

Université Assane Seck de Ziguinchor (U.A.S.Z)



U.F.R des Sciences et Technologies (UFR-ST)

Département de Mathématiques

Mémoire de Master

Domaine : Sciences et Technologies
Mention : Mathématiques et Applications
Spécialité : Mathématiques Appliquées
Option : Statistique

TITRE :

Tests d'adéquation des copules bivariées et applications

Présenté par

Lala DIEME

Sous la direction de : Dr Emmanuel Nicolas CABRAL

Sous la supervision de : Pr Alassane DIEDHIOU

Membres du Jury :

PRESIDENT :

Alassane DIEDHIOU, Professeur titulaire (UFR-ST/U.A.S.Z)

DIRECTEUR :

Emmanuel N. CABRAL, Maître de conférences titulaire
(UFR-ST/U.A.S.Z)

EXAMINATEURS :

Clément MANGA, Professeur assimilé (UFR-ST/U.A.S.Z)
Mor NDONGO, Maître de conférences titulaire (UFR-SES/U.A.S.Z)

Soutenu le 09 août 2022

Remerciements

Tout d'abord, je remercie Allah le tout puissant pour toute la volonté et le courage qu'Il m'a donné pour l'achèvement de ce travail.

Je voudrais exprimer ma profonde reconnaissance à mon directeur de mémoire, Mr Emmanuel Nicolas CABRAL pour avoir accepté de diriger ce travail. En plus de ses conseils et de ses suggestions, j'ai toujours bénéficié de ses encouragements et de sa disponibilité. Qu'il trouve ici, le témoignage de ma parfaite et profonde gratitude.

J'exprime ma gratitude envers le président du jury, le professeur Alasane DIEDHIOU d'avoir supervisé ce travail et d'avoir accepté de présider le jury de ce mémoire. J'exprime aussi ma gratitude envers les professeurs Clément MANGA, Mor NDONGO qui ont bien accepté d'examiner ce travail.

Je tiens à remercier l'Université Assane SECK de Ziguinchor, l'UFR des Sciences et Technologies, le Département de Mathématiques ainsi que tous les professeurs qui nous ont initié et donné les bases de l'étude mathématique pour exceller dans ce domaine.

Mes plus vifs remerciements à ma très chère mère, mon très cher père, à ma défunte grand-mère, à mes oncles Selle SARR et Henry NDIAYE, à mon défunt oncle Toumboulou DJIBA, à tout le personnel de la BU, en particulier à Mr Habib SAMB, à mes frères et sœurs pour tout le soutien qu'ils m'ont apporté.

Mes remerciements s'adressent spécialement aux docteurs Souhaibou SAMBOU et Marcel Sihintoé BADIANE pour leur aide et encouragement. Merci pour tout le temps que vous m'avez consacré.

Je remercie également les doctorants Ousmane BA, Bernard MANGA, Papa BADIANE, Fatou DIENG, Dieynaba SAMB et Daouda DIACK pour

leur soutien. Merci à vous!

Enfin, j'aimerais remercier mes camarades de promotion, plus particulièrement à Emile ndéné NDONG, Moussa Diatta, Abdourahmane MBAYE, Malick DIBA, Algassimou DIALLO, Mamadou Boye DIALLO, Amadou NDIAYE et Ibrahima DRAME. Sans oublier ceux de la deuxième promotion MPI, en particulier à Souadou DIALLO, Dieynaba SAMB, Amadou SEYDI, Daouda DIACK, Awa BARRY, Fatou DIENG...

Dédicaces

À ma très chère mère **Gaissiry SOUMARE.**

À mon très cher père **Souleymane DIEME.**

À ma défunte grand-mère **Lala CISSE.**

À mon grand frère **Pape DIEME.**

À ma grande sœur **Aissatou DIEME.**

À mes petits frères **Youssouph DIEME et Sémou DIEME.**

À mon papa **Sékou DIEME.**

À mon défunt oncle **Toumboulou DJIBA.**

À mon oncle **Selle SARR.**

À mon oncle **Henry NDIAYE.**

À mes petites sœurs **Adja Oulimatou DIEDHIOU et Mbayang DIOP.**

À mes sœurs **Anna KA, Anna SOW, Mariétou SOW, Souhadou DIALLO, Eliza DASYLVA et Awa CAMARA.**

À mes amis plus que frères **Moussa DIATTA et Emile ndéné NDONG.**

À mes filles chéries **Habibatou BATHILY, Aissatou BATHILY, Nany BATHILY et à ma maman Gaissiry BATHILY.**

Résumé

Dans la théorie moderne de la dépendance stochastique, les copules s'avèrent un outil de choix. En effet, la dépendance entre les variables est entièrement intégrée dans la copule, et permet une simple description de la structure de dépendance entre les variables indépendamment des marginales. Un aspect important de l'application de la théorie des copules concerne le choix d'une famille de modèles qui s'ajustent adéquatement à des observations multivariées. Dans le cas bivarié, ce problème a été abordé par Shih (1998) pour le modèle de fragilité gamma, alors que Genest et Rivest (1993) ont proposé une méthode de sélection graphique applicable à une classe plus large de copules, dites archimédiennes.

Des statistiques d'adéquation applicables à de nombreux modèles de copules à $d \geq 2$ variables sont développées.

Les résultats de Barbe et *al.* (1996) sur la convergence faible du processus de Kendall permettent de caractériser la limite sous l'hypothèse nulle d'un processus empirique proposé pour l'adéquation. Ceci justifie la définition des statistiques de type Cramer-von Mises et Kolmogorov-Smirnov pour l'adéquation dont les seuils s'obtiennent par bootstrap paramétrique. La conception de tests basés sur d'autres processus, en particulier le processus de la copule empirique proposé et étudié par Deheuvels (1981), ainsi que Gaenssler et Stute (1987), est aussi abordée.

Table des matières

Introduction générale	7
1 Notions préliminaires	8
1.1 Outils de probabilités	8
1.1.1 Vecteur aléatoire	8
1.1.2 Loi conjointe et loi marginale d'un vecteur aléatoire	8
1.1.3 Espérance, covariance et corrélation	9
1.2 Concept de dépendance	10
2 Théorie des copules bivariées	13
2.1 Propriétés des copules	13
2.1.1 Continuité d'une copule	14
2.1.2 Théorème de Sklar	14
2.1.3 Inversion du théorème de Sklar	14
2.1.4 Invariance de la fonction copule	15
2.1.5 Densité d'une copule	16
2.1.6 Copules usuelles	17
2.1.7 Bornes de Fréchet-Hoeffding	18
2.2 Notions de dépendance	19
2.2.1 Les mesures de la dépendance	19
2.2.2 Dépendance de queue	22
2.2.3 Estimation de la structure de dépendance	23
2.3 Familles de copules paramétriques	26
2.3.1 Les copules elliptiques	26
2.3.2 Les copules archimédiennes	28
2.3.3 Copules à valeurs extrêmes	31
3 Tests d'adéquation	34
3.1 Approche de Genest, Quessy et Rémillard	34
3.2 Approche par la copule empirique	40
3.3 Approche de Malevergne et Sornette	43

3.3.1	La copule gaussienne	43
3.3.2	Test de l'hypothèse de la copule gaussienne	44
3.4	Approche de Breymann, Dias et Embrechts	48
4	Mise en œuvre pratique	50
4.1	Analyse exploratoire des données	50
4.2	Modélisation de la dépendance	51
	Conclusion	56
	Bibliographie	57

Introduction générale

Le coefficient de corrélation de Pearson, le tau de Kendall et le rho de Spearman ont longtemps été considérés comme les principaux outils de quantification de la dépendance entre les variables aléatoires (voir [16]). Bien que ces mesures soient simples à calculer et peuvent être facilement interprétées, elles ne sont pas en mesure de détecter toutes les formes de dépendance. En effet, le coefficient de corrélation n'est pas défini si les moments simples des variables aléatoires ne sont pas finis. De plus, un coefficient de corrélation nul n'implique pas l'indépendance entre les variables étudiées. Notons aussi qu'on ne peut mesurer qu'une corrélation linéaire. Ajoutons à cela qu'il ne tient pas compte de la dépendance des valeurs extrêmes et que les marginales des variables sont entièrement déterminées par le modèle choisi ; par exemple dans le modèle Student d-varié à ν degrés de liberté, les marges sont nécessairement Student à ν degrés de liberté. L'usage de ces modèles dans plusieurs applications, notamment en finance, a montré que ces lois sont souvent inaptes à expliquer un phénomène (voir [17]).

Une nouvelle approche basée sur les copules s'est imposée dans le monde de la modélisation multivariée. Les principales utilités de cette théorie sont qu'elle permette de modéliser la structure de dépendance entre les variables aléatoires et de définir des lois à plusieurs variables avec les marges désirées (voir [17]). Dans ce cadre, il est possible de construire une distribution dont la structure de dépendance possède les caractéristiques d'une loi gaussienne par exemple, mais dont les marginales sont de loi de Student, ou même Gamma, pour ne nommer que quelques-unes des possibilités offertes par l'emploi des copules (voir [17]). Dans ce mémoire, nous introduisons au chapitre 1 quelques notions de probabilités et des notions de base du concept de dépendance. Ensuite le chapitre 2 sera consacré à la théorie des copules bivariées. Au chapitre 3, nous décrirons des tests d'adéquation permettant de sélectionner la copule la mieux adaptée aux données. Enfin, au chapitre 4, sera présentée une mise en œuvre pratique des tests d'adéquation sur des familles de copules bivariées.

Chapitre 1

Notions préliminaires

1.1 Outils de probabilités

1.1.1 Vecteur aléatoire

Définition 1.1.1 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On appelle variable aléatoire réelle (v.a.r) toute application $X : \Omega \rightarrow \mathbb{R}$ telle que pour tout intervalle I de \mathbb{R} , $X^{-1}(I) \in \mathcal{A}$.

Définition 1.1.2 On appelle vecteur aléatoire de dimension n tout n -uplet (X_1, \dots, X_n) où chaque X_i est une v.a.r définie sur le même espace.

-Vecteur aléatoire discret

Définition 1.1.3 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. On dit que X est un vecteur aléatoire discret s'il prend ses valeurs dans un sous-ensemble dénombrable de \mathbb{R}^n c'est-à-dire les X_1, \dots, X_n sont des variables aléatoires discrètes.

-Vecteur aléatoire à densité

Définition 1.1.4 Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est à densité s'il existe une fonction positive Lebesgue-intégrable $f_X : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que $\forall A \in \mathcal{B}(\mathbb{R}^n), \mathbb{P}(X \in A) = \int_A f_X(x_1, \dots, x_n) dx_1 \cdots dx_n$.

1.1.2 Loi conjointe et loi marginale d'un vecteur aléatoire

Comme pour les variables aléatoires, un vecteur aléatoire induit naturellement une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Définition 1.1.5 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On appelle fonction de répartition (f.d.r) conjointe de X la fonction $F_X : \mathbb{R}^n \rightarrow [0, 1]$ définie par :

$$F_X(x) = \mathbb{P}(X \leq x), \forall x \in \mathbb{R}^n.$$

La densité f de X existe si F_X est n fois différentiable et on a :

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n F_X(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}.$$

Propriété 1.1.1 Les fonctions de répartition conjointes possèdent des propriétés tout à fait analogues à celles des fonctions de répartition. La fonction de répartition conjointe F_X d'un vecteur aléatoire X satisfait :

1. $\lim_{x_i \rightarrow -\infty} F_X(x_i) = 0, \lim_{x_i \rightarrow +\infty} F_X(x_i) = 1 \quad \forall i = 1, \dots, n;$
2. si $x \leq y$ alors $F_X(x) \leq F_X(y)$.

Définition 1.1.6 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. Alors les f.d.r $F_{X_i}, i = 1, \dots, n$ sont appelées fonctions de répartition marginales de F_X . Il est possible de caractériser simplement l'indépendance de variables aléatoires en termes de leur fonction de répartition conjointe.

Lemme 1.1.1 La famille (X_1, \dots, X_n) de variables aléatoires est indépendante si et ssi $F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i), \forall (x_1, \dots, x_n) \in \mathbb{R}^n$.

1.1.3 Espérance, covariance et corrélation

Espérance d'un vecteur aléatoire.

Soit $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire (discret ou continu).

Définition 1.1.7 L'espérance d'un vecteur aléatoire de dimension n est le vecteur de \mathbb{R}^n constitué des espérances de chacune des coordonnées i.e, $\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))$.

Proposition 1.1.1 Soit $X = (X_1, \dots, X_n) \in \mathbb{R}^n$, on a : $\mathbb{E}(X_1 + X_2 + \dots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n)$. En particulier soit $b \in \mathbb{R}^m$ et $A \in \mathcal{M}_{m,n}(\mathbb{R})$ (l'ensemble des matrices à m lignes et n colonnes à coefficients dans \mathbb{R}), L'espérance du vecteur aléatoire $Y = AX + b \in \mathbb{R}^m$ est $\mathbb{E}(Y) = A\mathbb{E}(X) + b$.

Covariance et corrélation.

Soit $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire (discret ou continu).

Définition 1.1.8 La covariance de X_i et X_j est donnée par :

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \mathbb{E}\left((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right) \\ &= \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j). \end{aligned}$$

Définition 1.1.9 Le coefficient de corrélation linéaire entre X_i et X_j est donné par :

$$\rho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}} \in [-1, 1].$$

Avec $\text{Var}(X_i) = \mathbb{E}\left(X_i - \mathbb{E}(X_i)\right)^2$.

Proposition 1.1.2 Soit $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire (discret ou continu). Si X_i et X_j sont indépendantes alors $\text{Cov}(X_i, X_j) = 0$, car $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$.

Attention! La réciproque est en général fausse.

1.2 Concept de dépendance

Le terme copule (copula) vient du mot latin "copeulae" en anglais, qui signifie au sens figuré, lien, liaison, alliance ou union. Les copules représentent un outil innovant pour modéliser la structure de dépendance de plusieurs variables aléatoires. Elle est introduite par Sklar (1959) dans des études sur les espaces de probabilité métriques, ce dernier a élaboré un théorème fondamental dans la théorie des copules. Ce théorème permet la liaison entre la densité conjointe d'un vecteur aléatoire et ses densités marginales. Pour citer ce résultat, supposons un couple (X, Y) de variables aléatoires dont la loi conjointe est définie par $H(x, y) = \mathbb{P}(X \leq x, Y \leq y)$. 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

$$H(x, y) = C\left(F(x), G(y)\right)$$

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

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

On peut donc en déduire les copules de Student et gaussienne, en particulier. La représentation de Sklar met en lumière les trois parties d'une loi bivariée, à savoir :

- la loi de X , caractérisée par la fonction de répartition F ,
- la loi de Y , caractérisée par la fonction de répartition G ,
- la structure de dépendance C de (X, Y) .

