). A limitation for using these models is the many restrictions imposed on the parameters, which causes estimation problems. A promising way based on a pair-copula construction has been described by Aas, Czado, Frigessi & Bakken (2008), but little is known about the validity of the estimation and goodness-of-fit procedures they proposed.

A possibility for multivariate modelling, though not entirely satisfactory, consists in fitting appropriate bivariate models to all the possible pairs of variables. This method is illustrated here, where tests based on  $\mathcal{V}_n$  and  $\mathcal{W}_{n,N}$  are performed on a data set about six characteristics of 382 small and medium-size company (SME) in the province of Quebec. The variables, whose values come from the financial statements of each company, are  $X_1$ : *liquidity expressed as a percentage of the sales*,  $X_2$ : *rate of output of the credit*,  $X_3$ : *clear margin*,  $X_4$ : *gross margin*,  $X_5$ : *rate of export* and  $X_6$ : *average growth rate of the sales*. For the fifteen possible pairs, the quality of the fit to eight dependence models has been investigated. The results are presented in Table 4.7.

Table 4.7: Results of pairwise goodness-of-fit tests based on  $\mathcal{V}_n$  and  $\mathcal{W}_{n,N}$  for the SME data set. The upper triangle indicates the value of Spearman's rho for each pair and the lower triangle gives the models among a. Gaussian, b. Clayton, c. Frank, d. Durante A, e. Durante B, f. Cuadras–Augé, g. Fréchet and h. Ordinal sum for which the null hypothesis is accepted

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$
$X_1$	—	0.66	0.86	0.42	0.28	0.27
$X_2$	$\emptyset$	—	0.87	0.33	0.24	0.36
$X_3$	a	a	—	0.42	0.27	0.32
$X_4$	a,b,c,f,g	a,c,f,g	a,c,e,f,g	—	0.28	0.14
$X_5$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	—	0.29
$X_6$	a,b	a,b,c,g	a,b,c	a,b,c,d,g	$\emptyset$	—

For all the pairs involving  $X_5$  and the pair  $(X_1, X_2)$ , no copula in the list is accepted as a suitable model. For a more complete analysis, other dependence structure should be tested, e.g. extreme-value copulas. In all other cases, the Gaussian copula is accepted at the 5% level, which indicates that a multivariate Gaussian dependence structure could provide a good approximation to the distribution of  $(X_1, \dots, X_6)$ . This has to be tested separately for two main reasons : firstly, only the pairs are considered here, hence no general conclusions on the joint normality is possible and secondly, the size of the pairwise tests would have to be adjusted in order to achieve a prescribed global level. The assumption of multivariate normality could be however tested directly using a multivariate version of Spearman's goodness-of-fit process, which is the subject of the next section.

## 4.6 Multivariate extension of $\mathcal{S}_n$

A multivariate version of the process  $\mathcal{S}_n$  is straightforward. To this end, let  $(X_{1i}, \dots, X_{di}), 1 \leq i \leq n$  be independent copies of a random vector  $(X_1, \dots, X_d)$  with marginals  $F_1, \dots, F_d$  and copula  $C$ . The multivariate version of Spearman's dependence function can then be defined as  $S_d(t) = P \{F_1(X_1) \times \dots \times F_d(X_d) \leq t\}$ ,



whose empirical counterpart is based on the distribution  $S_{n,d}$  of the pseudo-observations  $\xi_{i,n,d} = F_{n1}(X_{1i}) \times \cdots \times F_{nd}(X_{di})$ . Then, under Assumptions  $\mathcal{A}_1$ – $\mathcal{A}_4$ , the process  $\mathcal{S}_{n,d} = \sqrt{n}(S_{n,d} - S_{\hat{\theta}_n})$  converges to a Gaussian limit  $\mathcal{S}_d$ . In the case when  $\theta$  is real-valued, the estimation can be based on an inversion of one of the multivariate extensions of Spearman's rho described by Schmid & Schmidt (2007).

To illustrate the method, take the random vector  $(X_1, X_3, X_4, X_6)$  in the context of the dependence modelling of the SME data set of subsection 5.2. As can be seen from Table 4.7, the null hypothesis of a Gaussian copula is accepted for each of the six possible pairs. Hence it is suspected that the vector has a multivariate normal dependence structure with some correlation matrix  $R$ . The empirical Spearman dependence function  $S_{n,d}$  based on the 382 independent copies of  $(X_1, X_3, X_4, X_6)$  was compared to the approximation of  $S_{\hat{R}_n}$  using  $N = 10\,000$  bootstrap samples. The entries of  $\hat{R}_n$  were based on the inversion of Spearman's rho for each pair. The value of the Cramér–von Mises statistic computed from  $\mathcal{S}_{n,d}$ , namely  $\mathcal{W}_{n,N} = 0.0215$ , had an estimated p-value of 0.5780 (based on 5 000 replicates), which suggests that the Gaussian copula is a good model for these data.

## CHAPITRE 5

### CONCLUSION ET DISCUSSION

L'approche par les copules dans le domaine de la modélisation multivariée s'impose depuis plus d'une quinzaine d'années. Quelques auteurs se sont intéressés plus particulièrement à la question du choix de la copule adéquate qui permet de modéliser la structure de dépendance entre des variables aléatoires. Citons notamment Genest & Rivest (1993), Chen & Fan (2005), Fermanian (2005) et Genest, Quessy & Rémillard (2006). Le but principal de ce mémoire fut de développer une nouvelle procédure d'adéquation pour les copules. À cette fin, la fonction de dépendance de Spearman a été introduite et la convergence en loi d'un processus empirique associé a été démontrée. Une statistique de test qui s'écrit comme fonctionnelle de ce processus a ensuite été définie et étudiée.

Le test proposé dans ce travail possède quelques avantages. Notamment, plusieurs familles populaires de copules peuvent être testées, incluant les copules gaussienne, Clayton, Frank, Farlie–Gumbel–Morgenstern ainsi que les modèles appartenant à la famille de Durante (2007). Le test est également facile d'application puisque la statistique associée est facilement calculable à partir des rangs des observations. De plus, la version empirique de la fonction de dépendance de Spearman fournit un outil graphique pour évaluer la qualité de l'ajustement d'un jeu de donnée à une famille de copules.

Malgré que l'étude de puissance montre que le test basé sur le processus de Kendall est souvent plus puissant, le test basé sur le processus de Spearman est néanmoins

efficace sous plusieurs contre-hypothèses. De plus, celui-ci permet facilement de vérifier l'adéquation des familles de copules qui peuvent difficilement l'être avec le test basé sur le processus de Kendall. On pense particulièrement à la copule normale et aux modèles de la famille générale de Durante (2007).

Les outils mathématiques utilisés pour montrer la convergence du processus proposé sont typiques du genre d'arguments nécessaires pour établir le comportement asymptotique de statistiques d'adéquation à des modèles semi-paramétriques. Ceux-ci pourraient être adaptés afin d'étudier d'autres processus d'adéquation. Par exemple, pour  $J[0, 1]^2 \rightarrow [0, 1]$ , soient les variables dépendantes

$$e_{i,n}^J = J\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right), \quad 1 \leq i \leq n,$$

auxquelles on associe la fonction de répartition empirique

$$S_n^J(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(e_{i,n}^J \leq t).$$

Cette fonction est un estimateur de  $S_\theta^J(t) = P\{J(U, V) \leq t\}$ , où  $(U, V) \sim C_\theta$ . Lorsque  $J(u, v) = uv$ , on retrouve le cas de la fonction de dépendance de Spearman mais en général,

$$\int_0^1 \{1 - S_n^J(t)\} dt = \frac{1}{n} \sum_{i=1}^n J\left(\frac{R_i}{n}, \frac{S_i}{n}\right)$$

est la définition d'une *statistique linéaire de rangs*. En particulier, quand  $J(u, v) = \Phi^{-1}(u)\Phi^{-1}(v)$ , on retrouve la statistique de Van der Waerden et pour  $J(u, v) = (\ln u + 1)(\ln v + 1)$ , on a la statistique de Savage. Plus généralement que dans le cadre de ce travail, on pourrait étudier le *processus linéaire de rangs*

$$\mathcal{S}_n^J(t) = \sqrt{n}\{S_n^J(t) - S_{\hat{\theta}_n}^J(t)\}$$

et considérer la statistique de test

$$V_n^J = \int_0^1 \{\mathcal{S}_n^J(t)\}^2 dS_{\hat{\theta}_n}^J(t),$$

qui s'écrira explicitement en fonction des pseudo-observations  $e_{1,n}^J, \dots, e_{n,n}^J$ .

