

UNIVERSITÉ DU QUÉBEC

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

COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES

PAR  
ALBERT FOLCHER

NOUVELLES MÉTHODES D'ESTIMATION SEMI-PARAMÉTRIQUES DE  
PÉRIODES DE RETOUR BIVARIÉES

DÉCEMBRE 2023

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

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

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

## AVANT-PROPOS

Je tiens à remercier sincèrement Monsieur Jean-François Quessy, mon directeur de recherche, qui en plus d'avoir fourni un enseignement de qualité grâce à sa pédagogie, ses connaissances et sa persévérance, m'a proposé un projet passionnant tout en m'accompagnant au long de celui-ci. Je tiens à exprimer mes remerciements à Monsieur Ahmat Boysara, qui a su m'inculquer l'amour des mathématiques.

Je remercie aussi mes parents, Magali et Olivier, dont la confiance et le soutien ont été indispensables. Je tiens à remercier toutes personnes ayant contribué de près ou de loin à ce mémoire.

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

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

# Table des matières

<b>Avant-propos</b>	ii
<b>Table des matières</b>	iii
<b>Liste des tableaux</b>	vii
<b>Table des figures</b>	viii
<b>Chapitre 1. Périodes de retour, courbes de niveau et plan du mémoire</b>	1
1.1 Périodes de retour . . . . .	1
1.1.1 Motivation et utilité . . . . .	1
1.1.2 Définition formelle . . . . .	2
1.1.3 Le cas à une variable . . . . .	2
1.1.4 Le cas à deux variables . . . . .	3
1.1.5 Autres événements bivariés . . . . .	4
1.2 Seuils critiques et courbes de niveau . . . . .	5
1.2.1 Seuils critiques unidimensionnels . . . . .	5
1.2.2 Ensembles de niveaux dans le cas bivarié . . . . .	5
1.2.3 Courbes de niveau bivariées . . . . .	6
1.3 Objectifs du mémoire . . . . .	7
<b>Chapitre 2. Intervalle de confiance sur une courbe de niveau</b>	9
2.1 Lien entre courbe de niveau et copules . . . . .	9
2.1.1 Théorème de Sklar . . . . .	9
2.1.2 Versions copules des ensembles de niveau . . . . .	11

2.1.3	Versions copules des courbes de niveau . . . . .	12
2.2	Intervalles de valeurs probables sur une courbe de niveau . . . . .	13
2.2.1	Description de l'objectif poursuivi . . . . .	13
2.2.2	Ensemble de confiance sur $L_p$ . . . . .	14
2.2.3	Ensemble de confiance sur $\bar{L}_p^C$ . . . . .	15
2.3	Distribution de $U$ sur la courbe de niveau $L_p^C$ . . . . .	15
2.3.1	Loi conjointe de $U$ et de $W = C(U, V)$ . . . . .	15
2.3.2	Loi de $W = C(U, V)$ . . . . .	17
2.3.3	Loi conditionnelle de $U$ sachant $W = C(U, V)$ . . . . .	18
2.3.4	Preuve alternative de la Proposition 2.3 . . . . .	19
2.4	Illustrations . . . . .	20
2.4.1	Copules symétriques . . . . .	20
2.4.2	Copule de l'indépendance . . . . .	21
2.4.3	Copule de Clayton . . . . .	22
2.4.4	Copule de Farlie–Gumbel–Morgenstern . . . . .	24
2.5	Projection de l'intervalle sur la courbe de niveau . . . . .	26
<b>Chapitre 3. Copules Archimédiennes</b>		<b>28</b>
3.1	Définition de la famille Archimédienne . . . . .	28
3.1.1	Expression générale d'une copule Archimédienne . . . . .	28
3.1.2	Courbe de niveau . . . . .	29
3.1.3	Tau de Kendall . . . . .	29
3.2	Expression générale de l'ensemble de confiance . . . . .	30
3.2.1	Détermination de $C_\Psi^{(01)}$ et $c_\Psi$ . . . . .	30
3.2.2	Détermination de $k_\Psi$ . . . . .	30
3.2.3	Distribution de $U$ sachant $C_\Psi(U, V)$ . . . . .	31
3.2.4	Ensemble de confiance . . . . .	33
3.3	Modèles Archimédiens particuliers . . . . .	33
3.3.1	Copule de Clayton . . . . .	33
3.3.2	Copule de Frank . . . . .	35

3.3.3	Copule de Gumbel . . . . .	36
3.3.4	Copule de Joe . . . . .	38
3.3.5	Copule de Ali–Mikhail–Haq . . . . .	38
<b>Chapitre 4. Copules Normale et Student</b>		<b>40</b>
4.1	Copule Normale . . . . .	40
4.1.1	Quelques préliminaires . . . . .	40
4.1.2	Obtention du quotient $c/C^{(01)}$ . . . . .	41
4.1.3	Obtention numérique de l'intervalle de confiance . . . . .	43
4.2	Copule de Student . . . . .	43
4.2.1	Quelques préliminaires . . . . .	43
4.2.2	Obtention du quotient $c/C^{(01)}$ . . . . .	44
4.3	Preuves . . . . .	47
4.3.1	Preuve du Lemme 4.3 . . . . .	47
4.3.2	Preuve du Lemme 4.4 . . . . .	48
<b>Chapitre 5. La grande famille des copules Beta</b>		<b>50</b>
5.1	Construction et propriétés de la copule Beta . . . . .	50
5.1.1	La loi MGB2 . . . . .	50
5.1.2	Copule et densité Beta . . . . .	51
5.1.3	Cas particuliers de la famille Beta . . . . .	52
5.2	Obtention de $C^{(01)}$ . . . . .	53
5.3	Illustration du cas particulier $p_1 = p_2 = 1$ : copule de survie Clayton . .	56
<b>Chapitre 6. Estimation semi-paramétrique des ensembles de confiance</b>		<b>58</b>
6.1	Version empirique des intervalles de confiance . . . . .	58
6.1.1	Contexte . . . . .	58
6.1.2	Objectif . . . . .	59
6.2	Description de la méthodologie proposée . . . . .	59
6.2.1	Estimer le paramètre d'une copule . . . . .	60
6.2.2	Choix formel d'une famille de copules . . . . .	61

6.2.3	Estimation non-paramétrique des lois marginales	63
6.3	Illustration sur des données financières	63
6.3.1	Présentation des données	63
6.3.2	Stationnarité et indépendance sérielle	64
6.3.3	Choix d'une copule	65
6.3.4	Intervalle de confiance semi-paramétrique	68
6.3.5	Interprétation des résultats	69
<b>Conclusion</b>		<b>70</b>
<b>Bibliographie</b>		<b>72</b>

# Liste des tableaux

6.1	Statistiques descriptives univariées pour le cours des actions de <i>Exxon-Mobil</i> et de <i>Schlumberger</i> entre le 1 <sup>er</sup> janvier et le 31 décembre 2021 . . .	64
6.2	P-valeurs (en pourcentage) des tests d'adéquation pour copules basés sur les statistiques $W_{n,N}^{\text{GRB}}$ et $W_{n,N}^{\text{GQR}}$ pour le prix des actions de <i>Exxon-Mobil</i> et de <i>Schlumberger</i> entre le 1 <sup>er</sup> janvier et le 31 décembre 2021 . . .	67

# Table des figures

3.1 Densité $f_p$ (à gauche) et courbe de niveau (à droite) de la copule de Clayton lorsque $p = 0, 2$ (en haut) et $p = 0, 8$ (en bas) lorsque $\tau = 0$ (rouge), $\tau = 1/4$ (bleu foncé), $\tau = 1/2$ (vert), $\tau = 2/3$ (bleu pâle) et $\tau = 3/4$ (magenta) . . . . .	34
3.2 Densité $f_p$ (à gauche) et courbe de niveau (à droite) de la copule de Frank lorsque $p = 0, 2$ (en haut) et $p = 0, 8$ (en bas) lorsque $\tau = 0$ (rouge), $\tau = 1/4$ (bleu foncé), $\tau = 1/2$ (vert), $\tau = 2/3$ (bleu pâle) et $\tau = 3/4$ (magenta) . . . . .	36
3.3 Densité $f_p$ (à gauche) et courbe de niveau (à droite) de la copule de Gumbel lorsque $p = 0, 2$ (en haut) et $p = 0, 8$ (en bas) lorsque $\tau = 0$ (rouge), $\tau = 1/4$ (bleu foncé), $\tau = 1/2$ (vert), $\tau = 2/3$ (bleu pâle) et $\tau = 3/4$ (magenta) . . . . .	37
6.4 Évolution du prix des actions de <i>ExxonMobil</i> et de <i>Schlumberger</i> entre le 1 <sup>er</sup> janvier et le 31 décembre 2021. En haut : séries originales ; en bas : séries différencierées à l'ordre un . . . . .	65
6.5 Sur la diagonale : histogrammes de la différence des prix des actions des compagnies <i>ExxonMobil</i> (en haut à gauche) et <i>Schlumberger</i> (en bas à droite) ; en haut à droite : nuage de points des données originales ; en bas à gauche : nuage de points des données standardisées . . . . .	66

# CHAPITRE 1

## PÉRIODES DE RETOUR, COURBES DE NIVEAU ET PLAN DU MÉMOIRE

### 1.1 Périodes de retour

#### 1.1.1 Motivation et utilité

La notion de période de retour d'un événement est utilisée fréquemment en hydrologie, en ingénierie civile et, plus généralement, dans les sciences de l'environnement. Elle sert notamment à prendre des décisions rationnelles vis-à-vis les impacts possibles d'événements naturels jugés dangereux. On s'en sert par exemple pour caractériser les risques naturels reliés aux tremblements de terre, à la crue des eaux et aux ouragans, pour n'en nommer que quelques-uns. Également, les autorités se basent souvent sur la valeur d'une période de retour pour planifier des infrastructures qui devront répondre aux risques encourus, par exemple les barrages hydro-électriques. Un autre domaine où l'évaluation des périodes de retour est cruciale est celui de l'assurance. En effet, les assureurs se basent sur la période de retour pour évaluer le niveau de risque d'un assuré et ainsi fixer la prime qu'il aura à débourser. Les lecteurs intéressés peuvent consulter les travaux de Klemeš [24] et Klemeš [25] en hydrologie, de même que Singh *et al.* [35] pour de plus amples détails concernant l'usage de la période de retour en ingénierie civile et environnementale.

### 1.1.2 Définition formelle

La période de retour d'un événement récurrent est définie comme la durée moyenne entre deux occurrences consécutives d'une même intensité. Formellement, soit un événement récurrent noté  $\mathcal{A}$  et  $\mathbb{P}(\mathcal{A})$  la probabilité que cet événement survienne. En posant  $\mu_{\mathcal{A}}$  le temps moyen qui s'écoule entre deux occurrences consécutives de cet événement, la période de retour se définit par

$$T_{\mathcal{A}} = \frac{\mu_{\mathcal{A}}}{\mathbb{P}(\mathcal{A})}. \quad (1.1)$$

Dans les applications les plus courantes, l'événement d'intérêt  $\mathcal{A}$  fait intervenir une variable aléatoire  $X$ . Dans d'autres contextes,  $\mathcal{A}$  est relié à une paire  $(X, Y)$  de variables aléatoires. Cette situation engendre une théorie plus riche que le cas à une seule variable, car elle fait intervenir la notion de *dépendance* entre  $X$  et  $Y$ .

**Exemple 1.1.** *Lors de la construction d'un barrage, on fixe habituellement la période de retour à  $T_{\mathcal{A}} = 1000$  ans, de sorte que celui-ci soit capable de résister à tous les événements météorologiques possibles, sauf ceux qui se produisent une fois par millénaire. Si on s'intéresse à des événements qui s'observent sur une échelle annuelle, alors  $\mu_{\mathcal{A}} = 1$  ; on peut mentionner par exemple les crues annuelles maximales et les débits totaux annuels. Dans ce cas, on aura  $\mathbb{P}(\mathcal{A}) = \mu_{\mathcal{A}}/T_{\mathcal{A}} = 1/1000 = 0,001$ . Ainsi, les ingénieurs et les architectes devront s'assurer de bâtir une structure capable de résister à des intempéries dont la probabilité d'occurrence annuelle est de 0,001.*

### 1.1.3 Le cas à une variable

Supposons que l'événement récurrent  $\mathcal{A}$  fasse intervenir une variable aléatoire  $X$  à valeurs dans un sous-ensemble  $\mathcal{X}$  de  $\mathbb{R}$ . Généralement, on s'intéressera à l'événement *extrême* de la forme  $\mathcal{A}^{\uparrow} = \{X > \zeta_X\}$ , où  $\zeta_X$  est un seuil jugé critique. Dans ce cas,

$\mathbb{P}(\mathcal{A}^\uparrow) = \mathbb{P}(X > \zeta_X) = 1 - F_X(\zeta_X)$ , où  $F_X$  est la fonction de répartition de  $X$ . La période de retour s'exprime alors par

$$T^\uparrow = \frac{\mu}{1 - F_X(\zeta_X)}.$$

Il arrive aussi que l'événement d'intérêt porte une attention aux valeurs extrêmement petites de  $X$ , c'est-à-dire  $\mathcal{A}^\downarrow = \{X \leq \zeta_X\}$ . Dans ce cas,  $\mathbb{P}(\mathcal{A}^\downarrow) = \mathbb{P}(X \leq \zeta_X) = F_X(\zeta_X)$  et la période de retour s'écrit

$$T^\downarrow = \frac{\mu}{F_X(\zeta_X)}.$$

#### 1.1.4 Le cas à deux variables

Ce mémoire portera une attention particulière aux événements qui font intervenir une paire  $(X, Y)$  de variables aléatoires à valeurs dans un sous-ensemble  $\mathcal{X} \times \mathcal{Y}$  du plan réel  $\mathbb{R}^2$ . Dans cette situation, il y a plusieurs événements qui peuvent être d'un intérêt particulier. Le domaine de l'hydrologie fait régulièrement usage de la notion de période de retour bivariée. Par exemple, les quantités de pluie et la hauteur de la crue des eaux d'un cours d'eau sont, en règle générale, fortement liées. Il en est de même de la pointe et du volume annuels d'un cours d'eau. Un recensement des méthodologies disponibles a été rédigé par Gräler *et al.* [21], alors que Brunner *et al.* [4] soulignent leur importance pour l'évaluation des risques reliés aux pointes et aux volumes des crues. Voir également l'article de Volpi et Fiori [38] concernant l'évaluation des structures hydrauliques à la lumière de périodes de retour pour des événements extrêmes.

Les événements les plus courants faisant intervenir deux variables aléatoires concernent les valeurs élevées de  $X$  et de  $Y$  jugées critiques selon des seuils  $\zeta_X$  et  $\zeta_Y$ , à savoir

$$\mathcal{A}_\cup^{\uparrow\uparrow} = \{X > \zeta_X \cup Y > \zeta_Y\} \quad \text{et} \quad \mathcal{A}_\cap^{\uparrow\uparrow} = \{X > \zeta_X \cap Y > \zeta_Y\}.$$

Si  $F_{X,Y}(x,y) = \mathbb{P}(X \leq x \cap Y \leq y)$  est la fonction de répartition conjointe de  $(X, Y)$ , alors une application de la formule de de Morgan  $\mathbb{P}(A \cup B) = 1 - \mathbb{P}(A^c \cap B^c)$  amène

$$\mathbb{P}(\mathcal{A}_{\cup}^{\uparrow\uparrow}) = \mathbb{P}(X > \zeta_X \cup Y > \zeta_Y) = 1 - \mathbb{P}(X \leq \zeta_X \cap Y \leq \zeta_Y) = 1 - F_{X,Y}(\zeta_X, \zeta_Y).$$

D'autre part, en posant  $\bar{F}_{X,Y}(x,y) = \mathbb{P}(X > x, Y > y)$  comme étant la fonction de survie de  $(X, Y)$ , on obtient aisément  $\mathbb{P}(\mathcal{A}_{\cap}^{\uparrow\uparrow}) = \bar{F}_{X,Y}(\zeta_X, \zeta_Y)$ . Les périodes de retour associées à ces deux événements sont alors

$$T_{\cup}^{\uparrow\uparrow} = \frac{\mu_{\cup}^{\uparrow\uparrow}}{1 - F_{X,Y}(\zeta_X, \zeta_Y)} \quad \text{et} \quad T_{\cap}^{\uparrow\uparrow} = \frac{\mu_{\cap}^{\uparrow\uparrow}}{\bar{F}_{X,Y}(\zeta_X, \zeta_Y)}.$$

### 1.1.5 Autres événements bivariés

Dans une moindre mesure, on pourrait aussi considérer des événements de la forme  $\mathcal{A}_{\cap}^{\uparrow\downarrow} = \{X > \zeta_X \cap Y \leq \zeta_Y\}$  et  $\mathcal{A}_{\cup}^{\uparrow\downarrow} = \{X > \zeta_X \cup Y \leq \zeta_Y\}$ . Puisque l'événement  $\{Y \leq \zeta_Y\}$  peut s'écrire comme l'union des ensembles disjoints  $\{X \leq \zeta_X, Y \leq \zeta_Y\}$  et  $\{X > \zeta_X, Y \leq \zeta_Y\}$ , on peut écrire  $\mathbb{P}(Y \leq \zeta_Y) = \mathbb{P}(X \leq \zeta_X, Y \leq \zeta_Y) + \mathbb{P}(X > \zeta_X, Y \leq \zeta_Y)$ . En posant  $F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y)$  comme la loi marginale de  $Y$ , on obtient

$$\begin{aligned} \mathbb{P}(\mathcal{A}_{\cap}^{\uparrow\downarrow}) &= \mathbb{P}(X > \zeta_X \cap Y \leq \zeta_Y) \\ &= \mathbb{P}(Y \leq \zeta_Y) - \mathbb{P}(X \leq \zeta_X, X \leq \zeta_Y) \\ &= F_Y(y) - F_{X,Y}(x, y). \end{aligned}$$

De manière similaire, on calcule  $\mathbb{P}(\mathcal{A}_{\cup}^{\uparrow\downarrow}) = 1 - F_X(x) + F_{X,Y}(x, y)$ , où  $F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$  est la loi marginale de  $X$ .

## 1.2 Seuils critiques et courbes de niveau

### 1.2.1 Seuils critiques unidimensionnels

Généralement, les seuils critiques sont déterminés par rapport à une probabilité d'occurrence. Autrement dit, au lieu de fixer l'événement  $\mathcal{A} \subset \mathcal{X}$ , on cherche plutôt à déterminer l'ensemble  $\mathcal{A}$  tel que  $\mathbb{P}(\mathcal{A}) = p$ , où  $p \in (0, 1)$  est une probabilité préatable que l'événement survienne. Pour être plus précis, prenons un événement de la forme  $\mathcal{A}^\uparrow = \{X > \zeta_X\}$  pour lequel on a vu que  $\mathbb{P}(\mathcal{A}^\uparrow) = 1 - F_X(x)$ . Alors pour faire en sorte que  $\mathbb{P}(\mathcal{A}^\uparrow) = p$ , il faut choisir  $\zeta_X$  de telle sorte que  $1 - F_X(\zeta_X) = p$ . Si  $X$  est une variable aléatoire continue, sa fonction de répartition admet un inverse unique et

$$\zeta_X = F_X^{-1}(1 - p).$$

Dans le cas de l'événement  $\mathcal{A}^\downarrow$ , on a simplement  $\zeta_X = F_X^{-1}(p)$ .

### 1.2.2 Ensembles de niveaux dans le cas bivarié

Les seuils critiques dans le cas à une seule variable sont basés sur les sous-ensembles  $S_p^\uparrow = \{x \in \mathcal{X} : x > p\}$  et  $S_p^\downarrow = \{x \in \mathcal{X} : x \leq p\}$ . Une façon de généraliser cette idée au cas à deux variables consiste à définir un sous-ensemble de  $\mathcal{X} \times \mathcal{Y}$  du plan réel  $\mathbb{R}^2$  correspondant à un couple de variables aléatoires  $(X, Y)$ . Une des possibilités consiste à poser

$$L_p^{\uparrow\downarrow} = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : F_{X,Y}(x, y) \leq p\}. \quad (1.2)$$

Le sous-ensemble  $L_p^{\uparrow\downarrow}$  s'appelle un *ensemble de niveau* (en anglais, *level set*). Coblenz et al. [7] montrent que  $\{L_p^{\uparrow\downarrow} : p \in [0, 1]\}$  caractérise entièrement le comportement

aléatoire de  $(X, Y)$ . Ils décrivent, en outre, une façon non-paramétrique d'estimer  $L_p^{\downarrow\downarrow}$ . On pourrait également définir

$$L_p^{\uparrow\uparrow} = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : F_{X,Y}(x, y) > p\}. \quad (1.3)$$

Le lecteur est dirigé vers l'article de Dung *et al.* [10] concernant les incertitudes quand à l'estimation d'un ensemble de niveau et pour une illustration concernant les risques de débordement du fleuve Mékong, en Asie du Sud-Est.

À noter que la probabilité qu'une paire  $(X, Y)$  prenne ses valeurs dans l'ensemble  $L_p^{\downarrow\downarrow}$  n'est pas égale à  $p$ . En effet, si on définit  $K$  comme la fonction de répartition de la variable aléatoire  $W = F_{X,Y}(X, Y)$ , alors

$$\mathbb{P}\{(X, Y) \in L_p^{\downarrow\downarrow}\} = \mathbb{P}\{F_{X,Y}(X, Y) \leq p\} = K(p).$$

Une manière simple d'avoir un ensemble de niveau dont la probabilité d'occurrence est  $p$  consiste donc simplement à redéfinir  $L_p^{\downarrow\downarrow}$  par

$$\tilde{L}_p^{\downarrow\downarrow} = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : F_{X,Y}(x, y) \leq K^{-1}(p)\},$$

tel que cela a été suggéré par Coblenz *et al.* [7]. De même, comme  $\mathbb{P}\{(X, Y) \in L_p^{\uparrow\uparrow}\} = \mathbb{P}\{F_{X,Y}(X, Y) > p\} = 1 - K(p)$ , il s'agit de poser  $\tilde{L}_p^{\uparrow\uparrow} = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : F_{X,Y}(x, y) > K^{-1}(1 - p)\}$  de manière à ce que  $\mathbb{P}\{(X, Y) \in \tilde{L}_p^{\uparrow\uparrow}\} = p$ .

