

Résumé

La copule représente d'une manière exhaustive la dépendance entre des variables aléatoires. L'information qu'elle délivre est malheureusement trop complexe pour pouvoir être interprétée en tant que lien de dépendance. Pour cette raison, des mesures d'associations numériques sont utilisées pour avoir un résumé interprétable de la copule un certain nombre de mesures d'association entre variables aléatoires (X_1, \dots, X_d) sont fonction uniquement de la copule \mathbb{C} . Les mesures les plus courantes sont le tau Kendall et le rhô de Spearman.

Dans ce présent travail, nous avons présenté des mesures d'association dans le cas continue, nous montrons que ses mesures peuvent se généraliser au cas multidimensionnelle, des mesures basées sur le rhô de Spearman, ainsi leurs propriétés.

L'intérêt de ce mémoire porté sur l'estimation non paramétrique du rhô de Spearman ainsi, trois estimateurs sont considérés $(\hat{\rho}_{i,n}, i = 1, 2, 3)$, nous montrons aussi que la copule empirique est utilisée pour tester l'indépendance et pour estimer un ensemble de mesures d'association, puis établir leur normalité asymptotique. Dans le dernier chapitre, nous avons proposé des distributions de Cauchy bivariées, ainsi leur utilisation pour obtenir la probabilité quadrant d'un couple de variable aléatoire $(Z_{\lambda_1}, Z_{\lambda_2})$ suit une loi normale asymétrique bivariée "Orthant probability of bivariate skew-normal distribution".

Abstract

The copula is show exhaustively the dependence between random variables. The information it delivers is unfortunately too complex to be interpreted as a relation of.

For this reason, measures of numerical associations are used to have an interpretable summary of the copula, a number of measures of association between random variables $\{X_1, \dots, X_d\}$, depend only on the copula \mathbb{C} . The most common measures are Kendall tau and Spearman's rho. In this work, we introduced measures of association in the continuous case, we show that its measures can be generalized to multidimensional cases, measures based on Spearman's rho, and their properties.

The interest of this dissertation focused on nonparametric estimating via the empirical copula of Spearman's rho and three estimators are considered ($\hat{\rho}_{i,n} = 1, 2, 3$), we also show that the empirical copula is used to test the independence and to estimate a set of measures of association and establish their asymptotic normality. In the last chapter, we proposed bivariate Cauchy distribution, and their use to show the formula of orthant probability of bivariate skew-normal distribution.

Introduction Générale

La copule est devenue un outil de base dans la modélisation des distributions multivariées la connaissance de cet outil probabiliste est essentielle à l'appréhension de nombreux domaines. Les concepts de copules et de structure de dépendance sont plutôt récents. Le mot même de copule n'était pas très présent dans la littérature jusqu'à peu.

Il s'agit, d'un outil de modélisation statistique mis en avant pour la première fois par Sklar (1959) de nombreux statisticiens se sont attachés à étudier les caractéristiques de ce nouvel outil. Nous présenterons un bref historique du développement et de l'étude des copules, ainsi que les principaux auteurs et résultats. Le mot copule, qui signifie lien en latin, a été employé pour la première fois par Sklar en 1959 dans le théorème décrivant la fonction qui joint ensembles des fonctions de distributions univariées à la forme multivariée de la fonction de distribution comme le note lui-même Sklar, avant d'être nommées ainsi, les fonctions copules apparaissaient déjà dans les travaux de Fréchet Dall'Aglio, Feron et un certain nombre d'autres sur l'étude des distributions multivariées avec fonctions marginales fixées. La majorité des propriétés basiques des copules peut être attribuée aux travaux d'Höfding (1940-1941).

Il étudia les propriétés d'invariance des mesures de dépendance, il a obtenu des inégalités optimales, fournissant des bornes supérieures et inférieures pour ces versions particulières des copules. Indépendamment de ces recherches, on trouve dans Fréchet (1951) plusieurs résultats voisins donnant lieu aujourd'hui, aux appellations comme les bornes de Fréchet-Hoeffding et les classes de Fréchet-Hoeffding (voir Nelsen, 1999) Cette notion de copule et ses propriétés ont été redécouvertes plus tard, vers le milieu des années 1970 par Kindorf et Sampson (1975) Galambos (1978) et Deheuvels (1978) qui leurs donnèrent le nom de (Fonction de dépendance). Il est possible de regrouper les champs d'utilisation de ce concept en plusieurs parties. Tout d'abord Sklar et Schweizer ont travaillé sur la théorie probabiliste des espaces métriques de 1958 à 1976 la plupart des résultats fondamentaux sur les copules ont été obtenus dans le cadre de l'étude des espaces métriques. D'autre part un deuxième champ d'utilisation pourrait porter le nom de mesure de dépendance Schweizer et Wolf (1981) ont pour la première fois présenté les correspondances entre copule et mesure de dépendance. Les copules étaient déjà présentes dans un grand nombre d'articles traitant de

la dépendance. Si nous devons en retenir qu'un, ce serait Höfding (1941) qui utilisait les copules pour étudier les mesures d'associations non paramétriques comme le tau de Kendall et le rhô de Spearman Deheuvels en 1979-1981 utilisa la fonction de dépendance empirique pour construire des tests non paramétriques d'indépendance. D'une façon explicite, les copules sont des fonctions de répartition particulières, qui lient les fonctions de répartition multivariées de lois de probabilités dans \mathbb{R}^d , pour $d \geq 2$ aux fonctions de répartition marginales de leurs coordonnées. Plus précisément pour étudier le comportement simultané des composantes d'un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$, on peut faire appel à la fonction de répartition jointe

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

qui caractérise totalement le comportement stochastique de \mathbf{X} par exemple, dans plusieurs applications de l'analyse multivariée classique, on émet souvent l'hypothèse que \mathbb{H} est la fonction d'une loi normale multivariée. Cet usage du modèle normal est souvent justifié par la simplicité relative des calculs à effectuer. Toutefois il est connue, notamment en économie, que la loi normale s'ajuste très mal à certains types de données.

Dans l'étude moderne de la dépendance, on utilise fréquemment le fait que toute l'information à propos de la structure de dépendance de \mathbf{X} est contenue dans une fonction appelée la copule. Le concept de copule est un outil puissant et flexible puis qu'il permet de modéliser la dépendance sans tenir compte de l'effet du comportement des marges, c'est à dire des fonctions de répartition F_1, \dots, F_d des variables X_1, \dots, X_d prises individuellement. Une partie importante des recherches effectuées sur les mesures d'association entre les variables aléatoires concerne le cas où ces variables sont continues. De plus, les copules permettent de résumer la structure de dépendance interne d'un vecteur. D'ailleurs les mesures les plus conventionnelles de dépendance peuvent être exprimées explicitement en fonction de la copule. En particulier le Tau de Kendall et le rhô de Spearman entre deux variables aléatoires X_1 et X_2 s'expriment, respectivement, par :

$$\tau(X_1, X_2) = \tau(\mathbb{C}) = 4 \int_{[0,1]^2} \mathbb{C}(u, v) d\mathbb{C}(u, v) - 1.$$

$$\rho(X_1, X_2) = \rho(\mathbb{C}) = 12 \int_{[0,1]^2} \mathbb{C}(u, v) dudv - 3.$$

Grâce à ces avantages, la théorie de copule trouve des applications dans de nombreux domaines, tels que ceux de la finance, de l'assurance et de l'actuariat. Le but de notre travail est concentré sur l'estimation non paramétrique de

mesure d'association multi-variée tel que rhô de Spearman $(\rho_i, i = 1, 2, 3)$, ainsi établir la normalité asymptotique.

Plan du mémoire

Les étapes de réalisations de ce mémoire sont énoncées dans la suite est organisé en deux parties, la première partie se divise en trois chapitres, la seconde partie en deux chapitre.

Dans le **premier chapitre**, après une introduction aux distributions multi-variées nous exposons les outils de base liés à la théorie des copules multivariées, nous fournissons un exposé de synthèse sur ce sujet, en y présentant les caractéristiques et les principaux théorèmes qui serviront de base aux chapitres suivants. Dans ce chapitre on y introduit aussi la familles des copules paramétriques, ce sont-là les copules les plus fréquemment retrouvées dans la littérature financier, et l'intérêt de chaque copule.

Le **deuxième chapitre**, présente deux autres coefficients de corrélation à savoir le tau de Kendall et le rhô de Spearman dans le cas bivariée, nous rappellerons les définitions du tau de Kendall et du rhô de Spearman dans le cas continu. Dans ce chapitre nous passerons en revue les différentes mesures de corrélation qui sont habituellement utilisées en littérature par exemple en finance de marché.

Le **troisième chapitre** (basé sur [14]) fera l'objet d'une extension multivariée du rhô de Spearman. Dans la section (3.1), nous allons présenter de l'extension des définitions de cette mesure au cas continu. Ce chapitre, décrit également une méthode d'estimation non paramétrique du rhô de Spearman via la copule empirique de plus nous énoncerons deux théorèmes de la convergence en lois du processus empirique de la copule

$$\mathbb{G}_n = \sqrt{n} \left(\tilde{\mathbb{C}}_n(\mathbf{u}) - \mathbb{C}(\mathbf{u}) \right)$$

vers un processus gaussien. En plus, nous étudions la normalité asymptotique des estimateurs $(\hat{\rho}_{i,n}, i = 1, 2, 3)$ ainsi de calculer leurs variances asymptotiques.

La dernière section contient des exemples qu'on utilise pour calculer les variances asymptotique et les mesures d'association dans les deux cas où la variance asymptotique d'une part est une constante d'autre part dépend d'un paramètre inconnue.