# ANNEXE A

## PROOF OF PROPOSITION 1

Proposition 1 will be established if Hypotheses II and III of Ghoudi & Rémillard (1998) are satisfied, since Hypothesis I is an immediate consequence of Assumption  $\mathcal{A}_1$ . To check Hypothesis II, let  $\tilde{F}_n$  and  $\tilde{G}_n$  be the empirical functions obtained from a linear interpolation between the steps of  $F_n$  and  $G_n$ . If  $I = [0, 1]$ , it is then clear that

$$\sup_{x \in I} \left| \tilde{F}_n(x) - F_n(x) \right| \leq \frac{1}{n} \quad \text{and} \quad \sup_{y \in I} \left| \tilde{G}_n(y) - G_n(y) \right| \leq \frac{1}{n}.$$

If one defines  $\tilde{H}_n(x, y) = \tilde{F}_n(x)\tilde{G}_n(y)$ , this version of  $H_n$  is continuous and

$$\begin{aligned} & \sqrt{n} \sup_{(x,y) \in I^2} \left| \tilde{H}_n(x, y) - H_n(x, y) \right| \\ & \leq \sqrt{n} \sup_{y \in I} \left| \tilde{G}_n(y) \right| \sup_{x \in I} \left| \tilde{F}_n(x) - F_n(x) \right| + \sqrt{n} \sup_{x \in I} \left| F_n(x) \right| \sup_{y \in I} \left| \tilde{G}_n(y) - G_n(y) \right| \\ & \leq \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} = \frac{2}{\sqrt{n}}, \end{aligned}$$

which establishes the first part of Hypothesis II.

Standard arguments that can be found in Billingsley (1968) enable to obtain that the following sequence is tight for any continuous function  $\varphi$  on  $\mathbb{R}$  :

$$\alpha_{n,\varphi}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\varphi(X_i, Y_i) \mathbf{1}(X_i Y_i \leq t) - \mathbb{E} \{ \varphi(X_i, Y_i) \mathbf{1}(X_i Y_i \leq t) \}].$$

In particular,

$$\sup_{t \in B} \left| \alpha_{n,\varphi} \left( t + \frac{s}{\sqrt{n}} \right) - \alpha_{n,\varphi}(t) \right|$$

converges in probability to 0 for all compact measurable sets  $B$  of  $(0, 1]$ . The next result establishes the joint behavior of  $(\alpha_n, \mathbb{H}_n)$ , where  $\alpha_n = \alpha_{n,1}$  and  $\mathbb{H}_n = \sqrt{n}(\tilde{F}_n \tilde{G}_n - FG)$ , which shows that the last part of Hypothesis II holds. Its proof, which is omitted, uses standard arguments in the theory of empirical processes based on i.i.d. observations. The computation of the covariance functions is lengthy but straightforward.

PROPOSITION 3. *The empirical processes*

$$\alpha_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\mathbf{1}(\xi_i \leq t) - S(t)\}$$

and

$$\mathbb{H}_n(x, y) = \sqrt{n} \left\{ \tilde{F}_n(x) \tilde{G}_n(y) - F(x)G(y) \right\}$$

converge jointly to a limit  $(\alpha, \mathbb{H})$  with covariance structure

$$\Gamma_\alpha(t_1, t_2) = \text{cov} \{ \alpha(t_1), \alpha(t_2) \} = S(t_1 \wedge t_2) - S(t_1)S(t_2),$$

$$\begin{aligned} \Gamma_{\alpha, \mathbb{H}}(t, x, y) &= \text{cov} \{ \alpha(t), \mathbb{H}(x, y) \} \\ &= F(x)P(UV \leq t, Y \leq y) + G(y)P(UV \leq t, X \leq x) \\ &\quad - 2F(x)G(y)S(t) \end{aligned}$$

and

$$\begin{aligned} \Gamma_{\mathbb{H}}(x, y, x', y') &= \text{cov} \{ \mathbb{H}(x, y), \mathbb{H}(x', y') \} \\ &= F(x)F(x')G(y \wedge y') + F(x)G(y')C\{F(x'), G(y)\} \\ &\quad + F(x')G(y)C\{F(x), G(y')\} + G(y)G(y')F(x \wedge x') \\ &\quad - 4F(x)F(x')G(y)G(y'), \end{aligned}$$

where  $U = F(X)$  and  $V = G(Y)$ .

Let  $\mathcal{Q}_*$  be the set of positive and increasing functions in a neighborhood of zero such that  $q_* \in \mathcal{Q}_*$  satisfy that  $q_*(t)/t$  is decreasing and  $q_*(2t)/q_*(t)$  is bounded

above. In order to verify Hypothesis III, one needs a function  $q_\star \in \mathcal{Q}_\star$  and a sequence  $(t_n)$  of positive numbers that converges to zero for which

$$\lim_{n \rightarrow \infty} \sqrt{n} S(t_n) = 0, \quad \lim_{n \rightarrow \infty} \frac{q_\star(t_n)}{t_n \sqrt{n}} = 0, \quad \lim_{t \rightarrow 0} s(t) q_\star(t) = 0 \quad (5.1)$$

and such that the sequence

$$\mathcal{B}_n = \left\{ \sup_{xy > t_n} \frac{|\mathbb{H}_n(x, y)|}{q_\star(xy)} \right\}_{n \geq 1}$$

is tight. To this end, the next lemma will be instrumental.

LEMMA A.1. *If  $s = s_\theta$  satisfies Assumptions  $\mathcal{A}_2$ ,  $\tilde{q}_\star(t) = t^{1/2} \log(1/t)$  belongs to  $\mathcal{Q}_\star$  and satisfies the relations (5.1) for  $t_n = (\log n)^3/n$ .*

If  $\beta_{n1}(x) = \sqrt{n} \{F_n(x) - F(x)\}$  and  $\beta_{n2}(y) = \sqrt{n} \{G_n(y) - G(y)\}$ , classical results in the theory of empirical processes imply that  $(\beta_{n1}, \beta_{n2})$  converges to  $(\beta_1, \beta_2)$ , where  $\beta_1, \beta_2$  are centered limiting processes such that  $\mathbb{E}\{\beta_1(s)\beta_1(t)\} = F(s \wedge t) - F(s)F(t)$ ,  $\mathbb{E}\{\beta_2(s)\beta_2(t)\} = G(s \wedge t) - G(s)G(t)$  and  $\mathbb{E}\{\beta_1(s)\beta_2(t)\} = C\{F(s), G(t)\} - F(s)G(t)$ . If  $\mathbb{H}(x, y) = G(y)\beta_1(x) + F(x)\beta_2(y)$ , then one can show that

$$\begin{aligned} \mathbb{H}_n(x, y) - \mathbb{H}(x, y) &= F_n(x) \{\beta_{n2}(y) - \beta_2(y)\} + \beta_2(y) \{\beta_{n1}(x)/\sqrt{n}\} \\ &\quad + y \{\beta_{n1}(x) - \beta_1(x)\}. \end{aligned}$$

An application of the triangle inequality yields

$$\sup_{(x, y) \in I^2} \left| \frac{\mathbb{H}_n(x, y) - \mathbb{H}(x, y)}{\tilde{q}_\star(xy)} \right| \leq \mathcal{B}_{n1} + \mathcal{B}_{n2} + \mathcal{B}_{n3},$$

where

$$\mathcal{B}_{n1} = \sup_{(x, y) \in I^2} \frac{F_n(x)}{x} \frac{r(y)}{r(xy)} \left| \frac{\beta_{n2}(y) - \beta_2(y)}{\tilde{q}_\star(y)} \right|,$$

$$\mathcal{B}_{n2} = \sup_{(x, y) \in I^2} \frac{r(y)}{r(xy)} \left| \frac{\beta_2(y)}{\tilde{q}_\star(y)} \right| \left| \frac{\beta_{n1}(x)/\sqrt{n}}{x} \right|,$$

$$\mathcal{B}_{n3} = \sup_{(x, y) \in I^2} \frac{r(x)}{r(xy)} \left| \frac{\beta_{n1}(x) - \beta_1(x)}{\tilde{q}_\star(x)} \right|,$$

with  $r(x) = \tilde{q}^*(x)/x$ . On the other side, one deduces from Shorack & Wellner (1986) that the processes  $\beta_{n1}$  and  $\beta_{n2}$  converge to  $\beta_1$  and  $\beta_2$  with respect to the weighted metric

$$d_q(g_1, g_2) = \sup_{x \in I} \left| \frac{g_1(x) - g_2(x)}{q(x)} \right|$$

if and only if

$$\lim_{x \rightarrow 0} \frac{q(x)}{\{x \log(1/x)\}^{1/2}} = +\infty.$$

Since  $\tilde{q}^*$  belongs to  $\mathcal{Q}^*$ , the function  $r$  is decreasing and  $r(x)/r(xy) \leq 1$ . By repeated applications of Shorack & Wellner's result using the fact that

$$\lim_{x \rightarrow 0} \frac{\tilde{q}^*(x)}{\{x \log(1/x)\}^{1/2}} = +\infty,$$

one can show that  $\mathcal{B}_{n1}$ ,  $\mathcal{B}_{n2}$  and  $\mathcal{B}_{n3}$  converge in probability to 0. This shows that  $\mathbb{H}_n(x, y)/\tilde{q}^*(xy)$  converges to  $\mathbb{H}(x, y)/q^*(xy)$  and ensures, in particular, the tightness of the sequence  $\mathcal{B}_n$ . Thus, Hypothesis III of Ghoudi & Rémillard (1998) is satisfied.