### 1.2.3 Courbes de niveau bivariées

Plutôt que de travailler sur des ensembles tels que  $\tilde{L}_p^{\downarrow\downarrow}$  et  $\tilde{L}_p^{\uparrow\uparrow}$ , soit la *courbe de niveau*

$$L_p = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : F_{X,Y}(x, y) = 1 - p\}. \quad (1.4)$$

L'ensemble  $L_p \subset \mathcal{X}$  est non-dénombrable et peut être décrit par la courbe dans  $\mathbb{R}^2$  telle que  $y = F_x^{-1}(1 - p)$ , où  $F_x^{-1}(1 - p) = \{y : F_{X,Y}(x, y) = 1 - p\}$ . Contrairement aux ensembles de niveau, il y a un lien direct entre la courbe de niveau  $L_p$  et les périodes de retour bivariées  $T^\uparrow$  et  $T^\downarrow$ . En effet, l'ensemble  $L_p$  contient les valeurs  $(\zeta_X, \zeta_Y) \in \mathcal{X} \times \mathcal{Y}$  telles que  $\mathbb{P}(\mathcal{A}_U^{\uparrow\uparrow}) = p$ . Pour faire le lien entre une courbe de niveau et  $T^\downarrow$ , il s'agit de définir  $\bar{L}_p = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : \bar{F}_{X,Y}(x, y) = p\}$ , qui contient les valeurs  $(\zeta_X, \zeta_Y) \in \mathcal{X} \times \mathcal{Y}$  telles que  $\mathbb{P}(\mathcal{A}_U^{\downarrow\downarrow}) = p$ .

### 1.3 Objectifs du mémoire

Tel que noté à la section précédente, la courbe de niveau  $L_p$  définie par l'Équation (1.4) contient les valeurs de  $\mathcal{X} \times \mathcal{Y}$  telles que  $\mathbb{P}(\mathcal{A}_U^{\uparrow\uparrow}) = p$ . Cependant, tel que l'ont bien noté Salvadori *et al.* [32] et Volpi et Fiori [37], ces valeurs du sous-ensemble de  $\mathcal{X} \times \mathcal{Y}$  décrites par  $L_p$  ne sont pas équiprobables au sens où la loi de  $(X, Y)$  sur cette courbe n'est pas uniforme. Ce commentaire s'applique également à la courbe de niveau  $\bar{L}_p$ .

La suggestion faite par Salvadori *et al.* [32] consiste à choisir la valeur de  $(x, y)$  sur la courbe  $L_p$  qui maximise une certaine fonction objectif. Plus formellement, pour  $\omega : L_p \rightarrow [0, \infty)$ , ces auteurs suggèrent de choisir

$$(x_0, y_0) = \arg \max_{(x, y) \in L_p} \omega(x, y).$$

Cette solution ne nous apparaît pas satisfaisante. Ce mémoire s'inspirera plutôt de l'approche de Volpi et Fiori [37], qui consiste à construire des ensembles de confiance probabilistes à partir de la loi de probabilité de  $(X, Y)$  sur la courbe  $L_p$ .

En se basant sur cette idée de Volpi et Fiori [37], ce mémoire de maîtrise s'attardera à construire des ensemble de confiance pour des événements concernant un couple

$(X, Y)$  distribué sur une courbe de niveau. De cette manière, on pourra identifier un sous-ensemble de valeurs possibles de  $(X, Y)$  sur la courbe  $L_p$  qui sont susceptibles de survenir à un certain seuil de probabilité  $1 - \alpha$ . Il est important de souligner ce qui distingue ce travail de recherche de l'article de Volpi et Fiori [37] :

- (i) On adopte une approche plus élégante qui permet d'obtenir une formule générale pour la loi de probabilité d'un couple aléatoire sur une courbe de niveau ;
- (ii) On déduit des formules explicites pour plusieurs familles de copules importantes, notamment les copules Archimédienne, Normale, Student et Beta ;
- (iii) L'approche par copules est privilégiée, permettant une estimation semi-paramétrique de l'ensemble de confiance libre du choix des lois marginales ;
- (iv) On étudie les probabilités de couverture des ensembles de confiance.

Le reste du mémoire est organisé de la manière suivante. Au Chapitre 2, une expression générale est déduite pour la loi de probabilité sur une courbe de niveau ; ce résultat permet de définir formellement l'ensemble de confiance. Le Chapitre 3 développe des formules explicites pour cet ensemble pour les copules qui appartiennent à la famille Archimédienne. Le Chapitre 4 déduit de telles formules pour les copules Normale et Student, alors que le Chapitre 5 est dédié à la famille des copules Beta. Enfin, il est montré au Chapitre 6 comment les ensembles de confiance peuvent être estimés et une illustration est effectuée sur des données financières.

## CHAPITRE 2

# INTERVALLE DE CONFIANCE SUR UNE COURBE DE NIVEAU

## 2.1 Lien entre courbe de niveau et copules

### 2.1.1 Théorème de Sklar

La théorie des copules permet d'analyser la structure de la dépendance entre plusieurs variables aléatoires. Le point de départ en est le Théorème de Sklar [36], qui est devenu célèbre pour l'étude de la dépendance entre deux ou plusieurs variables aléatoires. Afin de l'énoncer, soit un vecteur  $\mathbf{X} = (X_1, \dots, X_d)$  de variables aléatoires dont la fonction de répartition conjointe est donnée par

$$F_{\mathbf{X}}(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Les lois marginales de  $\mathbf{X}$  peuvent s'extraire de  $F_{\mathbf{X}}$  via

$$F_j(x) = \lim_{\substack{x_\ell \rightarrow \infty \\ \forall \ell \neq j}} F_{\mathbf{X}}(x_1, \dots, x_d).$$

**Théorème 2.1.** *Si  $F_{\mathbf{X}}$  est une fonction de répartition  $d$ -dimensionnelle dont les marges sont  $F_1, \dots, F_d$ , alors il existe une fonction  $C : [0, 1]^d \rightarrow [0, 1]$  telle que*

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

*Si les lois marginales  $F_1, \dots, F_d$  sont continues, alors la fonction  $C$  est unique et s'appelle la copule de la loi  $F$ .*

Dans la suite, on supposera toujours que les lois marginales sont continues, sans le spécifier à chaque fois. Dans ce cas, le Théorème 2.1 permet d'extraire la copule d'une loi multidimensionnelle donnée  $F_{\mathbf{X}}$  via la relation

$$C(u_1, \dots, u_d) = F_{\mathbf{X}} \{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\}. \quad (2.2)$$

Il est parfois utile de noter que la copule  $C$  associée à une loi  $F_{\mathbf{X}}$  correspond à la loi conjointe du vecteur  $\mathbf{U} = (F_1(X_1), \dots, F_d(X_d))$ , où  $\mathbf{X} = (X_1, \dots, X_d) \sim F_{\mathbf{X}}$ . En effet,

$$\begin{aligned} \mathbb{P} \{F_1(X_1) \leq u_1, \dots, F_d(X_d) \leq u_d\} &= \mathbb{P} \{X_1 \leq F_1^{-1}(u_1), \dots, X_d \leq F_d^{-1}(u_d)\} \\ &= F \{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\} \\ &= C \{F_1 \circ F_1^{-1}(u_1), \dots, F_d \circ F_d^{-1}(u_d)\} \\ &= C(u_1, \dots, u_d). \end{aligned}$$

Puisque pour chaque  $j \in \{1, \dots, d\}$ , la loi de  $U_j = F_j(X_j)$  est uniforme sur  $(0, 1)$ , une copule  $C$  est une loi  $d$ -dimensionnelle à marges uniformes. À noter qu'une version du Théorème de Sklar [36] pour la fonction de survie  $\bar{F}_{\mathbf{X}}(x_1, \dots, x_d) = \mathbb{P}(X_1 >$

$x_1, \dots, X_d > x_d$ ) est possible en remarquant que

$$\begin{aligned}\bar{F}_{\mathbf{X}}(x_1, \dots, x_d) &= \mathbb{P}\{F_1(X_1) > F_1(x_1), \dots, F_d(X_d) > F_d(x_d)\} \\ &= \mathbb{P}\{1 - F_1(X_1) \leq 1 - F_1(x_1), \dots, 1 - F_d(X_d) \leq F_d(x_d)\} \\ &= \mathbb{P}\{1 - U_1 \leq \bar{F}_1(x_1), \dots, 1 - U_d \leq \bar{F}_d(x_d)\} \\ &= \hat{C}\{\bar{F}_1(x_1), \dots, \bar{F}_d(x_d)\},\end{aligned}$$

où  $\hat{C} : [0, 1]^d \rightarrow [0, 1]$  est la loi conjointe de  $\bar{\mathbf{U}} = (1 - U_1, \dots, 1 - U_d)$ , où  $\mathbf{U} \sim C$ . Puisque pour chaque  $j \in \{1, \dots, d\}$ , la loi de  $1 - U_j$  est uniforme sur  $[0, 1]$ , la fonction  $\hat{C}$  est une copule ; on y réfère comme la *copule de survie*. Il y a un lien direct entre la copule  $C$  de  $F_{\mathbf{X}}$  et la copule de survie, puisque cette dernière est construite à partir d'un vecteur  $\mathbf{U}$  de loi  $C$ . Dans le cas  $d = 2$ , par exemple, on montre aisément que  $\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$ .

Lorsque la densité  $c(\mathbf{u}) = \partial C(\mathbf{u}) / \partial u_1 \cdots \partial u_d$  d'une copule  $C$  existe, la représentation de l'Équation (2.2) permet de déduire

$$c(\mathbf{u}) = f_{\mathbf{X}}\{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\} \left/ \prod_{\ell=1}^d f_{\ell}\{F_{\ell}^{-1}(u_{\ell})\}\right..$$

Inversement, en se basant plutôt sur l'expression d'une loi  $F_{\mathbf{X}}$  dans le Théorème 2.1, on peut exprimer la densité associée à  $F_{\mathbf{X}}$  via la formule

$$f_X(\mathbf{x}) = c\{F_1(x_1), \dots, F_d(x_d)\} \prod_{\ell=1}^d F_{\ell}(x_{\ell}). \quad (2.3)$$

### 2.1.2 Versions copules des ensembles de niveau

Il est clair que la prise en compte de la dépendance qui existe entre deux variables aléatoires est primordiale pour une analyse appropriée. En effet, supposer faussement

que deux variables sont indépendantes peut mener à de graves surestimation ou, surtout, à de graves erreurs de sous-estimation de risques encourus. À ce titre, l'usage des copules dans le domaine de l'hydrologie s'avère un outil puissant, voire essentiel, tel que l'ont souligné Salvadori et De Michele [31].

Plusieurs auteurs ont souligné et exploité le lien qui existe entre les périodes de retour bivariées et la copule sous-jacente à la loi d'une paire aléatoire. Par exemple, Salvadori [30], Salvadori *et al.* [33] et Requena *et al.* [29] s'en servent pour estimer des périodes de retour bivariées et multivariées, alors que Belzunce *et al.* [3] et Yin *et al.* [40] les exploitent pour évaluer des percentiles bidimensionnels. Gräler *et al.* [20] utilisent une construction de copules en vignes pour modéliser les périodes de retour multivariées.

L'idée maîtresse qui sera exploitée dans la suite est que l'ensemble de niveau de l'Équation (1.2) peut se ré-écrire sous la forme

$$L_p^{\downarrow\downarrow} = \{(F_X^{-1}(u), F_Y^{-1}(v)) \in \mathcal{X} \times \mathcal{Y} : (u, v) \in [0, 1]^2, C(u, v) \leq p\}.$$

On peut donc adjoindre à  $L_p^{\downarrow\downarrow}$  sa version *copule* en considérant l'ensemble  $\{(u, v) \in [0, 1]^2 : C(u, v) \leq p\}$ . On peut procéder similairement pour l'ensemble de niveau  $L_p^{\uparrow\uparrow}$  en définissant  $\{(u, v) \in [0, 1]^2 : C(u, v) > p\}$ .

### 2.1.3 Versions copules des courbes de niveau

Pour un couple  $(X, Y)$  de variables aléatoires dont la loi conjointe est  $F_{X,Y}$ , l'Équation (2.1) du Théorème de Sklar s'exprime par  $F_{X,Y}(x, y) = C\{F_X(x), F_Y(y)\}$ . À

partir de la définition de la courbe de niveau  $L_p$  à l'Équation (1.4), on a ainsi

$$\begin{aligned} L_p &= \{(x, y) \in \mathcal{X} \times \mathcal{Y} : F_{X,Y}(x, y) = 1 - p\} \\ &= \{(x, y) \in \mathcal{X} \times \mathcal{Y} : C\{F_X(x), F_Y(y)\} = 1 - p\} \\ &= \{(F_X^{-1}(u), F_Y^{-1}(v)) \in \mathcal{X} \times \mathcal{Y} : (u, v) \in [0, 1]^2, C(u, v) = 1 - p\}. \end{aligned}$$

Ainsi, on a que  $(u, v) \in L_p^C$  si et seulement si  $(F_X^{-1}(u), F_Y^{-1}(v)) \in L_p$ , où

$$L_p^C = \{(u, v) \in [0, 1]^2 : C(u, v) = 1 - p\}. \quad (2.4)$$

On procède de la même manière pour la courbe de niveau  $\bar{L}_p$  en notant d'abord que la version *survie* du Théorème de Sklar [36] permet d'écrire

$$\begin{aligned} \bar{L}_p &= \{(x, y) \in \mathcal{X} \times \mathcal{Y} : \bar{F}_{X,Y}(x, y) = p\} \\ &= \{(x, y) \in \mathcal{X} \times \mathcal{Y} : \bar{C}\{\bar{F}_X(x), \bar{F}_Y(y)\} = p\} \\ &= \{(\bar{F}_X^{-1}(u), \bar{F}_Y^{-1}(v)) \in \mathcal{X} \times \mathcal{Y} : (u, v) \in [0, 1]^2, \bar{C}(u, v) = p\}, \end{aligned}$$

de sorte que  $(u, v) \in \bar{L}_p^C$  si et seulement si  $(\bar{F}_X^{-1}(u), \bar{F}_Y^{-1}(v)) \in L_p$ , où  $\bar{L}_p^C = \{(u, v) \in [0, 1]^2 : \bar{C}(u, v) = p\}$ .

## 2.2 Intervalles de valeurs probables sur une courbe de niveau

### 2.2.1 Description de l'objectif poursuivi

Tel que noté par Volpi et Fiori [37], les valeurs d'une paire aléatoire  $(X, Y)$  sur la courbe de niveau  $L_p$  ne sont pas équiprobables. Autrement dit, il y a des valeurs de

$(X, Y)$  qui sont plus susceptibles de se produire que d'autres. Par exemple, les points limites  $(F_X^{-1}(p), \infty)$  et  $(\infty, F_Y^{-1}(p))$ , bien qu'ils appartiennent à la courbe de niveau  $L_p$ , sont peu susceptibles de se matérialiser.

### 2.2.2 Ensemble de confiance sur $L_p$

Afin d'identifier un ensemble de valeurs probables de  $(X, Y)$  sur une courbe de niveau, soit un niveau de confiance  $1 - \alpha$  et un sous-ensemble  $\mathcal{B}_{p,1-\alpha}$  de  $L_p$  tels que

$$\mathbb{P}\{(X, Y) \in \mathcal{B}_{p,1-\alpha} \mid F_{X,Y}(X, Y) = 1 - p\} = 1 - \alpha.$$

En définissant  $\mathcal{B}_{p,1-\alpha}^C$  tel que  $(u, v) \in \mathcal{B}_{p,1-\alpha}^C$  si et seulement si  $(F_X^{-1}(u), F_Y^{-1}(v)) \in \mathcal{B}_{p,1-\alpha}$ , on peut travailler au niveau de la copule  $C$  de  $(X, Y)$  en considérant

$$\mathbb{P}\{(U, V) \in \mathcal{B}_{p,1-\alpha}^C \mid (U, V) \in L_p^C\} = 1 - \alpha.$$

Puisque  $(u, v) \in L_p^C$  si et seulement si  $v = C_u^{-1}(1 - p)$ , les éléments de l'ensemble  $\mathcal{B}_{p,1-\alpha}^C$  peuvent être uniquement déterminés par la coordonnée  $u \in [1 - p, 1]$ . On peut alors définir un sous-ensemble  $\mathcal{I}_{p,1-\alpha}^C$  de  $[0, 1]$  tel que

$$\mathbb{P}\{U \in \mathcal{I}_{p,1-\alpha}^C \mid C(U, V) = 1 - p\} = 1 - \alpha.$$

Soit la fonction de répartition conditionnelle de  $U$  sachant  $C(U, V) = 1 - p$ , à savoir

$$F_p(u) = \mathbb{P}\{U \leq u \mid C(U, V) = 1 - p\}.$$

On définit alors l'intervalle de confiance de niveau  $1 - \alpha$  par

$$I_{p,1-\alpha}^C = [F_p^{-1}(\alpha_1), F_p^{-1}(1 - \alpha_2)], \quad (2.5)$$

où pour  $\alpha_1, \alpha_2 \in (0, 1)$  tels que  $\alpha_1 + \alpha_2 = \alpha$ .

### 2.2.3 Ensemble de confiance sur $\bar{L}_p^C$

On procède de la même façon pour ce qui est de la courbe de niveau  $\bar{L}_p$  en cherchant un sous-ensemble  $\bar{\mathcal{B}}_{p,1-\alpha}^C$  tel que  $\mathbb{P}\{(U, V) \in \bar{\mathcal{B}}_{p,1-\alpha}^C \mid \bar{C}(U, V) = p\} = 1 - \alpha$ , et en définissant l'ensemble  $\bar{\mathcal{B}}_{p,1-\alpha} \subset \bar{L}_p$  tel que  $(u, v) \in \bar{\mathcal{B}}_{p,1-\alpha}^C$  si et seulement si  $(\bar{F}_X^{-1}(u), \bar{F}_Y^{-1}(v)) \in \bar{\mathcal{B}}_{p,1-\alpha}$ .

## 2.3 Distribution de $U$ sur la courbe de niveau $L_p^C$

L'objectif de cette section est d'obtenir la loi conditionnelle de  $U$  sachant  $C(U, V) = 1 - p$ , ce qui correspond à la distribution de  $U$  sur la courbe de niveau  $L_p^C$ .

### 2.3.1 Loi conjointe de $U$ et de $W = C(U, V)$

**Proposition 2.1.** *Si la copule  $C$  possède une densité  $c$ , alors la densité conjointe de  $U$  et de  $W = C(U, V)$  est donnée par*

$$f_{U,W}(u, w) = \frac{c\{u, C_u^{-1}(w)\}}{C^{(01)}\{u, C_u^{-1}(w)\}}, \quad 0 \leq w \leq u \leq 1,$$

où  $C^{(01)}(u, v) = \partial C(u, v)/\partial v$  et  $C_u^{-1}(w) = \{t : C(u, t) = w\}$ .

**Démonstration.** La transformation inverse de  $Z = U$  et  $W = C(U, V)$  est donnée

par  $U = Z$  et  $V = C_Z^{-1}(W)$ . Le Jacobien de cette transformation est donné par

$$\mathbb{J} = \begin{pmatrix} 1 & 0 \\ \frac{\partial}{\partial z} C_z^{-1}(w) & \frac{\partial}{\partial w} C_z^{-1}(w) \end{pmatrix} = \frac{\partial}{\partial w} C_z^{-1}(w).$$

Si on pose  $r_z(w) = C_z^{-1}(w)$ , on a alors  $C\{z, r_z(w)\} = w$ . En dérivant par rapport à  $w$  de chaque côté de cette équation, on obtient  $C^{(01)}\{z, r_z(w)\} r'_z(w) = 1$ , et donc

$$\mathbb{J} = \frac{\partial}{\partial w} C_z^{-1}(w) = r'_z(w) = \frac{1}{C^{(01)}\{z, C_z^{-1}(w)\}}.$$

Puisque la densité conjointe de  $(U, V)$  est  $c(u, v)$  pour  $(u, v) \in [0, 1]^2$ , on déduit que

$$f_{U,W}(u, w) = c\{u, C_u^{-1}(w)\} \left| \frac{1}{C^{(01)}\{u, C_u^{-1}(w)\}} \right| = \frac{c\{u, C_u^{-1}(w)\}}{C^{(01)}\{u, C_u^{-1}(w)\}},$$

où  $0 \leq u \leq 1$  et  $0 \leq C_u^{-1}(w) \leq 1$ . Cette dernière restriction sur les valeurs de  $(u, v)$  revient à  $C(u, 0) \leq w \leq C(u, 1)$ , c'est-à-dire  $0 \leq w \leq u$ .  $\diamond$

**Remarque 2.1.** *Par définition, la loi de  $U$  est l'uniforme sur  $(0, 1)$ . Sans surprise, on retrouve ce résultat à partir de la Proposition 2.1. En effet, en intégrant la densité bivariée  $f_{U,W}$  sur toutes les valeurs de  $W$ , c'est-à-dire pour  $w \in [0, u]$ , alors*

$$f_U(u) = \int_0^u f_{U,W}(u, w) dw = \int_0^u \frac{c\{u, C_u^{-1}(w)\}}{C^{(01)}\{u, C_u^{-1}(w)\}} dw.$$

*En effectuant ensuite le changement de variable  $t = C_u^{-1}(w)$ , de sorte que  $dt = 1/C^{(01)}\{u, C_u^{-1}(w)\}$ , on obtient*

$$f_U(u) = \int_0^1 c(u, t) dt = 1.$$

### 2.3.2 Loi de $W = C(U, V)$

La variable aléatoire  $W = C(U, V)$  à valeurs dans  $[0, 1]$  s'appelle la *transformation intégrale de probabilité bivariée*. Sa fonction de répartition est notée

$$K_C(w) = \mathbb{P}\{C(U, V) \leq w\}, \quad \text{où } (U, V) \sim C.$$

En conditionnant par rapport à  $U = u$  et en définissant  $C_u^{-1}(t) = \{v \in [0, 1] : C(u, v) = t\}$ , on obtient alors

$$\begin{aligned} K_C(w) &= \int_0^1 \mathbb{P}\{C(U, V) \leq w | U = u\} du \\ &= w + \int_w^1 \mathbb{P}\{V \leq C_u^{-1}(w) | U = u\} du \\ &= w + \int_w^1 C^{(10)}\{u, C_u^{-1}(w)\} du. \end{aligned} \quad (2.6)$$

Cette expression générale a été déduite par Genest et Rivest [19], qui explorent également plusieurs cas particuliers de cette formule. La proposition suivante déduit une expression générale pour la densité de  $K_C$ .