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

La copule a la propriété souhaitable de découpler l'étude de la distribution marginale de chaque variable à partir de l'étude de leur comportement collectif ou de leur dépendance. En effet, la dépendance entre les variables est entièrement intégrée dans la copule. De sorte qu'une copule permet une simple description de la structure de dépendance entre les variables indépendamment des marginales. Par exemple, les variables peuvent avoir des marginales de loi non gaussienne et une copule gaussienne ou alternativement des marginales gaussiennes et une copule non gaussienne, et toute combinaison possible de celles-ci. Par conséquent, la détermination de la distribution multivariée des variables aléatoires peut se faire en deux étapes :

- une détermination indépendante des distributions marginales, en utilisant des techniques standards pour les distributions d'une seule variable,
- une étude de la nature de la copule caractérisant complètement la dépendance entre les variables.

Cette séparation exacte entre les distributions marginales et la dépendance est très utile pour la gestion des risques et l'analyse de sensibilité, car elle permet de tester plusieurs scénarios avec différents types de dépendances entre les variables. Les copules constituent un outil statistique qui présente de nombreux avantages. La dépendance entre les variables n'a aucune raison à priori d'être linéaire, encore moins monotone. Outre une grande souplesse dans la mise en œuvre de l'analyse multivariée, les copules autorisent une sélection plus étendue des distributions conjointes des séries financières. Les fonctions copules permettent une représentation moins naïve de la dépendance statistique en finance, fondée sur la mesure traditionnelle de corrélation qui présente des limites dans l'étude de l'interdépendance entre deux variables. Elles autorisent des distributions de probabilité jointes moins restrictives, prenant mieux en compte certains faits en finance. Elles permettent la construction de distributions multidimensionnelles assez générales, et ce, indépendamment des lois des marginales et peuvent avoir des lois différentes

et quelconques ce qui permet de s'affranchir de certaines hypothèses peu réalistes faites dans les études empiriques.

Par ailleurs l'approche par les copules a beaucoup contribué dans la modélisation statistique multivariée. En effet, la théorie des copules permet une décomposition de la loi multidimensionnelle en ses marginales univariées et en une fonction de dépendance, rendant possible des extensions naturelles de certains résultats obtenus dans le cas univarié au cas multivarié.

IL est clair maintenant 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. Les distributions multidimensionnelles obtenues des copules sont d'avantages en adéquation avec la réalité et surtout dans l'utilisation financière des statistiques. Par souci de simplicité et du fait que la théorie multivariée est une extension du cas bivarié, nous nous limiterons à la théorie bivariée des copules.

Chapitre 2

Théorie des copules bivariées

Les copules représentent un outil innovant pour modéliser la structure de dépendance de plusieurs variables aléatoires. Cette notion "copula" fut utilisée pour la première fois dans le travail de Sklar (1959) comme solution d'un problème de probabilité énoncé par Maurice Fréchet dans le cadre des espaces métriques aléatoires et dérive du mot latin "copulare" qui signifie lien. Sklar a élaboré un théorème fondamental dans la théorie des copules. Ce théorème permet de relier la densité conjointe d'un vecteur aléatoire et ses densités marginales. Le théorème de Sklar 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, ce résultat permet de construire des lois avec une copule et des marges choisies. On en arrive à la question de la construction de la copule.

2.1 Propriétés des copules

Définition 2.1.1 *Une copule bivariée est une fonction $C : [0, 1]^2 \rightarrow [0, 1]$ satisfaisant les conditions suivantes :*

1. $C(u, 0) = C(0, v) = 0$,
2. $C(u, 1) = u$ et $C(1, v) = v$,
3. $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ Pour tout $u_1 \leq u_2$ et pour tout $v_1 \leq v_2$.

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

2.1.1 Continuité d'une copule

Soit une copule C de dimension 2 telle que pour u et v dans $[0, 1]^2$,

$$\left| C(u) - C(v) \right| \leq \sum_{i=1}^2 |v_i - u_i|,$$

alors C est uniformément continue sur $[0, 1]^2$. Sklar (1959) établit un lien entre la marginale conjointe et les marginales individuelles. Ce lien est appelé la "fonction copule". Ce qui nous donne la possibilité d'utiliser la copule si l'on veut déterminer la distribution d'un ensemble de variables aléatoires.

2.1.2 Théorème de Sklar

Théorème 2.1.1 *Soit $H : \mathbb{R}^2 \rightarrow [0, 1]$ une fonction conjointe avec les marginales F et G , il existe toujours une fonction $C : [0, 1]^2 \rightarrow [0, 1]$ telle que*

$$H(x, y) = C(F(x), G(y)). \quad (2.1.1)$$

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

2.1.3 Inversion du théorème de Sklar

Corollaire 2.1.1 *Une copule peut être aussi vue comme une fonction qui associe les quantiles de l'une des variables aléatoires avec les quantiles des autres variables aléatoires. En réécrivant l'équation énoncée dans le théorème de Sklar, nous obtenons le résultat suivant :*

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)). \quad (2.1.2)$$

Avec F^{-1} l'inverse généralisée de F ou la fonction quantile de F définie pour tout $u \in]0, 1[$ par :

$$F^{-1}(u) = \inf \left\{ x \in \mathbb{R} : F(x) \geq u \right\}. \quad (2.1.3)$$

À partir de cette expression, nous déduisons que la copule n'est autre qu'une fonction de répartition dont les marginales sont uniformes. Elle peut être écrite comme suit :

soit U et V des variables aléatoires uniformes, alors :

$$C(u, v) = \mathbb{P}(U \leq u, V \leq v). \quad (2.1.4)$$

2.1.4 Invariance de la fonction copule

Une propriété très puissante des copules est leur invariance sous des transformations strictement croissantes des variables aléatoires.

Théorème 2.1.2 *Soient X et Y deux variables aléatoires continues de copule associée C_{XY} . Si f et g sont deux fonctions strictement croissantes alors les variables $f(X)$ et $g(Y)$ ont exactement la même copule C_{XY} .*

Preuve:

Notons d'une part H la distribution conjointe des variables aléatoires continues X et Y de lois respectives F et G , et d'autre part H' la distribution de la transformée $(f(X), g(Y))$ de marges respectives F' et G' , on a :

$$\forall x \in \mathbb{R}, F'(x) = \mathbb{P}(f(X) \leq x) = \mathbb{P}(X \leq f^{-1}(x)) = F(f^{-1}(x)).$$

f^{-1} existe, car f est strictement croissante.

De même, on a :

$$\forall y \in \mathbb{R}, G'(y) = \mathbb{P}(g(Y) \leq y) = \mathbb{P}(Y \leq g^{-1}(y)) = G(g^{-1}(y)).$$

Donc $F'^{-1}(u) = f(F^{-1}(u))$ et $G'^{-1}(v) = g(G^{-1}(v))$.

Il vient que $\forall u, v \in [0, 1]$,

$$\begin{aligned} C_{f(X)g(Y)}(u, v) &= H'(F'^{-1}(u), G'^{-1}(v)) \\ &= \mathbb{P}\left(f(X) \leq F'^{-1}(u), g(Y) \leq G'^{-1}(v)\right) \\ &= \mathbb{P}\left(f(X) \leq f(F^{-1}(u)), g(Y) \leq g(G^{-1}(v))\right) \\ &= \mathbb{P}\left(X \leq F^{-1}(u), Y \leq G^{-1}(v)\right) \\ &= H\left(F^{-1}(u), G^{-1}(v)\right) \\ &= C_{XY}(u, v). \end{aligned}$$

■

Ce résultat nous montre que la dépendance entre les variables aléatoires est complètement capturée par la copule, indépendamment de la forme des distributions marginales.

2.1.5 Densité d'une copule

La fonction de densité d'une copule est donnée par :

$$c(u_1, u_2, \dots, u_d) = \frac{\partial^d C(u_1, u_2, \dots, u_d)}{\partial u_1 \partial u_2 \cdots \partial u_d} = \frac{f(F_1^{-1}[u_1], \dots, F_d^{-1}[u_d])}{f_1(F_1^{-1}[u_1]) \cdots f_d(F_d^{-1}[u_d])}.$$

Avec f la fonction de densité conjointe de F , f_1, \dots, f_d les densités marginales de F_1, \dots, F_d et $F_1^{-1}, \dots, F_d^{-1}$ les fonctions de distributions marginales inverses; si et seulement si les hypothèses suivantes sont satisfaites :

1. la copule C est une copule absolument continue et différentiable,
2. F est une fonction de répartition jointe absolument continue,
3. les fonctions de répartition marginales $F_1[x_1], \dots, F_d[x_d]$ sont continues et strictement croissantes.

Denuit et *al.* (2005) ont introduit une propriété importante qui permet de déterminer la densité conjointe d'une suite de variables aléatoires à partir de la densité de la copule. Cette propriété s'énonce comme suit :

Propriété 2.1.1 Soient F_1, \dots, F_d des fonctions de répartition continues avec des densités marginales respectives f_1, \dots, f_d . La distribution conjointe de $(X_1, X_2, \dots, X_d)^t$ est donnée par :

$$f(x_1, x_2, \dots, x_d) = \prod_{i=1}^d f_i(x_i) \times c(u_1, u_2, \dots, u_d), \quad (2.1.5)$$

avec

$$c(u_1, u_2, \dots, u_d) = \frac{\partial^d C(u_1, u_2, \dots, u_d)}{\partial u_1 \partial u_2 \cdots \partial u_d}. \quad (2.1.6)$$

Interprétation de cette propriété :

cette équation signifie qu'en cas de dépendance, la densité conjointe est égale au produit de la densité conjointe en cas d'indépendance et de la densité de la copule C de dimension d . Autrement dit, si les variables X_1, X_2, \dots, X_d sont indépendantes, alors $c = 1$, ce qui implique que

$$f(x_1, x_2, \dots, x_d) = \prod_{i=1}^d f_i(x_i). \quad (2.1.7)$$

2.1.6 Copules usuelles

Copule indépendante.

Une des copules les plus commune 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 produit de ses marges F et G , i.e. $H(x, y) = F(x)G(y)$, nous pouvons montrer par une application de la formule (2.1.2), que la copule associée est $\Pi(u, v) = u \times v$. En effet,

$$\begin{aligned} \forall u, v \in [0, 1], C(u, v) &= H(F^{-1}(u), G^{-1}(v)) \\ &= F[F^{-1}(u)]G[G^{-1}(v)] \\ &= uv. \end{aligned}$$

$\forall u, v \in [0, 1]$, $\Pi(u, v) = u \times v$ définit une copule. En effet, soient $u, v \in [0, 1]$, on a $\Pi(u, 0) = \Pi(0, v) = 0 \Rightarrow \Pi$ vérifie la condition 1 de la définition 2.1.1 ; $\Pi(u, 1) = u$ et $\Pi(1, v) = v \Rightarrow \Pi$ vérifie aussi la condition 2 de la définition 2.1.1 ; de même $\forall u, u', v, v' \in [0, 1]$, avec $u \leq u'$ et $v \leq v'$ on a :

$$\begin{aligned} \Pi(u', v') - \Pi(u', v) - \Pi(u, v') + \Pi(u, v) &= u'v' - u'v - uv' + uv \\ &= u'(v' - v) - u(v' - v) \\ &= (v' - v)(u' - u) \\ &\geq 0 \end{aligned}$$

$\Rightarrow \Pi$ vérifie de plus la condition 3 de la définition 2.1.1.

Par conséquent Π est une copule.

Copule maximale.

$\forall u, v \in [0, 1]$, $M(u, v) = \min(u, v)$ définit une copule. En effet, soient $u, v \in [0, 1]$, on a $\min(u, 0) = \min(0, v) = 0 \Rightarrow M$ vérifie la condition 1 de la définition 2.1.1 ; $\min(u, 1) = u$ et $\min(1, v) = v \Rightarrow M$ vérifie aussi la condition 2 de la définition 2.1.1 ; de même $\forall u, u', v, v' \in [0, 1]$, avec $u \leq u'$ et $v \leq v'$ on a $M(u', v') - M(u', v) - M(u, v') + M(u, v) \geq 0 \Rightarrow M$ vérifie de plus la condition 3. En effet, pour $u' \geq v'$ on a les trois possibilités :

1. $u' \geq u \geq v' \geq v$,
2. $u' \geq v' \geq u \geq v$,
3. $u' \geq v' \geq v \geq u$.

$$\begin{aligned}
1. \Rightarrow \min(u', v') - \min(u', v) - \min(u, v') + \min(u, v) &= v' - v - v' + v \\
&= 0, \\
2. \Rightarrow \min(u', v') - \min(u', v) - \min(u, v') + \min(u, v) &= v' - v - u + v \\
&= v' - u \\
&\geq 0, \\
3. \Rightarrow \min(u', v') - \min(u', v) - \min(u, v') + \min(u, v) &= v' - v - u + u \\
&= v' - v \\
&\geq 0.
\end{aligned}$$

Dans tous les trois cas, on a $M(u', v') - M(u', v) - M(u, v') + M(u, v) \geq 0$.

Et avec la même démarche on retrouve que

$M(u', v') - M(u', v) - M(u, v') + M(u, v) \geq 0$ pour $v' \geq u'$. Par conséquent

M est une copule.

La copule maximale M, $M(u, v) = \min(u, v)$, est la borne sup de Fréchet.

Copule minimale.

$\forall u, v \in [0, 1]$, $W(u, v) = \max(u + v - 1, 0)$ définit aussi une copule, c'est la copule minimale. W est la borne inf de Fréchet.

2.1.7 Bornes de Fréchet-Hoeffding

Dès 1940, Hoeffding travailla sur les distributions multivariées. Certes, il n'a pas introduit cette notion de copule directement, mais ces travaux ont contribué à déduire beaucoup de résultats intéressants.

On a : $\forall u, v \in [0, 1]$, $C(u, v) \leq C(u, 1) = u$ et $C(u, v) \leq C(1, v) = v$.

$\Rightarrow C(u, v) \leq \min(u, v)$. (a)

La condition 3 de la définition 2.1.1

$\Rightarrow \forall u, v \in [0, 1]$, $C(u, v) \geq C(u, 1) + C(1, v) - C(1, 1)$

$\Rightarrow C(u, v) \geq u + v - 1$

or $\forall u, v \in [0, 1]$, $C(u, v) \geq 0$

donc $C(u, v) \geq \max(u + v - 1, 0)$. (b)

(a) et (b) $\Rightarrow \max(u + v - 1, 0) = W(u, v) \leq C(u, v) \leq M(u, v) = \min(u, v)$.

Les copules W et M sont appelées borne inférieure respectivement borne supérieure de Fréchet-Hoeffding ou copule minimale respectivement maximale.

Elles sont aussi notées C^- respectivement C^+ .

Pour toute copule C, on a la relation : $W(u, v) \leq C(u, v) \leq M(u, v)$.