## ANNEXE B

### PROOF OF LEMMA 4.1

Note that  $\partial\Lambda_\rho(t, \omega) = K_1 \{\gamma_1(\omega) - \gamma_1(t/\omega)\} + K_2 \{\gamma_2(\omega) - \gamma_2(t/\omega)\} / \partial\omega$ , where

$$\gamma_1(\omega) = \frac{\omega\Phi^{-1}(\omega)}{\phi\{\Phi^{-1}(\omega)\}}, \quad \gamma_2(\omega) = \frac{\phi\{\Phi^{-1}(\omega)\}\Phi^{-1}(\omega)}{\omega}$$

and

$$K_1 = \frac{2}{\omega} \geq 0, \quad K_2 = \frac{2(t/\omega)}{\rho\phi\{\Phi^{-1}(\omega)\}\phi\{\Phi^{-1}(t/\omega)\}} \geq 0.$$

As this derivative clearly vanishes at  $\omega = \sqrt{t}$ , it remains to show that  $\Lambda_\rho(t, \omega)$  is decreasing for  $t \leq \omega \leq \sqrt{t}$ . First note that  $\gamma_j(\omega) \leq 0$ ,  $j = 1, 2$ , for all  $\omega \leq 1/2$ , and then in particular for all  $\omega \leq \sqrt{t}$  since  $t < \{\Phi(-1)\}^2 < 1/4$ , by assumption.

For  $\omega \leq \sqrt{t}$ , one has to show that  $\gamma_j(\omega) \leq \gamma_j(t/\omega)$ ,  $j = 1, 2$ , for all  $\omega < \Phi(-1)$ , which is equivalent to proving that  $\gamma_1, \gamma_2$  are increasing functions since  $\omega \leq t/\omega$ .

Using the fact that  $\Phi(\cdot)$  is increasing, it is equivalent to proving that

$$\tilde{\gamma}_1(x) = \gamma_1\{\Phi(x)\} = \frac{x\Phi(x)}{\phi(x)} \quad \text{and} \quad \tilde{\gamma}_2(x) = \gamma_2\{\Phi(x)\} = \frac{x\phi(x)}{\Phi(x)}$$

are increasing for all  $x < -1$ . Since

$$\tilde{\gamma}'_1(x) = x + (x^2 + 1)\frac{\Phi(x)}{\phi(x)} \quad \text{and} \quad \tilde{\gamma}'_2(x) = -\left\{\frac{\phi(x)}{\Phi(x)}\right\}^2 \left\{x + (x^2 - 1)\frac{\Phi(x)}{\phi(x)}\right\},$$

a little algebra indicates that it is sufficient to show that

$$\frac{-x}{x^2 + 1} \leq \frac{\Phi(x)}{\phi(x)} \leq \frac{-x}{x^2 - 1},$$

which holds true for all  $x < -1$ . ◇



## ANNEXE C

### VERIFICATION OF ASSUMPTION $\mathcal{A}_2$ FOR MODELS IN DURANTE'S CLASS

The density associated to  $S_\psi$  given in (4.6) is

$$s_\psi(t) = 2 - 3t^{-1/2}\psi(t^{1/2}) - \frac{1}{2} \psi'(t^{1/2}) + 2 \int_{t^{1/2}}^1 \frac{\psi(\omega)}{\omega^2} d\omega. \quad (5.2)$$

As long as  $\psi(t)$  is  $o\{\log^{-2}(1/t)\}$  when  $t \rightarrow 0$ , which is true for the models in Table 8 (except in the particular case  $\psi(t) \equiv 1$ ), one can show that Assumption  $\mathcal{A}_2$  is satisfied. Indeed, first note that

$$\lim_{t \rightarrow 0} t^{-1/2} \psi(t^{1/2}) t^{1/2} \log(1/t) = \left\{ \lim_{t \rightarrow 0} \psi(t^{1/2}) \log^2(1/t) \right\} \left\{ \lim_{t \rightarrow 0} \log^{-1}(1/t) \right\} = 0.$$

For the second righthand term of (5.2), the fact that  $\psi(t)/t$  is non-increasing implies  $\psi'(t) \leq \psi(t)/t$  for all  $t \in (0, 1]$  and

$$\begin{aligned} 0 &\leq \lim_{t \rightarrow 0} \psi'(t^{1/2}) t^{1/2} \log(1/t) \\ &\leq \lim_{t \rightarrow 0} \psi(t^{1/2}) \log(1/t) \\ &= \lim_{t \rightarrow 0} \psi(t^{1/2}) \log^2(1/t) \lim_{t \rightarrow 0} \log^{-1}(1/t) \\ &= 0. \end{aligned}$$

Finally, note that  $\omega^{-1}\psi(\omega) \leq t^{-1/2}\psi(t^{1/2})$  for  $v \in [t^{1/2}, 1]$ . This enables to write

$$\int_{t^{1/2}}^1 \frac{\psi(\omega)}{\omega^2} d\omega \leq t^{-1/2} \psi(t^{1/2}) \int_{t^{1/2}}^1 \frac{d\omega}{\omega} = \frac{1}{2} t^{-1/2} \psi(t^{1/2}) \log(1/t)$$

and then to conclude that the last righthand term in (5.2) vanishes as  $t \rightarrow 0$ .

## ANNEXE D

### SOME MODELS IN DURANTE'S CLASS

Table 5.1: Generator, Spearman's dependence function and Spearman's rho for some members of Durante's family.

Model	$\psi(t)$	$S_\psi(t)$	$\rho_\psi$	$\mathcal{M}$
Durante A	$1 - (1 - t)^\alpha$	$\sqrt{t} \left\{ 1 + (1 - \sqrt{t})^\alpha \right\} + 2t \int_{t^{1/2}}^1 \omega^{-2} (1 - \omega)^\alpha d\omega$	$1 - \frac{24}{(\alpha + 1)(\alpha + 2)(\alpha + 3)}$	$[1, \infty)$
Durante B	$\frac{(\beta + 1)t}{\beta t + 1}$	$2(\beta + 1)t \log \left( \frac{\beta + t^{-1/2}}{\beta + 1} \right) + 2t - \frac{(\beta + 1)t}{\beta \sqrt{t} + 1}$	$1 - 2\beta^{-1} + 6\beta^{-2} + 12\beta^{-3} - \frac{12(\beta + 1)}{\beta^4} \log(\beta + 1)$	$[0, \infty)$
Cuadras– Augé	$t^{1-\gamma}$	$t + \left( 1 - \frac{2}{\gamma} \right) \left\{ 1 - (\sqrt{t})^{-\gamma} \right\} t$	$\frac{3\gamma}{4 - \gamma}$	$[0, 1]$
Fréchet	$(1 - \delta)t + \delta$	$(1 - \delta)(t - t \log t) + \delta \sqrt{t}$	$\delta$	$[0, 1]$
Ordinal sum	$(\eta t \wedge 1)$ ,	$\eta t \left\{ 1 - 2 \log \left( \eta \sqrt{t} \right) \right\}, t \leq \eta^{-2}$ $\sqrt{t}, t > \eta^{-2}$	$1 - \eta^{-3}$	$[1, \infty)$

## ANNEXE E

### QUELQUES AJOUTS À L'ARTICLE

#### E.1 Algorithme de simulation pour la famille de Durante

Pour simuler un couple aléatoire  $(U, V)$  dont la loi jointe est  $C$ , la méthode basée sur la fonction de répartition conditionnelle est souvent employée. À cette fin, définissons d'abord  $F_v(u) = P(U \leq u | V = v) = C_{01}(u, v)$ . Ainsi, si  $T$  et  $V$  sont des variables aléatoires indépendantes de loi uniforme sur  $(0, 1)$ , alors la loi jointe de  $U = F_V^{-1}(T)$  et  $V$  est  $C$ .

Pour un modèle dans la classe générale de Durante, on a

$$F_v(u) = \begin{cases} u\psi'(v), & u \leq v \\ \psi(u), & u > v, \end{cases}$$

de telle sorte que

$$F_v^{-1}(t) = \begin{cases} \frac{t}{\psi'(v)}, & 0 \leq t < v\psi'(v) \\ v, & v\psi'(v) \leq t < \psi(v) \\ \psi^{-1}(t), & \psi(v) \leq t \leq 1. \end{cases}$$

## E.2 Estimateurs déduits d'une inversion du rho de Spearman

Dans toutes les applications de ce travail, que ce soit lors des simulations et pour les analyses des jeux de données, seules des familles de copules à un paramètre ont été considérées. Ce paramètre inconnu a été estimé par l'inversion du rho de Spearman. Comme on peut le noter par les résultats présentés au Tableau 5.2, cette méthode fournit un estimateur avec des propriétés enviables en terme du biais ( $B$ ), du biais relatif ( $BR$ ) et de l'erreur quadratique moyenne ( $EQM$ ), même pour des tailles d'échantillon relativement petites.