**Proposition 2.2.** *Si  $C$  est une copule bivariée dont la densité  $c$  existe, alors la densité de la variable aléatoire  $W = C(U, V)$ , où  $(U, V) \sim C$ , est donnée par*

$$k_C(w) = \int_w^1 \frac{c\{u, C_u^{-1}(w)\}}{C^{(01)}\{u, C_u^{-1}(w)\}} du,$$

où  $C^{(01)}(u, v) = \partial C(u, v) / \partial v$  et  $C_u^{-1}(w) = \{v : C(u, v) = w\}$ .

**Démonstration.** En dérivant l'expression de  $K_C(w)$  donnée à l'Équation (2.6), on a

$$k_C(w) = 1 - C^{(10)}\{w, C_w^{-1}(t)\} + \int_w^1 \frac{d}{dw} C^{(10)}\{u, C_u^{-1}(w)\} du.$$

Puisque  $C_w^{-1}(w) = 1$ , on a  $C^{(10)}\{w, C_w^{-1}(w)\} = C^{(10)}(w, 1)$ . Comme  $C^{(10)}(w, 1) =$

$\mathbb{P}(V \leq 1|U = w) = 1$ , on a donc  $1 - C^{(10)}\{w, C_w^{-1}(w)\} = 0$ . Maintenant, en posant  $g_u(w) = C_u^{-1}(w)$ , on obtient

$$\frac{d}{dw} C^{(10)}\{u, C_u^{-1}(w)\} = \frac{d}{dw} C^{(10)}\{u, g_u(w)\} = c\{u, g_u(w)\} g'_u(w).$$

Pour déduire la dérivée de  $g_u$ , on part du fait que  $C\{u, g_u(w)\} = w$ , ce qui, dérivé de chaque côté par rapport à  $w$ , amène  $C^{(01)}\{u, g_u(w)\} g'_u(w) = 1$ . On a donc

$$g'_u(w) = \frac{1}{C^{(01)}\{u, g_u(w)\}} = \frac{1}{C^{(01)}\{u, C_u^{-1}(w)\}}.$$

Il s'ensuit que

$$\frac{d}{dw} C^{(10)}\{u, C_u^{-1}(w)\} = \frac{c\{u, C_u^{-1}(w)\}}{C^{(01)}\{u, C_u^{-1}(w)\}},$$

ce qui complète la démonstration.  $\diamond$

### 2.3.3 Loi conditionnelle de $U$ sachant $W = C(U, V)$

On peut désormais obtenir la loi de  $U$  sur la courbe de niveau  $L_p^C$ , c'est-à-dire la loi de  $U$  sachant que  $C(U, V) = 1 - p$ .

**Proposition 2.3.** *Si la copule  $C$  possède une densité  $c$ , alors la loi conditionnelle de  $U$  étant donné  $C(U, V) = 1 - p$  a comme densité*

$$f_p(u) = \frac{1}{k_C(q)} \frac{c\{u, C_u^{-1}(q)\}}{C^{(01)}\{u, C_u^{-1}(q)\}}, \quad u \in [q, 1],$$

où  $q = 1 - p$ ,  $C^{(01)}(u, v) = \partial C(u, v)/\partial v$  et  $C_u^{-1}(q) = \{t : C(u, t) = q\}$ .

**Démonstration.** La preuve est très simple, car par définition de la densité conditionnelle, on a  $f_p(u) = f_{U,W}(u, q)/k_C(q)$ , où  $0 \leq q \leq u \leq 1$ . Il s'agit donc de remplacer  $f_{U,W}$  par son expression obtenue à la Proposition 2.1 pour conclure.

À la lumière de la Proposition 2.3, la fonction de répartition de  $U$  étant donné  $C(U, V) = q$ , où  $q = 1 - p$ , peut s'exprimer pour  $u \in [q, 1]$  par

$$F_p(u) = \int_q^u f_p(s) \, ds = \frac{1}{k_C(q)} \int_q^u \frac{c\{s, C_s^{-1}(q)\}}{C^{(01)}\{s, C_s^{-1}(q)\}} \, ds. \quad (2.7)$$

En injectant l'expression de  $k_C$  obtenue à la Proposition 2.2, on obtient alors

$$F_p(u) = \int_q^u \frac{c\{s, C_s^{-1}(q)\}}{C^{(01)}\{s, C_s^{-1}(q)\}} \, ds \Big/ \int_q^1 \frac{c\{s, C_s^{-1}(q)\}}{C^{(01)}\{s, C_s^{-1}(q)\}} \, ds.$$

### 2.3.4 Preuve alternative de la Proposition 2.3

Par définition,  $f_p$  est la densité conditionnelle de  $U$  sachant  $W = 1 - p$ , où  $W = C(U, V)$ . Par conséquent,  $f_p(u) = f_{U|W}(u, p)$  et on peut écrire

$$f_p(u) = \frac{f_{W|U}(1 - p, u) f_U(u)}{f_W(1 - p)} = \frac{f_{W|U}(1 - p, u)}{k_C(1 - p)}, \quad (2.8)$$

car  $U$  est uniforme sur  $(0, 1)$  et  $f_W$  est la densité de  $W = C(U, V)$ , c'est-à-dire  $k_C$ . Il reste donc à déterminer la densité conditionnelle  $f_{W|U}$ . Pour ce faire, soit

$$\begin{aligned} F_{W|U}(w, u) &= \mathbb{P}(W \leq w | U = u) \\ &= \mathbb{P}\{C(U, V) \leq w | U = u\} \\ &= \mathbb{P}\{V \leq C_u^{-1}(w) | U = u\} \\ &= C^{(10)}\{u, C_u^{-1}(w)\}. \end{aligned}$$

On retrouve ainsi la densité conditionnelle  $f_{W|U}$  en faisant

$$\begin{aligned} f_{W|U}(w, u) &= \frac{\partial}{\partial w} C^{(10)} \{u, C_u^{-1}(w)\} \\ &= c \{u, C_u^{-1}(w)\} \frac{\partial}{\partial w} C_u^{-1}(w) \\ &= \frac{c \{u, C_u^{-1}(w)\}}{C^{(01)} \{u, C_u^{-1}(w)\}}. \end{aligned}$$

En injectant cette expression dans l'Équation (2.8) avec  $w = 1 - p = q$ , on retrouve l'expression obtenue à la Proposition 2.3

## 2.4 Illustrations

### 2.4.1 Copules symétriques

Considérons d'abord le résultat suivant concernant les copules symétriques. On rappelle que  $C$  est symétrique si  $C(u, v) = C(v, u)$  pour tout  $(u, v) \in (0, 1)^2$ .

**Lemme 2.1.** *Si  $C$  est une copule symétrique, alors  $F_p \{C_v^{-1}(q)\} = 1 - F_p(q)$ .*

**Démonstration.** On a d'abord

$$F_p \{C_v^{-1}(q)\} = \int_q^{C_v^{-1}(q)} \frac{c \{s, C_s^{-1}(q)\}}{C^{(01)} \{s, C_s^{-1}(q)\}} ds.$$

En effectuant le changement de variable  $t = C_s^{-1}(q)$ , pour lequel

$$ds = -\frac{C^{(01)} \{t, C_t^{-1}(q)\}}{C^{(10)} \{t, C_t^{-1}(q)\}} dt,$$

on obtient

$$\begin{aligned}
F_p\{C_v^{-1}(q)\} &= - \int_{C_q^{-1}(q)}^v \frac{c\{t, C_t^{-1}(q)\}}{C^{(01)}\{t, C_t^{-1}(q)\}} \frac{C^{(01)}\{t, C_t^{-1}(q)\}}{C^{(10)}\{t, C_t^{-1}(q)\}} dt \\
&= \int_v^1 \frac{c\{t, C_t^{-1}(q)\}}{C^{(10)}\{t, C_t^{-1}(q)\}} dt \\
&= \int_q^1 \frac{c\{t, C_t^{-1}(q)\}}{C^{(10)}\{t, C_t^{-1}(q)\}} dt - \int_q^v \frac{c\{t, C_t^{-1}(q)\}}{C^{(10)}\{t, C_t^{-1}(q)\}} dt \\
&= 1 - F_p(v). \quad \diamond
\end{aligned}$$

Une conséquence du Lemme 2.1 est que lorsque  $C$  est symétrique,

$$\begin{aligned}
\mathcal{I}_{p,1-\alpha}^C &= \{(u, v) \in [0, 1]^2 : C(u, v) = q, F_p^{-1}(\alpha_1) \leq C_v^{-1}(q) \leq F_p^{-1}(1 - \alpha_2)\} \\
&= \{(u, v) \in [0, 1]^2 : C(u, v) = q, \alpha_1 \leq F_p\{C_v^{-1}(q)\} \leq 1 - \alpha_2\} \\
&= \{(u, v) \in [0, 1]^2 : C(u, v) = q, \alpha_1 \leq 1 - F_p(v) \leq 1 - \alpha_2\} \\
&= \{(u, v) \in [0, 1]^2 : C(u, v) = q, \alpha_2 \leq F_p(v) \leq 1 - \alpha_1\} \\
&= \{(u, v) \in [0, 1]^2 : C(u, v) = q, F_p^{-1}(\alpha_2) \leq v \leq F_p^{-1}(1 - \alpha_1)\}.
\end{aligned}$$

## 2.4.2 Copule de l'indépendance

Supposons que  $U$  et  $V$  sont indépendantes, c'est-à-dire que leur loi conjointe est la copule d'indépendance  $C(u, v) = \Pi(u, v) = uv$ . Dans ce cas, Barbe *et al.* [2] et Genest et Rivest [19] ont déduit que  $k_C(w) = -\ln w$ . Ainsi, puisque  $c(u, v) = 1$  et  $C^{(01)}(u, v) = \partial uv / \partial v = u$ , une application de la Proposition 2.3 amène

$$f_p(u) = \frac{-1}{u \ln q}, \quad u \in [q, 1].$$

Par conséquent, la fonction de répartition est donnée par

$$F_p(u) = \int_q^u \frac{-1}{s \ln q} ds = \frac{\ln(q/u)}{\ln q} = 1 - \frac{\ln u}{\ln q}.$$

On obtient alors  $F_p^{-1}(\alpha) = e^{(1-\alpha)\ln q} = q^{1-\alpha}$  et l'intervalle de l'Équation (2.5) devient  $I_{p,1-\alpha}^{\Pi} = [(1-p)^{1-\alpha_1}, (1-p)^{\alpha_2}]$ . Ainsi, puisque  $\Pi_u^{-1}(1-p) = (1-p)/u$ , l'ensemble  $\mathcal{B}_{p,1-\alpha}^{\Pi}$  est constitué des paires de points  $(u, p/u)$  tel que  $(1-p)^{1-\alpha_1} \leq u \leq (1-p)^{\alpha_2}$ .

### 2.4.3 Copule de Clayton

La copule de Clayton [6] bivariée est définie pour  $\theta \in (0, \infty)$  par  $C_{\theta}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$ . On a d'abord

$$\begin{aligned} C_{\theta}^{(01)}(u, v) &= -\left(\frac{1}{\theta}\right) (-\theta v^{-(\theta+1)}) (u^{-\theta} + v^{-\theta} - 1)^{-(1+\frac{1}{\theta})} \\ &= v^{-(\theta+1)} (u^{-\theta} + v^{-\theta} - 1)^{-(1+\frac{1}{\theta})}. \end{aligned} \quad (2.9)$$

En dérivant ensuite cette expression par rapport à  $u$ , on obtient que la densité est

$$c_{\theta}(u, v) = (\theta + 1)(uv)^{-(\theta+1)} (u^{-\theta} + v^{-\theta} - 1)^{-(\frac{1}{\theta}+2)}. \quad (2.10)$$

On a donc

$$\frac{c_{\theta}(u, v)}{C_{\theta}^{(01)}(u, v)} = (\theta + 1) u^{-(\theta+1)} (u^{-\theta} + v^{-\theta} - 1)^{-1}.$$

Puisque  $C_u^{-1}(q) = (q^{-\theta} - u^{-\theta} + 1)^{-1/\theta}$ , où  $q = 1 - p$ , on trouve

$$\begin{aligned} \frac{c_{\theta}\{u, C_u^{-1}(q)\}}{C_{\theta}^{(01)}\{u, C_u^{-1}(q)\}} &= (\theta + 1) u^{-(\theta+1)} \left[ u^{-\theta} + \left\{ (q^{-\theta} - u^{-\theta} + 1)^{-1/\theta} \right\}^{-\theta} - 1 \right]^{-1} \\ &= (\theta + 1) u^{-(\theta+1)} (u^{-\theta} + q^{-\theta} - u^{-\theta} + 1 - 1)^{-1} \\ &= (\theta + 1) u^{-(\theta+1)} q^{\theta}. \end{aligned}$$

Il s'ensuit que

$$\begin{aligned}
 k_\theta(q) &= \int_q^1 \frac{c\{u, C_u^{-1}(q)\}}{C^{(01)}\{u, C_u^{-1}(q)\}} du \\
 &= (\theta + 1) q^\theta \int_q^1 u^{-(\theta+1)} du \\
 &= \left(\frac{\theta + 1}{\theta}\right) q^\theta (q^{-\theta} - q^0) \\
 &= \left(\frac{\theta + 1}{\theta}\right) (1 - q^\theta).
 \end{aligned}$$

Cette expression correspond à la densité  $k_C$  de la transformation intégrale de probabilité de la copule de Clayton telle qu'obtenue par Genest et Rivest [19]. Comme cas particulier de l'Équation (2.7), on obtient donc

$$\begin{aligned}
 F_p(u) &= \left(\frac{\theta}{\theta + 1}\right) \frac{1}{(1 - q^\theta)} \int_{1-p}^u (\theta + 1) s^{-(\theta+1)} q^\theta ds \\
 &= \frac{\theta q^\theta}{1 - q^\theta} \int_q^u s^{-(\theta+1)} ds \\
 &= \frac{q^\theta}{1 - q^\theta} s^{-\theta} \Big|_{s=u}^{s=q} \\
 &= \frac{q^\theta (q^{-\theta} - u^{-\theta})}{1 - q^\theta} \\
 &= \frac{1 - (q/u)^\theta}{1 - q^\theta}.
 \end{aligned}$$

Quelques manipulations permettent ensuite de montrer que

$$F_p^{-1}(\alpha) = \frac{q}{\{1 - \alpha(1 - q^\theta)\}^{1/\theta}} = \left\{ \frac{q^\theta}{1 - \alpha(1 - q^\theta)} \right\}^{1/\theta}.$$

L'ensemble  $\mathcal{B}_{p,1-\alpha}^C$  est donc constitué de  $(u, (q^{-\theta} - u^{-\theta} + 1)^{-1/\theta})$  tel que

$$\left\{ \frac{q^\theta}{1 - \alpha_1(1 - q^\theta)} \right\}^{1/\theta} \leq u \leq \left\{ \frac{q^\theta}{1 - (1 - \alpha_2)(1 - q^\theta)} \right\}^{1/\theta}.$$

**Remarque 2.2.** *On retrouve l'expression de  $F_p^{-1}$  dans le cas de la copule d'indépen-*

dance lorsque  $\theta \rightarrow 0$ . En effet, si on pose  $h(\theta) = \{1 - \alpha(1 - q^\theta)\}^{1/\theta}$ , alors

$$\log h(\theta) = \frac{\log \{1 - \alpha(1 - q^\theta)\}}{\theta}.$$

Par la règle de l'Hôpital, on obtient

$$\lim_{\theta \rightarrow 0} \log h(\theta) = \lim_{\theta \rightarrow 0} \frac{\alpha(1 - p)^\theta \log q}{1 - \alpha(1 - q^\theta)} = \alpha \log q.$$

Il s'ensuit que  $h(\theta) \rightarrow e^{\alpha \log q} = q^\alpha$  lorsque  $\theta \rightarrow 0$ , et donc

$$\lim_{\theta \rightarrow 0} F_p^{-1}(\alpha) = \lim_{\theta \rightarrow 0} \frac{q}{h(\theta)} = \frac{q}{q^\alpha} = q^{1-\alpha}.$$

#### 2.4.4 Copule de Farlie–Gumbel–Morgenstern

La copule de Farlie-Gumbel-Morgenstern est définie par  $C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v)$ , où  $|\theta| \leq 1$ . On obtient alors facilement  $C_\theta^{(01)}(u, v) = u + \theta u(1 - u)(1 - 2v)$  et  $c_\theta(u, v) = 1 + \theta(1 - 2u)(1 - 2v)$ . On a donc

$$\begin{aligned} \frac{c_\theta(u, v)}{C_\theta^{(01)}(u, v)} &= \frac{1 + \theta(1 - 2u)(1 - 2v)}{u + \theta u(1 - u)(1 - 2v)} \\ &= \frac{1 + \theta(1 - 2u)(1 - 2v)}{u \{1 + \theta(1 - u)(1 - 2v)\}} \\ &= \frac{1 + \theta(1 - u)(1 - 2v) - \theta u(1 - 2v)}{u \{1 + \theta(1 - u)(1 - 2v)\}} \\ &= \frac{1}{u} - \theta \left\{ \frac{1 - 2v}{1 + \theta(1 - u)(1 - 2v)} \right\}. \end{aligned}$$

Afin d'obtenir  $C_{\theta,u}^{-1}(q)$ , on note que  $C(u, v) = q$  si et seulement si  $v + \theta(1 - u)v(1 - v) = q/u$ , ceci pouvant être formulé de façon équivalente par

$$\theta(1 - u)v^2 - \{1 + \theta(1 - u)\}v + \frac{q}{u} = 0. \quad (2.11)$$

Si  $\theta = 0$ , on a  $v = q/u$ , et donc  $C_{0,u}^{-1}(q) = q/u$ . Si  $\theta \neq 0$ , les zéros du polynôme de l'Équation (2.11) sont

$$v = C_{\theta,u}^{-1}(q) = \frac{1 + \theta(1 - u) \pm \ell_{\theta,q}(u)}{2\theta(1 - u)},$$

où  $\ell_{\theta,q}(u) = \sqrt{\{1 + \theta(1 - u)\}^2 - 4\theta(1 - u)q/u}$ . On peut montrer que la seule des deux solutions ci-dessus qui est acceptable, c'est-à-dire telle que  $C_{\theta,u}^{-1}(q) \in [0, 1]$ , est

$$C_{\theta,u}^{-1}(q) = \frac{1 + \theta(1 - u) - \ell_{\theta,q}(u)}{2\theta(1 - u)}.$$

À ce point-ci, il est utile de noter que

$$1 - 2C_{\theta,u}^{-1}(q) = 1 - 2 \left\{ \frac{1 + \theta(1 - u) - \ell_{\theta,q}(u)}{2\theta(1 - u)} \right\} = \frac{-1 + \ell_{\theta,q}(u)}{\theta(1 - u)},$$

ce qui fait que

$$\begin{aligned} \frac{c_\theta \{u, C_{\theta,u}^{-1}(q)\}}{C_\theta^{(01)} \{u, C_{\theta,u}^{-1}(q)\}} &= \frac{1}{u} - \theta \left\{ \frac{1 - 2C_{\theta,u}^{-1}(q)}{1 + \theta(1 - u)(1 - 2C_{\theta,u}^{-1}(q))} \right\} \\ &= \frac{1}{u} - \theta \left[ \frac{\{-1 + \ell_{\theta,q}(u)\}/\theta(1 - u)}{1 + \theta(1 - u)\{-1 + \ell_{\theta,q}(u)\}/\theta(1 - u)} \right] \\ &= \frac{1}{u} - \theta \left\{ \frac{-1 + \ell_{\theta,q}(u)}{\theta(1 - u)\ell_{\theta,q}(u)} \right\} \\ &= \frac{1}{u} + \frac{1 - \ell_{\theta,q}(u)}{(1 - u)\ell_{\theta,q}(u)} \\ &= \frac{1}{u} - \frac{1}{1 - u} + \frac{1}{(1 - u)\ell_{\theta,q}(u)}. \end{aligned}$$

Ceci permet de conclure que pour  $q = 1 - p$ ,

$$F_p(u) = \ln \left\{ \frac{u(1 - u)}{q(1 - q)} \right\} + \int_q^u \frac{ds}{(1 - s)\ell_{\theta,q}(s)}.$$

L'inversion de cette fonction devra se faire de manière numérique.