Le **quatrième chapitre** : Ce chapitre commence par présenter des définitions des lois normales et des lois normales asymétrique (Skew-Normal Distribution) ces modeles de lois seront utilisés dans le chapitre 5, nous donnons dans la deuxième section une représentation stochastique de la distribution normale asymétrique bivariée

Le **cinquième chapitre** (basé sur [7]) de cette seconde partie porte sur la généralisation de la distribution de Cauchy bivariée introduite par Fang (1990)

ainsi nous étudions la probabilité d'un vecteur dont ses composantes sont tous de même signe ce vecteur suit une loi normale asymétrique bivariée (Orthant probability of bivariat skew-normal distribution), discuté par Azzalini et Dalla Valle (1996).

Table des matières

Introduction	i
Résumé	i
Abstract	ii
Introduction Générale	iii
I Copules et mesures d'association	2
1 Définitions et propriétés des copules multivariées	3
1.1 Quelques rappels concernant les distributions multivariées et lois uniformes	3
1.1.1 Notations et définitions	3
1.1.2 Définition de la fonction de répartition conjointe	4
1.1.3 Lois marginales	4
1.1.4 Notion d'indépendance	4
1.1.5 Les bornes de Fréchet-Hoeffding	4
1.1.6 Définitions et propriétés de la loi uniforme	5
1.2 Copules et leurs propriétés	6
1.2.1 Introduction aux copules	6
1.2.2 Théorème de Sklar	8
1.2.3 La densité de la copule	8
1.2.4 Copule de survie (cas $d=2$)	12
1.2.5 Copules de survie et fonction de survie multivariée ($d>2$) .	13
1.2.6 La copule radialement symétrique	15
1.3 Modeles de copules	16

1.3.1	La copule d'indépendance	16
1.3.2	La borne supérieure de Fréchet $M(\mathbf{u})$	18
1.3.3	La borne inférieure de Fréchet $W(u, v)$	19
1.3.4	La copule de Farlie-Gumbel-Morgenstern (FGM)	21
1.3.5	Copules empiriques d'une loi multivariée	21
1.4	Copules elliptiques	23
1.4.1	La copule Normale (Gaussienne)	23
1.4.2	La copule de Student	26
1.5	Copules Archimédiennes ($d > 2$)	28
1.5.1	Copules Archimédiennes 2-dimensions	29
1.5.2	Exemples de copules Archimédiennes	29
2	Mesures d'association bidimensionnelle	34
2.1	Introduction	34
2.2	Les mesures de dépendance	35
2.2.1	La corrélation linéaire de (X, Y)	35
2.2.2	Mesures d'association non paramétriques bivariées	37
2.2.3	Les mesures de concordance	38
2.2.4	Tau de Kendall	44
2.2.5	Le rhô de Spearman	49
2.2.6	Liens entre le τ de Kendall et le ρ de Spearman	54
3	Mesures d'association multivariées	56
3.1	Introduction	56
3.2	Extentions multivariées du rhô de Spearman	57
3.2.1	La relation existante entre $\rho_1(C)$ et $\rho_2(C)$	60
3.3	Estimation non paramétrique de ρ_i par la copule empirique	63
3.4	Convergence en loi du processus empirique de la copule	67
3.5	La normalité asymptotique des estimateurs	70
3.5.1	Processus Gaussiens	74
3.6	Exemples (Variance asymptotique)	75
3.6.1	Exemple de la copule d'indépendance	75
3.7	Stabilisation de la variance par transformation :	84
3.7.1	La copule FGM	85
3.7.2	La copule de Kotz-J.(Copule de deux paramètres)	87

II	Distributions Continues bivariées	88
4	Définition et représentation de la loi normale asymétrique	89
4.1	La loi normale	89
4.1.1	Distribution normale bidimensionnelle	90
4.1.2	Distribution normale standard de dimension trois	90
4.1.3	La probabilité quadrant (Orthant probability)	91
4.2	Loi normale asymétrique (Skew-Normal Distribution)	92
4.2.1	Représentation de la loi normale asymétrique univariée	93
4.2.2	Représentation de la distribution normale asymétrique bi- variée	96
5	Généralisation de la distribution de Cauchy bivariée	97
5.1	Loi de Cauchy unidimensionnelle	97
5.1.1	Fonction de répartition	98
5.1.2	Espérance et écart type	98
5.2	Loi de Cauchy bivariée	99
5.3	Généralisation	99
5.4	La probabilité quadrant de la loi normale bivariée asymétrique	105

Première partie

Copules et mesures d'association

Chapitre 1

Définitions et propriétés des copules multivariées

1.1 Quelques rappels concernant les distributions multivariées et lois uniformes

1.1.1 Notations et définitions

- Dans tout ce qui suit nous adopterons la notation $\mathbf{X} = (X_1, \dots, X_d)$, vecteur de variables aléatoires de dimension d , $d \geq 2$.
- L'inégalité $\mathbf{X} \leq \mathbf{Y}$ signifie que $\forall i \in \{1, \dots, d\} \quad x_i \leq y_i$.
- Nous noterons par la suite \mathbb{R}^d , l'espace d-dimensionnel $\mathbb{R} \times \dots \times \mathbb{R}$, où $\mathbb{R} =]-\infty, +\infty[$
- La fonction $\mathbf{1}_A$ désigne la fonction indicatrice de l'ensemble A définie par

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases} \quad (1.1)$$

- On définit l'espace $\ell^\infty([0, 1]^d)$ comme espace de toutes les fonctions bornées à valeurs réelles sur $[0, 1]^d$ muni de la métrique uniforme :

$$m_1(f_1, f_2) = \sup |f_1(t_1) - f_2(t_2)|, \quad t \in [0, 1]^d \text{ et } f_1, f_2 \in \ell^\infty([0, 1]^d)$$

- Le produit cartésien de d -intervalles unidimensionnels de la forme :

$$[a, b]^d = [a, b] \times \dots \times [a, b] \subset \mathbb{R}^d, \quad a \leq b.$$

1.1.2 Définition de la fonction de répartition conjointe

Soit un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$ de dimension d . La **fonction de répartition conjointe** de \mathbf{X} est la fonction de d -variables définie par :

$$\mathbb{F}(x_1, \dots, x_d) \mapsto \mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d]. \quad (1.2)$$

1.1.3 Lois marginales

Les **fonctions de répartition marginales** sont les fonctions de répartition F_1, \dots, F_d , des variables aléatoires X_1, \dots, X_d , et sont définies par :

$$F_i(x_i) = \mathbb{P}(X_i \leq x_i) = \mathbb{F}(\infty, \dots, x_i, \infty, \dots, \infty), \quad i = 1, \dots, d. \quad (1.3)$$

Cette définition doit se comprendre comme la limite de \mathbb{F} lorsque chacune des composantes x_j ($j \neq i$) tend vers l'infini. Les fonctions de densités, lorsqu'elles existent, sont données par :

$$f_i(x_i) = \frac{\partial F_i(x_i)}{\partial x_i},$$

et

$$f(x_1, \dots, x_d) = \frac{\partial \mathbb{F}(x_1, \dots, x_d)}{\partial x_1 \dots \partial x_d}, \quad i = 1, \dots, d. \quad (1.4)$$

1.1.4 Notion d'indépendance

Les composantes du vecteur \mathbf{X} sont mutuellement indépendantes si pour tout $(x_1, \dots, x_d) \in \mathbb{R}^d$:

$$\mathbb{F}(x_1, \dots, x_d) = \prod_{i=1}^d F_i(x_i).$$

Si \mathbf{X} a pour densité f , on aura :

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

1.1.5 Les bornes de Fréchet-Hoeffding

On définit la classe de Fréchet associée à F_1, \dots, F_d par l'ensemble :

$$\mathcal{F}(F_1, \dots, F_d) = \{\mathbb{F}, \mathbb{F} : \mathbb{R} \rightarrow [0, 1]^2\}.$$

Deux éléments de cette classe sont :

1. La borne supérieure de Fréchet définie par :

$$M(x_1, \dots, x_d) = \min \{F_1, \dots, F_d\}.$$

2. La borne inférieure de Fréchet définie par

$$W(x_1, \dots, x_d) = \max \left\{ \sum_{i=1}^d F_i(x_i) + 1 - d, 0 \right\}.$$

Toute fonction de répartition multivariée est comprise entre ces deux bornes

$$W(x_1, \dots, x_d) \leq \mathbb{F}(x_1, \dots, x_d) \leq M(x_1, \dots, x_d).$$

1.1.6 Définitions et propriétés de la loi uniforme

Soit U une variable aléatoire à valeur dans $[0, 1]$ est dite uniformément répartie sur $[0, 1]$, si elle est absolument continue et la densité est définie par :

$$f(x) = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 0 & \text{si } x \notin [0, 1] \end{cases}$$

L'esperance $E(U) = \frac{1}{2}$, la variance $V(U) = \frac{1}{12}$.

Définition 1.1.1 Soit \mathbb{F} une fonction de répartition de dimension 1. Le quantile de \mathbb{F} est une fonction, noté \mathbb{F}^{-1} , de domaine de définition $I = [0, 1]$, telle que :

1- Pour tout $t \in \text{Rang}\mathbb{F}$.

$$\mathbb{F}(x) = t \Leftrightarrow \mathbb{F}(\mathbb{F}^{-1}(t)) = t. \quad (1.5)$$

2- Dans tous les cas alors $\mathbb{F}^{-1}(t)$ est définie par

$$\mathbb{F}^{-1}(t) = \inf \{x : \mathbb{F}(x) \geq t\} = \sup \{x : \mathbb{F}(x) \leq t\}. \quad (1.6)$$

Si \mathbb{F} est strictement croissante, alors sa fonction de quantiles \mathbb{F}^{-1} (l'inverse généralisée) coïncide avec l'inverse habituelle de \mathbb{F} pour la composition des applications. Nous rappelons une proposition très utile, parmi les classiques de la théorie de quantile.