Table 5.2: Biais ( $B$ ), biais relatif ( $BR$ ) et erreur quadratique moyenne ( $EQM$ ) pour l'estimateur basé sur l'inversion du rho de Spearman

Modèle	$n$	$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
		$B$	$BR$	$EQM$	$B$	$BR$	$EQM$	$B$	$BR$	$EQM$
Normale	100	-0.004	-1.153	0.009	-0.005	-1.017	0.006	-0.007	-0.945	0.003
	250	-0.001	-0.406	0.004	-0.001	-0.255	0.003	-0.003	-0.345	0.001
Clayton	100	0.014	2.765	0.046	0.026	2.434	0.097	0.042	1.985	0.237
	250	0.009	1.790	0.018	0.012	1.084	0.037	0.019	0.874	0.091
Frank	100	-0.040	-2.083	0.428	-0.035	-1.027	0.511	0.063	1.084	0.953
	250	-0.055	-2.861	0.173	-0.033	-0.968	0.179	0.055	0.935	0.386
Durante A	100	0.016	1.148	0.025	0.019	1.115	0.050	0.035	1.450	0.123
	250	0.003	0.217	0.009	0.006	0.365	0.019	0.014	0.602	0.044
Durante B	100	0.045	6.027	0.149	0.114	6.407	0.562	0.364	8.475	3.585
	250	0.019	2.604	0.055	0.044	2.467	0.196	0.150	3.495	1.154
Cuadras- Augé	100	-0.004	-1.014	0.013	-0.004	-0.755	0.009	-0.004	-0.492	0.005
	250	-0.002	-0.630	0.005	-0.003	-0.520	0.003	-0.001	-0.117	0.002
Fréchet	100	-0.002	-0.577	0.011	-0.001	-0.216	0.009	-0.002	-0.260	0.006
	250	-0.000	-0.060	0.004	-0.002	-0.317	0.004	-0.001	-0.078	0.003
Somme ordinaire	100	0.004	0.360	0.003	0.005	0.404	0.007	0.012	0.768	0.016
	250	0.001	0.113	0.001	0.003	0.250	0.003	0.006	0.378	0.006

# ANNEXE F

## PROGRAMMES EN C

Toutes les simulations présentées dans ce mémoire ont été faites à partir d'ordinateurs de type Pentium 4 sous Windows XP. Les nombres aléatoires ont été générés à partir du compilateur de Visual Studio. Les programmes en C utilisés sont présentés ci-après.

```

/*****
** Programme : AdequationCopule.c
**
** Dernière modification : Octobre 2007
** Auteur : Marie-Hélène Toupin
** Courriel : toupimar@uqtr.ca
**
** Description :
**
** Permet de tester l'adéquation de familles de copules avec le
** test basé sur le processus de Spearman. Les familles testées
** sont :
** 1. Clayton
** 5. Frank
** 11. Normale
** 13. Fréchet
** 16. Cuadras-Augé
** 17. Somme ordinale
** 18. Durante A
** 19. Durante B
*****/
int main()
{
    long int *ii;
    unsigned long int i0, j00, k0;
    time_t *tloc;

```

```

ii = (long int *)malloc(sizeof(long int));
tloc=(time_t*)malloc(sizeof(time_t));
*ii=time(tloc);
j00=*ii*rand()*123;
k0=*ii*rand()*756;
i0=*ii*rand()*6675;
kisset(i0,j00,k0);

int it, i, j, k, m, v, w, n, N, model[8];
double **X1, **X2, *R, *S, *e1, *e2, **Y, *R_star, *S_star;
double *e_star, v1, v2;
double rho1, rho2, theta1, theta2, somme1, somme2, somme3;
double V1, V2, p_value;
FILE *fp1, *fp2;

// Les 8 familles de copules testées en H0
model[0] = 13; model[1] = 16; model[2] = 19; model[3] = 17;
model[4] = 1; model[5] = 5; model[6] = 11; model[7] = 18;

fp1 = fopen("Donnes.txt","r"); // Fichier contenant les données
fp2 = fopen("Tests_Copules.txt","w"); // Fichier dans lequel
// on écrit les résultats
fprintf(fp2,"Modèle theta Vn p-value \n");

n = 382; // La taille de l'échantillon
N = 500; // Taille d'échantillon pour le double-bootstrap
it=10000; // nombre d'itérations

X1 = (double **)malloc(sizeof(double)*2);
for(j=0;j<2;j++) { X1[j] = (double *)malloc(sizeof(double)*n); }
X2 = (double **)malloc(sizeof(double)*2);
for(j=0;j<2;j++) { X2[j] = (double *)malloc(sizeof(double)*n); }
R = (double*)malloc(sizeof(double)*n);
S = (double*)malloc(sizeof(double)*n);
e1 = (double*)malloc(sizeof(double)*n);
e2 = (double*)malloc(sizeof(double)*n);
R_star = (double*)malloc(sizeof(double)*N);
S_star = (double*)malloc(sizeof(double)*N);
e_star = (double*)malloc(sizeof(double)*N);
Y = (double **)malloc(sizeof(double)*2);
for(j=0;j<2;j++) { Y[j] = (double *)malloc(sizeof(double)*N); }

for(j=0;j<n;j++)

```

```

{
    fscanf(fp1, "%lf %lf \n", &v1, &v2);
    X1[0][j]= v1; X1[1][j]= v2
}
fclose(fp1);

rank(n, X1[0], R); rank(n, X1[1], S);
rho1 = rho(n, R, S);

for(j=0;j<n;j++) // Calcul des pseudo-observations
{
    e1[j] = (R[j]*S[j]/pow((double)n,2.0));
}

for(m=0;m<8;m++) // Pour les 8 familles en HO
{
    theta1 = rho_theta(model[m], rho1);

    // Pour la calcul de Vn (copules sans double-boostap, i.e.
    // Fréchet, Cuadras-Augé, somme ordinale et Durante B)
    if(modelHo == 13 or modelHo == 16 or modelHo == 17
        or modelHo == 19 )
    {
        V1 = Vn(model[m], n, X);
    }

    // Pour la calcul de Wn (copules avec double-boostap, i.e.
    // Clayton, Frank, normale, Student et Durante A)
    else
    {
        V1 = Wn(model[m], n, N, X);
    }

    p_value=0.0;
    for(k=0;k<it;k++)
    {
        sim_copula(n, model[m], theta1, 0.0, X2);
        rank(n, X2[0], R); rank(n, X2[1], S);
        rho2 = rho(n, R, S);

        for(j=0;j<n;j++)
        {
            e2[j] = (R[j]*S[j]/pow((double)n,2.0));
        }
    }
}

```

```

theta2 = rho_theta(model[m], rho2);

// Pour la calcul de Vn (copules sans double-boostap,
// i.e. Fréchet, Cuadras-Augé, somme ordinale
// et Durante B)
if(modelHo == 13 or modelHo == 16 or modelHo == 17
    or modelHo == 19 )
{
    V2 = Vn(model[m], n, X)
}

// Pour la calcul de Wn (copules avec double-boostap, i.e.
// Clayton, Frank, normale, Student et Durante A)
else
{
    V2 = Wn(model[m], n, N, X);
}
if(V2 > V1){p_value += 1.0/it;} // Calcul des p-values
}
fprintf(fp2,"%d %.4lf %.4lf %.4lf \n", model[m], theta1, V1,
    p_value);
}
fclose(fp2);
return 0;
}

/*****
** Programme : Puissance.c **
** Dernière modification : Octobre 2007 **
** Auteur : Marie-Hélène Toupin **
** Courriel : toupimar@uqtr.ca **
** **
** Description : **
** **
** Calcul de la puissance de la statistique Vn (ou Wn) **
*****/
int main()
{
    double **X, *R, *S, *e, **Y, *e_star, *R_star, *S_star;
    double theta_n, rho_n, a, c, t1, t2, b1, b2, r, theta, power;
    double somme1, somme2, somme3, t, temp, statistique;
    int i, j, k, l, m, it, n, N, modelH1, modelHo;
    FILE *fp1, *fp2;

```



```

long int *ii;
unsigned long int i0, j00, k0;
time_t *tloc;

ii = (long int *)malloc(sizeof(long int));
tloc=(time_t*)malloc(sizeof(time_t));
*ii=time(tloc);
j00=*ii*rand()*123;
k0=*ii*rand()*756;
i0=*ii*rand()*6675;
kisset(i0,j00,k0);

n = 250; // Taille d'échantillon
N = 500; // Taille d'échantillon pour le double-bootstrap
it = 10000; // nombre d'itérations
modelH1 = 18; // Modèle sous H1
theta = 0.50; // Paramètre de la famille sous H1
modelHo = 13; // Modèle sous H0

fp2 = fopen("Quantiles.txt", "r"); // Le fichier des quantiles de
// la copule sous H0
X = (double **)malloc(sizeof(double)*2);
for(j=0;j<2;j++) { X[j] = (double *)malloc(sizeof(double)*n); }
R = (double*)malloc(sizeof(double)*n);
S = (double*)malloc(sizeof(double)*n);
e = (double*)malloc(sizeof(double)*n);

R_star = (double*)malloc(sizeof(double)*N);
S_star = (double*)malloc(sizeof(double)*N);
e_star = (double*)malloc(sizeof(double)*N);
Y = (double **)malloc(sizeof(double)*2);
for(j=0;j<2;j++) { Y[j] = (double *)malloc(sizeof(double)*N); }

power = 0.0;
for(k=0;k<it;k++)
{
    // Simulation de la copule sous H1 de paramètre theta
    sim_copula(n, modelH1, theta, 0.0, X);

    // Calcul des rangs
    rank(n, X[0], R); rank(n, X[1], S);

    // Calcul du rho de Spearman
    rho_n = rho(n, R, S);
}