## 2.5 Projection de l'intervalle sur la courbe de niveau

Volpi et Fiori [37] ont considéré la loi de probabilité de la distance entre  $(X, Y) \in L_p$  et le point référence  $(x_0, y_0) \in L_p$ , où  $x_0 = F_X^{-1}\{\delta_C^{-1}(p)\}$  et  $y_0 = F_Y^{-1}\{\delta_C^{-1}(p)\}$ , avec  $\delta_C(u) = C(u, u)$  la diagonale de la copule  $C$  de  $(X, Y)$ . Dans la suite, on montrera le lien qui existe entre la densité de probabilité  $f_p$  obtenue à la Proposition 2.3 et l'approche de Volpi et Fiori [37]. À cette fin, le résultat suivant sera crucial.

**Lemme 2.2.** *Soit une fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  dérivable en tous points. Alors la distance entre les points  $(a, f(a))$  et  $(b, f(b))$  le long de la courbe définie par  $f$  est donnée par*

$$\mathcal{D}(a, b) = \int_a^b \sqrt{1 + \{f'(x)\}^2} dx.$$

**Démonstration.** Les points  $\mathcal{P}_1 = (a, f(a))$ ,  $\mathcal{P}_2 = (b, f(a))$  et  $\mathcal{P}_3 = (b, f(b))$  forment un triangle rectangle. Ainsi, par une application du Théorème de Pythagore,

$$\ell(\mathcal{P}_1, \mathcal{P}_3) = \sqrt{\ell^2(\mathcal{P}_1, \mathcal{P}_2) + \ell^2(\mathcal{P}_2, \mathcal{P}_3)} = \ell(\mathcal{P}_1, \mathcal{P}_2) \sqrt{1 + \left\{ \frac{\ell(\mathcal{P}_2, \mathcal{P}_3)}{\ell(\mathcal{P}_1, \mathcal{P}_2)} \right\}^2}.$$

En supposant que ces trois longueurs sont infinitésimales, on a alors  $\ell(\mathcal{P}_1, \mathcal{P}_2) = dx$ ,  $\ell(\mathcal{P}_2, \mathcal{P}_3) = dy$  et  $\ell(\mathcal{P}_1, \mathcal{P}_3) = ds$ . On a alors

$$ds = \sqrt{1 + \left( \frac{dy}{dx} \right)^2} dx = \sqrt{1 + \{f'(x)\}^2} dx.$$

On a donc

$$\mathcal{D}(a, b) = \int_a^b ds = \int_a^b \sqrt{1 + \{f'(x)\}^2} dx,$$

ce qui termine la preuve.  $\diamond$

**Exemple 2.2.** *Pour  $f(x) = \beta_0 + \beta_1 x$ , on a  $f'(x) = \beta_1$  et donc la distance sur cette courbe entre  $x = a$  et  $x = b$  est  $\mathcal{D}(a, b) = \int_a^b \sqrt{1 + \beta_1^2} dx = \sqrt{1 + \beta_1^2}(b - a)$ .*

**Exemple 2.3.** *Le Lemme 2.2 permet de calculer la longueur d'un demi-cercle de rayon unitaire dont les points  $(x, y)$  sont unis par la relation  $y^2 + x^2 = 1$ . Le demi-cercle des valeurs positives de  $y$  est donc défini par la fonction  $f(x) = \sqrt{1 - x^2}$ . Comme*

$f'(x) = -x/\sqrt{1-x^2}$ , la distance sur ce demi-cercle entre  $x = -1$  et  $x = 1$  est

$$\mathcal{D}(-1, 1) = \int_{-1}^1 \sqrt{1 + \left(\frac{-x}{\sqrt{1-x^2}}\right)^2} dx = \int_{-1}^1 \sqrt{\frac{1}{1-x^2}} dx = 2 \int_0^1 \frac{1}{\sqrt{1-x^2}} dx.$$

En notant que la primitive de l'intégrant dans la dernière expression est la fonction arc sinus, à savoir  $\sin^{-1}(x)$ ,  $\mathcal{D}(-1, 1) = 2\{\sin^{-1}(1) - \sin^{-1}(0)\} = 2(\pi/2) = \pi$ .

Maintenant, si  $(u, v) \in L_p^C$ , alors on a la relation  $v = C_u^{-1}(q)$ , où  $q = 1 - p$ . En posant  $f(u) = C_u^{-1}(q)$ , on a que  $C\{u, f(u)\} = q$ . En dérivant de chaque côté par rapport à  $u$ , on obtient  $C^{10}\{u, f(u)\} + C^{01}\{u, f(u)\}f'(u) = 0$ , et donc

$$f'(u) = -\frac{C^{10}\{u, f(u)\}}{C^{01}\{u, f(u)\}} = -\frac{C^{10}\{u, C_u^{-1}(q)\}}{C^{01}\{u, C_u^{-1}(q)\}}.$$

Du Lemme 2.2, la distance entre  $u = a$  et  $u = b$  sur la courbe de niveau  $L_p$  est

$$\mathcal{D}_p(a, b) = \int_a^b \sqrt{1 + \left\{ \frac{C^{10}\{u, C_u^{-1}(q)\}}{C^{01}\{u, C_u^{-1}(q)\}} \right\}^2} du.$$

À noter que cette expression a été obtenue indépendamment de Coblenz *et al.* [8]. Si on prend comme repère le point  $(\delta_C^{-1}(q), \delta_C^{-1}(q)) \in L_p^C$ , où  $\delta_C(u) = C(u, u)$ , alors la distance entre  $u \in [q, 1]$  et  $\delta_C^{-1}(q)$  le long de la courbe de niveau  $L_p^C$  est

$$\mathcal{H}_p(u) = \int_u^{\delta_C^{-1}(q)} \sqrt{1 + \left\{ \frac{C^{10}\{s, C_s^{-1}(q)\}}{C^{01}\{s, C_s^{-1}(q)\}} \right\}^2} ds.$$

Par une application du résultat classique concernant la transformation d'une variable aléatoire continue par une fonction monotone, on déduit que la densité de la variable  $S = \mathcal{H}_p(U)$  sur la courbe de niveau  $L_p^C$ , c'est-à-dire sachant  $C(U, V) = 1 - p$ , est

$$h_p(s) = \frac{f_p\{\mathcal{H}_p^{-1}(s)\}}{\mathcal{H}_p\{\mathcal{H}_p^{-1}(s)\}}, \quad \mathcal{H}_p(1) \leq s \leq \mathcal{H}_p(1 - p).$$

# CHAPITRE 3

## COPULES ARCHIMÉDIENNES

### 3.1 Définition de la famille Archimédienne

#### 3.1.1 Expression générale d'une copule Archimédienne

La famille des copules Archimédiennes comprend un grand nombre de modèles de dépendance, incluant les copules de Clayton [6], Frank [14] et Gumbel [22]. Une copule bivariée  $C$  est dite Archimédienne si elle peut s'écrire sous la forme

$$C_\Psi(u, v) = \Psi \left\{ \Psi^{-1}(u) + \Psi^{-1}(v) \right\}, \quad (3.1)$$

où  $\Psi : [0, \infty) \rightarrow [0, 1]$  s'appelle le *générateur*. Pour que l'expression (3.1) défuisse bel et bien une copule, il faut que  $\Psi$  soit une fonction décroissante et convexe, à savoir que pour tout  $t \in [0, \infty)$ , on a

$$\Psi'(t) < 0 \quad \text{et} \quad \Psi''(t) > 0.$$

### 3.1.2 Courbe de niveau

Il est facile de montrer que la courbe de niveau d'une copule Archimédienne est déterminée pour  $q = 1 - p$  par l'équation

$$C_u^{-1}(q) = \Psi \left\{ \Psi^{-1}(q) - \Psi^{-1}(u) \right\}.$$

Tel qu'énoncé au Théorème 4.3.2 de Nelsen [27], les courbes de niveau d'une copule Archimédienne sont convexes, c'est-à-dire que l'application  $q \mapsto C_u^{-1}(q)$  est convexe.

### 3.1.3 Tau de Kendall

Une façon de mesurer la dépendance entre deux variables aléatoires a été proposée par Kendall [23]. Ainsi, sur la base d'un échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  de paires, le tau de Kendall empirique est défini par

$$\tau_n = \frac{2}{n(n-1)} \sum_{i < j} \{C_{ij} - D_{ij}\},$$

où  $C_{ij} = I\{(X_i - X_j)(Y_i - Y_j) > 0\}$  et  $D_{ij} = I\{(X_i - X_j)(Y_i - Y_j) < 0\}$  sont les indicateurs, respectivement, de paires concordantes et discordantes. Le tau de Kendall empirique correspond ainsi à la différence entre les proportions de paires concordantes et discordantes dans un jeu de données bivariées. Il s'agit d'une alternative non-paramétrique au célèbre coefficient de corrélation de Pearson.

Il appert que la version théorique du tau de Kendall ne dépend pas des marges d'une population bivariée, mais seulement de sa copule (en autant que les marges sont continues). En fait, en se référant par exemple à Nelsen [27], le tau de Kendall d'une

paire aléatoire  $(X, Y)$  dont la copule est  $C$  s'exprime via la formule

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

Dans le cas d'une copule Archimédienne générée par  $\Psi$ , on a

$$\tau_\Psi = 1 - 4 \int_0^{\Psi^{-1}(0)} s \{ \Psi'(s) \}^2 \, ds.$$

## 3.2 Expression générale de l'ensemble de confiance

### 3.2.1 Détermination de $C_\Psi^{(01)}$ et $c_\Psi$

Afin de simplifier les expressions à venir, posons  $\phi = \Psi^{-1}$ . On obtient alors

$$C_\Psi^{(01)}(u, v) = \frac{\partial}{\partial v} C_\Psi(u, v) = \Psi' \{ \Psi^{-1}(u) + \Psi^{-1}(v) \} \phi'(v).$$

En dérivant cette dernière expression par rapport à  $u$ , on déduit que la densité de  $C_\Psi$  s'exprime par

$$c_\Psi(u, v) = \frac{\partial^2}{\partial u \partial v} C_\Psi(u, v) = \Psi'' \{ \Psi^{-1}(u) + \Psi^{-1}(v) \} \phi'(u) \phi'(v).$$

### 3.2.2 Détermination de $k_\Psi$

À partir des travaux de Genest et Rivest [18], Barbe *et al.* [2] et Genest et Rivest [19], on sait que la transformation intégrale de probabilité d'une copule Archimédienne s'exprime par

$$K_\Psi(w) = w - \frac{\phi(w)}{\phi'(w)}, \quad \text{où } \phi = \Psi^{-1}.$$

En dérivant, on déduit alors que la densité de  $K_\Psi$  est

$$k_\Psi(w) = \frac{\phi(w)\phi''(w)}{\{\phi'(w)\}^2}.$$

En notant que

$$\phi'(w) = \frac{d}{dw} \Psi^{-1}(w) = \frac{1}{\Psi'\{\phi(w)\}} \quad \text{et} \quad \phi''(w) = \frac{d^2}{dw^2} \Psi^{-1}(w) = -\frac{\Psi''\{\phi(w)\}}{\left(\Psi'\{\phi(w)\}\right)^3},$$

on peut écrire de façon équivalente que

$$k_\Psi(w) = -\frac{\phi(w)\Psi''\{\phi(w)\}}{\left(\Psi'\{\phi(w)\}\right)^3} \times (\Psi'\{\phi(w)\})^2 = -\frac{\phi(w)\Psi''\{\phi(w)\}}{\Psi'\{\phi(w)\}}. \quad (3.2)$$

### 3.2.3 Distribution de $U$ sachant $C_\Psi(U, V)$

Le résultat suivant établit la forme de la fonction de répartition et de la densité conditionnelles de  $U$  sachant  $W = C_\Psi(U, V)$ . Il s'agit d'un corollaire à la Proposition 2.2 dans le cas où  $C$  est une copule Archimédienne.

**Corollaire 3.1.** *Si  $C_\Psi$  est une copule Archimédienne bivariée dont le générateur est  $\Psi$ , alors la densité conditionnelle de  $X = U$  étant donné  $C_\Psi(U, V) = 1 - p$  et sa fonction de répartition sont données respectivement par*

$$f_p(u) = -\frac{\phi'(u)}{\phi(1-p)} \quad \text{et} \quad F_p(u) = 1 - \frac{\phi(u)}{\phi(1-p)}, \quad u \in [1-p, 1].$$

**Démonstration.** On remarque d'abord que

$$\frac{c_\Psi(u, v)}{C_\Psi^{(01)}(u, v)} = \frac{\Psi''\{\Psi^{-1}(u) + \Psi^{-1}(v)\}}{\Psi'\{\Psi^{-1}(u) + \Psi^{-1}(v)\}} \phi'(u).$$

Ensuite, puisque  $C_u^{-1}(q) = \Psi\{\phi(q) - \phi(u)\}$ , où  $q = 1 - p$ , on a

$$\begin{aligned} \frac{c_\Psi\{u, C_u^{-1}(q)\}}{C_\Psi^{(01)}\{u, C_u^{-1}(q)\}} &= \frac{\Psi''\{\Psi^{-1}(u) + \Psi^{-1}(\Psi\{\phi(q) - \phi(u)\})\}}{\Psi'\{\Psi^{-1}(u) + \Psi^{-1}(\Psi\{\phi(q) - \phi(u)\})\}} \phi'(u) \\ &= \frac{\Psi''\{\phi(u) + \phi(q) - \phi(u)\}}{\Psi'\{\phi(u) + \phi(q) - \phi(u)\}} \phi'(u) \\ &= \frac{\Psi''\{\phi(q)\}}{\Psi'\{\phi(q)\}} \phi'(u). \end{aligned}$$

Maintenant, en se basant sur l'Équation (3.2), on obtient comme cas particulier de la Proposition 2.1 que

$$f_p(u) = \frac{-\Psi'\{\phi(q)\}}{\phi(q) \Psi''\{\phi(q)\}} \frac{\Psi''\{\phi(q)\}}{\Psi'\{\phi(q)\}} \phi'(u) = -\frac{\phi'(u)}{\phi(q)} \quad u \in [q, 1].$$

La fonction de répartition de  $f_p$  est alors donnée par

$$F_p(u) = -\frac{1}{\phi(q)} \int_q^u \phi'(s) \, ds = \frac{\phi(q) - \phi(u)}{\phi(q)} = 1 - \frac{\phi(u)}{\phi(q)},$$

ce qui termine la preuve.  $\diamond$

Il est facile de montrer que la densité  $f_p$  est décroissante sur  $[1 - p, 1]$ . En effet,

$$f'_p(u) = -\frac{\phi''(u)}{\phi(1 - p)} < 0,$$

car par hypothèse,  $\phi$  est convexe et donc  $\phi''(u) > 0$  pour tout  $u \in [0, 1]$ . Par conséquent, le mode de la loi conditionnelle de  $U$  sachant  $C_\Psi(U, V) = 1 - p$  est  $M_0 = 1 - p$ . On montre aisément que la moyenne conditionnelle de  $U$  est

$$\mathbb{E}\{U | C_\Psi(U, V) = 1 - p\} = \int_{1-p}^1 u f_p(u) \, du = 1 - p + \frac{1}{\phi(1 - p)} \int_{1-p}^1 \phi(u) \, du.$$

Par exemple, dans le cas de la copule d'indépendance,  $\phi(u) = -\ln u$  et donc

$$\begin{aligned}\mathbb{E}(U|UV = 1-p) &= 1-p + \frac{1}{\ln(1-p)} \int_{1-p}^1 \ln u \, du \\ &= 1-p + \frac{1}{\ln(1-p)} (u \ln u - u)_{1-p}^1 \\ &= \frac{-p}{\ln(1-p)}.\end{aligned}$$

### 3.2.4 Ensemble de confiance

Le Corollaire 3.1 permet de déduire l'intervalle de confiance de niveau  $1 - \alpha$  sur la courbe de niveau  $L_p$  dans le cas d'une copule Archimédienne générée par  $\Psi$ . En effet, puisque  $F_p^{-1}(\alpha) = \phi^{-1}\{(1-\alpha)\phi(1-p)\} = \Psi\{(1-\alpha)\phi(1-p)\}$ , on obtient facilement

$$I_{p,1-\alpha}^\Psi = [\Psi\{(1-\alpha_1)\phi(1-p)\}, \Psi\{\alpha_2\phi(1-p)\}],$$

où  $\alpha_1, \alpha_2 \geq 0$  sont tels que  $\alpha_1 + \alpha_2 = \alpha$ .

## 3.3 Modèles Archimédiens particuliers

### 3.3.1 Copule de Clayton

La copule de Clayton de paramètre  $\theta > 0$  est générée par  $\Psi_\theta(t) = (\theta t + 1)^{-1/\theta}$ , de sorte que  $\phi_\theta(u) = (u^{-\theta} - 1)/\theta$ . On trouve, pour  $q = 1 - p$ ,

$$F_p(u) = 1 - \left( \frac{u^{-\theta} - 1}{q^{-\theta} - 1} \right) = \frac{q^{-\theta} - 1 - u^{-\theta} + 1}{q^{-\theta} - 1} = \frac{u^{-\theta} - q^{-\theta}}{1 - q^{-\theta}}.$$

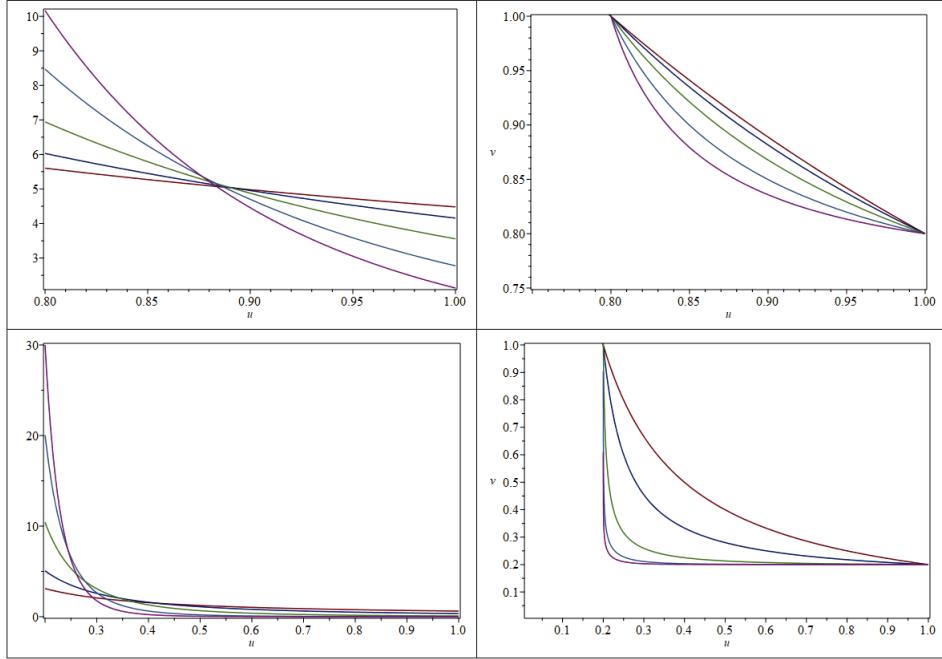


FIGURE 3.1 – Densité  $f_p$  (à gauche) et courbe de niveau (à droite) de la copule de Clayton lorsque  $p = 0, 2$  (en haut) et  $p = 0, 8$  (en bas) lorsque  $\tau = 0$  (rouge),  $\tau = 1/4$  (bleu foncé),  $\tau = 1/2$  (vert),  $\tau = 2/3$  (bleu pâle) et  $\tau = 3/4$  (magenta)

De là,  $F_p^{-1}(\alpha) = \{q^{-\theta} + \alpha(1 - q^{-\theta})\}^{-1/\theta}$ . L'intervalle de confiance est donc le même que celui obtenu à la Section 2.4.3.