La famille de Fréchet est formée par les combinaisons linéaires convexes de M, W et Π . Comme exemple, on a la copule de Mardia définit par $\forall \theta \in$

$[-1,1], \forall u, v \in [0,1],$

$$C_\theta(u, v) = \frac{\theta^2(1 + \theta)}{2}M(u, v) + (1 - \theta^2) \prod(u, v) + \frac{\theta^2(1 - \theta)}{2}W(u, v). \quad (2.1.8)$$

2.2 Notions de dépendance

On dit que deux variables X et Y sont indépendantes si la réalisation de l'une des variables n'a aucun effet sur la réalisation de l'autre variable. Soient F et G les fonctions de répartition de X et Y , H la loi conjointe.

X et Y sont dépendantes $\iff H(x, y) \neq F(x) \times G(y), \forall (x, y) \in \mathbb{R}^2$.

2.2.1 Les mesures de la dépendance

Planchet a défini une mesure de dépendance comme étant une application qui associe à deux variables aléatoires un réel permettant de quantifier la force de la dépendance qui les lie.

Propriété 2.2.1 *Une mesure de dépendance δ est une mesure de concordance si elle possède les propriétés suivantes :*

1. *symétrie* : $\delta(X, Y) = \delta(Y, X),$
2. *normalisation* : $-1 \leq \delta(X, Y) \leq 1,$
3. $\delta(X, Y) = 1 \iff (X, Y) =_{loi} (F_1^{-1}(U), F_2^{-1}(U))$ avec $U \rightsquigarrow \mathcal{U}_{[0,1]},$
4. $\delta(X, Y) = -1 \iff (X, Y) =_{loi} (F_1^{-1}(U), F_2^{-1}(U))$ avec $U \rightsquigarrow \mathcal{U}_{[0,1]},$
5. *si f est strictement monotone alors,*

$$\delta(f(X, Y)) = \begin{cases} \delta(X, Y) & \text{si } f \uparrow \\ -\delta(X, Y) & \text{si } f \downarrow. \end{cases}$$

Coefficient de corrélation de Pearson.

Le coefficient de corrélation de Pearson est une mesure d'association entre deux variables aléatoires permettant de capturer le degré de la relation linéaire qui existe entre celles-ci. Pour deux variables X et Y possédant des variances finies, il se définit comme suit :

$$\rho_{pears}(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)} \times \sqrt{Var(Y)}}. \quad (2.2.1)$$

Les variances de X et Y doivent être finies pour que le coefficient de corrélation de Pearson soit défini. Ceci est donc un inconvénient pour les mesures de dépendance et cause des problèmes lorsqu'on travaille avec des distributions à queue lourde.

Propriété 2.2.2 Soit t_1 et t_2 deux transformations strictement croissantes, alors :

$$\rho_{pears}(t_1(X), t_2(Y)) \neq \rho_{pears}(X, Y),$$

(pas d'invariance par transformation strictement croissante).

Soit F et G les fonctions de répartition de X et Y , et C une copule, alors :

$$\rho_{pears} = \frac{1}{\sqrt{Var(X)} \times \sqrt{Var(Y)}} \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) dF^{-1}(u_1) dG^{-1}(u_2), \quad (2.2.2)$$

ρ_{pears} dépend de la copule et des marginales dans cette expression.

Coefficient de corrélation de rang de Kendall.

C'est une mesure d'association non paramétrique basée sur le nombre de concordances et de discordances dans un échantillon d'observation paire.

Autrement dit, soit $X = (X_1, X_2)$ et $X' = (X'_1, X'_2)$ deux paires d'observations i.i.d. (indépendantes et identiquement distribuées), On a :

$$P(\text{concordance}) = \mathbb{P}([X_1 - X'_1][X_2 - X'_2] > 0). \quad (2.2.3)$$

Et

$$P(\text{discordance}) = \mathbb{P}([X_1 - X'_1][X_2 - X'_2] < 0). \quad (2.2.4)$$

Le coefficient de corrélation de rang de Kendall noté $\rho_{Kendall}$ se définit comme suit :

$$\rho_{Kendall} = P(\text{concordance}) - P(\text{discordance}). \quad (2.2.5)$$

Si les marginales de X_1 et X_2 sont continues alors le $\rho_{Kendall}$ peut être réécrite comme suit :

$$\rho_{Kendall}(X_1, X_2) = 2\mathbb{P}([X_1 - X'_1][X_2 - X'_2] > 0) - 1, \quad (2.2.6)$$

où (X'_1, X'_2) est une copie indépendante de (X_1, X_2) et les deux couples sont mutuellement indépendants.

Une autre propriété intéressante est l'invariance du coefficient de corrélation de rang de Kendall, c'est-à-dire $\forall t_1$ et t_2 deux transformations strictement croissantes,

$$\rho_{Kendall}(t_1(X_1), t_2(X_2)) = \rho_{Kendall}(X_1, X_2). \quad (2.2.7)$$

Comme

$$\begin{aligned}\rho_{Kendall}(X_1, X_2) &= 2\mathbb{P}([X_1 - X'_1][X_2 - X'_2] > 0) - 1 \\ &= 4\mathbb{P}(X_1 \leq X'_1, X_2 \leq X'_2) - 1,\end{aligned}$$

alors $\rho_{Kendall}(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1 = 4\mathbb{E}(C(U_1, U_2)) - 1$,
avec $U_i \rightsquigarrow \mathcal{U}_{[0,1]} \forall i = 1, 2$.

Soit φ le générateur associé à la copule archimédienne, alors :

$$\rho_{Kendall}^\varphi = 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt + 1. \quad (2.2.8)$$

La démonstration de ce résultat peut être vue dans Genest et Mackay (1986b).

En appliquant cette relation à la copule de Clayton de générateur $\varphi(t) = \frac{t^{-\alpha}-1}{\alpha}$, nous obtenons :

$$\rho_{Kendall}^\varphi = \frac{\alpha}{\alpha + 2}. \quad (2.2.9)$$

Avec α , le paramètre de la copule de Clayton.

Coefficient de corrélation de rang de Spearman.

Le coefficient de corrélation de Spearman $\rho_{Spearman}$ pour une paire de variables continue $X = (X_1, X_2)$ est identique au coefficient de corrélation de Pearson pour les rangs de F_1 et F_2 .

$$\rho_{Spearman}(X_1, X_2) = \rho_{pears}(F_1(X_1), F_2(X_2)). \quad (2.2.10)$$

Comme dans le cas du coefficient de corrélation de Kendall,

$$\rho_{Spearman} = 3 \left(\mathbb{P}([X_1 - X'_1][X_2 - X'_2] > 0) \right) - \mathbb{P}([X_1 - X'_1][X_2 - X'_2] < 0), \quad (2.2.11)$$

où $X' = (X'_1, X'_2)$ est une copie indépendante de (X_1, X_2) . En outre (X'_1, X'_2) est un vecteur aléatoire avec des composantes indépendantes telles que $X_1 \stackrel{loi}{=} X'_1$, $X_2 \stackrel{loi}{=} X'_2$ et, (X_1, X_2) et (X'_1, X'_2) sont mutuellement indépendants.

Une autre propriété intéressante est l'invariance du coefficient de corrélation de rang de Spearman c'est-à-dire, $\forall t_1$ et t_2 deux transformations strictement croissantes,

$$\rho_{Spearman}(t_1(X_1), t_2(X_2)) = \rho_{Spearman}(X_1, X_2). \quad (2.2.12)$$

De la relation qui lie le coefficient de corrélation de Spearman et le coefficient de corrélation de Pearson, nous avons :

$$\rho_{Spearman}(X_1, X_2) = \rho_{Pearson}(F_1(X_1), F_2(X_2)) \quad (2.2.13)$$

$$= 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2. \quad (2.2.14)$$

2.2.2 Dépendance de queue

Une question importante pour la gestion des risques concerne la détermination de la dépendance des distributions dans les queues. En pratique, la question est de savoir s'il est plus probable que des événements importants ou extrêmes se produisent simultanément ou au contraire, de manière plus ou moins indépendante. Ceci est noté comme la présence ou l'absence de "dépendance de queue".

Techniquement, nous devons déterminer la probabilité qu'une variable aléatoire X soit grande sachant que la variable aléatoire Y est grande.

Définition 2.2.1 *Soit X et Y des variables aléatoires avec des marginales F_X et F_Y continues. Le coefficient de dépendance de queue (supérieur) de X et Y est s'il existe,*

$$\lim_{u \rightarrow 1} \mathbb{P} \left\{ X > F_X^{-1}(u) / Y > F_Y^{-1}(u) \right\} = \lambda \in [0, 1]. \quad (2.2.15)$$

Autrement dit, étant donné que Y soit très grande (ce qui se produit avec la probabilité $1-u$), la probabilité que X soit très grande au même niveau de probabilité u , définit asymptotiquement le coefficient de dépendance de queue.

Il s'avère que cette dépendance de queue est une propriété de copule pure qui est indépendante des marginales. Soit C la copule des variables X et Y (supposées continues), alors :

Théorème 2.2.1 *Si la copule bivariée C est telle que*

$$\lim_{u \rightarrow 1} \frac{\bar{C}(u, u)}{1 - u} = \lambda, \quad (2.2.16)$$

existe (où $\bar{C}(u, u) = 1 - 2u + C(u, u)$), alors C a un coefficient de dépendance de queue supérieur λ .

Si $\lambda > 0$, la copule présente une dépendance de queue, et les grands événements ont tendance à se produire simultanément avec la probabilité λ . Au contraire, lorsque $\lambda = 0$, la copule n'a pas de dépendance de queue dans

ce sens et les grands événements semblent se produire essentiellement indépendamment. Il y a cependant une subtilité dans cette définition de la dépendance de queue. Pour être clair considérons d'abord le cas où pour les grands X et Y , la fonction de distribution $F(x, y)$ se factorise de telle sorte que

$$\lim_{x, y \rightarrow \infty} \frac{F(x, y)}{F_X(x)F_Y(y)} = 1. \quad (2.2.17)$$

Cela signifie que pour X et Y suffisamment grandes, ces deux variables peuvent être considérées comme indépendantes. On montre alors que

$$\begin{aligned} \lim_{u \rightarrow 1} \mathbb{P}\left\{X > F_X^{-1}(u)/Y > F_Y^{-1}(u)\right\} &= \lim_{u \rightarrow 1} 1 - F_X\left(F_X^{-1}(u)\right) \\ &= \lim_{u \rightarrow 1} 1 - u = 0. \end{aligned}$$

De sorte que les variables indépendantes n'ont vraiment aucune dépendance de queue, comme on peut s'y attendre. Malheureusement l'inverse n'est pas toujours vrai : une valeur $\lambda = 0$ n'implique pas automatiquement que l'indépendance est vraie, à savoir que

$$\lim_{x, y \rightarrow \infty} \frac{F(x, y)}{F_X(x)F_Y(y)} = 1.$$

En effet le critère d'indépendance de queue $\lambda = 0$ peut encore être dû à une absence de factorisation de la distribution multivariée pour les grandes X et Y . Dans un sens plus faible, il peut avoir une dépendance dans la queue même lorsque $\lambda = 0$. Un tel comportement est par exemple présenté par la copule gaussienne, qui a une dépendance de queue nulle, mais n'a pas de distribution multivariée factorisable, puisque le terme non diagonal de la forme quadratique dans la fonction exponentielle ne devient pas négligeable en général, car X et Y tendent vers l'infini (voir [14]). Pour résumer, l'indépendance de queue selon la définition, n'est pas équivalente à l'indépendance dans la queue telle que donnée par (2.2.17).

2.2.3 Estimation de la structure de dépendance

Ce paragraphe présente les principales méthodes permettant d'estimer le paramètre de la copule.

Estimation non paramétrique.

1. Fonction de répartition empirique.

Le cadre général de l'estimation non paramétrique d'une loi marginale s'appuie sur la fonction de répartition empirique, définie par :

$$\hat{F}(x) = \frac{1}{k} \sum_{i=1}^k \mathbf{1}_{\{X_i \leq x\}}, \quad (2.2.18)$$

pour un échantillon de taille k , (x_1, \dots, x_k) de la loi F .

Pour un d -échantillon d'un vecteur X de dimension k , on peut généraliser l'expression de la fonction de répartition empirique en posant :

$$\hat{F}_j(x) = \frac{1}{k} \sum_{i=1}^k \mathbf{1}_{\{X_{ij} \leq x\}}, \text{ avec } j \in \{1, \dots, d\}. \quad (2.2.19)$$

Cet estimateur conduit à un estimateur non paramétrique naturel d'une copule.

2. Copule empirique.

Les copules empiriques ont été initialement introduites par Deheuvels (1979) et connues sous l'appellation de fonction empirique de dépendance.

Cas des copules bivariées :

soient $\{(x_k, y_k)\}_{k=1}^n$ un échantillon de taille n de couples de variables aléatoires. La copule empirique est la fonction C_n définie par :

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\text{nombre de paires } (x, y) \text{ tels que } x \leq x_i, y \leq y_j}{n}. \quad (2.2.20)$$

Généralisation dans les cas de copules multivariées :

pour un p -échantillon du vecteur X (de dimension n), $(X_{1j}, \dots, X_{nj})_{j=1, \dots, p}$, la copule empirique est définie de façon analogue au cas bivarié, C_n est donc définie comme suit :

$$C_n\left(\frac{k_1}{n}, \dots, \frac{k_p}{n}\right) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_{ij} \leq k_j\}}. \quad (2.2.21)$$

Cette expression peut être également écrite à partir des rangs des observations puisque la copule est invariante pour toute transformation croissante des marginales. On a :

$$C_n\left(\frac{k_1}{n}, \dots, \frac{k_p}{n}\right) = \frac{1}{n} \text{card} \left\{ i / R_{i1} \leq k_1, \dots, R_{ip} \leq k_p \right\}, \quad (2.2.22)$$

avec R_{ij} , $j = \{1, \dots, p\}$ et $i = \{1, \dots, n\}$ le rang de X_{ij} .

Estimation paramétrique et semi-paramétrique.

1. Méthode des moments.

C'est une méthode simple qui permet d'exprimer les paramètres de la copule à partir de la relation qui existe entre cette dernière et les mesures de concordances. Par exemple, pour le tau de Kendall nous avons : $\rho_{Kendall} = \frac{c-d}{c+d}$, avec c (respectivement d) le nombre de paires concordantes (respectivement discordantes) dans l'échantillon. Pour les copules archimédiennes de paramètre α , nous avons :

- pour la Copule de Clayton, $\alpha = \frac{2\rho_{Kendall}}{1-\rho_{Kendall}}$,

- pour la Copule de Gumbel, $\alpha = \frac{1}{1-\rho_{Kendall}}$.