Proposition 1.1.1 Soit X une variable aléatoire de fonction de répartition \mathbb{F} alors

1. Si $U \sim U(0, 1)$ alors $\mathbb{F}^{-1}(U) \sim \mathbb{F}$;
2. Si \mathbb{F} est continue, alors $\mathbb{F}(X) \sim U(0, 1)$.

1.2 Copules et leurs propriétés

1.2.1 Introduction aux copules

La dépendance entre les variables aléatoires est parfaitement décrite par leur distribution conjointe. Nous pouvons cependant distinguer les comportements des distributions marginales de la structure de la dépendance; la copule est l'outil permettant d'extraire la dépendance d'une distribution conjointe et ainsi de séparer la dépendance et le comportement marginal.

Dans ce premier chapitre, nous introduisons quelques définitions de base qui nous permettent de mieux décrire le concept de copule. Nous présentons les propriétés les plus importantes de la fonction de copule et son rôle dans l'étude de l'indépendance des variables aléatoires; ainsi que son lien avec la théorie de la mesure. On peut trouver l'élaboration de la théorie des copules dans l'ouvrage de Nelsen (1999). «An introduction to copulas». Nous intéressons plus particulièrement à l'étude des copules en dimensions $d \geq 3$.

Définition 1.2.1 Une copule d -dimensionnelle est une fonction de répartition multivariée ayant les lois marginales uniforme sur $[0, 1]$ en d'autre terme, si les variables aléatoires U_1, \dots, U_d sont de loi $U_{[0,1]}$, alors la fonction $\mathbb{C} : [0, 1]^d \mapsto [0, 1]$ est une copule si :

$$\mathbb{C}(u_1, \dots, u_d) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d), \quad (1.7)$$

est vérifie les propriétés suivantes :

- (i) Pour tout $\mathbf{u} \in [0, 1]^d$, on a : $\mathbb{C}(\mathbf{u}) = 0$, si au moins une coordonnée de \mathbf{u} est égale à 0;
- (ii) Pour tout $u_i \in [0, 1]$, et pour $i = 1, \dots, d$, on a $\mathbb{C}(1, \dots, 1, u_i, 1, \dots, 1) = u_i$;
- (iii) \mathbb{C} est d -croissante; c'est à dire, pour tout $\mathbf{u} = (u_1, \dots, u_d)$, et $\mathbf{v} = (v_1, \dots, v_d)$ dans $[0, 1]^d$ tels que $u_i \leq v_i$, pour $i = 1, \dots, d$ on a :

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1 + \dots + i_d} \times \mathbb{C}(x_{1i_1}, \dots, x_{di_d}) \geq 0, \quad (1.8)$$

où $x_{1j} = u_j$ et $x_{2j} = v_j$ pour tout $j \in \{1, \dots, d\}$.

Remarque 1.2.1 Dans ce qui suit on va essayer de donner des explications concernant les propriétés de la définition (1.2.1)

1. La seconde propriété découle du fait que les fonctions marginales sont uniformes
2. Dans le cas bivarié, pour tout $\mathbf{u} = (u_1, u_2) \in [0, 1]^2$ et $\mathbf{v} = (v_1, v_2)^2 \in [0, 1]^2$, la propriété (iii) s'exprime par :

$$\begin{aligned}
(iii) &\Leftrightarrow \sum_{i_1=1}^2 \sum_{i_2=1}^2 (-1)^{i_1+i_2} \times \mathbb{C}(x_{1i_1}, x_{2i_2}) \geq 0 & (1.9) \\
&\Leftrightarrow \sum_{i_1=1}^2 \left[(-1)^{i_1+1} \times \mathbb{C}(x_{1i_1}, x_{21}) + (-1)^{i_1+2} \times \mathbb{C}(x_{1i_1}, x_{22}) \right] \geq 0 \\
&\Leftrightarrow (-1)^2 \mathbb{C}(x_{11}, x_{21}) + (-1)^{1+2} \mathbb{C}(x_{11}, x_{22}) + (-1)^{1+2} \mathbb{C}(x_{12}, x_{21}) \\
&\quad + (-1)^4 \mathbb{C}(x_{12}, x_{22}) \geq 0 \\
&\Leftrightarrow \mathbb{C}(x_{11}, x_{21}) - \mathbb{C}(x_{11}, x_{22}) - \mathbb{C}(x_{12}, x_{21}) + \mathbb{C}(x_{12}, x_{22}) \geq 0 \\
&\Leftrightarrow \mathbb{C}(u_2, v_2) - \mathbb{C}(u_2, v_1) - \mathbb{C}(u_1, v_2) + \mathbb{C}(u_1, v_1) \geq 0,
\end{aligned}$$

où $x_{1j} = u_j$ et $x_{2j} = v_j$, pour tout $j \in \{1, 2\}$

3. Soient $\mathbf{u} = (u_1, u_2) \in [0, 1]^2$ et $\mathbf{v} = (v_1, v_2) \in [0, 1]^2$, tels que :
 $0 \leq u_1 \leq u_2 \leq 1$ et $0 \leq v_1 \leq v_2 \leq 1$. D'après (1.9), on a

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

Ce qui prouve que les applications $y \mapsto \mathbb{C}(u_2, y) - \mathbb{C}(u_1, y)$ et $x \mapsto \mathbb{C}(x, v_2) - \mathbb{C}(x, v_1)$

sont croissantes sur $I = [0, 1]$

4. La connaissance de l'ensemble des copules des couples $(X_i, X_j)_{1 \leq i \neq j \leq d}$ ne fournit pas une information complète sur la copule d'un vecteur aléatoire $X = (X_1, \dots, X_d)$.

Il découle de la définition (1.2.1) que chaque marginale k -dimensionnelle d'une copule d -dimensionnelle est elle-même une k -copule.

1.2.2 Théorème de Sklar

Ce théorème est fondamental dans la théorie des copules, il explique l'intérêt de la notion de copule, il suffit de connaître les fonctions marginales, puis les injecter dans une copule \mathbb{C} , pour obtenir la répartition multidimensionnelle. Toute la théorie des copules est fondée sur le théorème suivant :

Théorème 1.2.1 *Soit \mathbb{F} une fonction de répartition d -dimensionnelle de fonction de répartition marginales F_1, \dots, F_d alors il existe une d -copule \mathbb{C} telle que, pour tout $x \in \mathbb{R}^d$,*

$$\mathbb{F}(x) = \mathbb{F}(x_1, \dots, x_d) = \mathbb{C}(F_1(x_1), \dots, F_d(x_d)). \quad (1.10)$$

Si les fonctions F_1, \dots, F_d sont continues, alors \mathbb{C} est unique. Réciproquement ; si \mathbb{C} est une copule et F_1, \dots, F_d sont les fonctions de répartition, alors la fonction \mathbb{F} définie par l'équation (1.10) est une fonction de répartition conjointe avec les marges F_1, \dots, F_d .

Corollaire 1.2.1 (Sklar) *Si \mathbb{F} est une fonction de répartition conjointe avec les marges F_1, \dots, F_d , alors $F_1^{-1}, \dots, F_d^{-1}$ sont les fonctions inverses de F_1, \dots, F_d , avec :*

$$F_j^{-1}(\mathbf{u}) = \inf \{ \mathbf{x}, F_j(\mathbf{x}) \geq \mathbf{u} \}, \quad u \in [0, 1].$$

Alors pour tout $(u_1, \dots, u_d) \in \text{Domaine de } \mathbb{C}$:

$$\mathbb{C}(u_1, \dots, u_d) = \mathbb{F}(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)). \quad (1.11)$$

Si $\mathbf{X} \sim \mathbb{F}$ et \mathbb{F} est continue alors $(F_1(x_1), \dots, F_d(x_d)) \sim \mathbb{C}$ et si $\mathbf{U} \sim \mathbb{C}$ alors : $(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \sim \mathbb{F}$.

1.2.3 La densité de la copule

Si nous supposons que les distributions marginales F_1, \dots, F_d , et \mathbb{C} différentiables, alors la densité jointe de la variable aléatoire $\mathbf{X} = (X_1, \dots, X_d)$, prend la forme suivante:

$$f(x_1, \dots, x_d) = f_1(x_1) \times \dots \times f_d(x_d) c(F_1(x_1), \dots, F_d(x_d)),$$

où, pour $1 \leq k \leq d$, f_k est la densité de probabilité dérivée de F_k , f est la densité jointe issue de \mathbb{F} et c est la densité de la copule \mathbb{C} définie par :

$$c(u_1, \dots, u_d) = \frac{\partial^d \mathbb{C}(\mathbf{u})}{\partial u_1 \partial u_2 \dots \partial u_d}.$$

En utilisant le théorème de Sklar et la relation (1.11) nous pouvons calculer l'expression de la densité c de la copule \mathbb{C} par :

$$c(u_1, \dots, u_d) = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{\prod_{i=1}^d f_i(F_i^{-1}(u_i))}.$$

Cette écriture va permettre d'estimer la densité d'une copule à partir de l'estimation des lois marginales et de loi jointe. Observons la d-croissance de la copule \mathbb{C} correspond à la positivité de la densité :

$$c(\mathbf{u}) = \frac{\partial^d \mathbb{C}(\mathbf{u})}{\partial u_1 \dots \partial u_d} \geq 0,$$

sont lorsque celle ci existe, voir la définition et propriétés des copules multivariées

Théorème 1.2.2 (*Continuité uniforme*) Une copule \mathbb{C} est uniformément continue sur son domaine. En particulier, pour tout \mathbf{u} et \mathbf{v} dans $[0, 1]^d$, nous avons :

$$|\mathbb{C}(\mathbf{v}) - \mathbb{C}(\mathbf{u})| \leq \sum_{k=1}^d |v_k - u_k|.$$

Théorème 1.2.3 (*Propriété d'invariance*) Soit (X_1, \dots, X_d) un vecteur de variables aléatoires continues, de fonctions de répartition \mathbb{F} associée à une copule \mathbb{C} et (T_1, \dots, T_d) est une suite de fonctions strictement croissantes. Alors, la fonction de répartition jointe du vecteur aléatoires $(T_1(X_1), \dots, T_d(X_d))$ est aussi associée à la même copule \mathbb{C} .