La Figure 3.1 montre la densité et la courbe de niveau de la copule de Clayton lorsque  $p = 0, 2$  (figures du haut) et  $p = 0, 8$  (figures du bas) pour des valeurs du tau de Kendall correspondant à  $\tau \in \{0, 1/4, 1/2, 2/3, 3/4\}$ . On obtient les valeurs de  $\theta$  qui correspondent à ces niveaux de dépendance via la formule  $\tau(\theta) = \theta/(\theta+2)$ . On obtient donc  $\theta(\tau) = 2\tau/(1 - \tau)$ . À noter que la densité et la courbe de niveau s'expriment respectivement par les formules

$$f_p(u) = \frac{\theta q^\theta u^{-\theta-1}}{1 - q^\theta} \quad \text{et} \quad C_u^{-1}(q) = (q^{-\theta} - u^{-\theta} + 1)^{-1/\theta}.$$

### 3.3.2 Copule de Frank

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

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

En posant  $\ell(u) = e^{-u} - 1$ , on montre que  $\phi_\theta(u) = -\ln\{\ell(\theta u)/\ell(\theta)\}$ . Quelques calculs permettent alors d'obtenir

$$F_p(u) = \ln \left\{ \frac{\ell(\theta q)}{\ell(\theta u)} \right\} \Big/ \ln \left\{ \frac{\ell(\theta q)}{\ell(\theta)} \right\}.$$

En posant  $F_p(u) = \alpha$ , on déduit alors que

$$\ell(\theta u) = \left\{ \frac{\ell(\theta)}{\ell(\theta q)} \right\}^\alpha \ell(\theta q) = \frac{\{\ell(\theta)\}^\alpha}{\{\ell(\theta q)\}^{\alpha-1}}.$$

Comme  $\ell^{-1}(t) = -\ln(1+t)$ , il s'ensuit que

$$F_p^{-1}(\alpha) = \frac{-1}{\theta} \ln \left( 1 + \frac{\{\ell(\theta)\}^\alpha}{\{\ell(\theta q)\}^{\alpha-1}} \right).$$

L'intervalle de confiance est ainsi donné par

$$I_{p,1-\alpha}^{\text{Fr}} = \left[ \frac{-1}{\theta} \ln \left( 1 + \frac{\{\ell(\theta)\}^{\alpha_1}}{\{\ell(\theta q)\}^{\alpha_1-1}} \right), \frac{-1}{\theta} \ln \left( 1 + \frac{\{\ell(\theta)\}^{1-\alpha_2}}{\{\ell(\theta q)\}^{-\alpha_2}} \right) \right].$$

La Figure 3.2 montre la densité et la courbe de niveau de la copule de Frank lorsque  $p = 0,2$  (figures du haut) et  $p = 0,8$  (figures du bas) pour des valeurs du tau de Kendall correspondant à  $\tau \in \{0, 1/4, 1/2, 2/3, 3/4\}$ . On obtient les valeurs de  $\theta$  qui correspondent à ces niveaux de dépendance via

$$\tau(\theta) = 1 + \frac{4}{\theta} \{D_1(\theta) - 1\},$$

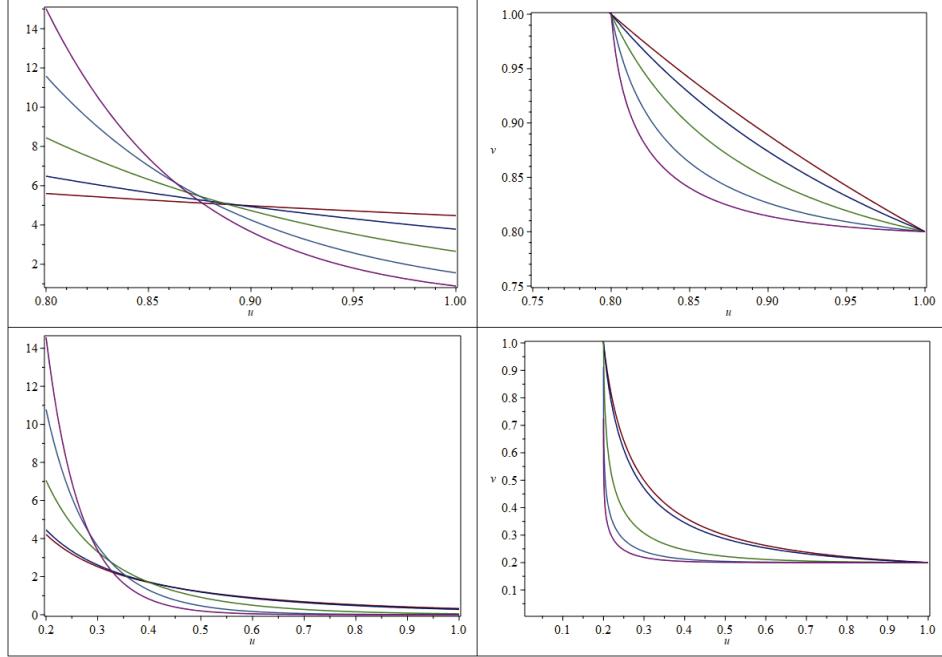


FIGURE 3.2 – Densité  $f_p$  (à gauche) et courbe de niveau (à droite) de la copule de Frank lorsque  $p = 0, 2$  (en haut) et  $p = 0, 8$  (en bas) lorsque  $\tau = 0$  (rouge),  $\tau = 1/4$  (bleu foncé),  $\tau = 1/2$  (vert),  $\tau = 2/3$  (bleu pâle) et  $\tau = 3/4$  (magenta)

où  $D_1(\theta) = \theta^{-1} \int_0^\theta t(e^t - 1)^{-1} dt$  est la fonction de Debye d'ordre un. À noter que la densité et la courbe de niveau s'expriment respectivement par les formules

$$f_p(u) = \frac{\theta e^{-\theta u}}{\ell(\theta u) \ln\{\ell(\theta q)/\ell(\theta)\}} \quad \text{et} \quad C_u^{-1}(q) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{\ell(\theta q)\ell(\theta)}{\ell(\theta u)} \right\}.$$

### 3.3.3 Copule de Gumbel

La copule de Gumbel est générée par  $\Psi(t) = \exp(-t^{1-\theta})$ , où  $\theta \in [0, 1]$ . On a donc  $\phi(u) = (-\ln u)^{1/(1-\theta)}$ , ce qui fait que pour  $q = 1 - p$ ,

$$F_p(u) = 1 - \frac{(-\ln u)^{1/(1-\theta)}}{(-\ln q)^{1/(1-\theta)}} = 1 - \left( \frac{\ln u}{\ln q} \right)^{1/(1-\theta)}.$$

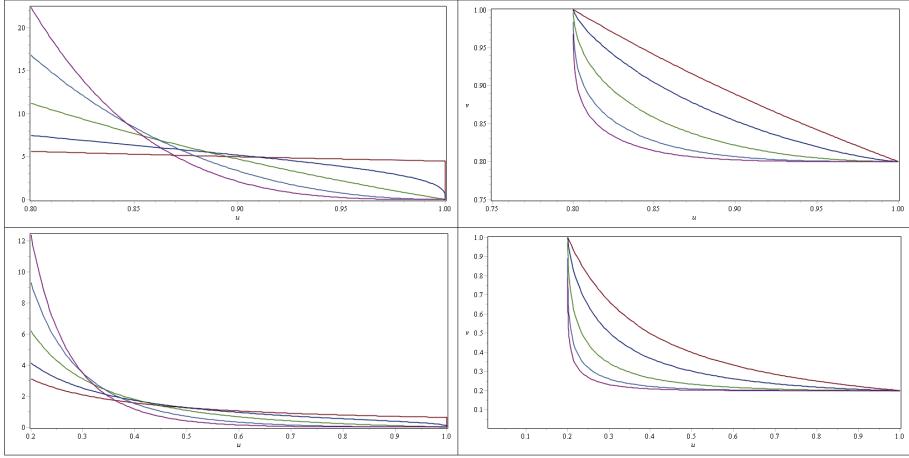


FIGURE 3.3 – Densité  $f_p$  (à gauche) et courbe de niveau (à droite) de la copule de Gumbel lorsque  $p = 0, 2$  (en haut) et  $p = 0, 8$  (en bas) lorsque  $\tau = 0$  (rouge),  $\tau = 1/4$  (bleu foncé),  $\tau = 1/2$  (vert),  $\tau = 2/3$  (bleu pâle) et  $\tau = 3/4$  (magenta)

On a alors que  $F_p(u) = \alpha$  si et seulement si  $\ln u = (1 - \alpha)^{1-\theta} \ln q$ , ce qui amène

$$F_p^{-1}(\alpha) = \exp \{ (1 - \alpha)^{1-\theta} \ln q \} = q^{(1-\alpha)^{1-\theta}}.$$

L'intervalle de confiance est alors donné par

$$I_{p,1-\alpha}^{\text{Gu}} = \left[ (1 - p)^{(1-\alpha_1)^{1-\theta}}, (1 - p)^{\alpha_2^{1-\theta}} \right].$$

La Figure 3.3 montre la densité et la courbe de niveau de la copule de Gumbel lorsque  $p = 0, 2$  (figures du haut) et  $p = 0, 8$  (figures du bas) pour des valeurs du tau de Kendall correspondant à  $\tau \in \{0, 1/4, 1/2, 2/3, 3/4\}$ . On obtient les valeurs de  $\theta$  qui correspondent à ces niveaux de dépendance via la formule  $\tau(\theta) = \theta$ , ainsi  $\theta(\tau) = \tau$ . À noter que la densité et la courbe de niveau s'expriment respectivement par les formules

$$f_p(u) = \frac{(\ln u / \ln q)^{\theta/(1-\theta)}}{u(\theta - 1) \ln q} \quad \text{et} \quad C_u^{-1}(q) = \exp \left[ - \{ (-\ln q)^{1/(1-\theta)} - (-\ln u)^{1/(1-\theta)} \}^{1-\theta} \right].$$

### 3.3.4 Copule de Joe

La copule de Joe est générée par  $\Psi(t) = 1 - (1 - e^{-t})^{1/\theta}$ , où  $\theta \in [0, 1]$ . On a donc  $\phi(u) = -\ln\{1 - (1 - u)^\theta\}$ , ce qui fait que pour  $q = 1 - p$ ,

$$F_p(u) = 1 - \frac{\ln\{1 - (1 - u)^\theta\}}{\ln\{1 - (1 - q)^\theta\}}.$$

On a alors que  $F_p(u) = \alpha$  si et seulement si  $\ln\{1 - (1 - u)^\theta\} = (1 - \alpha) \ln\{1 - (1 - q)^\theta\}$ , ce qui amène

$$F_p^{-1}(\alpha) = 1 - \{1 - (1 - p^\theta)^{1-\alpha}\}^{1/\theta}.$$

L'intervalle de confiance est alors donné par

$$I_{p,1-\alpha}^{\text{Joe}}(p) = \left[ 1 - \{1 - (1 - p^\theta)^{1-\alpha_1}\}^{1/\theta}, 1 - \{1 - (1 - p^\theta)^{\alpha_2}\}^{1/\theta} \right].$$

La densité et la courbe de niveau s'expriment respectivement par les formules

$$f_p(u) = \frac{\theta(1 - u)^{\theta-1}}{(1 - (1 - u)^\theta) \ln(1 - (1 - u)^\theta)} \quad \text{et} \quad C_u^{-1}(q) = 1 - \left(1 - \frac{1 - (1 - q)^\theta}{1 - (1 - u)^\theta}\right)^{1/\theta}.$$

### 3.3.5 Copule de Ali–Mikhail–Haq

La copule de Ali–Mikhail–Haq est générée par  $\Psi(t) = \frac{1-\theta}{e^t-\theta}$ , où  $\theta \in [0, 1]$ . On a donc  $\phi(u) = \ln\{(1 - \theta)u^{-1} + \theta\}$ , tel que pour  $q = 1 - p$ , on obtient

$$F_p(u) = 1 - \frac{\ln\{(1 - \theta)u^{-1} + \theta\}}{\ln\{(1 - \theta)q^{-1} + \theta\}}.$$

En posant  $\ell_{q,\theta} = (1 - \theta)q^{-1} + \theta$ , on aura  $F_p(u) = \alpha$  si et seulement si  $\ell_{q,\theta}^{1-\alpha} = (1 - \theta)u^{-1} + \theta$ . De là, on déduit aisément

$$F_p^{-1}(\alpha) = (1 - \theta)/(\ell_{q,\theta}^{1-\alpha} - \theta).$$

L'intervalle de confiance est alors donné par

$$I_{p,1-\alpha}^{\text{AMQ}}(p) = \left[ \frac{1 - \theta}{\ell_{q,\theta}^{1-\alpha_1} - \theta}, \frac{1 - \theta}{\ell_{q,\theta}^{\alpha_2} - \theta} \right].$$

La densité et la courbe de niveau s'expriment respectivement par les formules

$$f_p(u) = \frac{1 - \theta}{u^2 \{(1 - \theta)u^{-1} + \theta\} \ln \{(1 - \theta)u^{-1} + \theta\}} \quad \text{et} \quad C_u^{-1}(q) = \frac{1 - \theta}{\left\{ \frac{(1 - \theta)u^{-1} + \theta}{(1 - \theta)u^{-1} + \theta} \right\} + \theta}.$$

# CHAPITRE 4

## COPULES NORMALE ET STUDENT

### 4.1 Copule Normale

#### 4.1.1 Quelques préliminaires

Soient la densité et la fonction de répartition de la loi normale  $\mathbb{N}(0, 1)$ , à savoir

$$\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad \text{et} \quad \Phi(x) = \int_{-\infty}^x \phi(t) dt.$$

Soit également la densité normale bivariée de corrélation  $\rho \in [-1, 1]$  donnée par

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

ainsi que la fonction de répartition associée à  $\phi_\rho$ , qui s'écrit implicitement

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

Avec cette notation, la copule Normale s'exprime par

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

### 4.1.2 Obtention du quotient $c/C^{(01)}$

Premièrement, en dérivant  $C_\rho$  par rapport à  $v$ , on obtient

$$C_\rho^{(01)}(u, v) = \Phi_\rho^{(01)} \{ \Phi^{-1}(u), \Phi^{-1}(v) \} (\Phi^{-1}(v))', \quad (4.1)$$

où  $\Phi_\rho^{(01)}(x, y) = \partial \Phi_\rho(x, y) / \partial y$ . En dérivant maintenant l'expression (4.1) par rapport à  $u$ , on déduit la densité de la copule Normale, à savoir

$$c_\rho(u, v) = \phi_\rho \{ \Phi^{-1}(u), \Phi^{-1}(v) \} (\Phi^{-1}(u))' (\Phi^{-1}(v))'.$$

On a alors

$$\frac{c_\rho(u, v)}{C_\rho^{(01)}(u, v)} = \frac{\phi_\rho \{ \Phi^{-1}(u), \Phi^{-1}(v) \} (\Phi^{-1}(u))'}{\Phi_\rho^{(01)} \{ \Phi^{-1}(u), \Phi^{-1}(v) \}} = \frac{\phi_\rho \{ \Phi^{-1}(u), \Phi^{-1}(v) \}}{\phi \{ \Phi^{-1}(u) \} \Phi_\rho^{(01)} \{ \Phi^{-1}(u), \Phi^{-1}(v) \}},$$

où la deuxième égalité exploite le fait que  $dh^{-1}(x) = 1/h'(h^{-1}(x))$ . On a donc

$$\frac{c_\rho \{ \Phi(x), \Phi(y) \}}{C_\rho^{(01)} \{ \Phi(x), \Phi(y) \}} = \frac{\phi_\rho(x, y)}{\phi(x) \Phi_\rho^{(01)}(x, y)}. \quad (4.2)$$

À noter maintenant que si  $(Z_1, Z_2)$  est un couple dont la densité est  $\phi_\rho$ , alors

$$\mathbb{P}(Z_1 \leq x | Z_2 = y) = \frac{\partial \Phi_\rho(x, y) / \partial y}{\phi(y)} = \frac{\Phi_\rho^{(01)}(x, y)}{\phi(y)}.$$

D'un résultat classique sur les lois conditionnelles Normales que l'on peut retrouver par exemple dans Castillo et Galambos [5], on sait que la distribution conditionnelle de  $Z_1$  sachant  $Z_2 = y$  est la loi  $\mathbb{N}(\rho y, 1 - \rho^2)$ . On en déduit alors que

$$\begin{aligned} \Phi_\rho^{(01)}(x, y) &= \phi(y) \mathbb{P}(Z_1 \leq x | Z_2 = y) \\ &= \phi(y) \mathbb{P} \left( \frac{Z_1 - \rho y}{\sqrt{1 - \rho^2}} \leq \frac{x - \rho y}{\sqrt{1 - \rho^2}} \middle| Z_2 = y \right) \\ &= \phi(y) \Phi \{ z_\rho(x, y) \}, \end{aligned}$$

où on a posé  $z_\rho(x, y) = (x - \rho y) / \sqrt{1 - \rho^2}$ . En injectant cette expression dans la formule (4.2), on trouve donc

$$\frac{c_\rho \{\Phi(x), \Phi(y)\}}{C_\rho^{(01)} \{\Phi(x), \Phi(y)\}} = \frac{1}{\phi(x)} \frac{\phi_\rho(x, y)}{\phi(y)} \frac{1}{\Phi \{z_\rho(x, y)\}}.$$

Avant de poursuivre, soit le lemme suivant.

**Lemme 4.3.** *Si  $\phi$  est la densité de la loi Normale standard et  $\phi_\rho$  la densité de la loi Normale bivariée de moyennes nulles, de variances unitaires et de corrélation  $\rho$ , alors*

$$\frac{\phi_\rho(x, y)}{\phi(y)} = \frac{1}{\sqrt{1 - \rho^2}} \phi \{z_\rho(x, y)\},$$

Le Lemme 4.3 permet d'établir que

$$\frac{c_\rho \{\Phi(x), \Phi(y)\}}{C_\rho^{(01)} \{\Phi(x), \Phi(y)\}} = \frac{1}{\sqrt{1 - \rho^2}} \frac{1}{\phi(x)} \frac{\phi \{z_\rho(x, y)\}}{\Phi \{z_\rho(x, y)\}}.$$

On a donc

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

**Remarque 4.3.** *Dans le cas particulier où  $\rho = 0$ , on a  $z_\rho(x, y) = x$  et donc*

$$\frac{c_\rho \{\Phi(x), \Phi(y)\}}{C_\rho^{(01)} \{\Phi(x), \Phi(y)\}} = \frac{1}{\Phi(x)}.$$

On obtient alors facilement

$$\frac{c_\rho(u, v)}{C_\rho^{(01)}(u, v)} = \frac{1}{u}.$$

### 4.1.3 Obtention numérique de l'intervalle de confiance

On a

$$\begin{aligned}
 C_u^{-1}(q) &= \{v | C_\rho^N(u, v) = q\} \\
 &= \{v | \Phi_\rho \{\Phi^{-1}(u), \Phi^{-1}(v)\} = q\} \\
 &= \{\Phi(y) | \Phi_\rho \{\Phi^{-1}(u), y\} = q\} \\
 &= \Phi \left[ \Phi_{\rho, \Phi^{-1}(u)}^{-1}(q) \right].
 \end{aligned}$$

Ainsi,  $C_{\Phi(u)}^{-1}(q) = \Phi \left[ \Phi_{\rho, u}^{-1}(q) \right]$ . On cherche donc à résoudre, pour  $s$  et  $q$  fixés,  $\Phi_\rho(s, y) = q$ , ou encore  $\int_{-\infty}^y \phi(t) \Phi \left( \frac{s - \rho t}{\sqrt{1 - \rho^2}} \right) dt = q$ .

Le tau de Kendall est donné par  $\tau_{C_\rho^N}(\rho) = \frac{2}{\pi} \sin^{-1}(\rho)$ , permettant d'obtenir  $\rho_{C_\tau^N}(\tau) = \sin \left( \frac{\pi}{2} \tau \right)$ . On obtient, à partir du quotient  $\frac{c_\rho(u, v)}{C_\rho^{(01)}(u, v)}$  que  $f_p(x) = \frac{c\{s, C_s^{-1}(q)\}}{C^{(01)}\{s, C_s^{-1}(q)\}} \Big/ \int_q^1 \frac{c\{s, C_s^{-1}(q)\}}{C^{(01)}\{s, C_s^{-1}(q)\}} ds$ . La fonction de répartition est  $F_p(u) = \int_q^u \frac{c\{s, C_s^{-1}(q)\}}{C^{(01)}\{s, C_s^{-1}(q)\}} ds \Big/ \int_q^1 \frac{c\{s, C_s^{-1}(q)\}}{C^{(01)}\{s, C_s^{-1}(q)\}} ds$ . On obtient numériquement l'intervalle

$$I_C = [F^{-1}(\alpha_1); F^{-1}(1 - \alpha_2)].$$

## 4.2 Copule de Student

### 4.2.1 Quelques préliminaires

La densité et la fonction de répartition de la loi Student à  $\nu \in \mathbb{N}$  degrés de liberté sont données respectivement par les expressions

$$\omega_\nu(x) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\sqrt{\nu\pi} \Gamma \left( \frac{\nu}{2} \right)} \left( 1 + \frac{x^2}{\nu} \right)^{-\left( \frac{\nu+1}{2} \right)} \quad \text{et} \quad \Omega_\nu(x) = \int_{-\infty}^x \omega_\nu(t) dt, \quad (4.3)$$

où  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$  est la fonction Gamma. Soit également la densité de la loi de Student bivariée de paramètre  $\rho \in [-1, 1]$  donnée par

$$\omega_{\nu, \rho}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \left( 1 + \frac{x^2 - 2\rho xy + y^2}{\nu\sqrt{1-\rho^2}} \right)^{-(\frac{\nu}{2}+1)}. \quad (4.4)$$

À noter qu'à l'instar de la loi Normale bivariée, la fonction de répartition associée à  $\omega_{\nu, \rho}$  s'écrit implicitement

$$\Omega_{\nu, \rho}(x, y) = \int_{-\infty}^x \int_{-\infty}^y \omega_{\nu, \rho}(s, t) dt ds.$$

Avec cette notation, la copule Student s'exprime par

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

#### 4.2.2 Obtention du quotient $c/C^{(01)}$

En dérivant la copule de Student  $C_{\rho, \nu}$  par rapport à  $v$ , on obtient

$$C_{\rho, \nu}^{(01)}(u, v) = \Omega_{\rho, \nu}^{(01)} \{ \Omega_\nu^{-1}(u), \Omega_\nu^{-1}(v) \} (\Omega_\nu^{-1}(v))',$$

où  $\Omega_{\rho, \nu}^{(01)}(x, y) = \partial \Omega_{\rho, \nu}(x, y) / \partial y$ . En dérivant cette dernière expression par rapport à  $u$ , on déduit la densité de la copule Student, c'est-à-dire

$$c_{\rho, \nu}(u, v) = \omega_{\rho, \nu} \{ \Omega_\nu^{-1}(u), \Omega_\nu^{-1}(v) \} (\Omega_\nu^{-1}(u))' (\Omega_\nu^{-1}(v))'.$$