2. Maximum de vraisemblance.

La densité conjointe d'un vecteur aléatoire $X = (x_1, x_2, \dots, x_n)$ se définit comme suit :

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_i(x_i; \theta) \times c(u_1, u_2, \dots, u_n; \theta). \quad (2.2.23)$$

Avec $\theta \in \mathcal{O}$ (l'espace des paramètres). La fonction log-vraisemblance est donnée par :

$$l(\theta) = \sum_{i=1}^n \ln(f_i(x_i; \theta)) + \ln(c(u_1, u_2, \dots, u_n; \theta)). \quad (2.2.24)$$

Nous déduisons que pour un échantillon (x_{1j}, \dots, x_{nj}) avec $1 \leq j \leq d$, le log de vraisemblance est :

$$l(\theta) = \sum_{j=1}^d \sum_{i=1}^n \ln(f_i(x_{ij}; \theta)) + \sum_{j=1}^d \ln(c(F_j(x_{1j}), \dots, F_j(x_{nj}); \theta)). \quad (2.2.25)$$

$$\hat{\theta} = ArgMax(l(\theta)), \quad (2.2.26)$$

$\hat{\theta}$ étant un vecteur de paramètres composé des paramètres des lois marginales et du paramètre de la copule.

3. Méthode IFM (Inference Function for Marging).

Cette méthode est proposée par Joe et Xu (1996) pour palier à la complexité de la méthode du maximum de vraisemblance. Elle comprend deux étapes :

étape 1 : estimer les paramètres des distributions marginales

$$\hat{\theta}_i = ArgMax\left(\sum_{j=1}^d \ln(f_i(x_{ij}; \theta_i))\right), \forall i \in [1, n], \quad (2.2.27)$$

étape 2 : à partir des estimateurs obtenus à l'étape 1, nous déduisons l'estimateur du paramètre de la copule

$$\hat{\alpha} = \mathit{ArgMax} \left\{ \sum_{j=1}^d \ln \left(c \left(F_j(x_{1j}; \hat{\theta}_1), \dots, F_j(x_{nj}; \hat{\theta}_n); \alpha \right) \right) \right\}. \quad (2.2.28)$$

La méthode proposée par Joe et Xu est plus pratique que l'estimation directe par le maximum de la vraisemblance classique. En effet, elle permet de réduire considérablement la complexité des calculs et rend ainsi le calibrage de plusieurs modèles multivariés faisable.

Par ailleurs, il est important de préciser que la comparaison directe de l'IFM et de la méthode d'estimation par le maximum de vraisemblance classique est un point assez difficile à cause du temps de calcul nécessaire pour obtenir les estimations. Cependant, de nombreux tests comparatifs effectués dans Xu (1996) sur plusieurs modèles multivariés montrent que l'IFM est une méthode très efficace et donne des résultats proches de la méthode classique.

4. Méthode CML (Canonical Maximum Likelihood).

C'est une méthode recommandée par Bouye et *al.* (2000). Sa mise en œuvre ne nécessite pas l'estimation des paramètres des distributions marginales.

Il s'agira ici de transformer les observations $(x_{1j}, \dots, x_{nj})_{1 \leq j \leq d}$ en uniforme $(\hat{u}_{1j}, \hat{u}_{2j}, \dots, \hat{u}_{nj})_{1 \leq j \leq d}$ par le biais de la fonction de répartition empirique univariée. À partir de cette transformation, le paramètre de la copule se déduit comme suit :

$$\hat{\alpha} = \mathit{ArgMax} \left(\sum_{j=1}^d \ln \left(c(\hat{u}_{1j}, \hat{u}_{2j}, \dots, \hat{u}_{nj}; \alpha) \right) \right). \quad (2.2.29)$$

Cette méthode est la plus utilisée en pratique du fait de sa simplicité. Son grand avantage réside dans l'estimation paramétrique de la copule indépendamment de la forme paramétrique des marginales. De plus sa mise en œuvre demande un temps de calcul limité.

2.3 Familles de copules paramétriques

Nous exposons dans la présente section les principales copules paramétriques.

2.3.1 Les copules elliptiques

On dit qu'un vecteur aléatoire $X = (X_1, X_2, \dots, X_d)$ suit une loi elliptique si et seulement si il admet une représentation stochastique de la forme

$X = \mu + RAU$, où $\mu \in \mathbb{R}^d$ est le vecteur des moyennes, R est une variable aléatoire positive, U est un vecteur aléatoire uniformément distribué sur la sphère unité $\{z \in \mathbb{R}^d : z^t z = 1\}$ et $A \in \mathbb{R}^{d \times d}$ est une matrice telle que $\Sigma = AA^t$ est non singulière et définie positive. La fonction de répartition associée à X n'a généralement pas de forme explicite. Les copules elliptiques sont simplement les copules qui sont extraites des lois elliptiques via la représentation de Sklar inverse présentée à l'équation (2.1.2).

Copule normale :

la loi normale fait partie de la famille elliptique. Sa copule dans le cas bivarié est :

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} h_\rho(x, y) dx dy, \quad (2.3.1)$$

où

$$h_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\}, \quad (2.3.2)$$

est la densité d'une loi normale bivariée centrée, de variances unitaires et de corrélation $\rho \in [-1, 1]$, et Φ est la f.d.r d'une variable aléatoire de loi $\mathcal{N}(0, 1)$.

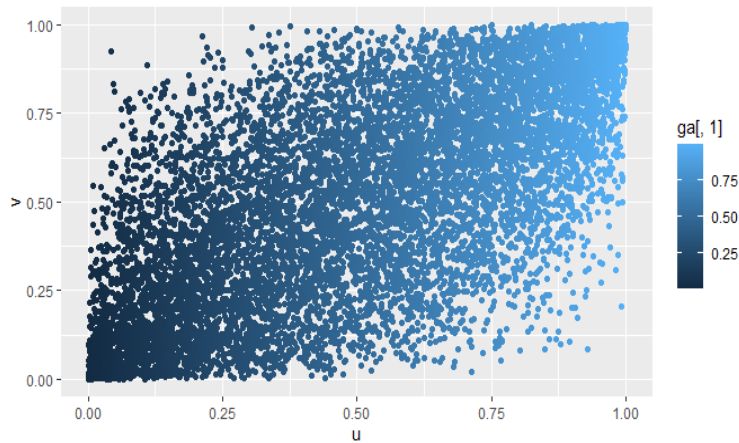


Figure 2.1 – Simulation de 10000 observations d'une copule gaussienne bivariée de paramètre $\theta = 0.7$.

Copule de Student :

la copule de Student est la copule qui dérive de la distribution multivariée de Student. La copule de Student est également un membre des modèles

elliptiques, elle est donnée par :

$$C_{\rho,\nu}(u, v) = \int_{-\infty}^{T_{\nu}^{-1}(u)} \int_{-\infty}^{T_{\nu}^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} dx dy, \quad (2.3.3)$$

où $\rho \in [-1, 1]$ et T_{ν} est la fonction de répartition de loi de Student à ν degré de liberté.

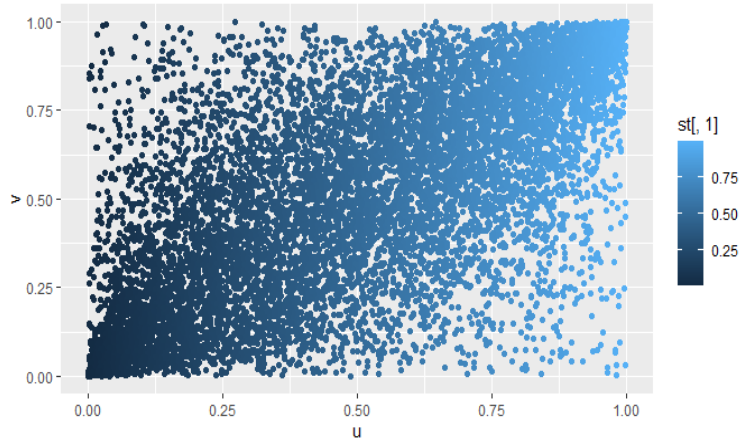


Figure 2.2 – Simulation de 10000 observations d'une copule de Student bivariée de paramètre $\theta = 0.7$.

2.3.2 Les copules archimédiennes

La théorie des copules archimédiennes a été introduite et développée par le professeur **Christian Genest** de l'université de Laval du Québec.

Définition 2.3.1 On appelle copule archimédienne toute copule de la forme

$$C(u, v) = \varphi^{-1}[\varphi(u) + \varphi(v)], \quad (2.3.4)$$

où la fonction $\varphi : [0, 1] \rightarrow \mathbb{R}^+$, appelée le générateur, est convexe, décroissante et telle que $\varphi(1) = 0$, et φ^{-1} la pseudo inverse de φ .

Ces copules possèdent des propriétés intéressantes telles que :

1. la symétrie, i.e. $C_{\varphi}(u, v) = C_{\varphi}(v, u)$, $\forall (u, v) \in [0, 1]^2$. En effet,

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

2. Si r est une constante strictement positive, alors $\varphi^*(t) := r\varphi(t)$ engendre la même copule que φ , i.e. $C_\varphi(u, v) = C_{\varphi^*}(u, v)$ pour tout $(u, v) \in [0, 1]^2$. En effet,

$$\begin{aligned} C_\varphi^*(u, v) &= \varphi^{*-1}[\varphi^*(u) + \varphi^*(v)] \\ &= \varphi^{-1}\left[\frac{1}{r}(r\varphi(u) + r\varphi(v))\right] \\ &= \varphi^{-1}[\varphi(u) + \varphi(v)] \\ &= C_\varphi(u, v). \end{aligned}$$

3. C_φ est la copule d'indépendance si et seulement si $\varphi(t) = -\ln(t)$. En effet on a $\varphi^{-1}(t) = \exp(-t)$, ce qui fait que

$$C(u, v) = \exp\left\{-\left(-\ln(u)\right) + \left(-\ln(v)\right)\right\} = uv.$$

Autrement dit, $\varphi(t) = -\ln(t)$ génère la copule d'indépendance.

Copule de Clayton :

parmi les modèles de copules archimédiennes, celle dont le générateur est :

$$\varphi_\theta(t) = \frac{t^{-\theta} - 1}{\theta}, \quad (2.3.5)$$

est la copule de Clayton. La loi jointe associée est :

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0. \quad (2.3.6)$$

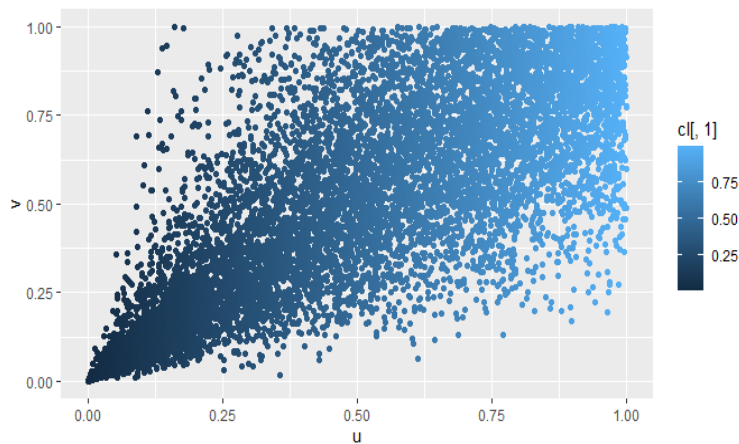


Figure 2.3 – Simulation de 10000 observations d’une copule de Clayton bivariée de paramètre $\theta = 3$ avec un tau de Kendall $\rho = 0.6$.

Copule de Frank :

un autre modèle populaire de copules archimédiennes est la copule de Frank dont le générateur est :

$$\varphi_\nu(t) = -\ln \left[\frac{\exp(-\vartheta t) - 1}{\exp(-\vartheta) - 1} \right], \quad \vartheta > 0. \quad (2.3.7)$$

L'expression de la copule engendrée est donc

$$C_\vartheta(u, v) = -\frac{1}{\vartheta} \ln \left[1 + \frac{(\exp(-\vartheta u) - 1)(\exp(-\vartheta v) - 1)}{\exp(-\vartheta) - 1} \right]. \quad (2.3.8)$$

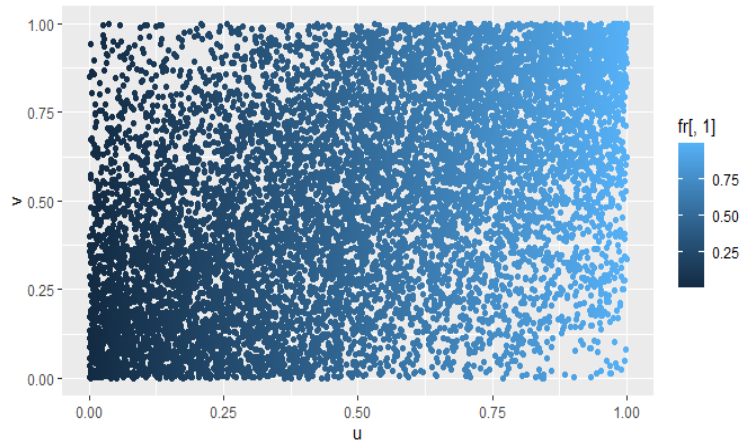


Figure 2.4 – Simulation de 10000 observations d'une copule de Frank bivariée de paramètre $\theta = 3$ avec un tau de Kendall $\rho = 0.3$.

Copule de Gumbel :

celle dont le générateur est :

$$\varphi_\theta(t) = \left(-(\ln(t))^\theta \right) \text{ avec } \theta \geq 1 \text{ et } t \in [0, 1], \quad (2.3.9)$$

est la copule de Gumbel. L'expression de la copule est donc

$$C_\theta(u, v) = \exp \left(- \left[(-\ln u)^\theta + (-\ln v)^\theta \right]^{\frac{1}{\theta}} \right). \quad (2.3.10)$$

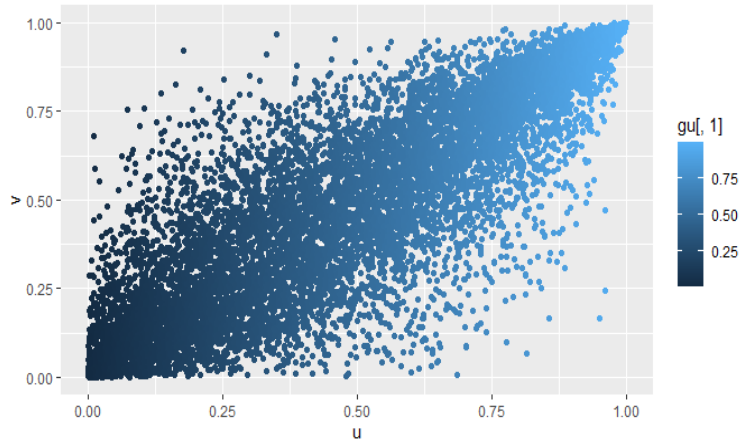


Figure 2.5 – Simulation de 10000 observations d’une copule de Gumbel bivariée de paramètre $\theta = 3$ avec un tau de Kendall $\rho = 0.66$.

Exemple : la borne inférieure de Fréchet est archimédienne. En effet, soit $\varphi(t) = 1 - t$, $\varphi^{-1}(t) = 1 - t$ si $t \in [0, 1]$ et $\varphi^{-1}(t) = 0$ si $t > 1$. Donc $\varphi^{-1}(t) = \max(1 - t, 0)$. La copule associée à ce générateur est :

$$\begin{aligned} C_\varphi(u, v) &= \varphi^{-1}[\varphi(u) + \varphi(v)] \\ &= \varphi^{-1}[2 - u - v] \\ &= \max[1 - (2 - u - v), 0] \\ &= \max[u + v - 1, 0]. \end{aligned}$$

Ainsi $C_\varphi = W$.