Ainsi, la copule est invariante par les transformations strictement croissantes de variables aléatoires.

Preuve. Notons par F_i et G_i les fonctions de répartition univariées, respectives, des variables aléatoires X_i et $T_i(X_i)$. Soient \mathbb{C} et \mathbb{D} les copules associées respectivement, aux vecteurs aléatoires \mathbf{X} et $T(\mathbf{X})$ donc pour tout $i = 1, \dots, d$, on a :

$$\begin{aligned} G_i(x_i) &= \mathbb{P}[T_i(X_i) \leq x_i] = \mathbb{P}[X_i \leq T_i^{-1}(x_i)] \\ &= F_i(T_i^{-1}(x_i)) = (F_i \circ T_i^{-1})(x_i) \end{aligned}$$

Car les transformations T_i sont croissantes. Compte tenu de (1.10), nous avons, pour tout $\mathbf{u} \in [0, 1]^d$:

$$\begin{aligned} \mathbb{C}(u_1, \dots, u_d) &= \mathbb{P}(X_1 \leq F_1^{-1}(u_1), \dots, X_d \leq F_d^{-1}(u_d)) \\ &= \mathbb{P}(T_1(x_1) \leq T_1 \circ F_1^{-1}(u_1), \dots, T_d \circ F_d^{-1}(u_d)) \\ &= \mathbb{P}(T_1(x_1) \leq G_1^{-1}(u_1), \dots, T_d(x_d) \leq G_d^{-1}(u_d)) \\ &= \mathbb{D}(u_1, \dots, u_d). \end{aligned}$$

Comme les variables X_1, \dots, X_d sont continues alors $\text{Rang}F_1 = \dots = \text{Rang}F_d = I$ et donc $\mathbb{C} = \mathbb{D}$ sur I^d . ■

La propriété d'invariance joue un rôle important dans la pratique. En effet, les statisticiens sont amenés par fois à transformer les données pour mieux les exploiter par exemple nous avons les transformations suivantes :

$$\mathbb{C}(x_1, x_2) = \mathbb{C}(\ln(x_1), x_2), \quad \mathbb{C}(x_1, x_2) = \mathbb{C}(x_1, \sqrt{x_2}), \quad \mathbb{C}(x_1, x_2) = \mathbb{C}(\ln(x_1), \ln(x_2)),$$

donc l'application de transformation croissante ne modifie pas la copule mais seulement les marges

Théorème 1.2.4 Soit $\mathbb{C} : [0, 1]^d \rightarrow [0, 1]$ une copule. Les dérivées partielles de \mathbb{C} existent presque sûrement, pour tout $i = 1, \dots, d$ et pour tout $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$, on a :

$$0 \leq \frac{\partial \mathbb{C}(\mathbf{u})}{\partial u_i} \leq 1.$$

De plus, les fonctions

$$\mathbf{u} \mapsto \frac{\partial \mathbb{C}(\mathbf{u})}{\partial u_i},$$

sont non décroissantes presque par tout.

Théorème 1.2.5 (Les bornes de Fréchet-Hoeffding) Soit \mathbb{F} une fonction de répartition d -dimensionnelle d'un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$ et de fonctions de répartition marginales F_1, \dots, F_d pour toute d -copules \mathbb{C} associée à \mathbb{F} et pour tout $\mathbf{u} \in [0, 1]^d$, on a :

$$W(\mathbf{u}) = \max\left(\sum_{i=1}^d u_i - d + 1, 0\right) \leq \mathbb{C}(\mathbf{u}) \leq M(\mathbf{u}) = \min(u_1, \dots, u_d).$$

W est une copule dans le cas où $d = 2$.

Preuve. Soit \mathbf{U} un vecteur de variables aléatoires uniforme de fonction de répartition \mathbb{C} on a :

$$\begin{aligned}\mathbb{C}(u_1, \dots, u_d) &= \mathbb{P}\{U_1 \leq u_1, \dots, U_d \leq u_d\} \\ &= \mathbb{P}\left\{\bigcap_{i=1}^{i=d} (U_i \leq u_i)\right\} \leq \mathbb{P}\{U_i \leq u_i\} = u_i, \text{ pour tout } i = 1, \dots, d.\end{aligned}$$

La seconde inégalité provient du fait que :

$$\bigcap_{i=1}^{i=d} (U_i \leq u_i) \subset (U_i \leq u_i), \text{ pour tout } i = 1, \dots, d.$$

Par conséquent :

$$\mathbb{C}(\mathbf{u}) \leq \min(u_1, \dots, u_d).$$

Pour la première inégalité provient du fait que :

$$\begin{aligned}\mathbb{P}\left\{\bigcap_{i=1}^d (U_i \leq u_i)\right\} &= 1 - \mathbb{P}\left\{\bigcup_{i=1}^d (U_i \geq u_i)\right\} \\ &= 1 - \mathbb{P}\{U_1 \geq u_1, \text{ ou } U_2 \geq u_2, \dots, \text{ ou } U_d \geq u_d\} \\ &\geq 1 - \sum_{i=1}^d \mathbb{P}[U_i \geq u_i] \geq 1 - \sum_{i=1}^d [1 - \mathbb{P}\{(U_i \leq u_i)\}] \\ &= \sum_{i=1}^d u_i - d + 1\end{aligned}$$

D'autre part comme \mathbb{C} est défini sur $[0,1]^d$, alors

$$W(\mathbf{u}) = \max\left(\sum_{i=1}^d u_i - d + 1, 0\right) \leq \mathbb{C}(\mathbf{u}).$$

■

Remarque 1.2.2 (Sur les copules)

1. La fonction M est une d -copule pour $d > 2$.
2. La fonction W n'est pas une d -copule pour $d > 2$.
3. Le théorème de Fréchet-Hoeffding implique qu'une copule est bornée, W est la borne inférieure M est la borne supérieure.

4. Pour les distributions conjointes bivariées, les bornes inférieure et supérieure de Fréchet-Hoeffding W et M sont les copules des vecteurs. $(U, 1 - U)$ et (U, U) , U étant de loi uniforme sur $[0, 1]$ M représente la parfaite dépendance positive.
5. Le théorème de Sklar et le théorème “Borne de Fréchet-Hoeffding” ont de nombreuses implications pratiques, vu que les mesures obtenues par les copules sont bornées, cela implique que ces mesures existent toujours .
6. Les copules sont de plus invariantes sous les transformations monotones, croissantes de leurs marges.

Dans cette section nous introduisons la copule de survie et sa relation avec le concept de symétrie.

1.2.4 Copule de survie (cas $d=2$)

Soit le vecteur (X, Y) de variables aléatoires de distribution jointe \mathbb{H} de copule \mathbb{C} et soient \mathbb{F} et \mathbb{G} les fonctions marginales de X et Y respectivement, notons par $\bar{\mathbb{H}}$ la fonction jointe de survie et par $\bar{\mathbb{F}}$ et $\bar{\mathbb{G}}$ les marges univariées.

Définition 1.2.2 La copule noté $\tilde{\mathbb{C}}$ définie par:

$$\tilde{\mathbb{C}}(u, v) = u + v - 1 + \mathbb{C}(1 - u, 1 - v), \quad (u, v) \in [0, 1]^2,$$

dite copule de survie. Il est facile de montrer que $\tilde{\mathbb{C}}$ est bien une copule satisfaisant

$$\bar{\mathbb{H}}(x, y) = \tilde{\mathbb{C}}(\bar{\mathbb{F}}(x), \bar{\mathbb{G}}(y)) \quad (\text{C.f. Nelsen 1999}).$$

Preuve. Soient les deux événements $A = \{X \leq x\}$, $B = \{Y \leq y\}$, d’après les formules des probabilités on a :

$$\mathbb{P}(\bar{A} \cap \bar{B}) = \mathbb{P}(\overline{A \cup B}) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B).$$

On a donc :

$$\bar{\mathbb{H}}(x, y) = \bar{\mathbb{F}}(x) + \bar{\mathbb{G}}(y) - 1 + \bar{\mathbb{C}}(\mathbb{F}(x), \mathbb{G}(y)) = \tilde{\mathbb{C}}(\bar{\mathbb{F}}(x), \bar{\mathbb{G}}(y)).$$

■

Il ne faut pas confondre la copule de survie $\tilde{\mathbb{C}}$ et la fonction de survie $\bar{\mathbb{C}}$ associée à \mathbb{C} :

$$\forall (u, v) \in I^d : \bar{\mathbb{C}}(u, v) = \bar{H}(x, y) = \mathbb{P}[U > u, V > v] = 1 - u - v + \mathbb{C}(u, v)$$

donc on peut trouver une relation entre $\tilde{\mathbb{C}}$ et \mathbb{C} , nous obtenons :

$$\tilde{\mathbb{C}}(u, v) = \bar{\mathbb{C}}(1 - u, 1 - v).$$

1.2.5 Copules de survie et fonction de survie multivariée ($d > 2$)

Théorème 1.2.6 *La relation existante entre $\tilde{\mathbb{C}}$ et $\bar{\mathbb{C}}$ est donnée par :*

$$\tilde{\mathbb{C}}(u_1, \dots, u_d) = \bar{\mathbb{C}}(1 - u_1, \dots, 1 - u_d) \quad (1.12)$$

$$= \mathbb{C}(1 - u_1, \dots, 1 - u_d) + \sum_{i=1}^d u_i - 1 \quad (1.13)$$

avec la fonction de survie multivariée $\bar{\mathbb{C}}$, donnée en fonction de \mathbb{C} par :

$$\begin{aligned} \bar{\mathbb{C}}(u_1, \dots, u_d) &= \mathbb{P}(U_1 > u_1, \dots, U_d > u_d) \\ &= 1 + \sum_{k=1}^d (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq d} C_{i_1 \dots i_k}(u_{i_1}, \dots, u_{i_k}), \end{aligned} \quad (1.14)$$

où $C_{i_1 \dots i_k}$ sont les $k^{\text{ème}}$ marges de la copule \mathbb{C} . Nous obtenons ce résultat en utilisant la formule des probabilités de Poincaré.