On a alors que le quotient  $c/C^{(01)}$  s'exprime par

$$\begin{aligned} \frac{c_{\rho,\nu}(u, v)}{C_{\rho,\nu}^{(01)}(u, v)} &= \frac{\omega_{\rho,\nu}\{\Omega_{\nu}^{-1}(u), \Omega_{\nu}^{-1}(v)\}(\Omega_{\nu}^{-1}(u))'}{\Omega_{\rho,\nu}^{(01)}\{\Omega_{\nu}^{-1}(u), \Omega_{\nu}^{-1}(v)\}} \\ &= \frac{\omega_{\rho,\nu}\{\Omega_{\nu}^{-1}(u), \Omega_{\nu}^{-1}(v)\}}{\omega_{\nu}\{\Omega_{\nu}^{-1}(u)\}\Omega_{\rho,\nu}^{(01)}\{\Omega_{\nu}^{-1}(u), \Omega_{\nu}^{-1}(v)\}}. \end{aligned}$$

Par conséquent,

$$\frac{c_{\rho,\nu}\{\Omega_{\nu}(x), \Omega_{\nu}(y)\}}{C_{\rho,\nu}^{(01)}\{\Omega_{\nu}(x), \Omega_{\nu}(y)\}} = \frac{\omega_{\rho,\nu}(x, y)}{\omega_{\nu}(x)\Omega_{\rho,\nu}^{(01)}(x, y)}. \quad (4.5)$$

Maintenant, tel que démontré par Ding [9], si  $(Z_1, Z_2)$  est un couple de loi de Student a  $\nu$  degrés de liberté, alors

$$\mathbb{P}\left(\frac{\sqrt{\nu+1}(Z_1 - \rho y)}{\sqrt{(\nu+y^2)(1-\rho^2)}} \leq x \middle| Z_2 = y\right) = \Omega_{\nu+1}(x).$$

Ce résultat sur la loi conditionnelle de la Student bivariée permet de déduire

$$\begin{aligned} \mathbb{P}(Z_1 \leq x | Z_2 = y) &= \mathbb{P}\left(\frac{\sqrt{\nu+1}(Z_1 - \rho y)}{\sqrt{(\nu+y^2)(1-\rho^2)}} \leq \frac{\sqrt{\nu+1}(x - \rho y)}{\sqrt{(\nu+y^2)(1-\rho^2)}} \middle| Z_2 = y\right) \\ &= \Omega_{\nu+1}\{z_{\rho,\nu}(x, y)\}, \end{aligned}$$

où on a posé

$$z_{\rho,\nu}(x, y) = \sqrt{\frac{\nu+1}{\nu+y^2}} \frac{x - \rho y}{\sqrt{1-\rho^2}}.$$

À partir de là, on obtient

$$\Omega_{\rho,\nu}^{(01)}(x, y) = \omega_{\nu}(y) \mathbb{P}(Z_1 \leq x | Z_2 = y) = \omega_{\nu}(y) \Omega_{\nu+1}\{z_{\rho,\nu}(x, y)\}.$$

En injectant cette dernière expression dans la formule (4.5), on trouve

$$\frac{c_{\rho,\nu}\{\Omega_{\nu}(x), \Omega_{\nu}(y)\}}{C_{\rho,\nu}^{(01)}\{\Omega_{\nu}(x), \Omega_{\nu}(y)\}} = \frac{1}{\omega_{\nu}(x)} \frac{\omega_{\rho,\nu}(x, y)}{\omega_{\nu}(y)} \frac{1}{\Omega_{\nu+1}\{z_{\rho,\nu}(x, y)\}}.$$

Le résultat suivant est une version du Lemme 4.3 adapté à la copule de Student.

**Lemme 4.4.** *Si  $\omega_\nu$  est la densité de la loi de Student à  $\nu$  degrés de liberté et  $\omega_{\rho,\nu}$  la densité de la loi de Student bivariée à  $\nu$  degrés de liberté et de corrélation  $\rho$ , alors*

$$\frac{\omega_{\rho,\nu}(x, y)}{\omega_\nu(y)} = \frac{1}{\sqrt{1 - \rho^2}} \sqrt{\frac{\nu + 1}{\nu + y^2}} \omega_{\nu+1} \{z_{\rho,\nu}(x, y)\}. \quad (4.6)$$

Une application du Lemme 4.4 amène alors

$$\frac{c_{\rho,\nu} \{\Omega_\nu(x), \Omega_\nu(y)\}}{C_{\rho,\nu}^{(01)} \{\Omega_\nu(x), \Omega_\nu(y)\}} = \frac{1}{\sqrt{1 - \rho^2}} \frac{1}{\omega_\nu(x)} \sqrt{\frac{\nu + 1}{\nu + y^2}} \frac{\omega_{\nu+1} \{z_{\rho,\nu}(x, y)\}}{\Omega_{\nu+1} \{z_{\rho,\nu}(x, y)\}}.$$

**Remarque 4.4.** *Lorsque  $\nu \rightarrow \infty$ , on a  $z_{\rho,\nu}(x, y) \rightarrow z_\rho(x, y)$ ,  $\omega_{\nu+1}(x) \rightarrow \phi(x)$  et  $\Omega_{\nu+1}(x) \rightarrow \Phi(x)$ . On retrouve alors l'expression obtenue dans le cas de la copule Normale, à savoir que*

$$\lim_{\nu \rightarrow \infty} \frac{c_{\rho,\nu} \{\Omega_\nu(x), \Omega_\nu(y)\}}{C_{\rho,\nu}^{(01)} \{\Omega_\nu(x), \Omega_\nu(y)\}} = \frac{1}{\sqrt{1 - \rho^2}} \frac{1}{\phi(x)} \frac{\phi \{z_\rho(x, y)\}}{\Phi \{z_\rho(x, y)\}}.$$

**Remarque 4.5.** *Dans le cas particulier où  $\rho = 0$ , on a*

$$z_{0,\nu}(x, y) = x \sqrt{\frac{\nu + 1}{\nu + y^2}}$$

et donc

$$\frac{c_{0,\nu} \{\Omega_\nu(x), \Omega_\nu(y)\}}{C_{0,\nu}^{(01)} \{\Omega_\nu(x), \Omega_\nu(y)\}} = \frac{1}{\omega_\nu(x)} \sqrt{\frac{\nu + 1}{\nu + y^2}} \omega_{\nu+1} \left( x \sqrt{\frac{\nu + 1}{\nu + y^2}} \right) / \Omega_{\nu+1} \left( x \sqrt{\frac{\nu + 1}{\nu + y^2}} \right).$$

*Puisque quand  $\nu \rightarrow \infty$ , on a  $z_{0,\nu}(x, y) \rightarrow x$ ,  $\omega_{\nu+1}(x) \rightarrow \phi(x)$  et  $\Omega_{\nu+1}(x) \rightarrow \Phi(x)$ , on retrouve l'expression déjà obtenue pour la copule de l'indépendance, à savoir que*

$$\lim_{\nu \rightarrow \infty} \frac{c_{0,\nu} \{\Omega_\nu(x), \Omega_\nu(y)\}}{C_{0,\nu}^{(01)} \{\Omega_\nu(x), \Omega_\nu(y)\}} = \frac{1}{\Phi(x)}.$$

## 4.3 Preuves

### 4.3.1 Preuve du Lemme 4.3

La densité Normale bivariée est donnée par l'expression

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

Avec  $\phi(y) = e^{-y^2/2}/\sqrt{2\pi}$  la densité de la loi Normale centrée réduite,

$$\begin{aligned} \frac{\phi_\rho(x, y)}{\phi(y)} &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ - \left( \frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)} \right) \right\} \times \frac{\sqrt{2\pi}}{e^{-y^2/2}} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp \left\{ - \left( \frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)} \right) + \frac{y^2}{2} \right\} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp \left\{ - \left( \frac{x^2 + y^2 - 2\rho xy - y^2 + \rho^2 y^2}{2(1-\rho^2)} \right) \right\} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp \left\{ - \left( \frac{x^2 - 2\rho xy + \rho^2 y^2}{2(1-\rho^2)} \right) \right\} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp \left\{ - \frac{1}{2} \left( \frac{x - \rho y}{\sqrt{1-\rho^2}} \right)^2 \right\} \\ &= \frac{1}{\sqrt{1-\rho^2}} \phi \{ z_\rho(x, y) \}. \end{aligned}$$

### 4.3.2 Preuve du Lemme 4.4

À partir de la définition de la densité de la loi de Student univariée énoncée à l'Équation (4.3) et de celle de la Student bivariée à l'Équation (4.4), on a

$$\begin{aligned}
\frac{\omega_{\rho,\nu}(x, y)}{\omega_\nu(y)} &= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \left\{ 1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)} \right\}^{-(\frac{\nu}{2}+1)}}{\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left( 1 + \frac{y^2}{\nu} \right)^{-(\frac{\nu+1}{2})}} \\
&= \frac{\sqrt{\nu}\Gamma(\frac{\nu}{2})}{2\sqrt{\pi}\sqrt{1-\rho^2}\Gamma(\frac{\nu+1}{2})} \left( 1 + \frac{y^2}{\nu} \right)^{-\frac{1}{2}} \left\{ \frac{1 + \left( \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)} \right)}{1 + \frac{y^2}{\nu}} \right\}^{-(\frac{\nu}{2}+1)} \\
&= \frac{\sqrt{\nu}\Gamma(\frac{\nu}{2})}{2\sqrt{\pi}\sqrt{1-\rho^2}\Gamma(\frac{\nu+1}{2})} \left( 1 + \frac{y^2}{\nu} \right)^{-\frac{1}{2}} \left\{ \frac{(x - \rho y)^2 + (y^2 + \nu)(1 - \rho^2)}{(y^2 + \nu)(1 - \rho^2)} \right\}^{-(\frac{\nu}{2}+1)} \\
&= \frac{\sqrt{\nu}\Gamma(\frac{\nu}{2})}{2\sqrt{\pi}\sqrt{1-\rho^2}\Gamma(\frac{\nu+1}{2})} \left( 1 + \frac{y^2}{\nu} \right)^{-\frac{1}{2}} \left\{ 1 + \frac{z_\rho^2(x, y)}{y^2 + \nu} \right\}^{-(\frac{\nu}{2}+1)},
\end{aligned}$$

où on rappelle que  $z_\rho(x, y) = (x - \rho y)/\sqrt{1 - \rho^2}$ . À noter maintenant que la densité de la loi de Student univariée à  $\nu + 1$  degrés de liberté s'exprime par

$$\omega_{\nu+1}(x) = \frac{\Gamma(\frac{\nu}{2} + 1)}{\sqrt{\nu + 1}\sqrt{\pi}\Gamma(\frac{\nu+1}{2})} \left( 1 + \frac{x^2}{\nu + 1} \right)^{-(\frac{\nu}{2}+1)},$$

ce qui permet d'écrire

$$\begin{aligned}
\left\{ 1 + \frac{z_\rho^2(x, y)}{y^2 + \nu} \right\}^{-(\frac{\nu}{2}+1)} &= \left\{ 1 + \frac{\left( \sqrt{\frac{\nu+1}{\nu+y^2}} z_\rho(x, y) \right)^2}{\nu+1} \right\}^{-(\frac{\nu}{2}+1)} \\
&= \frac{\sqrt{\nu+1}\sqrt{\pi}\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2} + 1)} \omega_{\nu+1} \left\{ \sqrt{\frac{\nu+1}{\nu+y^2}} z_\rho(x, y) \right\}.
\end{aligned}$$

Il s'ensuit enfin que

$$\begin{aligned}
 \frac{\omega_{\rho,\nu}(x, y)}{\omega_\nu(y)} &= \frac{\sqrt{\nu} \Gamma\left(\frac{\nu}{2}\right)}{2\sqrt{\pi}\sqrt{1-\rho^2}\Gamma\left(\frac{\nu+1}{2}\right)} \left(1 + \frac{y^2}{\nu}\right)^{-\frac{1}{2}} \\
 &\quad \times \frac{\sqrt{\nu+1}\sqrt{\pi}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}+1\right)} \omega_{\nu+1}\left\{\sqrt{\frac{\nu+1}{\nu+y^2}} z_\rho(x, y)\right\} \\
 &= \frac{1}{\sqrt{1-\rho^2}} \sqrt{\frac{\nu+1}{\nu+y^2}} \omega_{\nu+1}\left\{\sqrt{\frac{\nu+1}{\nu+y^2}} z_\rho(x, y)\right\}.
 \end{aligned}$$

La démonstration est donc complétée.

# CHAPITRE 5

## LA GRANDE FAMILLE DES COPULES BETA

### 5.1 Construction et propriétés de la copule Beta

#### 5.1.1 La loi MGB2

Soient d'abord la densité et la fonction de répartition de la loi Beta de paramètres  $p, q > 0$  définies respectivement pour  $x \in [0, 1]$ , par les expressions

$$b_{p,q}(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1} \quad \text{et} \quad B_{p,q}(x) = \int_0^x b_{p,q}(s) \, ds,$$

où  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$  est la fonction Gamma. La fonction de densité de la loi MGB2 donnée par Yang *et al.* [39] est

$$f_X(x) = \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)\sqrt{\nu}} \left\{ 1 + \left( \frac{x}{\sqrt{\nu}} \right)^2 \right\}^{-\left(\frac{\nu+1}{2}\right)}.$$

La fonction de densité bivariée de la loi MGB2 donnée par Yang *et al.* [39] est

$$f_{X,Y}(x, y) = \frac{4\Gamma\left(\frac{\nu}{2} + 1\right)}{\Gamma\left(\frac{\nu}{2}\right)\{\Gamma\left(\frac{1}{2}\right)\}^2\nu} \left\{ 1 + \left( \frac{x+y}{\sqrt{\nu}} \right)^2 \right\}^{-\left(\frac{\nu}{2}+1\right)}.$$

On peut montrer que la fonction de répartition associée est

$$F_{X,Y}(x, y) = B_{p,q} \left\{ \frac{(x+y)^a}{(x+y)^a + b^a} \right\}.$$

Dans le but de construire une expression de la copule Beta, définissons les lois de densité des variables de loi *InvGa*, *GG* et *GB2*.

### 5.1.2 Copule et densité Beta

La copule Beta bivariée s'exprime par

$$C_{U,V}(u, v) = \int F_{U|\theta} \{F_U^{-1}(u)\} F_{V|\theta} \{F_V^{-1}(v)\} dG(\theta).$$

Tel que démontré par Yang *et al.* [39], une expression pour la copule Beta est

$$C_{p_1, p_2, q}(u, v) = \int_0^\infty G_{p_1} \left\{ \frac{x_1(u)}{\theta} \right\} G_{p_2} \left\{ \frac{x_2(v)}{\theta} \right\} \frac{\theta^{-q-1} e^{-1/\theta}}{\Gamma(q)} d\theta,$$

où  $G_p$  est la fonction de répartition de la loi Gamma de paramètres  $p$  et 1, alors que

$$x_1(u) = \frac{B_{p_1, q}^{-1}(u)}{1 - B_{p_1, q}^{-1}(u)} \quad \text{et} \quad x_2(v) = \frac{B_{p_2, q}^{-1}(v)}{1 - B_{p_2, q}^{-1}(v)}.$$

Yang *et al.* [39] montrent que la densité de la copule  $C_{p_1, p_2, q}$  peut s'exprimer par

$$c_{p_1, p_2, q}(u, v) = \mathcal{K}_{p_1, p_2, q} \frac{\{1 + x_1(u)\}^{p_1+q} \{1 + x_2(v)\}^{p_2+q}}{\{1 + x_1(u) + x_2(v)\}^{p_1+p_2+q}}, \quad (5.1)$$

où  $\mathcal{K}_{p_1, p_2, q} = \Gamma(q) \Gamma(p_1 + p_2 + q) / \Gamma(p_1 + q) \Gamma(p_2 + q)$ .

### 5.1.3 Cas particuliers de la famille Beta

#### Copule d'indépendance

Lorsque  $q \rightarrow \infty$  et que  $p_1, p_2$  demeurent fixés.

#### Copule de survie Clayton

La copule de survie de Clayton apparaît lorsque  $p_1 = p_2 = 1$ .

#### Copule Normale

Tel que mentionné par Yang *et al.* [39], la copule Normale est un cas particulier de la famille Beta lorsque  $p_1, p_2, q \rightarrow \infty$  de telle sorte que

$$\frac{p_1}{q} \rightarrow c > 0 \quad \text{et} \quad \frac{p_2}{q} \rightarrow d > 0.$$

Le paramètre de la copule Normale est alors

$$\rho = \sqrt{\frac{cd}{(1+c)(1+d)}} > 0.$$

#### Copule inconnue (néanmoins intéressante...)

Lorsque  $p_1 \rightarrow \infty$  et que  $p_2, q$  sont fixes, on retrouve une copule de la forme

$$C_{p_2,q}(u, v) = \frac{\Gamma(q)}{\Gamma(p_2 + q)} \frac{(1+y)^{p_2+q}}{x^{p_2}} e^{-y/x},$$

où  $x = 1/G_q^{-1}(1 - u)$  et  $y = B_{p_2,q}^{-1}(v)/\{1 - B_{p_2,q}^{-1}(v)\}$ .

## 5.2 Obtention de $C^{(01)}$

On note en premier lieu que pour une copule générale  $C$  dont la densité  $c$  existe, alors

$$C^{(01)}(u, v) = \int_0^u c(t, v) dt. \quad (5.2)$$

En vertu de cette expression et à la lumière de l'Équation (5.1), on arrive à démontrer le joli résultat qui suit concernant la loi conditionnelle  $C_{p_1,p_2,q}^{(01)}$  d'une copule Beta.

**Proposition 5.4.** *On a*

$$C_{p_1,p_2,q}^{(01)}(u, v) = 1 - B_{p_2+q,p_1} \left\{ \frac{1 + x_2(v)}{1 + x_1(u) + x_2(v)} \right\}. \quad (5.3)$$

**Démonstration.** En utilisant la formule (5.2) dans le cas de la densité Beta, on a

$$C_{p_1,p_2,q}^{(01)}(u, v) = \mathcal{K}_{p_1,p_2,q} \{1 + x_2(v)\}^{p_2+q} \int_0^u \frac{\{1 + x_1(t)\}^{p_1+q}}{\{1 + x_1(t) + x_2(v)\}^{p_1+p_2+q}} dt.$$

Soit maintenant le changement de variable  $s = 1 + x_1(t) + x_2(v)$ . À partir de la définition de  $x_1$ , on a  $1 + x_1(t) = \{1 - B_{p_1,q}^{-1}(t)\}^{-1}$ , ce qui fait que

$$s - x_2(v) = \frac{1}{1 - B_{p_1,q}^{-1}(t)} \iff t = B_{p_1,q} \left\{ 1 - \frac{1}{s - x_2(v)} \right\}.$$

On en déduit donc que

$$\begin{aligned}
 dt &= b_{p_1, q} \left\{ 1 - \frac{1}{s - x_2(v)} \right\} \frac{1}{\{s - x_2(v)\}^2} ds \\
 &= \frac{\Gamma(p_1 + q)}{\Gamma(p_1) \Gamma(q)} \left\{ 1 - \frac{1}{s - x_2(v)} \right\}^{p_1-1} \left\{ \frac{1}{s - x_2(v)} \right\}^{q+1} ds \\
 &= \frac{\Gamma(p_1 + q)}{\Gamma(p_1) \Gamma(q)} \frac{\{s - x_2(v) - 1\}^{p_1-1}}{\{s - x_2(v)\}^{p_1+q}} ds.
 \end{aligned}$$

Il s'ensuit alors que pour  $s_1 = x_2(v) + 1$  et  $s_2 = x_2(v) + \{1 - B_{p_1, q}^{-1}(u)\}^{-1}$ ,

$$\begin{aligned}
 \int_0^u \frac{\{1 + x_1(t)\}^{p_1+q} dt}{\{1 + x_1(t) + x_2(v)\}^{p_1+p_2+q}} &= \frac{\Gamma(p_1 + q)}{\Gamma(p_1) \Gamma(q)} \int_{s_1}^{s_2} \frac{\{s - x_2(v)\}^{p_1+q} \{s - x_2(v) - 1\}^{p_1-1}}{s^{p_1+p_2+q} \{s - x_2(v)\}^{p_1+q}} ds \\
 &= \frac{\Gamma(p_1 + q)}{\Gamma(p_1) \Gamma(q)} \int_{s_1}^{s_2} \frac{\{s - x_2(v) - 1\}^{p_1-1}}{s^{p_1+p_2+q}} ds \\
 &= \frac{\Gamma(p_1 + q)}{\Gamma(p_1) \Gamma(q)} \int_{s_1}^{s_2} \left( \frac{1}{s} \right)^{p_2+q+1} \left\{ 1 - \frac{1 + x_2(v)}{s} \right\}^{p_1-1} ds.
 \end{aligned}$$

On effectue ensuite le changement de variable  $y = \{1 + x_2(v)\}/s$ . On a donc  $s = \{1 + x_2(v)\}/y$  et les bornes d'intégration deviennent alors  $y_1 = \{1 + x_2(v)\}/s_1 = 1$  et

$$y_2 = \frac{1 + x_2(v)}{s_2} = \frac{1 + x_2(v)}{x_2(v) + \{1 - B_{p_1, q}^{-1}(u)\}^{-1}} = \frac{1 + x_2(v)}{1 + x_1(u) + x_2(v)}.$$