2.3.3 Copules à valeurs extrêmes

Soit une suite X_1, \dots, X_n de variables aléatoires i.i.d. de fonction de répartition F , et considérons leur maximum, à savoir la statistique

$$M_n = \max(X_1, \dots, X_n). \quad (2.3.11)$$

Supposons qu’il existe deux suites de nombres réels (a_n) et (b_n) telles que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} \left(F(a_n x + b_n)\right)^n = F^*(x), \quad (2.3.12)$$

où F^* est une fonction de répartition non dégénérée. Tel que démontrer par Fisher et Tippett (1928), il y a uniquement trois formes possibles pour F^* , à savoir les lois de Gumbel, Fréchet et Weibull, dont les fonctions de répartition sont données respectivement par :

1. $F^*(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$.
2. $F^*(x) = \exp(-x^{-a})$, $x > 0$ et $a > 0$.
3. $F^*(x) = \exp(-(-x)^a)$, $x \leq 0$ et $a > 0$.

La théorie des valeurs extrêmes a aussi été décrite dans le cas bivariable. Pour la décrire soit une suite de vecteurs aléatoires indépendants $(X_1, Y_1), \dots, (X_n, Y_n)$ dans \mathbb{R}^2 de fonction de répartition H , on considère les statistiques d'ordre maximales

$$M_{n,X} = \max(X_1, \dots, X_n) \text{ et } M_{n,Y} = \max(Y_1, \dots, Y_n), \quad (2.3.13)$$

leur distribution conjointe est :

$$\begin{aligned} \mathbb{P}(M_{n,X} \leq x, M_{n,Y} \leq y) &= \mathbb{P}(X_i \leq x, Y_i \leq y, \forall i \in 1, \dots, n) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \leq x, Y_i \leq y) \\ &= (H(x, y))^n. \end{aligned}$$

Pour étudier le comportement asymptotique de $(M_{n,X}, M_{n,Y})$ quand $n \rightarrow \infty$, il faut considérer une normalisation de ces variables aléatoires car

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_{n,X} \leq x, M_{n,Y} \leq y) = \lim_{n \rightarrow \infty} (H(x, y))^n = 0.$$

Pour tout $(x, y) \in \mathbb{R}^2$, soient les suites $a_{n,X}$, $b_{n,X}$, $a_{n,Y}$, $b_{n,Y}$ et posons

$$M_{n,X}^* = \frac{M_{n,X} - b_{n,X}}{a_{n,X}} \text{ et } M_{n,Y}^* = \frac{M_{n,Y} - b_{n,Y}}{a_{n,Y}}.$$

On calcule alors,

$$\begin{aligned} \mathbb{P}(M_{n,X}^* \leq x, M_{n,Y}^* \leq y) &= \mathbb{P}(M_{n,X} \leq a_{n,X}x + b_{n,X}, M_{n,Y} \leq a_{n,Y}y + b_{n,Y}) \\ &= (H(a_{n,X}x + b_{n,X}, a_{n,Y}y + b_{n,Y}))^n. \end{aligned}$$

Si les suites sont choisies de telle sorte que la limite quand $n \rightarrow \infty$ de la probabilité précédente converge vers une loi bivariable non dégénérée, Pickands (1981) a démontré que cette distribution limite est nécessairement de la forme

$$H^*(x, y) = \exp\left(- (x + y)A\left(\frac{y}{x + y}\right)\right), \quad (2.3.14)$$

où $A : [0, 1] \rightarrow [0, 1]$ est une fonction convexe qui satisfait

$$\max(t, 1 - t) \leq A(t) \leq 1. \quad (2.3.15)$$

De l'équation (2.1.2), on déduit que les copules à valeurs extrêmes s'écrivent :

$$C_A(u, v) = \exp \left(\log uv A \left(\frac{\log u}{\log uv} \right) \right). \quad (2.3.16)$$

Parmi les copules à valeurs extrêmes, on retrouve la famille de Gumbel-Hougaard, dont la fonction de dépendance est :

$$A_\theta(t) = \left((1 - t)^\theta + t^\theta \right)^{\frac{1}{\theta}}. \quad (2.3.17)$$

La copule associée est :

$$C_\theta(u, v) = \exp \left[- \left((-\log u)^\theta + (-\log v)^\theta \right)^{\frac{1}{\theta}} \right], \theta \geq 1. \quad (2.3.18)$$

Alors que pour la copule de Galambos on a :

$$A_\theta(t) = 1 - \left(t^{-\theta} + (1 - t)^{-\theta} \right)^{-1/\theta}. \quad (2.3.19)$$

La copule associée est,

$$C_\theta(u, v) = uv \exp \left[- \left[(-\log u)^{-\theta} + (-\log v)^{-\theta} \right]^{\frac{-1}{\theta}} \right]. \quad (2.3.20)$$

Chapitre 3

Tests d'adéquation

Le choix de la meilleure copule permettant de modéliser le mieux la dépendance entre des variables aléatoires est d'importance capitale. La question qui se pose est quelle est la meilleure structure de dépendance pouvant être adaptée au phénomène étudié. Modéliser efficacement la dépendance dans un phénomène donné requiert quelques problèmes liés à l'inférence. Et une des voies actuelles et prometteuses réside dans le choix d'une copule adéquate pour modéliser la dépendance dans un vecteur. Dans cette section sont développés des tests d'adéquation dont le but consiste à vérifier si la copule sous-jacente à une population appartient à une certaine famille paramétrique $\mathcal{F} = (C_\theta; \theta \in \mathcal{O})$, avec $\mathcal{O} \subset \mathbb{R}^p$ et $p \geq 1$. En d'autres termes, les hypothèses

$$\mathcal{H}_0 : C \in \mathcal{F} \text{ et } \mathcal{H}_1 : C \notin \mathcal{F}$$

seront confrontées.

Toutes les statistiques de test proposées dans ce mémoire sont des fonctions de processus empiriques bâtis à partir des rangs des observations.

Une première solution peu puissante consiste à calculer la différence en valeur absolue entre la copule empirique définie d'une façon similaire à la formule (2.2.20) et la copule paramétrique estimée C_{θ_n} . La dépendance retenue est donc celle qui assure que C_{θ_n} soit la plus proche possible de C_n .

3.1 Approche de Genest, Quessy et Rémillard

Genest et Rivest (1993) ont introduit le processus de Kendall pour l'utiliser comme outil de sélection de copules archimédiennes. Ce dernier est défini

pour tout $t \in [0,1]$ par :

$$\mathcal{K}_{n,\theta}(t) = \sqrt{n}\{K_n(t) - K(\theta, t)\}, \quad (3.1.1)$$

où

$$K(\theta, t) = \mathbb{P}[C_\theta(U_1, \dots, U_d) \leq t], \quad (3.1.2)$$

est la transformation intégrale de probabilité définie pour chaque membre C_θ de la famille \mathcal{F} , où $U_i = F_i(X_i)$ pour $i \in \{1, \dots, d\}$ et (U_1, \dots, U_d) est un vecteur aléatoire de loi C_θ . Genest et Rivest (1993) ont proposé comme estimateur de $K(\theta, t)$ la fonction non paramétrique

$$K_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{V_{i,n} \leq t\}}, \quad (3.1.3)$$

où les $V_{i,n}$ sont des pseudo-observations définies par :

$$V_{i,n} = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_{1j} \leq X_{1i}, \dots, X_{dj} \leq X_{di}\}} = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{R_{1j} \leq R_{1i}, \dots, R_{dj} \leq R_{di}\}}, \quad (3.1.4)$$

avec R_{ij} représentant le rang de X_{ij} . K_n est un estimateur convergent de K . Le processus de Kendall proposé par Genest et Rivest a l'avantage d'être de dimension 1, ce qui facilite le choix de la meilleure copule.

Barbe et *al.* (1996) ont étudié plus à fond le processus de Kendall. Mais l'idée de tester l'adéquation d'une copule en utilisant ce processus a été développée par Wang et Well (2000). Ces derniers établissent 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_θ calculée sous l'hypothèse nulle \mathcal{H}_0 . Pour éliminer les difficultés pratiques issues des problèmes des censures de la densité $k(\theta, \cdot)$ de $K(\theta, \cdot)$, la statistique de Wang et Well présente un seuil $\xi > 0$. Leur statistique de test est toutefois basée sur une version tronquée du processus de Kendall à savoir

$$S_{\xi n} = n \int_{\xi}^1 \{K_n(t) - K_{\theta_n}(t)\}^2 dt, \quad (3.1.5)$$

qui est une fonction continue du processus

$$\mathcal{K}_n(t) = \sqrt{n}\{K_n(t) - K(\theta_n, t)\}. \quad (3.1.6)$$

En imitant l'approche de Barbe et *al.* (1996) ils ont pu identifier la limite de \mathcal{K}_n , et par conséquent celle de $S_{\xi n}$, même en présence de censure. Cependant,

en raison d'un biais observé dans une procédure bootstrap paramétrique qu'ils décrivent pour approximer la variance de S_{ξ_n} , Wang et Well (2000) ont fini par recommander que la sélection d'un modèle à partir d'un ensemble de familles de copules Archimédiennes soit basée sur une comparaison des valeurs non traitées de S_{ξ_n} . Par la suite Genest, Quessy et Rémillard (2006) ont élargi les travaux de Wang et Well en suggérant deux nouveaux tests d'adéquation basés sur une version non tronquée du processus \mathcal{K}_n , soit

$$S_n = \int_0^1 \{\mathcal{K}_n(t)\}^2 dK_{\theta_n}(t) \text{ et } T_n = \sup_{0 \leq t \leq 1} |\mathcal{K}_n(t)|. \quad (3.1.7)$$

La statistique S_n est basée sur la distance de Cramer-von Mises, tandis que T_n est basée sur la distance de Kolmogorov-Smirnov.

De grandes valeurs de l'une ou l'autre de ces statistiques conduisent au rejet de H_0 . Les p-valeurs approximatives peuvent être déduites de leurs distributions limites, qui dépendent du comportement asymptotique du processus \mathcal{K}_n . On déduit de la convergence de \mathcal{K}_n vers \mathcal{K} que :

$$S_n \xrightarrow[n \rightarrow \infty]{loi} S = \int_0^1 \{\mathcal{K}(t)\}^2 dK_{\theta}(t) \text{ et } T_n \xrightarrow[n \rightarrow \infty]{loi} T = \sup_{0 \leq t \leq 1} |\mathcal{K}(t)|. \quad (3.1.8)$$

Contrairement au test de Wang et well (2000), des formules simples sont disponibles pour S_n et T_n en fonction des rangs des observations. Ces tests s'appliquent à une grande variété de copules, pas nécessairement Archimédiennes; il n'y a pas à s'embêter d'un paramètre ξ dont la sélection est arbitraire, et s'appliquent à des familles de dimension $d \geq 2$. Soient $(X_{11}, X_{21}), \dots, (X_{1n}, X_{2n})$ n copies indépendantes du vecteur $X = (X_1, X_2)$ à partir d'une famille de copules continues bivariées $\mathcal{F} = (C_{\theta}; \theta \in \mathcal{O})$ avec des marginales continues, inconnues F_1, F_2 . En d'autres termes, supposons que la fonction de distribution cumulative H de X soit de la forme

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)), \quad (3.1.9)$$

pour une copule $C = C_{\theta} \in \mathcal{F}$.

Dans la suite on ne suppose pas que C_{θ} soit archimédienne.

La proposition 3.1.1 ci-dessous identifie la limite faible \mathcal{K} du processus \mathcal{K}_n . Les conditions sont suffisantes pour que les statistiques S_n et T_n soient des fonctions continues de \mathcal{K}_n , dont les limites sont donc données respectivement par

$$S = \int_0^1 \{\mathcal{K}(t)\}^2 dK_{\theta}(t) \text{ et } T = \sup_{0 \leq t \leq 1} |\mathcal{K}(t)|. \quad (3.1.10)$$

Hypothèse 1. Pour tout $\theta \in \mathcal{O}$, la fonction de distribution $K(\theta, t)$ de $H(X)$ admet une densité $k(\theta, t)$ qui est continue sur $\mathcal{O} \times (0, 1]$.

Hypothèse 2. Pour tout $\theta \in \mathcal{O}$, il existe une version de la distribution conditionnelle du vecteur $X = (X_1, X_2)$ étant donné $H(X) = t$ tel que, pour toute fonction continue à valeurs réelles f sur $[0, 1]^2$, l'application

$$t \longmapsto \mu(t, f) = k(\theta, t) E\{f(X_1, X_2) | H(X) = t\}$$

est continue sur $(0, 1]$ avec $\mu(1, f) = k(\theta, 1)f(1, 1)$.

Comme l'ont noté Barbe, Genest, Ghoudi et Rémillard (1996), ces deux hypothèses sont déjà suffisantes pour impliquer la faible convergence du processus de Kendall, à savoir

$$\mathcal{K}_{n,\theta}(t) = \sqrt{n}\{K_n(t) - K(\theta, t)\}, \quad (3.1.11)$$

vers un processus gaussien centré, continu \mathcal{K}_θ dont la fonction de covariance asymptotique $\Gamma_\theta(s, t)$ est identifié dans le théorème 1 dans Barbe et *al.* (1996) (page 5). Pour garantir la faible convergence du processus \mathcal{K}_n , cependant, il faut aussi restreindre le comportement sur grand échelle de $\Theta_n = \sqrt{n}(\theta_n - \theta)$ de telle sorte que θ_n soit un bon estimateur du paramètre θ .

Hypothèse 3. On a $(\mathcal{K}_{n,\theta}, \Theta_n) \rightarrow (\mathcal{K}_\theta, \Theta)$ dans $D[0, 1] \times \mathcal{R}^m$, et la limite est gaussienne et centrée.

La dernière hypothèse est une condition technique concernant l'existence et la régularité du gradient de $K(\theta, t)$ par rapport à θ , formellement défini par

$$\dot{K}(\theta, t) = \left(\frac{\partial}{\partial \theta_1} K(\theta, t), \dots, \frac{\partial}{\partial \theta_m} K(\theta, t) \right)^t. \quad (3.1.12)$$

Hypothèse 4. Pour tout $\theta \in \mathcal{O}$ donné, $\dot{K}(\theta, t)$ existe et est continu pour tout $t \in [0, 1]$.

De plus,

$$\sup_{\|\theta^* - \theta\| < \epsilon} \sup_{t \in [0, 1]} \left| \dot{K}(\theta^*, t) - \dot{K}(\theta, t) \right| \rightarrow 0 \text{ lorsque } \epsilon \rightarrow 0. \quad (3.1.13)$$

Cette condition est nécessaire en particulier pour montrer que S_n est bien une fonction continue de \mathcal{K}_n .