Preuve. Posons l'évènement

$$A_i = \{U_i \leq u_i, 1 \leq i \leq d\}.$$

Alors on a :

$$\begin{aligned} \bar{\mathbb{C}}(u_1, \dots, u_d) &= \mathbb{P}\{U_1 > u_1, \dots, U_d > u_d\} \\ &= \mathbb{P}\{\bar{A}_1 \cap \bar{A}_2 \dots \cap \bar{A}_d\} = \mathbb{P}\left\{\bigcap_{i=1}^d \bar{A}_i\right\} \\ &= \mathbb{P}\left\{\overline{\bigcup_{i=1}^d A_i}\right\} = 1 - \mathbb{P}\left\{\bigcup_{i=1}^d A_i\right\}, \quad 1 \leq i \leq d. \end{aligned}$$

D'après la formule (Inclusion-exclusion) des probabilités de Poincaré :

$$\mathbb{P} \left\{ \bigcup_{i=1}^d A_i \right\} = \sum_{k=1}^d (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq d} \mathbb{P}(A_{i_1} \cap A_{i_2} \dots \cap A_{i_k}).$$

Nous avons :

$$\bar{\mathbb{C}}(u_1, \dots, u_d) = 1 + \sum_{k=1}^d (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq d} C_{i_1 \dots i_k}(u_{i_1}, \dots, u_{i_k}),$$

où $C_{i_1 \dots i_k}$ désigne la copule marginale de \mathbb{C} reliée à (i_1, \dots, i_k) . ■

Définition 1.2.3 Soit \mathbf{X} une variable aléatoire et $\mathbf{a} \in \mathbb{R}$ tel que $\mathbf{X} - \mathbf{a} \stackrel{d}{=} \mathbf{a} - \mathbf{X}$ alors on dit que \mathbf{X} est symétrique autour de \mathbf{a} c'est-à-dire pour tout $\mathbf{x} \in \mathbb{R}$ on a :

$$\mathbb{P}(\mathbf{X} - \mathbf{a} \leq \mathbf{x}) = \mathbb{P}(\mathbf{a} - \mathbf{X} \leq \mathbf{x}). \quad (1.15)$$

Dans le cas où \mathbb{F} est continue l'équation (1.15) équivalente à :

$$\mathbf{X} \text{ est symétrique autour de } \mathbf{a} \Leftrightarrow \bar{\mathbb{F}}(\mathbf{a} - \mathbf{x}) = \mathbb{F}(\mathbf{a} + \mathbf{x}),$$

généralisons le cas

Définition 1.2.4 Soient X_1, \dots, X_d des variables aléatoires et $(a_1, \dots, a_d) \in \mathbb{R}^d$, le vecteur (X_1, \dots, X_d) est dite radialement symétrique autour de (a_1, \dots, a_d) si la fonction de distribution jointe de $(X_1 - a_1, \dots, X_d - a_d)$ à la même que $a_1 - X_1, \dots, a_d - X_d$ c'est-à-dire ;

$$(X_1 - a_1, \dots, X_d - a_d) \sim (a_1 - X_1, \dots, a_d - X_d)$$

Théorème 1.2.7 Soit le vecteur (X_1, \dots, X_d) de variables aléatoires avec la fonction de distribution jointe \mathbb{H} on dit que (X_1, \dots, X_d) est radialement symétrique autour de $(a_1, \dots, a_d) \in \mathbb{R}^d$ si et seulement si pour tout $i \in \{1, \dots, d\}$, et $x_i \in \mathbb{R}$

$$\mathbb{H}(a_1 + x_1, \dots, a_d + x_d) = \bar{\mathbb{H}}(a_1 - x_1, \dots, a_d - x_d), \quad (1.16)$$

où

$$\bar{\mathbb{H}}(x_1, \dots, x_d) = \mathbb{P}(X_1 > x_1, \dots, X_d > x_d).$$

et $\bar{\mathbb{H}}$ est la fonction de survie.

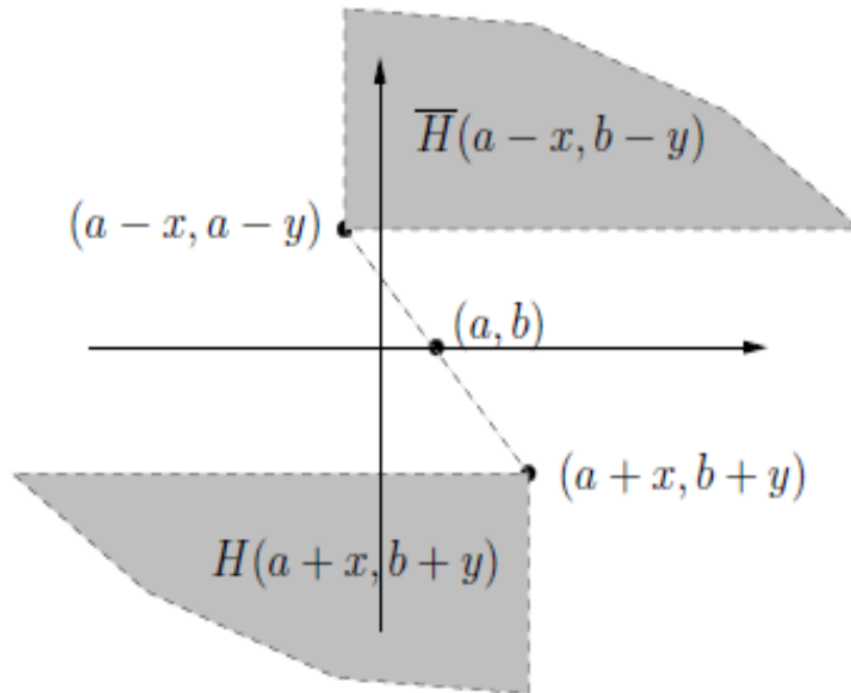


FIG. 1.1 – Régions de probabilités de $\mathbb{H}(a+x, b+y)$ et $\bar{\mathbb{H}}(a-x, b-y)$

1.2.6 La copule radialement symétrique

Définition 1.2.5 Soit \mathbb{C} la copule de dimension d et $\tilde{\mathbb{C}}$ sa copule de survie alors on dit que \mathbb{C} est radialement symétrique (radially symmetric) si et seulement si :

$$\tilde{\mathbb{C}}(\mathbf{u}) = \mathbb{C}(\mathbf{u}) \text{ pour tout } \mathbf{u} \in I^d, \quad (1.17)$$

ceci implique aussi que :

$$\begin{aligned} \mathbb{C} \text{ est radialement symétrique} &\Leftrightarrow \mathbb{C}(\mathbf{u}) = \mathbb{P}(\mathbf{U} \leq \mathbf{u}) \\ &= \mathbb{P}(\mathbf{U} > 1 - \mathbf{u}) \\ &= \tilde{\mathbb{C}}(1 - \mathbf{u}), \forall \mathbf{u} \in [0, 1]^d. \end{aligned} \quad (1.18)$$

1.3 Modeles de copules

Nous remarquons donc qu'une copule permet d'exprimer une fonction de répartition multivariée en fonction des fonctions de distributions marginales et que celle-ci contient toute l'information sur la structure de dépendance, les principales propriétés des copules ayant été définies, nous pouvons dès à présent donner quelques exemples de celles-ci, nous présentons les copules usuelles les plus célèbres et les principaux types de copules utilisées dans la pratique, ainsi que leurs propriétés, nous allons premièrement définir la copule la plus simple copule produit.

1.3.1 La copule d'indépendance

Définition 1.3.1 *Les composantes du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$ sont indépendantes si et seulement si :*

$$\mathbb{F}(x_1, \dots, x_d) = F_1(x_1) \times \dots \times F_d(x_d).$$

Nous définissons donc la copule d'indépendance par

$$\Pi(\mathbf{u}) = u_1 \times u_2 \times \dots \times u_d.$$

Cas où $d = 2$, alors :