```

```

// Calcul des pseudo-observations
for(j=0;j<n;j++) { e[j] = (R[j]*S[j]/pow((double)n,2.0)); }

// Inversion du rho de Spearman pour estimer le paramètre
theta_n = rho_theta(modelHo, rho_n);

// Pour la calcul de Vn (copules sans double-boostap, i.e.
// Fréchet, Cuadras-Augé, somme ordinale et Durante B)
if(modelHo == 13 or modelHo == 16 or modelHo == 17
    or modelHo == 19 )
{
    statistique = Vn(modelHo, n, X);
}

// Pour la calcul de Wn (copules avec double-boostap, i.e.
// Clayton, Frank, normale, Student et Durante A)
else
{
    statistique = Wn(modelHo, n, N, X);
}

// Trouver le bon quantiles dans le fichier
b2 = -1.0; t2 = 0.0;
do { t1 = t2; b1 = b2;
fscanf(fp2, "%lf %lf %lf %lf %lf \n", &r, &b2, &a, &t2, &c); }
while (theta_n > b2);
fclose(fp2);

// Rejet de H0 si la statistique est supérieure au quantile trouvé
if (statistique > t2) { power += 1.0/(double)it; }
}
fprintf(fp1, "%.4lf \n", power);

fclose(fp1);
for(i=0;i<2;i++) { free(X[i]); } free(X);
for(i=0;i<2;i++) { free(Y[i]); } free(Y);
free(R); free(S); free(e); free(R_star); free(S_star); free(e_star);
return 0;
}

```

```

/*****
** Programme : Quantiles.c                                     **
** Dernière modification : Octobre 2007                       **
** Auteur : Marie-Hélène Toupin                               **
** Courriel : toupimar@uqtr.ca                                **
**                                                            **
** Description :                                              **
**                                                            **
** Création d'un fichier des quantiles (90e, 95e et 99e) de la **
** statistique Vn (sans double-bootstrap), pour des rho de   **
** Spearman entre 0 et 1 avec un pas de 0.01, pour les familles de **
** copules suivantes :                                       **
**                                                            **
**     13. Fréchet                                           **
**     16. Cuadras-Augé                                       **
**     17. Somme ordinale                                     **
**     19. Durante B                                         **
*****/
double Vn(int model, int n, double **X);

int main()
{
    double **X, *V, theta, rho;
    int i, j, it, n, model, M, Q1, Q2, Q3;
    FILE *fp1;

    long int *ii;
    unsigned long int i0, j00, k0;
    time_t *tloc;

    ii = (long int *)malloc(sizeof(long int));
    tloc=(time_t*)malloc(sizeof(time_t));
    *ii=time(tloc);
    j00=*ii*rand()*123;
    k0=*ii*rand()*756;
    i0=*ii*rand()*6675;
    kisset(i0,j00,k0);

    fp1 = fopen("Quantiles.txt", "w");

    n = 250; // Taille des échantillons
    model = 13; // Famille de copule choisie
    it = 10000; // Nombre d'itérations
    M = 100;

```

```

X = (double **)malloc(sizeof(double)*2);
for(j=0;j<2;j++) { X[j] = (double *)malloc(sizeof(double)*n); }

V = (double *)malloc(sizeof(double)*it);

for(i=0;i<M;i++)
{
    rho = (double)i/(double)M; // entre 0 et 1 avec un pas de 0.01
    theta = rho_theta(model, rho); // Inversion du rho de Spearman

    for(j=0;j<it;j++)
    {
        sim_copula(n, model, theta, 0.0, X);

        V[j] = Vn(model, n, X);
    }

    quick_sort(V,0,it-1);

    Q1 = (int)(0.90*it); Q2 = (int)(0.95*it); Q3 = (int)(0.99*it);

    fprintf(fp1, "%.2f %.2f %.6lf %.6lf %.6lf \n", rho, theta,
            V[Q1-1], V[Q2-1], V[Q3-1]);
}
fclose(fp1);
free(V); for(i=0;i<2;i++) { free(X[i]); } free(X);
return 0;
}

/*****
/***** Calcul de la statistique Vn*****/
/*****/
double Vn(int model, int n, double **X)
{
    int i, j, k;
    double theta_n, rho_n, *R, *S, *e, t, temp, somme1, somme2;

    R = (double*)malloc(sizeof(double)*n);
    S = (double*)malloc(sizeof(double)*n);
    e = (double*)malloc(sizeof(double)*n);

    rank(n, X[0], R); rank(n, X[1], S);

```

```
rho_n = rho(n, R, S);

theta_n = rho_theta(model, rho_n);

for(j=0;j<n;j++)
{
    e[j] = (R[j]*S[j]/pow((double)n,2.0));
}

somme1 = 0.0;
for(i=0;i<n;i++)
{
    for(j=0;j<n;j++)
    {
        t = max(e[i],e[j]);
        somme1 += 1.0 - S(model, theta_n, t);
    }
}

somme2 = 0.0;
for(i=0;i<n;i++)
{
    temp = S(model, theta_n, e[i]);
    somme2 += 1.0 - pow(temp, 2.0);
}

free(R); free(S); free(e);

return somme1/(double)n - .somme2 + (double)n/3.0;
}
```

```

/*****
** Programme : Quantiles_Double_Bootstrap.c                               **
** Dernière modification : Octobre 2007                                   **
** Auteur : Marie-Hélène Toupin                                          **
** Courriel : toupimar@uqtr.ca                                           **
**                                                                         **
** Description :                                                         **
**                                                                         **
** Création d'un fichier des quantiles (90e, 95e et 99e) de la          **
** statistique Wn (avec double-bootstrap), pour des rho de Spearman    **
** entre 0 et 1 avec un pas de 0.01, pour les familles de copules     **
** suivantes :                                                           **
**                                                                         **
**     1. Clayton                                                         **
**     5. Frank                                                           **
**    11. Normale                                                         **
**    12. Student                                                         **
**    18. Durante A                                                      **
*****/
double Wn(int model, int n, double **X);

int main()
{
    double **X, *W, theta, rho;
    int i, j, it, n, N, model, M, Q1, Q2, Q3;
    FILE *fp1;

    long int *ii;
    unsigned long int i0, j00, k0;
    time_t *tloc;

    ii = (long int *)malloc(sizeof(long int));
    tloc=(time_t*)malloc(sizeof(time_t));
    *ii=time(tloc);
    j00=*ii*rand()*123;
    k0=*ii*rand()*756;
    i0=*ii*rand()*6675;
    kisset(i0,j00,k0);

    fp1 = fopen("Quantiles.txt", "w");

    n = 250; // Taille des échantillons
    N = 500; // Taille d'échantillon pour le double-bootstrap
    model = 13; // Famille de copule choisie

```

```

it = 10000; // Nombre d'itérations
M = 100;

X = (double **)malloc(sizeof(double)*2);
for(j=0;j<2;j++) { X[j] = (double *)malloc(sizeof(double)*n); }

W = (double *)malloc(sizeof(double)*it);

for(i=0;i<M;i++)
{
    rho = (double)i/(double)M;
    theta = rho_theta(model, rho);

    for(j=0;j<it;j++)
    {
        sim_copula(n, model, theta, 0.0, X);
        W[j] = Wn(model, n, X);
    }

    quick_sort(W,0,it-1);

    Q1 = (int)(0.90*it); Q2 = (int)(0.95*it); Q3 = (int)(0.99*it);
    fprintf(fp1, "%.2f %.2f %.6lf %.6lf %.6lf \n", rho, theta,
            V[Q1-1], V[Q2-1], V[Q3-1]);
}

fclose(fp1);
free(W); for(i=0;i<2;i++) { free(X[i]); } free(X);
return 0;
}

/*****
/***** Calcul de la statistique Wn*****/
/*****
double Wn(int model, int n, int N, double **X)
{
    int i, j, k;
    double theta_n, rho_n, *R, *S, *e, t, temp, somme1, somme2, somme3;
    double **Y, *e_star, *R_star, *S_star;

    R = (double*)malloc(sizeof(double)*n);
    S = (double*)malloc(sizeof(double)*n);
    e = (double*)malloc(sizeof(double)*n);

    R_star = (double*)malloc(sizeof(double)*N);