Proposition 3.1.1 *Sous les Hypothèses 1 à 4, $\mathcal{K}_n \rightarrow \mathcal{K}$ dans $D[0,1]$, où la limite faible \mathcal{K} est un processus gaussien continu centré, ayant la représentation*

$$\mathcal{K}(t) = \mathcal{K}_\theta(t) - \dot{K}^t(\theta, t)\Theta, \quad t \in [0, 1]. \quad (3.1.14)$$

Preuve:

Écrivons $\mathcal{K}_n(t) = \mathcal{K}_{n,\theta}(t) - L_n(t)$, avec $L_n(t) = \sqrt{n}\{K(\theta_n, t) - K(\theta, t)\}$ pour tout $t \in [0, 1]$. Comme mentionné déjà, il découle des hypothèses 1 et 2 que $\mathcal{K}_{n,\theta} \rightarrow \mathcal{K}_\theta$ dans $D[0,1]$. Où \mathcal{K}_θ est un processus gaussien centré, continu. Montrons que sous les hypothèses 3 et 4, le processus $L_n(t)$ est telle que

$$\sup_{t \in [0,1]} |L_n(t) - \dot{K}^t(\theta, t)\Theta_n| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Pour le voir, soit $\lambda > 0$ arbitraire. Par l'hypothèse 3, la suite (Θ_n) est serrée comme elle converge en loi vers Θ . Alors pour tout $\gamma > 0$, il existe $M = M_\gamma \in \mathbb{R}^+$ et N_0 tel que $\mathbb{P}\{\|\Theta_n\| > M\} < \gamma$ pour tout $n \geq N_0$. Pour un tel n , on a :

$$\begin{aligned} & \mathbb{P}\left\{ \sup_{t \in [0,1]} |L_n(t) - \dot{K}^t(\theta, t)\Theta_n| > \lambda \right\} \\ & \leq \mathbb{P}\left\{ \sup_{t \in [0,1]} |L_n(t) - \dot{K}^t(\theta, t)\Theta_n| > \lambda, \|\Theta_n\| \leq M \right\} + \mathbb{P}\{\|\Theta_n\| > M\} \\ & < \mathbb{P}\left\{ \sup_{t \in [0,1]} |L_n(t) - \dot{K}^t(\theta, t)\Theta_n| > \lambda, \|\Theta_n\| \leq M \right\} + \gamma. \end{aligned}$$

Ensuite le théorème de la valeur moyenne implique que pour toute réalisation de Θ_n , il existe $\theta - n^*$ avec $|\theta_n^* - \theta| \leq \|\Theta_n\|/\sqrt{n}$ tel que $L_n(t) = \dot{K}^t(\theta_n^*, t)\Theta_n$. Alors, sous l'hypothèse 4 on obtient :

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}\left\{ \sup_{t \in [0,1]} |L_n(t) - \dot{K}^t(\theta, t)\Theta_n| > \lambda, \|\Theta_n\| \leq M \right\} \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P}\left\{ \|\Theta_n\| \sup_{t \in [0,1]} \|\dot{K}(\theta_n^*, t) - \dot{K}(\theta, t)\| > \lambda, \|\Theta_n\| \leq M \right\} \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P}\left\{ \sup_{\|\theta^* - \theta\| \leq \frac{M}{\sqrt{n}}} \sup_{t \in [0,1]} \|\dot{K}(\theta^*, t) - \dot{K}(\theta, t)\| > \frac{\lambda}{M} \right\} = 0. \end{aligned}$$

Comme γ peut être choisi arbitrairement petit, le résultat s'en suit. ■

Genest et al. (2006) ont démontré que :

$$S_n = \frac{n}{3} + n \sum_{j=1}^{n-1} K_n^2\left(\frac{j}{n}\right) \left(K\left(\theta_n, \frac{j+1}{n}\right) - K\left(\theta_n, \frac{j}{n}\right) \right) \quad (3.1.15)$$

$$- n \sum_{j=1}^{n-1} K_n\left(\frac{j}{n}\right) \left(K^2\left(\theta_n, \frac{j+1}{n}\right) - K^2\left(\theta_n, \frac{j}{n}\right) \right) \quad (3.1.16)$$

et

$$T_n = \sqrt{n} \max_{i=0,1; 0 \leq j \leq n-1} \left\{ \left| K_n\left(\frac{j}{n}\right) K\left(\theta_n, \frac{j+1}{n}\right) \right| \right\}. \quad (3.1.17)$$

Une procédure de sélection basée uniquement sur la comparaison des valeurs non traitées de S_{ξ_n} peut parfois conduire à des modèles qui devraient être rejetés sur la base de leur p-valeurs. Bien sûr, la comparaison des valeurs non traitées de S_n et T_n pourrait être tout aussi trompeuse, d'où l'importance de la procédure de bootstrap paramétrique valide proposée ici.

Les procédures de test formelles basées sur ces statistiques consisteraient à rejeter H_0 lorsque la valeur observée de S_n ou T_n est supérieure aux quantiles d'ordre $1 - \alpha$ de sa distribution sous l'hypothèse nulle. Cependant, cette distribution dépend du paramètre inconnu θ , même dans la limite. Bien entendu, la statistique S_{en} de Wang et Wells (2000) souffre de la même limitation, sans parler de sa dépendance à l'arbitraire point de coupure ξ .

Pour contourner ces problèmes méthodologiques et obtenir une p-valeur approximative pour S_n ou T_n , on peut faire appel à une méthode bootstrap paramétrique basée sur C_{θ_n} .

À fin de calculer les p-valeurs pour toutes statistiques basées sur le processus empirique \mathcal{K}_n , il faut générer un grand nombre N échantillons indépendants, de taille n de C_{θ_n} et calculer les valeurs correspondantes de la statistique sélectionnée, comme S_n ou T_n .

La procédure fonctionnerait comme suit :

1. estimer le paramètre θ de la copule par un estimateur consistant θ_n ,
2. à la deuxième étape, nous devons générer N échantillons de taille n de loi C_θ et pour chacun de ces échantillons, estimer le paramètre θ par la même méthode et déterminer la valeur de la statistique du test.
3. si $S_{1,N}^* \leq \dots \leq S_{N,N}^*$ désigne les valeurs ordonnées croissantes des statistiques de test calculer à l'étape 2, un estimateur de la valeur critique du test à un seuil de risque α basé sur S_n est donnée par

$$S_{[(1-\alpha)N],N}^*, \quad (3.1.18)$$

et

$$\frac{j}{N} \#\{j : S_j^* \geq S_n\} \quad (3.1.19)$$

donne une estimation de la p-valeur associée à la valeur observée de la statistique S_n .

L'hypothèse H_0 est retenue si et seulement si l'estimation de la p-valeur associée à la valeur observée de la statistique S_n est supérieure au seuil de risque fixé au niveau α (voir Genest et Rémillard (2005)).

3.2 Approche par la copule empirique

Une étude de simulation a montré la bonne performance des procédures d'adéquation basées sur \mathcal{K}_n sous plusieurs modèles de contre-hypothèses exprimés à l'aide de copules. Cependant, malgré la puissance élevée des méthodes d'adéquation basées sur S_n et T_n sous plusieurs variétés de contre-hypothèses, quelques critiques concernant les tests étudiés à la section 3.1 peuvent être soulevées :

- (a) puisque la transformation intégrale de probabilité est une projection unidimensionnelle d'une fonction de dépendance à d dimensions, les tests développés à la section 3.1 ne seront pas toujours convergents ;
- (b) il se pourrait que les hypothèses qui assurent la convergence du processus d'adéquation ne soient pas satisfaites pour certaines classes importantes de copules ;
- (c) pour certaines familles de modèles, la transformation intégrale de probabilité n'admet pas de forme explicite.

Au vu du point (a), on s'attend à ce que l'efficacité de S_n et T_n décroisse, à mesure que la dimension du vecteur augmente. On retrouve la copule normale comme cas particulier de la remarque (c).

Pour pallier ces limitations, la section suivante procure les outils théoriques permettant de développer des tests d'ajustement basés sur une estimation directe et complètement non paramétrique de la copule sous-jacente à une population. Des tests pouvant rivaliser avec S_n et T_n , ainsi qu'une procédure d'adéquation applicable à la copule normale pourront alors être envisagés. Il est proposé ici d'appuyer une procédure de test sur une estimation non paramétrique C_n de C . Comme mentionné plus haut, les pseudo-observations $\hat{U}_1, \dots, \hat{U}_n$ constituent les statistiques maximales invariantes sur lesquelles tester H_0 . Spécifiquement, si $(X_{11}, X_{12}), \dots, (X_{n1}, X_{n2})$ est un échantillon aléatoire d'une population bivariable, la copule empirique est définie par

$$C_n(u) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^2 \mathbf{1}_{\{\hat{F}_j(X_{ij}) \leq u_j\}}, \quad (3.2.1)$$

où

$$\hat{F}_j(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_{ij} \leq x\}}, 1 \leq j \leq 2 \quad (3.2.2)$$

sont les fonctions de répartition empiriques marginales. Bien que ce ne soit ni une copule, ni exactement la même que celle définie à l'origine par Deheuvels (1979). Ganssler et Stute (1987), Fermanian et *al.* (2004) et Tsukahara (2005)

donnent diverses conditions dans lesquelles C_n est un estimateur consistant de la vraie copule sous-jacente C . Étant donné qu'il est entièrement non paramétrique, C_n est sans doute le plus objectif référence pour tester $H_0 : C \in \mathcal{F}$. Par conséquent, un test d'adéquation naturel consiste à comparer une distance entre C_n et un estimateur C_{θ_n} de C obtenue sous H_0 . L'estimateur C_n est entièrement fondé sur les rangs des observations, puisque $n\hat{F}_j(X_{ij}) = R_{ij}$ quels que soient $i \in \{1, \dots, n\}$ et $j \in \{1, 2\}$.

Les tests d'adéquation basés sur le processus empirique

$$\mathcal{C}_n = \sqrt{n}\{C_n - C_{\theta_n}\}, \quad (3.2.3)$$

où C_n est la copule empirique et C_{θ_n} est un estimateur de la copule C sous l'hypothèse H_0 avec θ_n un estimateur de θ basé sur les rangs sont brièvement considérés par Fermanian (2005), qui commente qu'ils "semblent être peu pratique, sauf par bootstrap".

Genest et Rémillard examinent la mise en œuvre du problème en détail. En particulier, ils considèrent les versions basées sur le classement des familles statistiques de type Cramer-Von Mises et Kolmogorov-Smirnov, à savoir

$$S_n^* = \int_0^1 \{\mathcal{C}_n(u)\}^2 dC_n(u) \text{ et } T_n^* = \sup_{u \in [0,1]^d} |\mathcal{C}_n(u)|. \quad (3.2.4)$$

De grandes valeurs de ces statistiques conduisent au rejet de H_0 .

Les p-valeurs approximatives peuvent être déduites de leurs distributions limites, qui dépendent du comportement asymptotique du processus \mathcal{C}_n . Genest et Rémillard montrent que les tests basés sur S_n^* et T_n^* sont cohérent ; c'est-à-dire que si $C \notin \mathcal{F}$, alors H_0 est rejetée avec une probabilité 1 quand $n \rightarrow \infty$.

Soit maintenant une famille paramétrique $\{C_\theta; \theta \in \mathcal{O}\}$. Afin d'obtenir la loi asymptotique d'un processus d'adéquation basé sur C_n , les trois hypothèses suivantes seront nécessaires.

Hypothèse 1. Pour tout $\theta \in \mathcal{O}$, les dérivées partielles $\frac{\partial C_\theta(u)}{\partial u_j}$, pour $j \in \{1, 2\}$, existent et sont continues sur $[0, 1]^2$.

Soit θ_n , un estimateur du paramètre inconnu θ .

Hypothèse 2. La suite $\Theta_n = \sqrt{n}(\theta_n - \theta)$ converge en loi vers Θ .

Hypothèse 3. Pour tout $\theta \in \mathcal{O}$,

$$\sup_{\|\theta^* - \theta\| < \epsilon} \sup_{u \in [0,1]^2} |\dot{C}_{\theta^*}(u) - \dot{C}_\theta(u)| \rightarrow 0, \quad (3.2.5)$$

quand $\epsilon \rightarrow 0$, où

$$\dot{C}_\theta = \left(\frac{\partial}{\partial \theta_1} C_\theta, \dots, \frac{\partial}{\partial \theta_m} C_\theta \right)^t, \quad (3.2.6)$$

est le gradient de C_θ par rapport à θ .

L'hypothèse 1 assure la convergence faible du processus empirique

$$\mathcal{C}_{n,\theta} = \sqrt{n}\{C_n - C_\theta\}, \quad (3.2.7)$$

vers une limite gaussienne \mathcal{C}_θ . Ce résultat dû à Gaenssler et Stute (1987), a récemment été reconsidéré par Ghoudi et Rémillard (2004) et par Fermanian et al. (2004).

Pour tester l'adéquation à une famille paramétrique, on considère le processus d'adéquation

$$\mathcal{C}_n = \sqrt{n}\{C_n - C_{\theta_n}\}.$$

Les hypothèses 2 et 3 permettent de caractériser la limite de ce processus.

Proposition 3.2.1 *Supposons que les hypothèses 1-3 soient vérifiées pour une famille de copules $\mathcal{F} = \{C_\theta, \theta \in \mathcal{O}\}$. Alors le processus empirique*

$$\mathcal{C}_n = \sqrt{n}\{C_n - C_{\theta_n}\}, \quad (3.2.8)$$

converge dans l'espace $\mathcal{D}([0, 1]^2)$ vers un processus gaussien et centré

$$\mathcal{C} = \mathcal{C}_\theta - \dot{C}_\theta^t \Theta. \quad (3.2.9)$$

Preuve:

Considérons la décomposition $\mathcal{C}_n = \mathcal{C}_{n,\theta} - L_{n,\theta}$, avec $L_{n,\theta} = \sqrt{n}(C_{\theta_n} - C_\theta)$ et $\mathcal{C}_{n,\theta} = \sqrt{n}(C_n - C_\theta)$.