$$\Pi(u, v) = u \times v.$$

La densité de cette copule est une constante

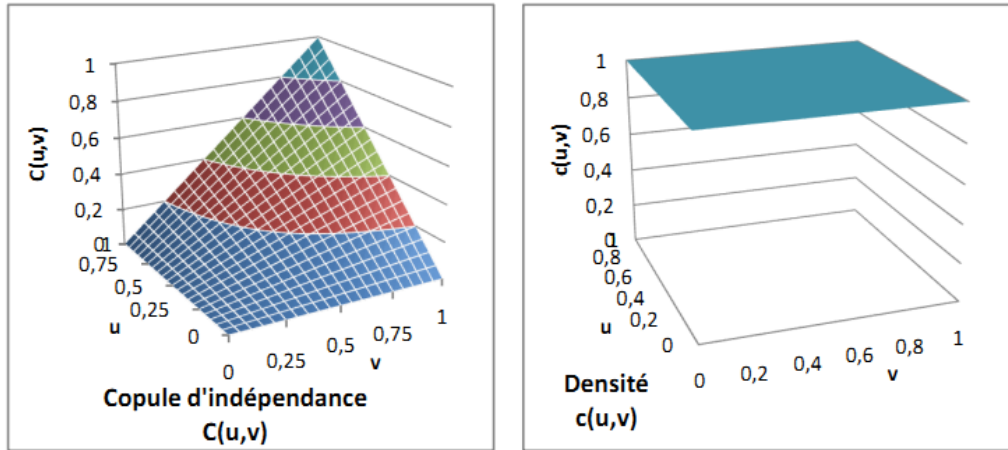


FIG. 1.2 – Représentation de la copule produit

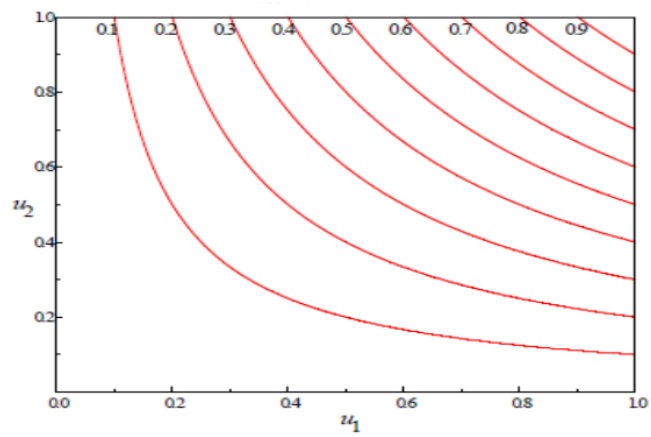


FIG. 1.3 – Contour plot de la copule produit

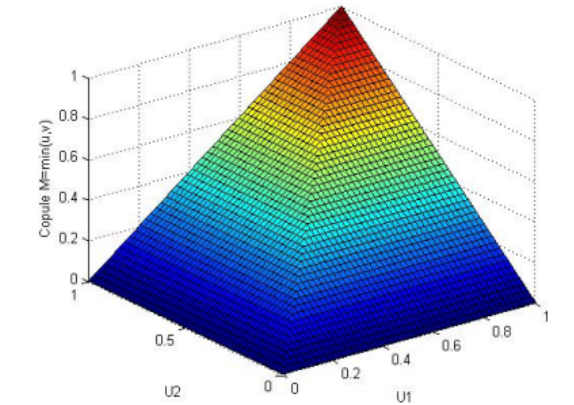


FIG. 1.4 – La représentation de la copule $M = \min(u, v)$

1.3.2 La borne supérieure de Fréchet $M(\mathbf{u})$

Définition 1.3.2 La copule comonote est définie comme suit :

$$M(\mathbf{u}) = M(u_1, \dots, u_d) = \min(u_1, \dots, u_d).$$

M et la fonction de répartition du vecteur (U, \dots, U) où U est uniforme

Preuve. Puisque :

$$P\{U \leq u_1, \dots, U \leq u_d\} = P\{U \leq \min(u_1, \dots, u_d)\} = \min(u_1, \dots, u_d).$$

Alors M est bien la copule associée au vecteur (U_1, \dots, U_d) , tel que

$$U_1 = U_2 = \dots = U_d = U. \quad \blacksquare$$

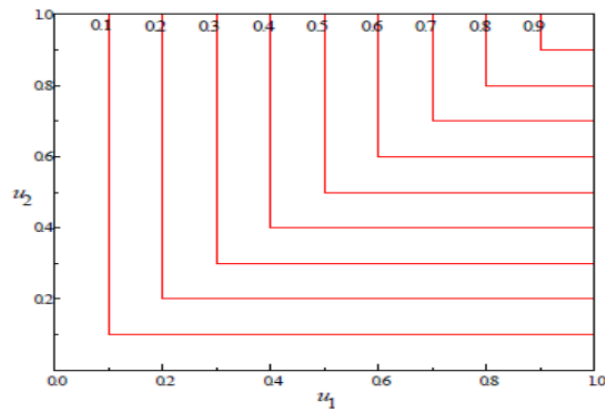


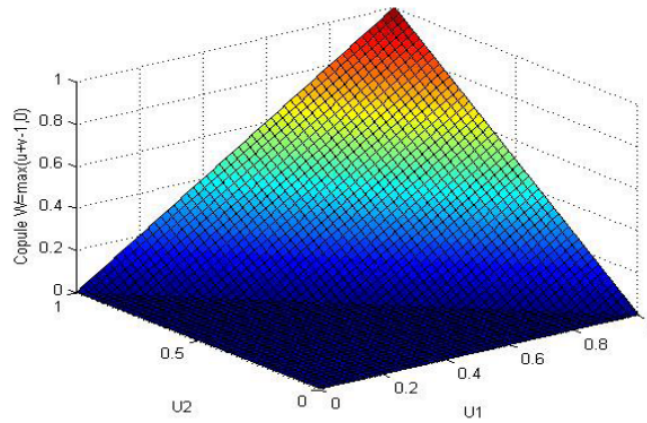
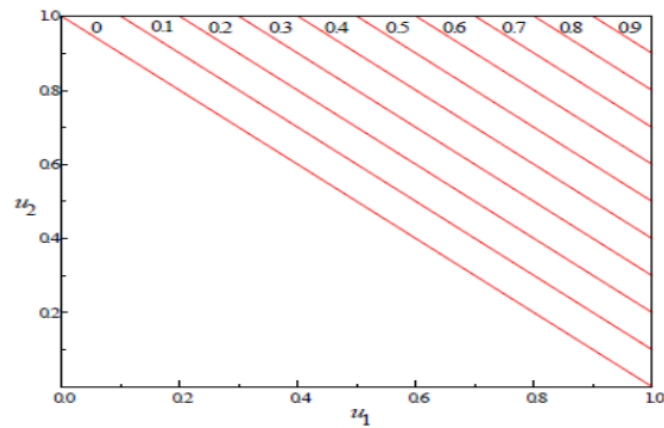
FIG. 1.5 – Contour plot de la copule $M = \min(u, v)$

1.3.3 La borne inférieure de Fréchet $W(u, v)$

Définition 1.3.3 W est définie pour tout u, v de $[0, 1]$ par : $W(u, v) = \max\{u + v - 1, 0\}$.

Remarque 1.3.1 Les bornes de Fréchet admet les propriétés suivantes :

1. La borne inférieure de Fréchet n'est une copule que dans le cas où $d = 2$
2. Les bornes de Fréchet n'admettent pas de densité.
3. Toutes les copules sont comprises entre ces deux bornes.
4. Les copules W et M sont appelées aussi copule minimale (respectivement copule maximale) et elle sont aussi notées respectivement C^- et C^+

FIG. 1.6 – La représentation de la copule $W = \max(u, v)$ FIG. 1.7 – Contour plot de la copule $W = \max(u, v)$

1.3.4 La copule de Farlie-Gumbel-Morgenstern (FGM)

Définition 1.3.4 La copule de Farlie-Gumbel-Morgenstern bivariée de paramètre α est une copule définie par la formule suivante :

$$\mathbb{C}_\alpha(u, v) = uv [1 + \alpha(1 - u)(1 - v)], \quad -1 \leq \alpha \leq 1,$$

dans le cas où $d = 3$ la copule FGM est donnée par la formule :

$$\mathbb{C}_\alpha(u, v, w) = uvw [1 + \alpha(1 - u)(1 - v)(1 - w)], \quad -1 \leq \alpha \leq 1.$$

Cette copule admet des copules marginales bivariées définies par la suite :

$$\begin{aligned} C_{12}(u, v) &= \mathbb{C}(u, v, 1) = uv, & C_{13}(u, w) &= \mathbb{C}(u, 1, w) = uw \\ C_{23}(v, w) &= \mathbb{C}(1, v, w) = vw. \end{aligned}$$

1.3.5 Copules empiriques d'une loi multivariée

La définition d'une fonction de répartition empirique est la suivante :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^d \mathbf{1}_{\{X_i \leq x\}} = \frac{\text{card}\{x_i \leq x\}}{n},$$

c'est dire le nombre d'observations d'un échantillon comprenant n élément notés (x_1, \dots, x_n) inférieures à x divisé par n . Dans le cas d'une distribution bivariée, en supposant l'échantillon noté $\{(x_1, y_1), \dots, (x_n, y_n)\}$, celle ci s'écrit :

$$\begin{aligned} F_n(x, y) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i \leq x\}} \mathbf{1}_{\{y_i \leq y\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i \leq x, y_i \leq y\}} \end{aligned}$$

En dimension d , si on se donne n -échantillons de taille d . Soit $X_j = \{X_{1j}, \dots, X_{dj}\}$ pour tout $j = 1, \dots, n$, une suite de vecteurs aléatoires indépendantes, identiquement distribuées de fonction de répartition jointe \mathbb{F} et de marges F_1, \dots, F_d nous définissons pour $d \geq 1$, la fonction de répartition empirique de cet échantillon et les fonctions de répartition empiriques marginales, respectivement, par

$$\begin{aligned} F_n(x_1, \dots, x_d) &= \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_{j1} \leq x_1, \dots, X_{jd} \leq x_d\}} \\ &= \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d \mathbf{1}_{\{X_{ij} \leq x_i\}}, \end{aligned}$$

et

$$\hat{F}_{i,n}(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{x_{ij} \leq x\}}, \text{ pour tout } 1 \leq i \leq d. \quad (1.19)$$

Les rangs des observations sont liés aux fonctions de répartition empirique marginales, pour $j \in \{1, \dots, n\}$, et $i \in \{1, \dots, d\}$, par :

$$n\hat{F}_{i,n} = R_{ij}$$

R_{ij} les rangs des X_{ij} dans $X_{1,n}, \dots, X_{i,n}$, l'estimateur F_n conduit à un estimateur non paramétrique naturel d'une copule. Nous pouvons introduire la copule empirique basé sur un échantillon multivariée $(X_{11}, \dots, X_{d1}), \dots, (X_{1n}, \dots, X_{dn})$, comme une fonction de répartition \mathbb{C}_n à lois marginales uniformes dans $[0, 1]^d$, qui peut s'exprimer en fonction des statistiques de rangs.

$$\mathbb{C}_n(u_1, \dots, u_d) = \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d \mathbf{1}_{\{\hat{U}_{ij} \leq u_i\}} = F_n \left\{ \hat{F}_{1,n}^{-1}(u_1), \dots, \hat{F}_{d,n}^{-1}(u_d) \right\}. \quad (1.20)$$

On trouvera une description sur ce sujet dans les travaux de Deheuvels (1979a, b, 1980, 1981), et connues sous le nom "Fonction empirique de dépendance"

Définition 1.3.5 Toute copule $\tilde{\mathbb{C}}_n$ définie sur l'ensemble :

$$\mathfrak{S} = \left\{ \left(\frac{k_1}{n}, \dots, \frac{k_d}{n} \right) : 1 \leq i \leq d, k_i = 0, \dots, n \right\}$$

par la fonction suivante :

$$\tilde{\mathbb{C}}_n \left(\frac{k_1}{n}, \dots, \frac{k_d}{n} \right) = \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d \mathbf{1}_{(R_{ij} \leq k_i)},$$

et une copule empirique multivariée.