Puisque  $ds = -\{1 + x_2(v)\}/y^2$ ,

$$\begin{aligned}
 \int_0^u \frac{\{1 + x_1(t)\}^{p_1+q} dt}{\{1 + x_1(t) + x_2(v)\}^{p_1+p_2+q}} &= \frac{\Gamma(p_1 + q)}{\Gamma(p_1) \Gamma(q)} \int_{y_2}^{y_1} \left\{ \frac{y}{1 + x_2(v)} \right\}^{p_2+q+1} (1 - y)^{p_1-1} \\
 &\quad \times \left\{ \frac{1 + x_2(v)}{y^2} \right\} dy \\
 &= \frac{\Gamma(p_1 + q)}{\Gamma(p_1) \Gamma(q)} \frac{1}{\{1 + x_2(v)\}^{p_2+q}} \int_{y_2}^{y_1} y^{p_2+q-1} (1 - y)^{p_1-1} dy \\
 &= \frac{\Gamma(p_1 + q) \Gamma(p_2 + q)}{\Gamma(q) \Gamma(p_1 + p_2 + q)} \frac{1}{\{1 + x_2(v)\}^{p_2+q}} \int_{y_2}^{y_1} b_{p_2+q, p_1}(y) dy \\
 &= \frac{B_{p_2+q, p_1}(y_1) - B_{p_2+q, p_1}(y_2)}{\mathcal{K}_{p_1, p_2, q} \{1 + x_2(v)\}^{p_2+q}}.
 \end{aligned}$$

L'insertion de cette dernière expression dans l'Équation (5.3) amène finalement

$$\begin{aligned}
 C_{p_1,p_2,q}^{(01)}(u,v) &= B_{p_2+q,p_1}(y_1) - B_{p_2+q,p_1}(y_2) \\
 &= B_{p_2+q,p_1}(1) - B_{p_2+q,p_1} \left\{ \frac{1+x_2(v)}{1+x_1(u)+x_2(v)} \right\} \\
 &= 1 - B_{p_2+q,p_1} \left\{ \frac{1+x_2(v)}{1+x_1(u)+x_2(v)} \right\},
 \end{aligned}$$

ce qui termine la démonstration.  $\diamond$

Si  $X$  est distribuée selon une loi Gamma de paramètres  $p$  et  $q$ , alors la loi de  $Y = X/(X + 1)$  est une Beta de paramètres  $p$  et  $q$ . Cette caractérisation de la loi Beta permet de déduire

$$B_{p,q}(s) = \mathbb{P} \left( \frac{X}{X+1} \leq s \right) = \mathbb{P} \left( X \leq \frac{s}{1-s} \right) = G_{p,q} \left( \frac{s}{1-s} \right),$$

où  $G_{p,q}$  est la fonction de répartition de la loi Gamma, à savoir

$$G_{p,q}(x) = \int_0^x \frac{t^{p-1} e^{-t/q}}{\Gamma(p) q^p} dt.$$

On a donc  $B_{p,q}^{-1}(u) = G_{p,q}^{-1}(u)/\{1 - G_{p,q}^{-1}(u)\}$ , ce qui amène

$$x_j(u) = \frac{B_{p_j,q}^{-1}(u)}{1 - B_{p_j,q}^{-1}(u)} = G_{p_j,q}^{-1}(u).$$

Quelques calculs permettent alors d'obtenir l'expression alternative

$$C_{p_1,p_2,q}^{(01)}(u,v) = 1 - G_{p_2+q,p_1} \left\{ \frac{1 + G_{p_2,q}^{-1}(v)}{G_{p_1,q}^{-1}(u)} \right\}. \quad (5.4)$$

### 5.3 Illustration du cas particulier $p_1 = p_2 = 1$ : copule de survie Clayton

Comme  $B_{1,q}(x) = 1 - (1 - x)^q$ , on a  $B_{1,q}^{-1}(u) = 1 - (1 - u)^{1/q}$  et donc

$$x_1(u) = x_2(u) = \frac{B_{1,q}^{-1}(u)}{1 - B_{1,q}^{-1}(u)} = \frac{1 - (1 - u)^{1/q}}{(1 - u)^{1/q}} = (1 - u)^{-1/q} - 1.$$

Ainsi, comme cas particulier de l'Équation (5.3) de la Proposition 5.4, on obtient

$$\begin{aligned} C_{1,1,q}^{(01)}(u, v) &= 1 - B_{q+1,1} \left\{ \frac{1 + x_2(v)}{1 + x_1(u) + x_2(v)} \right\} \\ &= 1 - \left\{ \frac{(1 - v)^{-1/q}}{(1 - u)^{-1/q} + (1 - v)^{-1/q} - 1} \right\}^{q+1} \\ &= 1 - \left\{ \frac{(1 - u)^{-1/q} + (1 - v)^{-1/q} - 1}{(1 - v)^{-1/q}} \right\}^{-(q+1)} \\ &= 1 - (1 - v)^{-(1+\frac{1}{q})} \left\{ (1 - u)^{-1/q} + (1 - v)^{-1/q} - 1 \right\}^{-(q+1)}. \end{aligned}$$

En posant  $q = 1/\theta$ , on trouve alors

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

Cette dernière expression correspond à la dérivée par rapport à  $v$  de la copule de survie Clayton, c'est-à-dire  $\widehat{C}_\theta^{\text{Cl}}(u, v) = u + v - 1 + C_\theta^{\text{Cl}}(1 - u, 1 - v)$ .

Maintenant, on déduit comme cas particulier de l'Équation (5.1) que

$$c_{1,1,q}(u, v) = \left( 1 + \frac{1}{q} \right) \frac{\{(1 - u)^{-1/q}\}^{q+1} \{(1 - v)^{-1/q}\}^{q+1}}{\{(1 - u)^{-1/q} + (1 - v)^{-1/q} - 1\}^{q+2}},$$

où on a utilisé le fait que

$$\mathcal{K}_{1,1,q} = \frac{\Gamma(q) \Gamma(q+2)}{\{\Gamma(q+1)\}^2} = \frac{q(q+1) \Gamma(q) \Gamma(q)}{\{q \Gamma(q)\}^2} = 1 + \frac{1}{q}.$$

On retrouve la densité de la copule de survie Clayton en posant  $q = 1/\theta$ . En effet,

$$\begin{aligned}
 c_{1,1,1/\theta}(1-u, 1-v) &= (1+\theta) \frac{\{(1-(1-u))^{-\theta}\}^{\frac{1}{\theta}+1} \{(1-(1-v))^{-\theta}\}^{\frac{1}{\theta}+1}}{\{(1-(1-u))^{-\theta} + (1-(1-v))^{-\theta} - 1\}^{\frac{1}{\theta}+2}} \\
 &= (1+\theta) \frac{(u^{-\theta})^{\frac{1}{\theta}+1} (v^{-\theta})^{\frac{1}{\theta}+1}}{(u^{-\theta} + v^{-\theta} - 1)^{\frac{1}{\theta}+2}} \\
 &= (1+\theta)(uv)^{-(\theta+1)}(u^{-\theta} + v^{-\theta} - 1)^{-(\frac{1}{\theta}-2)},
 \end{aligned}$$

ce qui est l'expression de la densité de la copule de Clayton de l'Équation (2.10).

# CHAPITRE 6

## ESTIMATION SEMI-PARAMÉTRIQUE DES ENSEMBLES DE CONFIANCE

### 6.1 Version empirique des intervalles de confiance

#### 6.1.1 Contexte

On a vu au Chapitre 2 que l'ensemble de confiance  $\mathcal{B}_{p,1-\alpha}$  d'une paire  $(X, Y)$  sur la courbe de niveau  $L_p$  est constitué des paires de points

$$(F_X^{-1}(u), F_Y^{-1}(v)) \quad (6.1)$$

telles que  $(u, v) \in \mathcal{B}_{p,1-\alpha}^C = \{(u, v) \in [0, 1]^2 : u \in I_{p,1-\alpha}^C, v = C_u^{-1}(1 - p)\}$ , où pour  $\alpha_1, \alpha_2 \geq 0$  tels que  $\alpha_1 + \alpha_2 = \alpha \in (0, 1)$ ,

$$I_{p,1-\alpha}^C = [F_p^{-1}(\alpha_1), F_p^{-1}(1 - \alpha_2)].$$

On rappelle que  $F_p$  est la loi conditionnelle de  $U$  sachant que  $C(U, V) = 1 - p$ . À la lumière de (6.1), l'intervalle de confiance pour les valeurs de  $(X, Y)$  est alors

$$I_{p,1-\alpha} = [F_X^{-1} \circ F_p^{-1}(\alpha_1), F_Y^{-1} \circ F_p^{-1}(1 - \alpha_2)].$$

### 6.1.2 Objectif

L'objectif de ce chapitre est d'estimer l'intervalle de confiance  $I_{p,1-\alpha}$  sur la base de paires indépendantes  $(X_1, Y_1), \dots, (X_n, Y_n)$  provenant d'une loi bivariée  $F_{X,Y}$ . On supposera que les marges  $F_X$  et  $F_Y$  de  $F_{X,Y}$  sont continues, de sorte que  $F_{X,Y}$  possède une unique copule  $C$  qui contient toute la dépendance entre  $X$  et  $Y$ . Pour ce faire, on accomplira les deux étapes suivantes :

- (1) Choisir un modèle paramétrique adéquat pour la copule de  $F_{X,Y}$ ;
- (2) Estimer les lois marginales  $F_X$  et  $F_Y$  avec les marges empiriques.

Cette approche repose à la fois sur un aspect paramétrique (sélection d'une famille de copules) et entièrement non-paramétrique (utilisation des marges empiriques). Notre stratégie est donc de nature semi-paramétrique.

## 6.2 Description de la méthodologie proposée

On suppose dans la suite que les lois marginales de la loi bivariée  $F_{X,Y}$  sont inconnues et que sa copule  $C$  appartient à une famille paramétrique  $\mathcal{F} = \{C_\theta; \theta \in \Theta \subset \mathbb{R}^p\}$ , où  $\Theta$  est l'espace des paramètres. Trois facettes sont à considérer dans ce contexte, à savoir (i) estimer le paramètre  $\theta$ , (ii) tester l'hypothèse que la copule  $C$  appartient à la famille  $\mathcal{F}$  et (iii) estimer  $F_X$  et  $F_Y$  de manière non-paramétrique. Ces trois aspects sont traités dans la suite.

### 6.2.1 Estimer le paramètre d'une copule

Lorsque la loi est à deux dimensions, c'est-à-dire que  $d = 2$ , et qu'il n'y a qu'un seul paramètre à estimer, on peut employer la méthode de l'inversion du tau de Kendall. Soit une famille de copules bivariées  $\mathcal{F} = \{C_\theta; \theta \in \Theta \subseteq \mathbb{R}^p\}$  dont le tau de Kendall est une fonction de  $\theta$  défini par

$$\tau(\theta) = 4 \int_0^1 \int_0^1 C_\theta(u, v) dC_\theta(u, v) - 1.$$

L'idée consiste alors à estimer  $\theta$  par le nombre  $\theta_{\tau,n}$  tel que  $\tau(\theta_{\tau,n}) = \tau_n$ , où  $\tau_n$  est le tau de Kendall empirique. Si la fonction  $\tau(\theta)$  est inversible, alors  $\theta_{\tau,n} = \tau^{-1}(\tau_n)$ . Par exemple, on a  $\tau(\theta) = 2/9$  pour la copule FGM, si bien que  $\theta_{\tau,n} = 9\tau_n/2$ . Dans les cas où  $\tau(\theta)$  n'est pas inversible, il faut résoudre l'équation  $\tau(\theta) - \tau_n = 0$  numériquement.

L'estimateur dit à *maximum de pseudo-vraisemblance* est défini sous l'hypothèse que la densité  $c_\theta$  de  $C_\theta$  existe. Dans ce cas, en se rapportant à la représentation de la densité conjointe d'un vecteur aléatoire de l'équation (2.3), la fonction de log-vraisemblance de  $(X_1, Y_1), \dots, (X_n, Y_n)$  s'écrit

$$L^*(\theta) = \sum_{i=1}^n [\ln c_\theta \{F_X(X_i), F_Y(Y_i)\} + \ln f_X(X_i) + \ln f_Y(Y_i)].$$

Dans un contexte où les lois marginales  $F_X$  et  $F_Y$  sont supposées inconnues, on considère plutôt, tel que suggéré par Shih et Louis [34] et Genest *et al.* [15], la fonction de pseudo-vraisemblance définie par

$$L(\theta) = \sum_{i=1}^n \ln c_\theta \{F_{Xn}(X_i), F_{Yn}(Y_i)\}.$$

L'estimateur à pseudo-maximum de vraisemblance  $\theta_{PV,n}$  est défini comme la valeur de  $\theta \in \Theta$  qui maximise  $L(\theta)$ .

### 6.2.2 Choix formel d'une famille de copules

Pour une certaine famille paramétrique de copules  $\mathcal{F} = \{C_\theta; \theta \in \Theta \subseteq \mathbb{R}^p\}$ , un test d'adéquation est une procédure statistique qui permet de choisir entre les hypothèses nulle et alternative

$$\mathbb{H}_0^{\mathcal{F}} : C \in \mathcal{F} \quad \text{et} \quad \mathbb{H}_1^{\mathcal{F}} : C \notin \mathcal{F}.$$

Plusieurs méthodes ont été proposées dans la littérature. On peut citer par exemple les travaux de Fermanian [13], Mesfioui *et al.* [26] et Bahraoui *et al.* [1]. Dans la suite, on s'attardera cependant à deux procédures qui se sont avérées très populaires, à savoir les méthodes de Genest *et al.* [16] et Genest *et al.* [17].

Les tests proposés par Genest *et al.* [16] sont basés sur la transformation intégrale de probabilité  $K_\theta(w) = \mathbb{P}\{C_\theta(U, V) \leq w\}$  associée à la copule  $C_\theta \in \mathcal{F}$ . La version empirique de  $K_\theta$  est définie par

$$K_n(w) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(W_{n,i} \leq w),$$

où  $W_{n,1}, \dots, W_{n,n}$  sont les pseudo-observations définies par

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

En supposant que  $\theta_n$  est un estimateur convergent pour  $\theta$ , une des statistiques de test proposée par Genest *et al.* [16] est

$$\widetilde{W}_n^{\text{GQR}} = n \int_0^1 \{K_n(w) - K_{\theta_n}(w)\}^2 dw.$$

Pour englober les situations où la fonction  $K_\theta$  ne possède pas de forme explicite, ou encore que  $K_\theta$  est explicite mais de forme compliquée, on utilisera plutôt la version

suggérée par Genest *et al.* [17], à savoir

$$\begin{aligned}
 W_{n,N}^{\text{GQR}} &= n \int_0^1 \{K_n(w) - K_N(w)\}^2 dw \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \max(W_{i,n}, W_{j,n}) - \frac{2}{N} \sum_{i=1}^n \sum_{j=1}^N \max(W_{i,n}, W_{j,N}^*) \\
 &\quad + \frac{n}{N^2} \sum_{i=1}^N \sum_{j=1}^N \max(W_{i,N}^*, W_{j,N}^*) ,
 \end{aligned}$$

où  $K_N$  est une version estimée de  $K_{\theta_n}$  basée sur des paires simulées  $(X_1, Y_1), \dots, (X_N, Y_N)$  i.i.d.  $C_{\theta_n}$  et  $W_{1,N}^*, \dots, W_{N,N}^*$  sont les pseudo-observations calculées à partir de cet échantillon. Le calcul de la p-valeur d'un test basé sur  $W_{n,N}^{\text{GQR}}$  utilise la méthode de ré-échantillonnage du bootstrap paramétrique.

Plutôt que de se baser sur  $K_C$ , on peut travailler directement au niveau de la copule d'un couple aléatoire en considérant la copule empirique

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}\{F_{Xn}(X_i) \leq u, F_{Yn}(Y_i) \leq v\}.$$

Une statistique de test dans la lignée de  $W_{n,N}^{\text{GQR}}$  est alors

$$W_{n,N}^{\text{GRB}} = n \int_0^1 \int_0^1 \{C_n(u, v) - C_N(u, v)\}^2 du dv.$$

Cette statistique fut proposée par Genest *et al.* [17], où on montre que

$$\begin{aligned}
 W_{n,N}^{\text{GRB}} &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \{1 - \max(U_{i,n}, V_{i,n})\} \{1 - \max(U_{j,n}, V_{j,n})\} \\
 &\quad - \frac{2}{N} \sum_{i=1}^n \sum_{j=1}^N \{1 - \max(U_{i,n}, V_{i,n})\} \{1 - \max(U_{j,N}^*, V_{j,N}^*)\} \\
 &\quad + \frac{n}{N} \sum_{i=1}^N \sum_{j=1}^N \{1 - \max(U_{i,N}^*, V_{i,N}^*)\} \{1 - \max(U_{j,N}^*, V_{j,N}^*)\} ,
 \end{aligned}$$

où  $(U_{i,n}, U_{j,n}) = (F_{Xn}(X_i), F_{Yn}(Y_i))$  et  $(U_{i,N}^*, U_{j,N}^*) = (F_{XN}(X_i^*), F_{YN}(Y_i^*))$ .

### 6.2.3 Estimation non-paramétrique des lois marginales

Il est possible d'estimer une distribution marginale en supposant qu'elle appartient à une certaine famille paramétrique, comme les lois normale, Gamma, Student, etc. Cependant, cette approche a le grand défaut de créer une énorme distortion lorsque la famille est mal spécifiée. Pour cette raison, les marges  $F_X$  et  $F_Y$  de  $F_{X,Y}$  seront estimées par les marges empiriques, à savoir

$$\widehat{F}_X(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(X_i \leq x) \quad \text{et} \quad \widehat{F}_Y(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(Y_i \leq y).$$

Les inverses  $F_X^{-1}(u)$  et  $F_Y^{-1}(v)$  sont alors estimés par

$$\widehat{F}_X^{-1}(u) = \inf \left\{ x \in \mathbb{R} : \widehat{F}_X(x) = u \right\} \quad \text{et} \quad \widehat{F}_Y^{-1}(v) = \inf \left\{ y \in \mathbb{R} : \widehat{F}_Y(y) = v \right\}.$$

## 6.3 Illustration sur des données financières

### 6.3.1 Présentation des données

Cette section vise à illustrer la construction d'un intervalle de confiance sur une courbe de niveau. Pour ce faire, les données qui seront analysées concernent le cours des actions (en dollars américains) des entreprises *ExxonMobil* et *Schlumberger* entre le 1<sup>er</sup> janvier et le 31 décembre 2021. Les données ont été récoltées sur *Google Sheet* et comprennent  $n = 251$  paires d'observations. Quelques statistiques descriptives sont répertoriées au Tableau 6.1.

La compagnie *ExxonMobil* compte parmi les plus grandes sociétés pétrolières au monde. Ses principales activités comprennent l'exploration, la production, le transport

TABLE 6.1 – Statistiques descriptives univariées pour le cours des actions de *Exxon-Mobil* et de *Schlumberger* entre le 1<sup>er</sup> janvier et le 31 décembre 2021

Compagnie	Moyenne	Médiane	Écart-type
<i>ExxonMobil</i>	57,95	58,82	5,02
<i>Schlumberger</i>	29,33	29,01	3,05

et la commercialisation de pétrole brut, de gaz naturel et de produits pétrochimiques. Elle est issue de la fusion, en 1999, de *Exxon Corporation* et *Mobil Corporation*. Elle possède des activités dans plus de 200 pays et territoires à travers le monde. En 2021, sa capitalisation boursière frôlait les 240 milliards de dollars, ce qui représente environ 2% de la capitalisation totale du *New York Stock Exchange*.

La compagnie *Schlumberger* est une entreprise multinationale de services pétroliers fondée en 1926 et présente dans plus de 120 pays. Ses principales activités sont la recherche et l'exploration de pétrole et de gaz, la conception et la construction de plates-formes de forage, la gestion de la production de pétrole et de gaz, la fourniture de services de forage et d'achèvement de puits, ainsi que la production de technologies de mesure, de contrôle et de gestion des données pour l'industrie pétrolière et gazière. En 2021, sa capitalisation boursière était d'environ 50 milliards de dollars, ce qui représente environ 0,4% de la capitalisation du *New York Stock Exchange*.

### 6.3.2 Stationnarité et indépendance sérielle

Les deux graphiques en haut de la Figure 6.4 montrent l'évolution des deux séries chronologiques. En les examinant, il apparaît assez clair qu'il y a présence d'une forte corrélation sérielle. Ce pressentiment est confirmé formellement par le test de Durbin–Watson, qui conclue dans les deux cas à la présence significative de dépendance sérielle, au seuil  $\alpha = 0,01$ . Cette présence de dépendance sérielle dans les données rend inadéquate l'analyse de la dépendance entre les paires via les copules. Afin de s'en affranchir, une idée consiste à travailler avec les données *différenciées*.

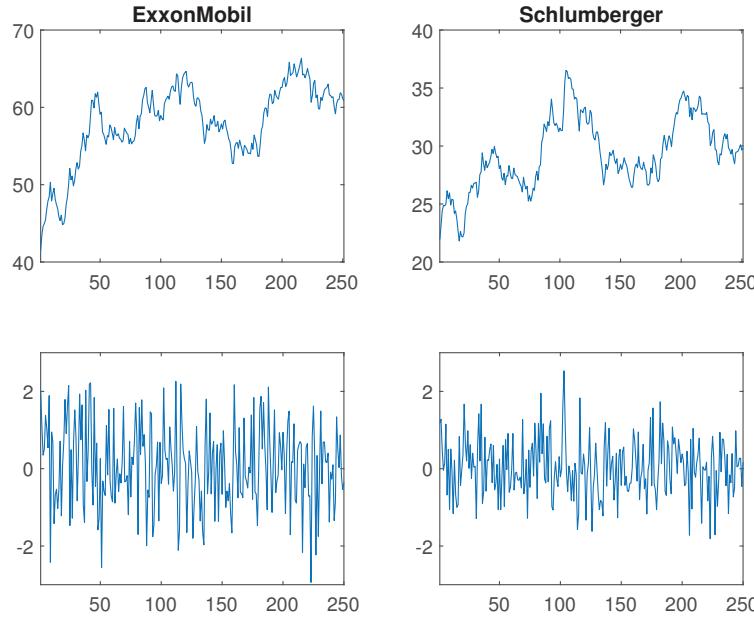


FIGURE 6.4 – Évolution du prix des actions de *ExxonMobil* et de *Schlumberger* entre le 1<sup>er</sup> janvier et le 31 décembre 2021. En haut : séries originales ; en bas : séries différencierées à l'ordre un

Spécifiquement, si  $W_1, \dots, W_n$  est une série chronologique, la série *différenciée* est donnée par  $W_1^*, \dots, W_{n-1}^*$ , où  $W_i^* = W_{i+1} - W_i$ . Cette opération a été effectuée sur les données *ExxonMobil* et *Schlumberger*. Les séries résultantes se retrouvent en bas de la Figure 6.4. Le test de Durbin–Watson confirme le succès de l'opération, car il conclut au non-rejet de l'hypothèse nulle d'indépendance sérielle dans les deux cas.