Dans un premier temps, le processus $\mathcal{C}_{n,\theta}$ converge vers \mathcal{C}_θ , puisque l'hypothèse 1 est vérifiée. Dans un second temps, il sera démontré que $L_{n,\theta}$ peut devenir arbitrairement près de $\dot{C}_\theta^t \Theta_n$, pour n suffisamment grand. De l'hypothèse 2, Θ_n converge en loi vers Θ , ce qui implique que la suite (Θ_n) est tendue. Ainsi soit $\gamma > 0$ arbitraire, alors pour tout $\delta > 0$ donné, il existe un $M = M_\delta$ tel que $\mathbb{P}\{\|\Theta_n\| > M\} < \delta$. Donc,

$$\begin{aligned} & \mathbb{P}\left\{ \sup_{u \in [0,1]^d} |L_{n,\theta}(u) - \dot{C}_\theta^t(u) \Theta_n| > \gamma \right\} \\ & \leq \mathbb{P}\left\{ \sup_{u \in [0,1]^d} |L_{n,\theta}(u) - \dot{C}_\theta^t(u) \Theta_n| > \gamma, \|\Theta_n\| \leq M \right\} + \mathbb{P}\left\{ \|\Theta_n\| > M \right\} \\ & < \mathbb{P}\left\{ \sup_{u \in [0,1]^d} |L_{n,\theta}(u) - \dot{C}_\theta^t(u) \Theta_n| > \gamma, \|\Theta_n\| \leq M \right\} + \delta. \end{aligned}$$

Nous déduisons alors du théorème de la valeur moyenne que $L_{n,\theta} = \dot{C}_{\theta_n^*}^t \Theta_n$, où $\theta_n^* = \theta + \epsilon n^{-1/2} \Theta_n$, $0 < \epsilon < 1$, en employant l'hypothèse 3,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{u \in [0,1]^d} |L_{n,\theta}(u) - \dot{C}_{\theta_n^*}^t(u) \Theta_n| > \gamma, \|\Theta_n\| \leq M \right\} \\ &= \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \|\Theta_n\| \sup_{u \in [0,1]^d} |\dot{C}_{\theta_n^*}^t(u) - \dot{C}_\theta(u)| > \gamma, \|\Theta_n\| \leq M \right\} \\ &\leq \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{\|\theta_n^* - \theta\| \leq n^{-1/2} M} \sup_{u \in [0,1]^d} |\dot{C}_{\theta_n^*}^t(u) - \dot{C}_\theta(u)| > \frac{\gamma}{M} \right\} \\ &= 0. \end{aligned}$$

Par conséquent, la limite de $L_{n,\theta}$ est la même que celle de $\dot{C}_\theta^t \Theta$, à savoir $\dot{C}_\theta^t \Theta$, ce qui complète la démonstration. ■

En pratique, les distributions limites de S_n^* et T_n^* dépendent de la famille de copules sous l'hypothèse nulle composite et de la valeur inconnue du paramètre θ en particulier. En conséquence, les distributions asymptotiques des statistiques de test et les p-valeurs approximatives ne peuvent être obtenues que par des méthodes Monte Carlo adaptées.

3.3 Approche de Malevergne et Sornette

Dans cette section nous testons l'hypothèse nulle selon laquelle la dépendance entre les variables peut être décrite par la copule gaussienne.

3.3.1 La copule gaussienne

La copule gaussienne est la copule dérivée de la distribution gaussienne multivariée. Soit Φ la distribution normale standard et $\Phi_{\rho,n}$ la distribution gaussienne à n dimensions avec matrice de corrélation ρ . Alors, la n -copule gaussienne avec matrice de corrélation ρ est

$$C_\rho(u_1, \dots, u_n) = \Phi_{\rho,n}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)). \quad (3.3.1)$$

Comme nous l'avons dit précédemment, la copule gaussienne n'a pas de dépendance de queue :

$$\lim_{u \rightarrow 1} \frac{\bar{C}(u, u)}{1 - u} = 0, \forall \rho \in [-1, 1]. \quad (3.3.2)$$

Par contre, la structure de dépendance décrite par la copule gaussienne est très différente de la structure de dépendance de la copule indépendante, sauf pour $\rho = 0$.

La copule gaussienne est entièrement déterminée par la connaissance de la matrice de corrélation.

3.3.2 Test de l'hypothèse de la copule gaussienne

Malgré le fait que la normalité multivariée est plus souvent une exception à la norme dans beaucoup de phénomènes naturels et financiers, la structure de dépendance normale demeure un choix populaire lors de la modélisation multivariée. Dans le cas de la normalité multivariée classique, plusieurs tests existent. Pour un article de synthèse sur le sujet, on peut consulter [13].

Dans un contexte semi-paramétrique où l'on désire établir que la distribution des variables a une structure de dépendance normale, mais pas nécessairement des marges normales, nous noterons H_0 l'hypothèse nulle selon laquelle la dépendance entre les variables aléatoires peut être décrite par la copule gaussienne. Une telle procédure est proposée par Malevergne et Sornette (2003).

Statistiques de test

Nous dérivons maintenant les statistiques de test qui nous permettront de rejeter ou non notre hypothèse nulle H_0 et dénoncer la proposition suivante :

Proposition 3.3.1 *Sous l'hypothèse que la structure de dépendance sous-jacente à la population est la copule normale de matrice de corrélation ρ , avec fonction de distribution F et que les marges sont F_1, \dots, F_n , la variable*

$$z^2 = \sum_{i,j=1}^n \Phi^{-1}(F_i(x_i))(\rho^{-1})_{ij}\Phi^{-1}(F_j(x_j)) \quad (3.3.3)$$

où la matrice ρ est

$$\rho_{ij} = Cov\left[\Phi^{-1}(F_i(x_i)), \Phi^{-1}(F_j(x_j))\right], \quad (3.3.4)$$

suit une loi du Khi-deux à n degrés de liberté.

Preuve:

considérons d'abord un vecteur aléatoire à n dimensions $\mathbf{x} = (x_1, \dots, x_n)$. Notons F sa fonction de distribution et F_i la distribution marginale de chaque x_i . Nous supposons maintenant que la fonction de distribution F satisfait H_0 , de sorte que F a une copule gaussienne, avec matrice de corrélation ρ , alors

que les F_i peuvent être n'importe quelle fonction de distribution. D'après le théorème de Sklar, la distribution F peut être représentée par :

$$F(x_1, \dots, x_n) = \Phi_{\rho, n}(\Phi^{-1}(F_1(x_1)), \dots, \Phi^{-1}(F_n(x_n))). \quad (3.3.5)$$

Transformons maintenant les variables x_i en variables aléatoires normales y_i :

$$y_i = \Phi^{-1}(F_i(x_i)). \quad (3.3.6)$$

L'application $\Phi^{-1}(F_i(\cdot))$ étant croissante, le théorème 2.1.2 permet de conclure que la copule des variables y_i est identique à celle des variables x_i . Par conséquent, les variables y_i ont des distributions marginales normales et une copule gaussienne avec une matrice de corrélation ρ .

$$G(\mathbf{y}) = \Phi_{\rho, n}(\Phi^{-1}(F_1(x_1)), \dots, \Phi^{-1}(F_n(x_n))) \quad (3.3.7)$$

$$= \Phi_{\rho, n}(y_1, \dots, y_n), \quad (3.3.8)$$

et \mathbf{y} est un vecteur aléatoire gaussien.

Considérons maintenant la variable aléatoire

$$z^2 = \mathbf{y}^t \rho^{-1} \mathbf{y} = \sum_{i,j=1}^n y_i (\rho^{-1})_{ij} y_j, \quad (3.3.9)$$

où \mathbf{y}^t désigne l'opérateur transposé de \mathbf{y} . Il est bien connu que la variable z^2 suit une distribution du Khi-deux à n degrés de liberté. En effet, puisque \mathbf{y} est un vecteur aléatoire gaussien, de matrice de covariance ρ , il s'ensuit que les composantes du vecteur

$$\tilde{\mathbf{y}} = \mathbf{A} \mathbf{y},$$

sont des variables aléatoires normales indépendantes. Ici \mathbf{A} désigne la racine carrée de la matrice ρ^{-1} , obtenue par la décomposition de Choleski, de sorte que $\mathbf{A}^t \mathbf{A} = \rho^{-1}$. Ainsi la somme $\tilde{\mathbf{y}}^t \tilde{\mathbf{y}} = z^2$ est la somme des carrés de n variables aléatoires normales indépendantes, donc z^2 suit une distribution du Khi-deux à n degrés de liberté. ■

Procédure de test :

nous considérons deux séries financières ($d = 2$) de taille T :

$\{x_1(1), \dots, x_1(t), \dots, x_1(T)\}$ et $\{x_2(1), \dots, x_2(t), \dots, x_2(T)\}$. Nous supposons que les vecteurs $\mathbf{x}(t) = (x_1(t), x_2(t))$, $t \in \{1, \dots, T\}$ sont i.i.d. avec la distribution F , ce qui implique que les variables $x_1(t)$ (respectivement $x_2(t)$), $t \in \{1, \dots, T\}$, sont aussi i.i.d. avec la distributions F_1 (respectivement F_2).

Dans le cas le plus général où la copule de \mathbf{x} est normale, mais dont les lois marginales sont inconnues, Malevergne et Sornette (2003) ont proposé de remplacer chaque F_i par son homologue empirique \hat{F}_i donné par

$$\hat{F}_i(x_i) = \frac{1}{T} \sum_{k=1}^T \mathbf{1}_{\{x_i(k) \leq x_i\}}, \quad (3.3.10)$$

ici $\mathbf{1}_{\{\cdot\}}$ est la fonction indicatrice, qui vaut 1 si son argument est vrai et zéro sinon. Nous utilisons ces distributions empiriques pour obtenir les variables gaussiennes \hat{y}_i telle que

$$\hat{y}_i(k) = \Phi^{-1}(\hat{F}_i(x_i(k))), k \in \{1, \dots, T\}. \quad (3.3.11)$$

La matrice de covariance d'échantillon ρ est estimée par l'expression

$$\hat{\rho} = \frac{1}{T} \sum_{i=1}^T \hat{\mathbf{y}}(i) \cdot \hat{\mathbf{y}}(i)^t, \quad (3.3.12)$$

qui permet de calculer la variable

$$\hat{z}^2(k) = \sum_{i,j=1}^2 \hat{y}_i(k) (\hat{\rho}^{-1})_{ij} \hat{y}_j(k), \quad (3.3.13)$$

tel que défini précédemment pour $k \in \{1, \dots, T\}$, doit être distribuée selon une distribution du Khi-deux si l'hypothèse de la copule gaussienne est correcte.

La manière habituelle de comparer une distribution empirique avec une distribution théorique est de mesurer la distance entre ces deux distributions et d'effectuer le test de Kolmogorov ou le test d'Anderson-Darling (voir Anderson et Darling (1952)).

La distance de Kolmogorov est la distance locale maximale le long du quantile qui se produit le plus souvent dans l'essentiel de la distribution, tandis que la distance d'Anderson-Darling met l'accent sur les queues des deux distributions par une normalisation appropriée.

$$\text{Kolmogorov} : D_{KS} = \max_z \left| F_{z^2}(z^2) - F_{\chi^2}(z^2) \right|, \quad (3.3.14)$$

$$\text{Anderson - Darling} : D_{AD} = \max_z \frac{\left| F_{z^2}(z^2) - F_{\chi^2}(z^2) \right|}{\sqrt{F_{\chi^2}(z^2)[1 - F_{\chi^2}(z^2)]}}. \quad (3.3.15)$$

Nous présentons nos tests statistiques avec la distance de Kolmogorov. Cependant, nous ne pouvons pas utiliser directement les résultats de ces

tests statistiques standards, car la matrice de covariance n'est pas connue et doit être estimée à partir des données empiriques. Comme remède, une méthode bootstrap est proposée (voir Efron et Tibshirani), dont la précision est prouvée par Chen et Lo (1997) comme étant au moins aussi bonne que celle donnée par les méthodes asymptotiques utilisées pour dériver les distributions théoriques. Pour le présent travail, Malevergne et Sornette (2003) déterminent que la génération de 10000 séries temporelles synthétiques est suffisante pour obtenir une bonne approximation de la répartition des distances décrites ci-dessus.

Pour résumer, notre procédure de test est la suivante :

1. étant donné la série temporelle originale $\mathbf{x}(t)$, $t \in \{1, \dots, T\}$, nous générons les variables gaussiennes $\hat{\mathbf{y}}(t)$, $t \in \{1, \dots, T\}$,
2. nous estimons ensuite la matrice de covariance $\hat{\rho}$, des variables aléatoires gaussiennes $\hat{\mathbf{y}}$, ce qui nous permet de calculer les variables \hat{z}^2 et alors mesurer la distance de sa distribution estimé au χ^2 -distribution,
3. étant donné cette matrice de covariance $\hat{\rho}$, nous générons numériquement une série temporelle de T vecteurs aléatoires gaussiens de même matrice de covariance $\hat{\rho}$,
4. pour la série temporelle de vecteurs gaussiens générés synthétiquement avec la matrice de covariance $\hat{\rho}$, nous estimons sa matrice de covariance d'échantillon $\tilde{\rho}$,
5. à chacun des T vecteurs de la série temporelle gaussienne synthétique, on associe la réalisation correspondante de la variable aléatoire z^2 , appelée $\tilde{z}^2(t)$,
6. on peut alors construire la distribution empirique de la variable \tilde{z}^2 et mesurer la distance entre cette distribution et la distribution du Khi-deux, avec la distance D_{KS} de Kolmogorov définie ci-dessus. Puis on compare D_{KS} à la valeur critique $D_\alpha(n)$ fournie par les tables de Kolmogorov-Smirnov. Si $D_{KS} > D_\alpha(n)$, on rejette l'hypothèse nulle que z^2 suit une distribution du χ^2 avec un risque α de se tromper. Sinon, l'hypothèse est retenue,
7. En répétant 10000 fois les étapes 3 à 6, on obtient une estimation précise de la distribution cumulative des distances entre la distribution des variables gaussiennes synthétiques et la distribution théorique du χ^2 . Cette

distribution cumulative représente la statistique de test, qui nous permettra de rejeter ou non l'hypothèse nulle H_0 à un niveau de signification donné.

3.4 Approche de Breymann, Dias et Embrechts

Soit $X = (X_1, X_2, \dots, X_d)$ un vecteur aléatoire avec une fonction de distribution absolument continue F . Soit $F_i(x_i) = \mathbb{P}(X_i \leq x_i)$ la fonction de distribution des marges univariées X_i , pour $i \in \{1, \dots, d\}$.

Nous définissons pour $2 \leq i \leq d$ les fonctions de répartition marginales conditionnelles

$$F_{i/1, \dots, i-1}(x_i/x_1, \dots, x_{i-1}) = \mathbb{P}(X_i \leq x_i / X_1 = x_1, \dots, X_{i-1} = x_{i-1}). \quad (3.4.1)$$

La transformation de Roseblatt des variables aléatoires X_1, \dots, X_d est définie par :

$$Z_1 = F_1(X_1), \quad (3.4.2)$$

$$Z_i = F_{i/1, \dots, i-1}(X_i/X_1, \dots, X_{i-1}), 2 \leq i \leq d. \quad (3.4.3)$$

L'utilité de cette transformation tient du 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 et Embrechts ont exploité cette transformation en remarquant que la loi de

$$Y_G = \sum_{i=1}^d \{\Phi^{-1}(Z_i)\}^2 \quad (3.4.4)$$

est une loi du Khi-deux à d degrés de liberté. En effet si $U \sim \mathcal{U}(0,1)$ et Φ est la fonction de répartition d'une variable de loi $\mathcal{N}(0,1)$, alors :

$\Phi^{-1}(U) \sim \mathcal{N}(0,1)$. Donc $\Phi^{-1}(Z_i)$, $i=1, \dots, d$ sont i.i.d suivant la $\mathcal{N}(0,1)$. Par conséquent $Y_G = \sum_{i=1}^d \{\Phi^{-1}(Z_i)\}^2$ suit une distribution du Khi-deux à d degrés de liberté.