```

```
S_star = (double*)malloc(sizeof(double)*N);
e_star = (double*)malloc(sizeof(double)*N);
Y = (double **)malloc(sizeof(double)*2);
for(j=0;j<2;j++) { Y[j] = (double *)malloc(sizeof(double)*N); }

rank(n, X[0], R); rank(n, X[1], S);
rho_n = rho(n, R, S);
theta_n = rho_theta(model, rho_n);

for(j=0;j<n;j++)
{
    e[j] = (R[j]*S[j]/pow((double)n,2.0));
}

sim_copula(N, model, theta_n, 0.0, Y);
rank(N, Y[0], R_star); rank(N, Y[1], S_star);
for(j=0;j<N;j++)
{
    e_star[j] = (R_star[j]*S_star[j]/pow((double)N,2.0));
}

somme1 = 0.0;
for(i=0;i<n;i++)
{
    for(j=0;j<n;j++)
    {
        somme1 += 1.0 - max(e[i],e[j]);
    }
}

somme2 = 0.0;
for(i=0;i<n;i++)
{
    for(j=0;j<N;j++)
    {
        somme2 += 1.0 - max(e[i],e_star[j]);
    }
}

somme3 = 0.0;
for(i=0;i<N;i++)
{
    for(j=0;j<N;j++)
    {
        somme3 += 1.0 - max(e_star[i],e_star[j]);
    }
}
```



```

    }

    free(R); free(S); free(e); free(R_star); free(S_star); free(e_star);
    for(i=0;i<2;i++) { free(Y[i]); } free(Y);

    return (somme1/(double)n) - (2.0*somme2/(double)N) +
           ((double)n*somme3/(pow((double)N,2.0)));
}
/*****
** Programme : S.c                                         **
** Dernière modification : Octobre 2007                 **
** Auteur : Marie-Hélène Toupin                         **
** Courriel : toupimar@uqtr.ca                           **
**                                                       **
** Description :                                         **
**                                                       **
** Donne la valeur de la fonction de dépendance de Spearman S(t) **
** pour les familles de copules suivantes :              **
**                                                       **
** 13. Fréchet                                           **
** 16. Cuadras-Augé                                     **
** 17. Somme ordinale                                    **
** 19. Durante B                                         **
*****/
double S(int model, double theta, double t)
{
    double S;

    if (t == 0) { S = 0.0; }
    else if (t == 1) { S = 1.0; }
    else
    {
        /***** Copule de Fréchet *****/
        if (model == 13)
        {
            S = (1.0-theta)*(t-t*log(t)) + theta*sqrt(t);
        }
        /***** Copule de Cuadras-Augé *****/
        if (model == 16)
        {
            if (theta == 0.0) { S = t - t*log(t); }
            else
            {
                S = t + (1.0 - (2.0/theta))*(1.0 - pow(sqrt(t),-theta))*t;
            }
        }
    }
}

```

```

    }
}
/***** Somme ordinale *****/
if (model == 17)
{
    if ( t > pow(theta,-2.0) ) { S = sqrt(t); }
    else { S = theta*t*(1.0-2.0*log(theta*sqrt(t))); }
}
/***** Copule de Durante B *****/
if (model == 19)
{
    if (theta == 0.0) { S = t - t*log(t); }
    else
    {
        S = 2.0*(theta+1.0)*t*log((theta*sqrt(t)+1.0)
            / (theta*sqrt(t) + sqrt(t))) + 2.0*t -
            (((theta+1.0)*t)/(theta*sqrt(t) + 1.0));
    }
}
}
return S;
}
/*****
** Programme : rho_theta.c **
** Dernière modification : Octobre 2007 **
** Auteur : Marie-Hélène Toupin **
** Courriel : toupimar@uqtr.ca **
** **
** Description : **
** **
** Estimation du paramètre par l'inversion du rho de Spearman **
** pour une famille de copule parmi les suivantes : **
** **
** 1. Clayton **
** 5. Frank **
** 11. Normale **
** 12. Student **
** 13. Fréchet **
** 16. Cuadras-Augé **
** 17. Somme ordinale **
** 18. Durante A **
** 19. Durante B **
*****/
#define pi 3.141592654

```

```

double rho_theta(int model, double rho)
{
    double theta;
    FILE *fp1;

    /***** Copule de Clayton *****/
    if (model == 1)
    {
        fp1 = fopen("../Librairie/rho_Clayton.txt", "r");
        double r1, r2, t1, t2;
        t2 = -1.0; r2 = -1.0;

        do { t1 = t2; r1 = r2; fscanf(fp1, "%lf %lf", &t2, &r2); }
        while (rho > r2);
        fclose(fp1);
        theta = ((t1 - t2)*rho + (r1*t2 - r2*t1))/(r1 - r2);
    }

    /***** Copule de Frank *****/
    if (model == 5)
    {
        fp1 = fopen("../Librairie/rho_Frank.txt", "r");
        double r1, r2, t1, t2;
        t2=0.0; r2=0.0;
        do { t1 = t2; r1 = r2; fscanf(fp1, "%lf %lf", &t2, &r2); }
        while (rho > r2);
        fclose(fp1);
        theta = ((t1 - t2)*rho + (r1*t2 - r2*t1))/(r1 - r2);
    }

    /***** Copule normale *****/
    if (model == 11)
    {
        theta = 2.0*sin((pi/6.0)*rho);
    }

    /***** Copule de Fréchet *****/
    if (model == 13)
    {
        theta = rho;
    }

    /***** Copule de Cuadras-Augé *****/
    if (model == 16)
    {
        if (rho<=0){ theta = 0.0; }
    }
}

```

```

        else { theta = 4.0*rho/(3.0+rho); }
    }
    /***** Somme ordinale *****/
    if (model == 17)
    {
        if(rho<=0) { theta = 1.0; }
        else { theta = pow(1.0-rho,-1.0/3.0); }
    }
    /***** Copule de Durante A *****/
    if (model == 18)
    {
        if (rho<=0) {theta = 1.0; }
        else
        {
            fp1 = fopen("../Librairie/rho_DuranteA.txt", "r");
            double r1, r2, t1, t2;
            t2 = 1.0; r2 = 0.0;
            do { t1 = t2; r1 = r2; fscanf(fp1, "%lf %lf", &t2, &r2); }
            while (rho > r2);
            fclose(fp1);
            theta = ((t1 - t2)*rho + (r1*t2 - r2*t1))/(r1 - r2);
        }
    }
    /***** Copule de Durante B *****/
    if (model == 19)
    {
        if (rho<=0) {theta = 0.0; }
        else
        {
            fp1 = fopen("../Librairie/rho_DuranteB.txt", "r");
            double r1, r2, t1, t2;
            t2 = 0.0; r2 = 0.0;
            do { t1 = t2; r1 = r2; fscanf(fp1, "%lf %lf", &t2, &r2); }
            while (rho > r2);
            fclose(fp1);
            theta = ((t1 - t2)*rho + (r1*t2 - r2*t1))/(r1 - r2);
        }
    }
    return theta;
}

```

```

/*****
** Programme : sim_copula.c **
** Dernière modification : Octobre 2007 **
** Auteur : Marie-Hélène Toupin **
** Courriel : toupimar@uqtr.ca **
** **
** Description : **
** **
** Génère des échantillons bivariés de taille n provenant des **
** familles de copules suivantes : **
** **
** 1. Clayton **
** 5. Frank **
** 11. Normale **
** 12. Student **
** 13. Fréchet **
** 16. Cuadras-Augé **
** 17. Somme ordinale **
** 18. Durante A **
** 19. Durante B **
*****/
void biv_student(int n, double nu, double rho, double **X);
void biv_normal(int n, double rho, double **X);

void sim_copula(int n, int model, double theta, double beta, double **X)
{
    int i;
    double u, v, s, t, a, b, c, d, p, U0, U1, U2, U3, z, z1;
    double Z1, T1, w, W, t1, t2, t3, t4, t5, racine;

    /***** Copule de Clayton *****/
    if (model == 1)
    {
        if (theta == 0.0)
        {
            for(i=0;i<n;i++) { X[0][i] = kiss(); X[1][i] = kiss(); }
        }
        else
        {
            for(i=0;i<n;i++)
            {
                v = kiss(); t = kiss();
                a = pow(v, -theta);
                b = pow(t, -theta/(theta+1.0)) - 1.0;
            }
        }
    }
}

```

```

        X[0][i] = pow(a*b+1.0,-1.0/theta);
        X[1][i] = v;
    }
}
/***** Copule de Frank *****/
if (model == 5)
{
    if (theta == 0.0)
    {
        for(i=0;i<n;i++) { X[0][i] = kiss(); X[1][i] = kiss(); }
    }
    else
    {
        for(i=0;i<n;i++)
        {
            u = kiss(); t = kiss();

            t1 = t*exp(-1.0*theta) + (1.0-t)*exp(-1.0*theta*u);
            t2 = t + (1.0-t)*exp(-1.0*theta*u);

            X[0][i] = u;
            X[1][i] = log(t2/t1)/theta;
        }
    }
}
/***** Copule normale *****/
if (model == 11)
{
    if (theta == 0.0)
    {
        for(i=0;i<n;i++) { X[0][i] = kiss(); X[1][i] = kiss(); }
    }
    else { biv_normal(n, theta, X); }
}
/***** Copule de Student *****/
if (model == 12)
{
    biv_student(n, beta, theta, X);
}
/***** Copule de Fréchet *****/
if (model == 13)
{