Remarque 1.3.2 Deheuvels (2009) atteste que :

$$\text{Sup}_{\mathbf{u} \in [0,1]^d} \left| \tilde{\mathbb{C}}_n(\mathbf{u}) - \mathbb{C}_n(\mathbf{u}) \right| = \frac{1}{n}.$$

De ce fait, les resultats établis pour le processus empirique de copule $\mathbb{G}_n = \sqrt{n}(\mathbb{C}_n(\mathbf{u}) - \mathbb{C}(\mathbf{u}))$ restent aussi valable pour le processus $\tilde{\mathbb{G}}_n = \sqrt{n}(\tilde{\mathbb{C}}_n(\mathbf{u}) - \mathbb{C}(\mathbf{u}))$ par exemple la convergence faible du \mathbb{G}_n entraîne la convergence faible du $\tilde{\mathbb{G}}_n$.

Les copules empiriques que nous avons introduites dans ce chapitre sont notamment utiles pour fournir des estimateurs non paramétriques de mesures de dépendance à savoir, le tau de Kendall et le rhô de Spearman dans le cas où $d \geq 2$ nous revenons à ce sujet d'estimations non paramétriques dans le troisième chapitre, (**Mesures d'association multivariée du rhô de Spearman**)

1.4 Copules elliptiques

Dans cette partie nous étudions quelques copules paramétriques, c'est à dire des fonctions qui dépend d'un paramètre ou plusieurs paramètres ces copules sont d'un grand intérêt pour la gestion des risques car elle permet de construire des modèles paramétriques ou semi paramétriques, dans littérature financier plusieurs fonctions de copules ont été utilisées, les principales sont la copule normale la copule de Student et les copules dites copules Archimédiennes comme la famille de Clayton, de Frank et de Gumbel.

1.4.1 La copule Normale (Gaussienne)

Cette copule fait partie de la famille des copules elliptiques de même que la copule de Student. La copule Gaussienne ne présente pas de dépendance de queue et n'est donc pas adaptée à des valeurs extrêmes. Elles sont en effet moins bien adaptées en assurance car elles s'appliquent à des distributions symétriques .

Soit $\mathbf{X} = (X_1, \dots, X_d)$ un vecteur de variables aléatoires Gaussien de loi jointe Φ , de dimension d , de matrice de covariance Σ et de marginales Φ_1, \dots, Φ_d . Pour définir la copule normale, nous supposons que Φ , est centrée réduite ceci signifie que les marges de X_1, \dots, X_d ont toutes une loi normale $\mathcal{N}(0, 1)$. Nous considérons la matrice des coefficients de corrélation linéaire usuels Σ , définie par :

$$\Sigma = \begin{pmatrix} \rho_{11} & \dots & \rho_{1d} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \rho_{d1} & \dots & \rho_{dd} \end{pmatrix}$$

avec

$$\rho_{ij} = \frac{Cov(x_i, x_j)}{\sqrt{Var(x_i)}\sqrt{Var(x_j)}}, \quad |\rho_{ij}| \leq 1.$$

La copule Gaussienne multivariée s'applique à une fonction de répartition conjointe avec la matrice des corrélations Σ elle est définie par :

$$\begin{aligned} \mathbb{C}(u_1, \dots, u_d, \rho) &= \Phi [\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)] \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) dx_1 \dots dx_d, \end{aligned}$$

pour $d > 2$.

Proposition 1.4.1 Dans le cas bivarié, la copule \mathbb{C}_G associée à une loi Gaussienne jointe de matrice de corrélation Σ est donnée par :

$$\mathbb{C}_G(u, v) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho_{12}^2}} \exp\left(-\frac{s^2 + t^2 + 2\rho_{12}st}{2(1-\rho_{12}^2)}\right) ds dt.$$

La fonction de densité est :

$$c(u, v) = \frac{1}{\sqrt{1-\rho_{12}^2}} \exp\left(-\frac{x_1^2 + x_2^2 - 2\rho_{12}x_1x_2}{2(1-\rho_{12}^2)} + \frac{x_1^2 + x_2^2}{2}\right).$$

La fonction de distribution Φ^{-1} est l'inverse généralisée de la distribution normale centrée réduite Φ univariée avec $\Phi^{-1}(u_1) = x_1$, $\Phi^{-1}(u_2) = x_2$

Remarque 1.4.1 Pour la copule Gaussienne de dimension 2 nous avons

1. Le cas où $\rho = 0$ correspond au cas d'indépendance et la copule $\mathbb{C}_0^G = \Pi$
2. Si $\rho = -1$ alors $\mathbb{C}_{-1}^G = W$.
3. Si $\rho = +1$ alors $\mathbb{C}_1^G = M$

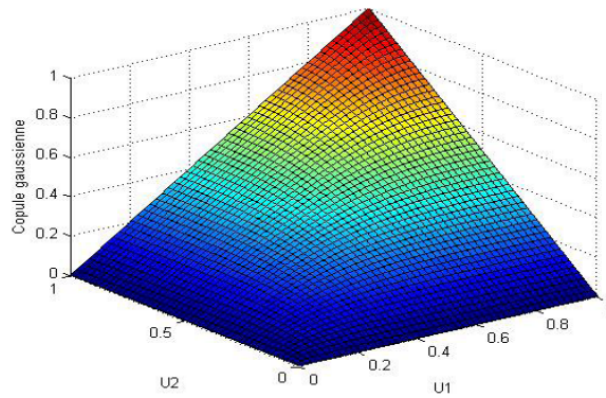


FIG. 1.8 – La représentation de la copule Gaussienne

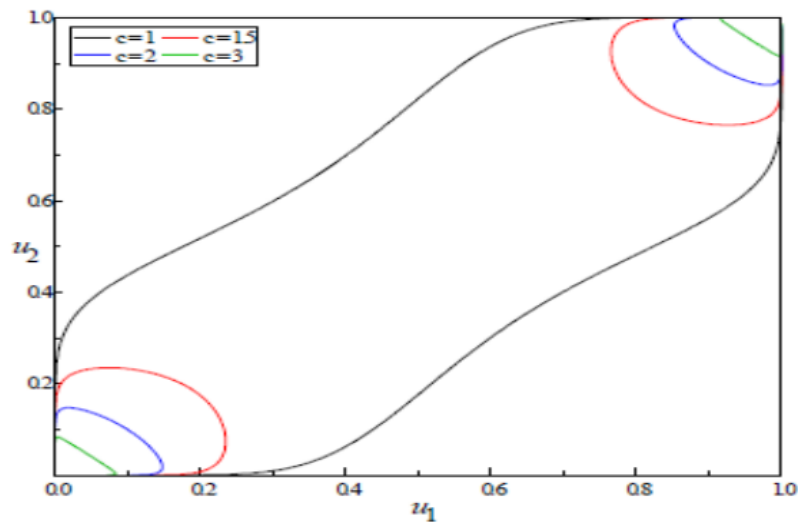


FIG. 1.9 – Contour plot de la copule Gaussienne du paramètre $\rho = 0,5$

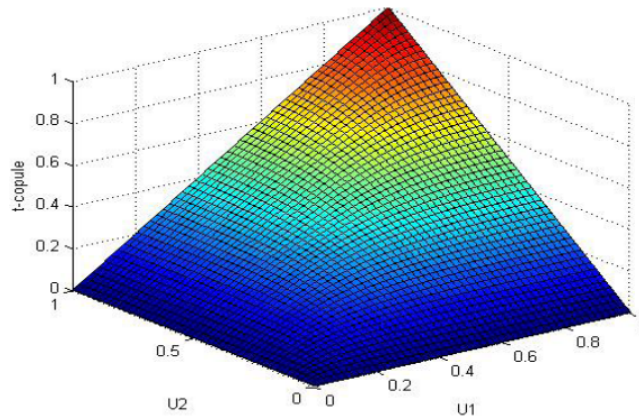


FIG. 1.10 – La représentation de la copule de Student

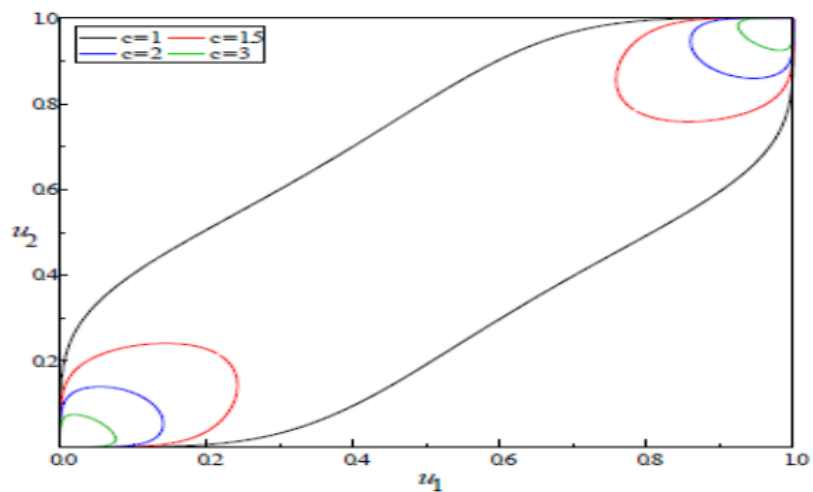
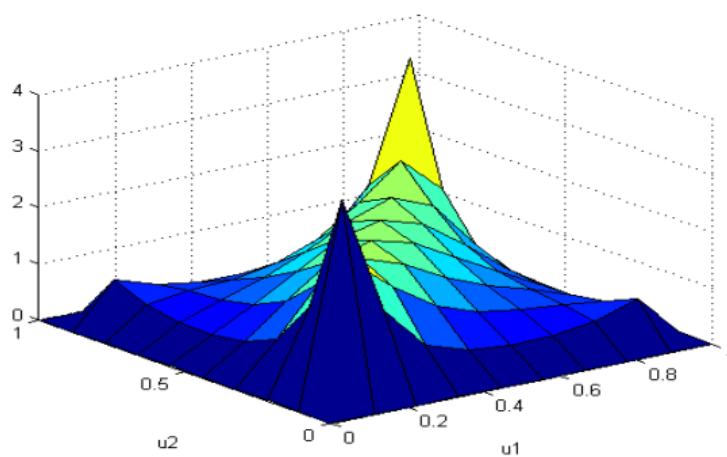
1.4.2 La copule de Student