Sous l'hypothèse \mathcal{H}_0 , $W_G = F_{\chi^2_d}(Y_G)$ est de loi $\mathcal{U}(0,1)$, i.e.

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

$$\hat{F}_G(w) = \frac{1}{n+1} \sum_{j=1}^n \mathbf{1}_{\{W_{G_j} \leq w\}}, w = \frac{1}{n+1}, \dots, \frac{n}{n+1}. \quad (3.4.5)$$

Breymann, Dias et Embrechts proposèrent d'utiliser la statistique d'Anderson-Darling appliquée à F_G , i.e.

$$\hat{\mathcal{T}}^{AD} = \frac{n}{n+1} \sum_{j=1}^n \frac{\{\hat{F}_G(\frac{j}{n+1} - \frac{j}{n+1})\}^2}{\frac{j}{n+1}(1 - \frac{j}{n+1})}. \quad (3.4.6)$$

En s'inspirant de ces travaux, Chen et 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

$$\hat{\mathcal{T}}^{L2KDE} = \frac{n}{q} \sum_{i=1}^q \{\hat{f}_G(\frac{i-1/2}{q}) - 1\}^2, \quad (3.4.7)$$

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

Chapitre 4

Mise en œuvre pratique

Cette section aborde l'aspect pratique de notre travail. On y présentera les données étudiées. On mettra ensuite en évidence un certain nombre de dépendances entre l'altitude maximale et la température moyenne, et on sélectionnera les copules les représentant le mieux .

4.1 Analyse exploratoire des données

Nos données portent sur l'altitude maximale et la température moyenne (obtenue à partir de la température minimale et maximale) de 100 villes de France (avec la Belgique et la suisse). La période de l'étude s'étend de 1981 à 2010. Nous avons considéré un échantillon de 100 couples (X_i, Y_j) avec X_i représentant l'altitude maximale et Y_j représentant la température moyenne. Les statistiques sont présentées dans le tableau suivant :

Statistique	Altitude max. (en m)	Température (en °C)
Minimum	4.0	3.80
1st Quantile	121.8	10.70
Mediane	214.0	11.45
Moyenne	394.6	11.77
3rd Quantile	437.8	12.82
Maximum	3816.0	16.45

TABLE 4.1 – Les statistiques descriptives de l'altitude maximale et la température moyenne.

4.2 Modélisation de la dépendance

On s'intéresse principalement à l'étude de la structure de dépendance entre l'altitude maximale et la température moyenne sur la période de 1981 à 2010. Une loi de distribution marginale a pu être ajustée pour l'altitude maximale et pour la température moyenne. La loi gamma présente le meilleur ajustement pour les variables à étudier, et ce après avoir fait plusieurs tests d'ajustement. La distribution observée de l'altitude maximale est en adéquation avec la distribution théorique d'une répartition exponentielle et celle de la température moyenne est en adéquation avec la distribution théorique d'une répartition gamma.

On donne ci-dessous un exemple graphique de l'ajustement de la distribution de l'altitude maximale et de la température moyenne à une distribution exponentielle et gamma respectivement en utilisant deux approches graphiques, sous le progiciel R, dites respectivement méthode de l'histogramme et la méthode de l'approximation de la densité.

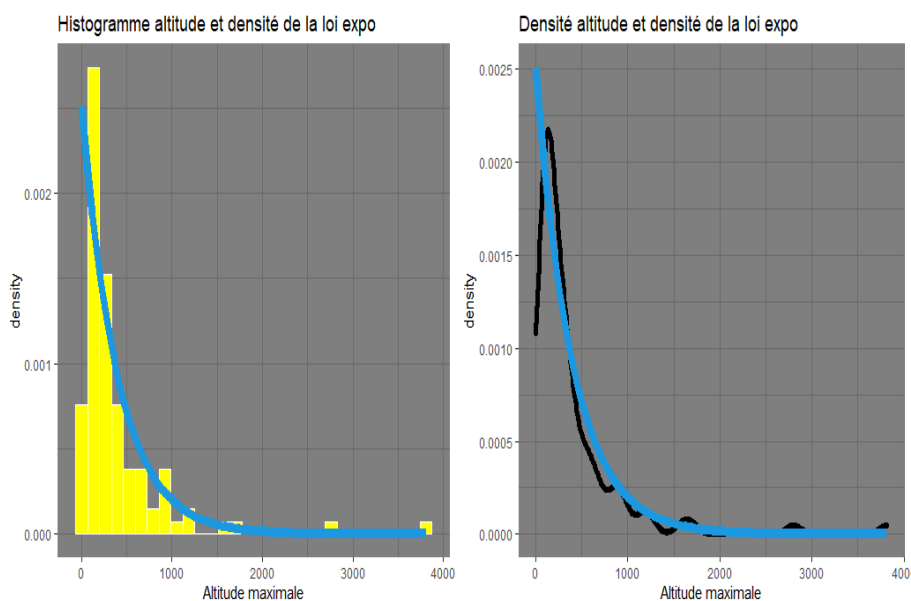


Figure 4.1 – Ajustement de la distribution de l'altitude à une distribution exponentielle.

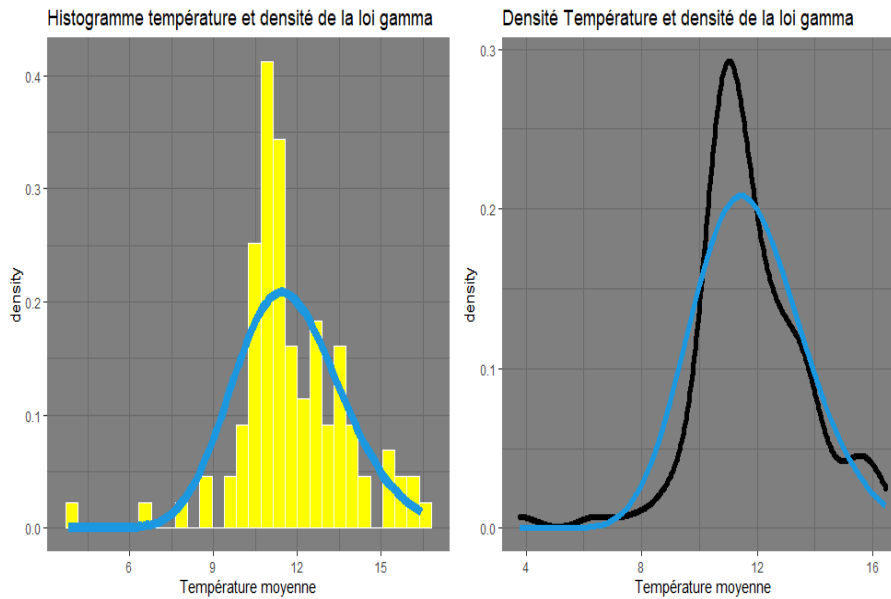


Figure 4.2 – Ajustement de la distribution de la température à une distribution gamma.

Le choix de la copule représente la première difficulté dans la mise en pratique de la modélisation de la dépendance. La figure 4.3 suivante permet d’appréhender la forme de la dépendance qui existe entre l’altitude maximale et la température moyenne.

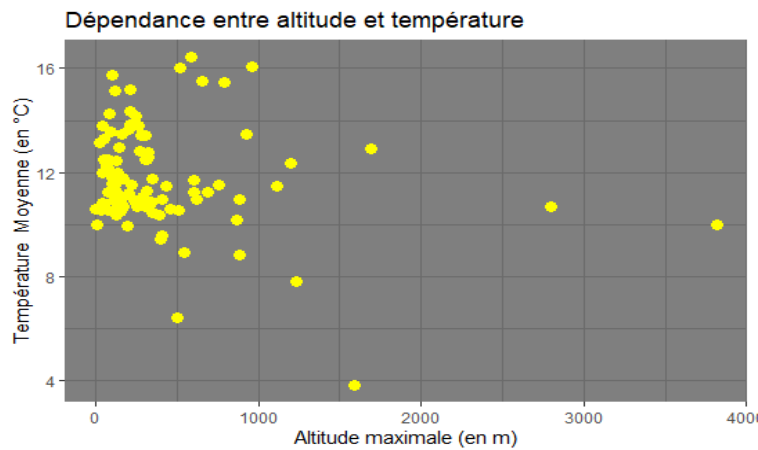


Figure 4.3 – Visualisation de la dépendance entre altitude maximale et température moyenne.

On note une dépendance négative entre l’altitude maximale et la température

moyenne.

La courbe d'ajustement lowess de la figure 4.4 s'ajuste mieux aux nuages de points. Il y a une dépendance négative entre l'altitude maximale et la température moyenne, mais elle n'est parfaitement linéaire qu'à partir d'une certaine altitude élevée comme le montre notre courbe lowess de la figure 4.4.

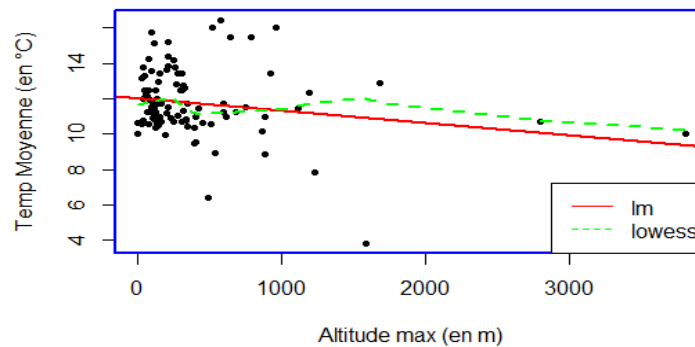


Figure 4.4 – Visualisation de la dépendance entre altitude maximale et température moyenne.

En utilisant la statistique de type Cramer-von Mises basée sur le processus empirique décrite par l'approche 3.2, nous obtenons une bonne adéquation des copules Frank, Gumbel et gaussienne aux données. Nous obtenons pour toutes ces copules une p-valeur supérieure au seuil de risque fixé à 0.05. Une comparaison graphique entre les copules paramétriques Frank et Gumbel et la copule empirique donne la figure 4.5 suivante :

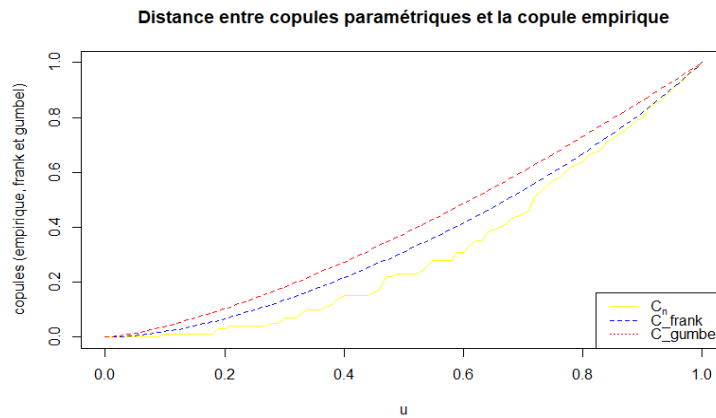


Figure 4.5 – Distance entre les copules Frank et Gumbel et la copule empirique.

La dépendance retenue est donc la copule de Frank.

La copule Frank a été retenue pour appréhender la nature de la dépendance entre l'altitude maximale et la température moyenne.

La figure 4.6 suivante présente la copule empirique modélisant la dépendance entre l'altitude maximale et la température moyenne, et la copule paramétrique permettant l'ajustement de la copule empirique à celle de Frank de paramètre $\alpha = -0.86798$ estimé par la méthode du maximum de vraisemblance.



Figure 4.6 – Copule paramétrique et copule empirique pour l'altitude maximale et la température moyenne.

La figure 4.7 suivante illustre la densité de la copule modélisant la dépendance entre l'altitude maximale et la température moyenne.

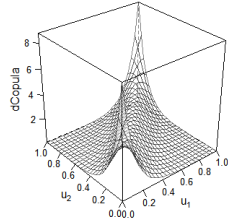


Figure 4.7 – Densité de la copule altitude maximale-température moyenne.

Conclusion

Dans ce travail, des tests d'adéquation pour des modèles de copules sont développés. Les statistiques de tests proposées sont des fonctionnelles de types Cramer-von Mises et Kolmogorov-Smirnov calculées sur des processus empiriques dont la convergence a été obtenue sous l'hypothèse nulle d'une appartenance à une famille donnée. Les hypothèses nécessaires à l'établissement de ce résultat asymptotique ont été vérifiées pour un grand nombre de familles de copules. La technique du bootstrap paramétrique a été employée pour calculer des seuils asymptotiquement exacts de ces statistiques.

En perspective, nous pouvons élargir ce travail sur les tests d'adéquation des copules de dimension $d > 2$.

Bibliographie

- [1] Bahraoui T. (2012). *Fonctions de c-puissance et application à l'inférence de copules*. Mémoire Trois-Rivières, Université du Québec à Trois-Rivières.
- [2] Barbe P., Genest C., Ghoudi K., Rémillard B. (1996). On Kendall's process. *Journal of Multivariate Analysis* 58, 197-229.
- [3] Bedoui R., Ben Dbabis M. (2009). Copulas and bivariate risk measures : an application to hedge funds. *26ème journée internationale d'économie monétaire et financière*.
- [4] Berg D. (2009). Copula goodness-of-fit testing : An overview and power comparison. *The European Journal of Finance*, vol. 15, issue 7-8, 675-701
- [5] Breyermann 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] Fermanian J.-D., Radulovic D., Wegkamp M. (2004). Weak convergence of empirical copula processes. *Bernoulli* 10, 847-860.
- [9] Ganssler P., Stute W. (1987). *Seminar on empirical processes*. DMV Seminar, 9. Birkhauser Verlag, Basel.
- [10] Genest C., Rémillard B., Beaudoin D. (2009). Goodness-of-fit tests for copulas : a review and a power study. *Insurance Math. Econom.* 44, 199-21.
- [11] Genest C., Rivest L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* 88, 1034-1043.

- [12] Genest C., Quessy J.-F., Remillard B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.* 33,337-366.
- [13] Henze N. (2002). Invariant tests for multivariate normality : a critical review. *Statistical Papers* 43, 467-506.
- [14] Malevergne Y., Sornette D. (2003). Testing the gaussian copula hypothesis for financial assets dependences. *Quantitative Finance* 3, 231-250.
- [15] 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.
- [16] Samira T., Razika H. (2018). *Les copules dans \mathbb{R}^2* . Mémoire de fin de cycle de Master. Université Mohammed Seddik Ben Yahai - Jijel.
- [17] Toupin M. H. (2008). *Nouveau test d'adéquation pour les copules basé sur le processus de Spearman*. Mémoire de fin de cycle de Master. Université du Québec à trois-rivières.
- [18] Tsukahara H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.* 33, 357-375.
- [19] 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.