### 6.3.3 Choix d'une copule

La suite de l'analyse se fera sur les paires de données différencierées concernant les cours des actions de *ExxonMobil* et *Schlumberger*. Le nuage de points de ces observations se retrouve sur le graphique en haut à droite de la Figure 6.5. Sur la diagonale, on peut observer les histogrammes du cours des actions de *ExxonMobil* et de *Schlumberger*. Dans le cas de *ExxonMobil*, il est clair que la distribution n'appartient pas à la famille des lois normales.

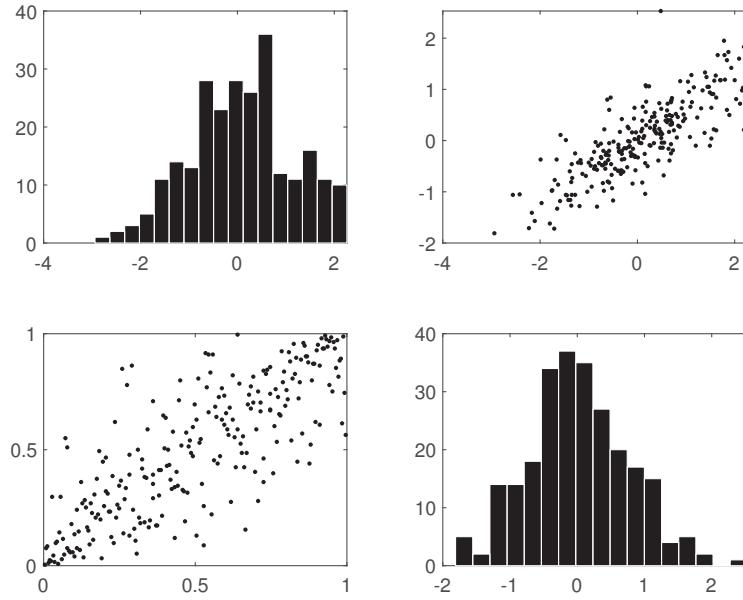


FIGURE 6.5 – Sur la diagonale : histogrammes de la différence des prix des actions des compagnies *ExxonMobil* (en haut à gauche) et *Schlumberger* (en bas à droite) ; en haut à droite : nuage de points des données originales ; en bas à gauche : nuage de points des données standardisées

En ce qui nous concerne, la figure qui revêt le plus d’importance est le graphique en bas à gauche de la Figure 6.5. En effet, ce nuage de points concerne les paires  $(\widehat{F}_X(X_i), \widehat{F}_Y(Y_i))$ ,  $i \in \{1, \dots, n-1\}$ . Puisque  $(\widehat{F}_X(X_i), \widehat{F}_Y(Y_i)) \approx (F_X(X_i), F_Y(Y_i)) = (U_i, V_i) \sim C$ , ces paires de points sont en quelque sorte des observations tirées de la copule sous-jacente à cette population. Autrement dit, le graphique en bas à gauche de la Figure 6.5 permet d’observer la forme de la copule entre le cours des actions.

Les tests d’adéquation pour copules basés sur les statistiques  $W_{n,N}^{\text{GRB}}$  et  $W_{n,N}^{\text{GQR}}$  ont été effectués pour vingt-neuf modèles de copules. Les résultats se retrouvent au Tableau 6.2. Les modèles testés concernent les familles usuelles Normale, Student, Clayton, Frank et Gumbel, ainsi que les extensions des copules Normale et Student fournies par les modèles Khi-deux et Fisher. En plus, les versions survies des copules qui sont radiallement asymétriques, à savoir Clayton, Gumbel, Fisher et Khi-deux, ont été considérées dans l’étude. On rappelle que si  $C$  est une copule, alors sa version survie est la copule  $\widehat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$ . La copule  $\widehat{C}$  correspond à la loi de

$(1 - U, 1 - V)$  pour  $(U, V) \sim C$ . Les résultats du Tableau 6.2 permettent de constater que les copules qui sont radialement asymétriques sont clairement rejetées par les deux tests d'adéquation, à l'exception de la copule de survie de Gumbel. *A contrario*, les familles de copules Normale, Frank et Student, qui sont symétriques, sont toutes acceptées au seuil de signification 5%.

TABLE 6.2 – P-valeurs (en pourcentage) des tests d'adéquation pour copules basés sur les statistiques  $W_{n,N}^{\text{GRB}}$  et  $W_{n,N}^{\text{GQR}}$  pour le prix des actions de *ExxonMobil* et de *Schlumberger* entre le 1<sup>er</sup> janvier et le 31 décembre 2021

Copule sous $\mathbb{H}_0$	$W_{n,N}^{\text{GRB}}$	$W_{n,N}^{\text{GQR}}$
Normale	41,9	50,8
Clayton	0,1	0,1
Survie Clayton	< 0,1	< 0,1
Frank	38,9	15,3
Gumbel	2,41	0,4
Survie Gumbel	31,4	17,7
Student $\nu = 1$	6,1	0,8
Student $\nu = 3$	37,2	14,6
Student $\nu = 5$	56,2	19,7
Student $\nu = 7$	45,1	45,1
Student $\nu = 10$	39,7	37,3
Fisher $\nu = 1$	< 0,1	< 0,1
Fisher $\nu = 3$	< 0,1	< 0,1
Fisher $\nu = 5$	< 0,1	< 0,1
Fisher $\nu = 7$	< 0,1	< 0,1
Fisher $\nu = 10$	< 0,1	< 0,1
Survie Fisher $\nu = 1$	0,7	< 0,1
Survie Fisher $\nu = 3$	0,8	0,2
Survie Fisher $\nu = 5$	0,6	0,1
Survie Fisher $\nu = 7$	0,6	0,1
Survie Fisher $\nu = 10$	0,8	0,1
Khi-deux $a = 0$	< 0,1	< 0,1
Khi-deux $a = 1/4$	< 0,1	< 0,1
Khi-deux $a = 1/2$	< 0,1	< 0,1
Khi-deux $a = 3/4$	0,1	< 0,1
Survie Khi-deux $a = 0$	0,7	< 0,1
Survie Khi-deux $a = 1/4$	0,3	0,1
Survie Khi-deux $a = 1/2$	0,5	0,2
Survie Khi-deux $a = 3/4$	1,1	0,4

### 6.3.4 Intervalle de confiance semi-paramétrique

À la lumière des tests d'adéquation dont les résultats sont présentés à la sous-section précédente, la copule de Frank est un modèle de dépendance en adéquation avec le jeu de données *ExxonMobil–Schlumberger*. Bien que d'autres modèles ont également été acceptés sous l'hypothèse nulle, la copule de Frank sera utilisée dans la suite pour calculer l'intervalle de confiance sur une courbe de niveau. On se basera donc sur les calculs effectuée à la section 3.3.2.

Premièrement, puisque le tau de Kendall observé est  $\tau_n = 0,61$ , la valeur estimée du paramètre de la copule de Frank selon la méthode de l'inversion du tau de Kendall est  $\hat{\theta} = 8,24$ . Il s'ensuit que  $\ell(\theta) = e^{-8,24} - 1 = -0,9997$ . En prenant  $\alpha_1 = \alpha_2 = 0,025$ , on a pour  $p = 0,1$  que  $\ell(\theta q) = e^{-8,24 \times 0,9} - 1 = -0,9994$ , et donc

$$\begin{aligned} I_{0,1,0,95}^{\text{Fr}} &= \left[ \frac{-1}{8,24} \ln \left( 1 + \frac{(-0,9997)^{0,025}}{(-0,9994)^{0,025-1}} \right), \frac{-1}{8,24} \ln \left( 1 + \frac{(-0,9997)^{1-0,025}}{(-0,9994)^{-0,025}} \right) \right] \\ &= [0,9017, 0,9962]. \end{aligned}$$

Dans le cas  $p = 0,05$ ,  $\ell(\theta q) = e^{-8,24 \times 0,95} - 1 = -0,9996$  et

$$\begin{aligned} I_{0,05,0,95}^{\text{Fr}} &= \left[ \frac{-1}{8,24} \ln \left( 1 + \frac{(-0,9997)^{0,025}}{(-0,9996)^{0,025-1}} \right), \frac{-1}{8,24} \ln \left( 1 + \frac{(-0,9997)^{1-0,025}}{(-0,9996)^{-0,025}} \right) \right] \\ &= [0,9510, 0,9985]. \end{aligned}$$

Ces intervalles de confiance concernent les valeurs de  $U$  sur la courbe de niveau  $C(U, V) = 1 - p$ . Pour obtenir les résultats à l'échelle des valeurs observées, il s'agit d'appliquer les marges inverses. On obtient alors

$$\begin{aligned} I_{0,1,0,95}^X &= [F_{n,X}^{-1}(0,9017), F_{n,X}^{-1}(0,9962)] = [1,5793, 2,2418] \\ \text{et } I_{0,05,0,95}^X &= [F_{n,X}^{-1}(0,9510), F_{n,X}^{-1}(0,9985)] = [1,8751, 2,2634]. \end{aligned}$$

Ces intervalles sont donc les valeurs des actions de *ExxonMobil* sur les courbes de

niveau  $p = 0,1$  et  $p = 0,05$ . Les valeurs correspondantes pour le cours des actions de *Schlumberger* sont

$$I_{0,1,0,95}^Y = [F_{n,Y}^{-1}(0,9962), F_{n,Y}^{-1}(0,9017)] = [2,2658, 1,0493]$$

et  $I_{0,05,0,95}^Y = [F_{n,Y}^{-1}(0,9985), F_{n,Y}^{-1}(0,9510)] = [2,5345, 1,2477]$ .

### 6.3.5 Interprétation des résultats

Pour les valeurs  $p = 0,1$  et  $p = 0,05$ , tel que l'on a respectivement une probabilité  $p$  de 10% et 5%, on observe, avec un taux de confiance de 95%, une montée des prix de l'action d'*ExxonMobil* supérieure ou égale à une valeur comprise respectivement entre 1,5793 et 2,2418, et entre 1,8751 et 2,26, ou une montée des prix de l'action de *Schlumberger* supérieure ou égale à une valeur associée comprise respectivement entre 1,0493 et 2,2658, et entre 1,2477 et 2,53. Ainsi, cela correspond respectivement, comparativement aux derniers prix des actions observés en 2021, à une montée de 2,6% à 3,69%, et de 3.08% à 3.72%, pour *ExxonMobil* et de 3,52% à 7,6%, et de 4.18% à 8.48%, pour *Schlumberger*.

Ces résultats suivent notre intuition, à savoir que plus la probabilité  $p$  est faible, donc plus l'événement est rare, plus la montée des prix est forte. Cela se comprend puisqu'il est rare d'observer une forte montée des prix. De plus, on observe des valeurs similaires pour les bornes maximales, celles-ci correspondent aux asymptotes des courbes de niveau qui ont été coupées au même endroit, tel que les données soient possiblement observables, avec un taux de confiance de 95%.

## CONCLUSION

En hydrologie, les périodes de retour bivariées sont employées fréquemment dans la conception d'infrastructures, notamment les barrages hydroélectriques. La notion de période de retour est basée sur le concept d'ensembles de niveau. Ces derniers définissent une plage de valeurs pour une paire aléatoire  $(X, Y)$  telle que la probabilité de leur apparition est supérieure à une certaine probabilité fixée  $p$ . Bien que plusieurs événements partagent le même niveau de risque  $p$ , tous ces événements n'ont pas la même probabilité d'occurrence. Sur la base de cette constatation, Volpi et Fiori [37] proposent de construire un intervalle de confiance afin d'identifier un ensemble de valeurs de  $(X, Y)$  dont le niveau de risque est  $p$  et dont la probabilité d'occurrence appartient à un ensemble de confiance de niveau  $1 - \alpha$ . Cette stratégie permet, entre autres, d'exclure les valeurs hautement improbables de  $X$  et de  $Y$ .

Les travaux de Volpi et Fiori [37] impliquent de connaître les marges uniformes, ce qui n'est pas toujours réalisable. De plus, le cheminement et les formules utilisées permettant d'obtenir l'intervalle de confiance ne sont pas explicités. En se basant sur leurs idées, une approche semi-paramétrique basée sur les copules est proposée, dans laquelle les marges empiriques sont utilisées. En isolant une variable dans l'équation de la courbe de niveau et en utilisant les distributions conditionnelles, un intervalle de confiance explicite et souvent simple apparaît alors pour des copules explicites tel que la famille des copules Archimédiennes. Un intervalle de confiance pour les copules implicites, telles que la Normale, Student ou encore Beta, peut être obtenu en recourant à l'intégration numérique. Ces intervalles sont généraux et leur inversion par les marges empiriques permettent d'obtenir un intervalle de confiance applicable. Une illustration a été effectuée sur le cours des actions de *ExxonMobil* et *Schlumberger* afin de montrer une des utilisations de la méthode introduite.

Dans le futur, il serait intéressant d'étendre la méthodologie développée dans ce mémoire pour des dimensions supérieures à  $d = 2$ , c'est-à-dire pour des vecteurs aléatoires  $\mathbf{X} = (X_1, \dots, X_d)$ . En complémentarité, il serait également souhaitable d'adapter la méthode dans le cas de copules dont les intervalles de confiance n'admettent pas de forme explicite. On peut penser notamment aux copules de type *squared* introduites par Durocher et Quessy [11], qui comprennent entre autres les familles des copules Khi-deux (voir Quessy *et al.* [28]) et Fisher (voir Favre *et al.* [12]). Enfin, en cette ère des données massives, il serait de mise d'étudier la fiabilité de la méthodologie pour des jeux de données de grande taille et comprenant un nombre élevé de variables.

# Bibliographie

- [1] Tarik BAHRAOUI, Taoufik BOUEZMARNI et Jean-François QUESSY : A family of goodness-of-fit tests for copulas based on characteristic functions. *Scand. J. Stat.*, 45(2):301–323, 2018.
- [2] Philippe BARBE, Christian GENEST, Kilani GHOUIDI et Bruno RÉMILLARD : On Kendall’s process. *J. Multivariate Anal.*, 58(2):197–229, 1996.
- [3] F. BELZUNCE, A. CASTAÑO, A. OLVERA-CERVANTES et A. SUÁREZ-LLORENS : Quantile curves and dependence structure for bivariate distributions. *Comput. Statist. Data Anal.*, 51(10):5112–5129, 2007.
- [4] Manuela Irene BRUNNER, Jan SEIBERT et Anne-Catherine FAVRE : Bivariate return periods and their importance for flood peak and volume estimation. *Wiley Interdisciplinary Reviews : Water*, 3(6):819–833, 2016.
- [5] E. CASTILLO et J. GALAMBOS : Conditional distributions and the bivariate normal distribution. *Metrika*, 36(3-4):209–214, 1989.
- [6] D. G. CLAYTON : A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151, 1978.
- [7] M. COBLENZ, R. DYCKERHOFF et O. GROTHE : Nonparametric estimation of multivariate quantiles. *Environmetrics*, 29(2):e2488, 23, 2018.

- [8] Maximilian COBLENZ, Oliver GROTHE, Manuela SCHREYER et Wolfgang TRUTSCHNIG : On the length of copula level curves. *J. Multivariate Anal.*, 167:347–365, 2018.
- [9] Peng DING : On the conditional distribution of the multivariate  $t$  distribution. *Amer. Statist.*, 70(3):293–295, 2016.
- [10] Nguyen Viet DUNG, B MERZ, A BÁRDOSSY et H APEL : Handling uncertainty in bivariate quantile estimation—an application to flood hazard analysis in the mekong delta. *Journal of Hydrology*, 527:704–717, 2015.
- [11] Martin DUROCHER et Jean-François QUÉSSY : Goodness-of-fit tests for copula-based spatial models. *Environmetrics*, 28(5):e2445, 19, 2017. ISSN 1180-4009.
- [12] Anne-Catherine FAVRE, Jean-François QUÉSSY et Marie-Hélène TOUPIN : The new family of Fischer copulas to model upper tail dependence and radial asymmetry : properties and application to high-dimensional rainfall data. *Environmetrics*, 29(3):e2494, 17, 2018. ISSN 1180-4009.
- [13] Jean-David FERMANIAN : Goodness-of-fit tests for copulas. *J. Multivariate Anal.*, 95(1):119–152, 2005.
- [14] M. J. FRANK : On the simultaneous associativity of  $F(x, y)$  and  $x + y - F(x, y)$ . *Aequationes Math.*, 19(2-3):194–226, 1979.
- [15] C. GENEST, K. GHOUIDI et L.-P. RIVEST : A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995.
- [16] Christian GENEST, Jean-François QUÉSSY et Bruno RÉMILLARD : Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.*, 33(2):337–366, 2006.
- [17] Christian GENEST, Bruno RÉMILLARD et David BEAUDOIN : Goodness-of-fit tests for copulas : a review and a power study. *Insurance Math. Econom.*, 44(2):199–213, 2009.

- [18] Christian GENEST et Louis-Paul RIVEST : Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.*, 88(423):1034–1043, 1993.
- [19] Christian GENEST et Louis-Paul RIVEST : On the multivariate probability integral transformation. *Statist. Probab. Lett.*, 53(4):391–399, 2001.
- [20] B GRÄLER, S VANDENBERGHE, MJ van den BERG, S GRIMALDI, A PETROSELLI, B DE BAETS et NEC VERHOEST : Multivariate return periods based on vine copulas. In *EGU General Assembly Conference Abstracts*, page 4097, 2012.
- [21] Benedikt GRÄLER, MJ VAN DEN BERG, Sander VANDENBERGHE, Andrea PETROSELLI, Salvatore GRIMALDI, Bernard DE BAETS et NEC VERHOEST : Multivariate return periods in hydrology : a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4):1281–1296, 2013.
- [22] Emil J. GUMBEL : Bivariate logistic distributions. *J. Amer. Statist. Assoc.*, 56:335–349, 1961.
- [23] M G. KENDALL : A new measure of rank correlation. *Biometrika*, 30:81–93, 1938.
- [24] Vít KLEMEŠ : Tall tales about tails of hydrological distributions. i. *Journal of Hydrologic Engineering*, 5(3):227–231, 2000.
- [25] Vít KLEMEŠ : Tall tales about tails of hydrological distributions. ii. *Journal of Hydrologic Engineering*, 5(3):232–239, 2000.
- [26] Mhamed MESFIOUI, Jean-François QUESSY et Marie-Hélène TOUPIN : On a new goodness-of-fit process for families of copulas. *Canad. J. Statist.*, 37(1):80–101, 2009.
- [27] Roger B. NELSEN : *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second édition, 2006.

- [28] Jean-François QUÉSSY, Louis-Paul RIVEST et Marie-Hélène TOUPIN : On the family of multivariate chi-square copulas. *J. Multivariate Anal.*, 152:40–60, 2016.
- [29] AI REQUENA, L MEDIERO et L GARROTE : A bivariate return period based on copulas for hydrologic dam design : accounting for reservoir routing in risk estimation. *Hydrology and Earth System Sciences*, 17(8):3023–3038, 2013.
- [30] G. SALVADORI : Bivariate return periods via 2-copulas. *Stat. Methodol.*, 1(1-2):129–144, 2004. ISSN 1572-3127.
- [31] G SALVADORI et Carlo DE MICHELE : On the use of copulas in hydrology : theory and practice. *Journal of Hydrologic Engineering*, 12(4):369–380, 2007.
- [32] Gianfausto SALVADORI, Carlo DE MICHELE et Fabrizio DURANTE : On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences*, 15(11):3293–3305, 2011.
- [33] Gianfausto SALVADORI, Fabrizio DURANTE et Elisa PERRONE : Semi-parametric approximation of Kendall’s distribution function and multivariate return periods. *J. SFdS*, 154(1):151–173, 2013.
- [34] Joanna H. SHIH et Thomas A. LOUIS : Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384–1399, 1995.
- [35] Vijay P SINGH, Sharad K JAIN et Aditya TYAGI : Risk and reliability analysis : a handbook for civil and environmental engineers. American Society of Civil Engineers, 2007.
- [36] Abe SKLAR : Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.
- [37] E VOLPI et A FIORI : Design event selection in bivariate hydrological frequency analysis. *Hydrological Sciences Journal*, 57(8):1506–1515, 2012.

- [38] E VOLPI et A FIORI : Hydraulic structures subject to bivariate hydrological loads : Return period, design, and risk assessment. *Water Resources Research*, 50(2):885–897, 2014.
- [39] Xipei YANG, Edward W. FREES et Zhengjun ZHANG : A generalized beta copula with applications in modeling multivariate long-tailed data. *Insurance Math. Econom.*, 49(2):265–284, 2011.
- [40] Jiabo YIN, Shenglian GUO, Shaokun HE, Jiali GUO, Xingjun HONG et Zhangjun LIU : A copula-based analysis of projected climate changes to bivariate flood quantiles. *Journal of hydrology*, 566:23–42, 2018.