```

```

    if (theta == 0.0)
    {
        for(i=0;i<n;i++) { X[0][i] = kiss(); X[1][i] = kiss(); }
    }
    else
    {
        for(i=0;i<n;i++)
        {
            t = kiss(); s = kiss(); X[1][i] = s;
            if (t <= theta) { X[0][i] = s; }
            else { X[0][i] = kiss(); }
        }
    }
}
/***** Copule de Cuadras-Augé *****/
if (model == 16)
{
    if (theta == 0.0)
    {
        for(i=0;i<n;i++) { X[0][i] = kiss(); X[1][i] = kiss(); }
    }
    else
    {
        for(i=0;i<n;i++)
        {
            t = kiss(); v = kiss(); X[1][i] = v;
            if (t < (1.0-theta)*pow(v,1.0 - theta))
            {
                X[0][i] = (t*pow(v,theta))/(1.0-theta);
            }
            else if (t < pow(v, 1.0 - theta)) { X[0][i] = v; }
            else { X[0][i] = pow(t,(1.0/(1.0-theta))); }
        }
    }
}
/***** Somme ordinale *****/
if (model == 17)
{
    for(i=0;i<n;i++)
    {
        t = kiss(); v = kiss(); X[1][i] = v;
        if ( v <= 1.0/theta ) { X[0][i] = t/theta; }
        else { X[0][i] = v; }
    }
}

```

```

}
/***** Copule de Durante A *****/
if (model == 18)
{
    if (theta == 1.0)
    {
        for(i=0;i<n;i++) { X[0][i] = kiss(); X[1][i] = kiss(); }
    }
    else
    {
        for(i=0;i<n;i++)
        {
            t = kiss(); v = kiss(); X[1][i] = v;
            if ( t < theta*v*pow(1.0-v,theta-1.0) )
            {
                X[0][i] = t/(theta*pow(1.0-v,theta-1.0));
            }
            else if ( t < 1.0 - pow(1.0-v,theta) ){ X[0][i] = v; }
            else { X[0][i] = 1.0 - pow(1.0-t,1.0/theta); }
        }
    }
}
/***** Copule de Durante B *****/
if (model == 19)
{
    if (theta == 0.0)
    {
        for(i=0;i<n;i++) { X[0][i] = kiss(); X[1][i] = kiss(); }
    }
    else
    {
        for(i=0;i<n;i++)
        {
            t = kiss(); v = kiss(); X[1][i] = v;
            if ( t < (theta+1.0)*v/pow(theta*v+1.0,2.0) )
            {
                X[0][i] = pow(theta*v+1.0,2.0)*t/(theta+1.0);
            }
            else if ( t < ((1.0+theta)*v)/(theta*v+1.0) )
            {
                X[0][i] = v;
            }
            else { X[0][i] = t/(1.0+theta-theta*t); }
        }
    }
}

```



```

    }
  }
}

/*****
/***** Génération d'un vecteur de loi normale bivariée *****/
/*****/
void biv_normal(int n, double rho, double **X)
{
  int i;
  double z1, z2, u1, u2, temp1, temp2;

  for(i=0;i<n;i++)
  {
    u1 = kiss(); u2 = kiss();
    z1 = cos((double)2.0*pi*u1)
        * pow((double)-2.0*log(u2), (double)0.5);
    z2 = sin((double)2.0*pi*u1)
        * pow((double)-2.0*log(u2), (double)0.5);

    temp1 = sqrt(0.5+0.5*sqrt(1.0-rho*rho));
    temp2 = sqrt(0.5-0.5*sqrt(1.0-rho*rho));
    X[0][i] = temp1*z1 + temp2*z2;
    if (rho >= 0.0) { X[1][i] = temp2*z1 + temp1*z2; }
    else { X[1][i] = -1.0*(temp2*z1 + temp1*z2); }
  }
}

/*****
/***** Génération d'un vecteur de loi Student bivariée *****/
/*****/
void biv_student(int n, double nu, double rho, double **X)
{
  int i, j;
  double z1, z2, u1, u2, temp1, temp2, khi, V1, V2;

  for(i=0;i<n;i++)
  {
    u1 = kiss(); u2 = kiss();
    z1 = cos((double)2.0*pi*u1)
        * pow((double)-2.0*log(u2), (double)0.5);
    z2 = sin((double)2.0*pi*u1)
        * pow((double)-2.0*log(u2), (double)0.5);

    temp1 = sqrt(0.5+0.5*sqrt(1.0-rho*rho));

```

```

temp2 = sqrt(0.5-0.5*sqrt(1.0-rho*rho));
V1 = temp1*z1 + temp2*z2;
if (rho >= 0.0) { V2 = temp2*z1 + temp1*z2; }
else { V2 = -1.0*(temp2*z1 + temp1*z2); }

khi = 0.0;
for(j=0;j<(int)nu;j++)
{
    u1 = kiss(); u2 = kiss();
    z1 = cos((double)2.0*pi*u1)
        * pow((double)-2.0*log(u2),(double)0.5);
    khi += z1*z1;
}
X[0][i] = sqrt(nu)*V1/sqrt(khi);
X[1][i] = sqrt(nu)*V2/sqrt(khi);
}
}

/*****
** Programme : rho.c **
**
** Dernière modification : Octobre 2007 **
** Auteur : Marie-Hélène Toupin **
** Courriel : toupimar@uqtr.ca **
**
** Description : **
**
** Calcul du rho de Spearman empirique pour un échantillon bivarié **
** de taille n à partir des deux vecteurs des rangs R et S **
*****/
double rho(int n, double *R, double *S)
{
    int j;
    double somme = 0.0;

    for(j=0; j<n; j++)
    {
        somme += pow((R[j] - S[j]),2);
    }

    return 1.0 - (6.0/((double)n*(pow((double)n,2.0)-1.0)))*somme;;
}

```

```

/*****
** Programme : rank.c                                     **
** Dernière modification : Octobre 2007                 **
** Auteur : Marie-Hélène Toupin                       **
** Courriel : toupimar@uqtr.ca                         **
**                                                     **
** Description :                                       **
**                                                     **
** Retourne le vecteur des rangs pour un échantillon de taille n **
*****/
void rank(int n, double *X, double *r)
{
    int i, j;

    for(i=0;i<n;i++)
    {
        r[i] = 0;
        for(j=0;j<n;j++)
        {
            if (X[j] <= X[i]) { r[i] += 1; }
        }
    }
}

/*****
** Programme : quick_sort.c                             **
** Dernière modification : Octobre 2007                 **
** Auteur : Marie-Hélène Toupin                       **
** Courriel : toupimar@uqtr.ca                         **
**                                                     **
** Description :                                       **
**                                                     **
** Effectue un tri du tableau t entre les éléments situés entre **
** lo et hi.                                           **
*****/
void quick_sort(double *t, int lo, int hi)
{
    int i,j;
    double mid;
    double tmp;

    i=lo; j=hi;
    mid = t[ (lo + hi) >> 1 ]; /* On choisit le pivot. */

```

```

do
{
    while (t[i] < mid) i++;
    while (mid < t[j]) j--;

    if (i <= j)
    {
        /* On echange deux elements. */
        tmp = t[i];
        t[i++] = t[j];
        t[j--] = tmp;
    }
} while (i <=j);

if (lo < j) quick_sort(t, lo, j); /* On trie la partie gauche. */
if (i < hi) quick_sort(t, i, hi); /* On trie la partie droite. */
}

```

```

/*****
** Programme : phi.c **
** Dernière modification : Octobre 2007 **
** Auteur : Marie-Hélène Toupin **
** Courriel : toupimar@uqtr.ca **
** **
** Description : **
** **
** Calcul de phi(x) **
*****/

```

```

double gammln(double xx)
double gammp(double a, double x)

```

```

double phi(double x)
{
    double temp;

    temp = gammp(0.5, x*x/2.0);

    if (x<0) { return 0.5 - 0.5*temp; }
    else return 0.5 + 0.5*temp;
}

```

```

double gammp(double a, double x)
{

```

```

double gamser, gammcf, gln, eps, sum, del, ap, fpmin;
double an, b, c, d, h, value;
int n, it;

it = 100; eps = 3.0e-7; fpmin = 1.0e-30;

if (x == 0.0) { return 0.0; }

else if (x < (a+1.0))
{
    gln = gammln(a); ap = a; del = sum = 1.0/a;

    for(n=1;n<=it;n++)
    {
        ++ap; del *= x/ap; sum += del;
        if (fabs(del)<fabs(sum)*eps)
        {
            gamser = sum*exp(-x + a*log(x) - gln);
            value = gamser;
        }
    }
}
else
{
    gln = gammln(a);
    b = x+1.0-a; c = 1.0/fpmin; d = 1.0/b; h = d;

    for(n=1;n<=it;n++)
    {
        an = -n*(n-a); b += 2.0; d = an*d + b;
        if (fabs(d) < fpmin) { d = fpmin; }
        c = b + an/c;
        if (fabs(c) < fpmin) {c = fpmin; }
        d = 1.0/d; del = d*c; h *= del;
        if (fabs(del-1.0)<eps) { break; }
    }
    gammcf = exp(-x + a*log(x) - gln)*h;
    value = 1.0 - gammcf;
}
return value;
}
double gammln(double xx)
{
    double x, y, temp, ser;

```

```

static double cof[6]={76.1800917294, -86.5053203294,
                    24.0140982408, -1.23173957245,
                    0.12086509738e-2, -0.539523938e-5};
int j;

y = x = xx;
temp = x + 5.5;
temp -= (x+0.5)*log(temp);
ser = 1.000000000190015;

for(j=0;j<=5;j++) { ser += cof[j]/++y; }

return -temp + log(2.506628274631*ser/x);
}

/*****
** Programme : phi_inverse.c **
** Dernière modification : Octobre 2007 **
** Auteur : Marie-Hélène Toupin **
** Courriel : toupimar@uqtr.ca **
** **
** Description : **
** Calcul de phi-1(p) pour p pris dans l'intervalle [0,1], **
** en utilisant l'algorithme de Kennedy & Gentle, STATISTICAL **
** COMPUTING, p. 95 **
*****/
double phi_inverse(double p)
{
    double p0, p1, p2, p3, p4, q0, q1, q2, q3, q4, x, y, value;

    p0 = - 0.322232431088;          q0 = 0.0993484626060;
    p1 = - 1.0;                    q1 = 0.588581570495;
    p2 = - 0.342242088547;          q2 = 0.531103462366;
    p3 = - 0.0204231210245;         q3 = 0.103537752850;
    p4 = - 0.453642210148*1e-4;     q4 = 0.38560700634*1e-2;

    if (p == 0.5) { value = 0.0; }
    if (p == 0.0) { value = -999999; }
    if (p == 1.0) { value = 999999; }
    else
    {
        if (p < 0.5)
        {
            y = sqrt(-2.0*log(p));

```

```

        x = y + (((y*p4 + p3)*y + p2)*y + p1)*y + p0)
/(((y*q4 + q3)*y + q2)*y + q1)*y + q0);
        value = -1.0*fabs(x);
    }
    else if (p > 0.5)
    {
        y = sqrt(-2.0*log(1.0-p));
        x = y + (((y*p4 + p3)*y + p2)*y + p1)*y + p0)
/(((y*q4 + q3)*y + q2)*y + q1)*y + q0);
        value = fabs(x);
    }
}
return value;
}

/*****
** Programme : kiss.c                                     **
** Dernière modification : Octobre 2007                 **
** Auteur : Marie-Hélène Toupin                         **
** Courriel : toupimar@uqtr.ca                          **
**                                                       **
** Description :                                         **
**                                                       **
** Retourne un nombre aléatoire situé dans l'intervalle [0,1]. **
*****/
struct seed_type { unsigned long int i; unsigned long int j;
                  unsigned long int k; }

seed = {274159265L, 392436169L, 721287629L};

void kisset(unsigned long int ii,unsigned long int jj,
             unsigned long int kk)
{
    seed.i=ii; seed.j=jj; seed.k=kk;
}

double kiss()
{
    seed.j = seed.j ^ (seed.j<<17);
    seed.k = (seed.k ^ (seed.k<<18)) & 0x7FFFFFFFL;
    return (1 + (seed.i = 69069L*seed.i + 23606797L) +
           (seed.j ^= (seed.j>>15)) + (seed.k ^= (seed.k>>13)))
           / 4294967296.;
}

```

# Bibliographie

- [1] AAS, K., CZEDO, C., FRIGESSI, A. & BAKKEN, H. (2007). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, sous presse.
- [2] BARBE, P., GENEST, C., GHOUDI, K. & RÉMILLARD, B. (1996). On Kendall's process. *Journal of Multivariate Analysis* **58**, 197–229.
- [3] BERG, D. (2007). Copula goodness-of-fit testing : An overview and power comparison. *Working paper, University of Oslo & Norwegian computing center*.
- [4] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- [5] BREYMAN, W., DIAS, A. & EMBRECHTS, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* **3**, 1–14.
- [6] CHEN, X. & FAN, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *La Revue Canadienne de Statistique* **33**, 389–414.
- [7] DEHEUVELS, P. (1979). La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendance. *Académie Royale de Belgique. Bulletin de la Classe des Sciences* **65**, 274–292.



- [8] DOBRIĆ, J. & SCHMID, F. (2007). A goodness of fit test for copulas based on Rosenblatt's transformation. *Computational Statistics and Data Analysis* **51**, 4633–4642.
- [9] DURANTE, F. (2007). A new family of symmetric bivariate copulas. *Comptes Rendus Mathématique. Académie des Sciences. Paris* **344**, 195–198.
- [10] DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics* **1**, 279–290.
- [11] FERMANIAN, J.-D. (2005). Goodness-of-fit tests for copulas. *Journal of Multivariate Analysis* **95**, 119–152.
- [12] FERMANIAN, J.-D., RADULOVIĆ, D. & WEGKAMP, M. (2004). Weak convergence of empirical copula processes. *Bernoulli* **10**, 847–860.
- [13] FREES, E. W. & VALDEZ, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal* **2**, 1–25.
- [14] GÄNSSLER, P. & STUTE, W. (1987). *Seminar on empirical processes*. DMV Seminar, **9**. Birkhäuser Verlag, Basel.
- [15] GENEST, C., GHOUDI, K. & RIVEST, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552.
- [16] GENEST, C. & MACKAY, R. J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *La Revue Canadienne de Statistique* **14**, 145–159.
- [17] GENEST, C., QUESSY, J.-F. & RÉMILLARD, B. (2002). Tests of serial independence based on Kendall's process. *La Revue Canadienne de Statistique* **30**, 441–461.

- [18] GENEST, C., QUESSY, J.-F. & RÉMILLARD, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* **33**, 337-366.
- [19] GENEST, C. & RIVEST, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* **88**, 1034-1043.
- [20] GENEST, C. & RÉMILLARD, B. (2007). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré* **37**, sous presse.
- [21] GENEST, C., RÉMILLARD, B. & BEAUDOIN, D. (2008). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* **42**, sous presse.
- [22] GHOUDI, K. & RÉMILLARD, B. (2004). Empirical processes based on pseudo-observations. II. The multivariate case. *Asymptotic methods in stochastics, 381-406, Fields Institute Communications* **44**, American Mathematical Society, Providence, RI.
- [23] GHOUDI, K. & RÉMILLARD, B. (1998). Empirical processes based on pseudo-observations. *Asymptotic Methods in Probability and Statistics (Ottawa, Ontario, 1997)*. North-Holland, Amsterdam, 171-197.
- [24] HENZE, N. (2002). Invariant tests for multivariate normality: a critical review. *Statistical Papers* **43**, 467-506.
- [25] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* **19**, 293-325.
- [26] LEE, A. J., *U-statistics: Theory and Practice*. Statistics: Textbooks and Monographs **110**, Marcel Dekker Inc., New York.

- [27] MALEVERGNE, Y. & SORNETTE, D. (2003). Testing the gaussian copula hypothesis for financial assets dependences. *Quantitative finance* **3**, 231–250.
- [28] QUESSY, J.-F. (2005). Théorie et application des copules : tests d'adéquation, tests d'indépendance et bornes pour la valeur-à-risque. *Thèse de doctorat, Université Laval*.
- [29] SAVU, C. & TREDE, M. (2006). Hierarchical Archimedean copulas. Dans *International conference on high frequency finance*, Konstanz, Germany.
- [30] SCAILLET, O. (2006). Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. *Journal of Multivariate Analysis* **98**, 533–543.
- [31] SCHMID, F. & SCHMIDT, R. (2007). Multivariate extensions of Spearman's rho and related statistics. *Statistics and Probability Letters* **77**, 407–416.
- [32] SHORACK, G. R. & WELLNER, J. A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.
- [33] SKLAR, A. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8**, 229–231.
- [34] STEPHENS, M. A. (1974). Components of goodness-of-fit statistics. *Annales de l'Institut Henri Poincaré* **10**, 37–54.
- [35] TSUKAHARA, H. (2005). Semiparametric estimation in copula models. *La Revue Canadienne de Statistique* **33**, 357–375.
- [36] WANG, W. & WELLS, M. A. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association* **95**, 62–72.
- [37] WHELAN, N. (2005). Sampling from Archimedean copulas. *Quantitative finance* **4**, 339–352.