La copule de Student est construite de la même manière que la copule Gaussienne mais à cette fois ci à partir de la distribution de Student t_ν multidimensionnelle standard à ν degrés de liberté et de matrice de corrélation ρ . Elle est donnée par la fonction

$$C(u_1, \dots, u_d, \rho, \nu) = t_\nu \{t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \dots, t_\nu^{-1}(u_d)\}.$$

Pour $d = 2$ on a :

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

FIG. 1.11 – Contour plot de la copule de Student avec $\rho = 0,5$ et $\nu = 8$ FIG. 1.12 – Densité de la copule de Student avec $\rho = 0,5$ et $\nu = 1$

1.5 Copules Archimédiennes ($d > 2$)

La famille des copules Archimédiennes est définie pour la première fois dans Genest et Mackay (1986 a) et ont l'avantage sur les copules elliptique. Ce sont des classes très populaires et celle sont utilisées en sciences actuarielle et en finance de marchés pour modéliser la dépendance entre les marchés financiers ou les classes d'actifs.

La classe de copules Archimédiennes est une classe importante cette famille de copules joue un rôle très important dans la théorie de statistiques paramétriques, il y a plusieurs raisons qui justifient leurs utilisations,

1. Grande variété de familles paramétriques.
2. Une grande possibilité de dépendance.
3. Elles peuvent être construite et simulées assez facilement.
4. Décrire des structures de dépendance très diverses dont notamment les dépendances dites asymétriques où les coefficients de queue inférieure et de queue supérieure différent.

Une copule Archimédiennes est construite à l'aide d'une fonction génératrice.

Définition 1.5.1 *On dit que ϕ est la fonction génératrice d'une copule à d -dimensions si :*

1. $\phi : [0, 1] \mapsto [0, \phi(0)]$ est décroissante avec $\phi(1) = 0$ et si $\phi(0) = \infty$, ϕ sera une fonction génératrice stricte.
2. Pour tout $0 < u < \phi(0)$ et pour tout $1 \leq j \leq d$,

$$(-1)^j \frac{\partial^j}{\partial u^j} \phi^{-1}(u) > 0. \quad (1.21)$$

Définition 1.5.2 *Étant donnée la définition ci dessus, une copule Archimédienne est construite de la façon suivante :*

$$\mathbb{C}(u_1, \dots, u_d) = \begin{cases} \phi^{[-1]} \left(\sum_{i=1}^d \phi(u_i) \right), & \text{si } \sum_{i=1}^d \phi(u_i) \leq \phi(0) \\ 0 & \text{sinon.} \end{cases} \quad (1.22)$$

1.5.1 Copules Archimédiennes 2-dimensions

Définition 1.5.3 Soit φ une fonction strictement décroissante, continue définie sur $\text{Dom}\varphi = [0, 1]$ et à valeurs dans $[0, \infty]$, telle que $\varphi(1) = 0$. La pseudo-inverse de φ est la fonction $\varphi^{[-1]}$ de domaine de définition $[0, \infty]$ et d'image $\text{Ran}(\varphi^{[-1]}) = [0, 1]$, donnée par :

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{(-1)}(t) = \inf \{x : \varphi(x) \leq t\}, & 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases} \quad (1.23)$$

Remarquons que $\varphi^{[-1]}$ est continue et non, décroissante sur $[0, \infty]$, et strictement croissante sur $[0, \varphi(0)]$. En outre $\varphi^{[-1]}(\varphi(u)) = u$ sur I et que :

$$\begin{aligned} \varphi(\varphi^{[-1]}(t)) &= \begin{cases} t, & \text{si } 0 \leq t \leq \varphi(0) \\ \varphi(0), & \text{si } \varphi(0) \leq t \leq \infty \end{cases} \\ &= \min(t, \varphi(0)). \end{aligned}$$

Ainsi si $\varphi(0) = \infty$, alors $\varphi^{-1} = \varphi^{[-1]}$

La méthodologie pour la construction des copules Archimédiennes est exposée dans l'ouvrage de Nelsen (1999). Il donne une importante liste de familles des copules dont la plus importante est la famille des copules Archimédiennes auxquelles, par exemple, les copules de Clayton, de Frank et de Gumbel, de plus en plus utilisées en finance, appartiennent.

1.5.2 Exemples de copules Archimédiennes

Les principales propriétés des copules ayant été définies, nous pouvons désormais présenter quelques exemples. Nous tâcherons de présenter les Copules les plus courantes. Dans ce qui suit, l'attention sera portée sur trois copules de la famille Archimédienne : Clayton, Frank, et Gumbel.

Exemple 1.5.1 Soit $\varphi(t) = -\ln t$, φ est continue strictement décroissante et convexe $\varphi(0) = +\infty$ et $\varphi(1) = 0$. Alors, son inverse est $\varphi^{-1}(t) = \exp(-t)$. La copule est définie par :

$$\mathbb{C}(u_1, u_2) = \exp[-(-\ln u_1 - \ln u_2)] = u_1 u_2 = \Pi(u_1, u_2)$$

ainsi la copule d'indépendance est Archimédienne, la formule généralisée est :

$$\prod_{i=1}^d (u_1, \dots, u_d) = u_1 \times \dots \times u_d = \exp[-(-\ln u_1 - \dots - \ln u_d)].$$

Exemple 1.5.2 Considérons la fonction génératrice $\varphi(t) = 1 - t$ on a $\varphi(0) = 1$ et son inverse :

$$\begin{aligned}\varphi^{-1}(t) &= \begin{cases} 1 - t, & t \in [0, 1] \\ 0, & t > 1 \end{cases} \\ &= \max\{1 - t, 0\}\end{aligned}$$

ainsi

$$\begin{aligned}\mathbb{C}(u_1, u_2) &= \max(1 - (1 - u_1 + 1 - u_2), 0) \\ &= \max\{u_1 + u_2 - 1, 0\} = W(u_1, u_2).\end{aligned}$$

Exemple 1.5.3 On peut définir des familles de copules à un paramètre

- **Famille de Frank** : Considérons $\varphi(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$, pour $\theta \in \mathbb{R} - \{0\}$, φ est continue et $\varphi(0) = +\infty$, $\varphi(1) = 0$, nous avons $\varphi'(t) = \frac{1 - \exp(\theta t)}{\theta} < 0$, sur $[0, 1]$ puisque si $\theta > 0$ alors $1 - \exp(\theta t) \leq 0$ et si $\theta < 0$, on a $1 - \exp(\theta t) > 0$.

Finalement $\varphi''(t) = \frac{\theta^2 \exp(\theta t)}{(1 - \exp(\theta t))^2} > 0$ donc φ est bien convexe, alors φ définit la fonction génératrice de Frank. La copule de Frank est définie par :

$$\mathbb{C}(u_1, u_2) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right), \quad \theta \in \mathbb{R} \setminus \{0\}.$$

Cette copule peut généraliser de la façon suivante :

$$\mathbb{C}(u_1, \dots, u_d) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1) \times \dots \times (e^{-\theta u_d} - 1)}{(e^{-\theta} - 1)^{d-1}} \right).$$

- **Famille de Clayton** : Considérons la fonction génératrice $\varphi(u) = (u^{-\theta} - 1)$ avec $\theta > 0$ et son inverse

$$\varphi^{-1}(s) = (s + 1)^{-\frac{1}{\theta}},$$

alors la copule associée sera pour $\theta > 0$

$$\mathbb{C}_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}.$$

- **Copule de Gumbel** : Soit la fonction $\varphi(u) = (-\ln(u))^\theta$, avec $\theta \geq 0$ et $u \in [0, 1]$. On remarque que φ est continue et $\varphi(1) = 0$, on a :

$$\begin{aligned} \varphi'(u) &= \frac{-\theta}{u} (-\ln(u))^{\theta-1} \quad \text{et} \quad \varphi'(u) < 0. \\ \varphi''(u) &= \frac{\theta}{u^2} (-\ln(u))^{\theta-2} [\theta - 1 - \ln(u)] \quad \text{et} \quad \varphi''(u) > 0. \end{aligned}$$

φ est une fonction continue strictement décroissante définie de $[0, 1]$ dans $[0, +\infty]$, convexe et elle est un générateur strict. De l'expression (1.20) on trouve :

$$\mathbb{C}_\theta(u, v) = \phi^{(-1)}[\varphi(u) + \varphi(v)] = \exp\left(-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{\frac{1}{\theta}}\right).$$

Cette copule peut généraliser de la façon suivante :

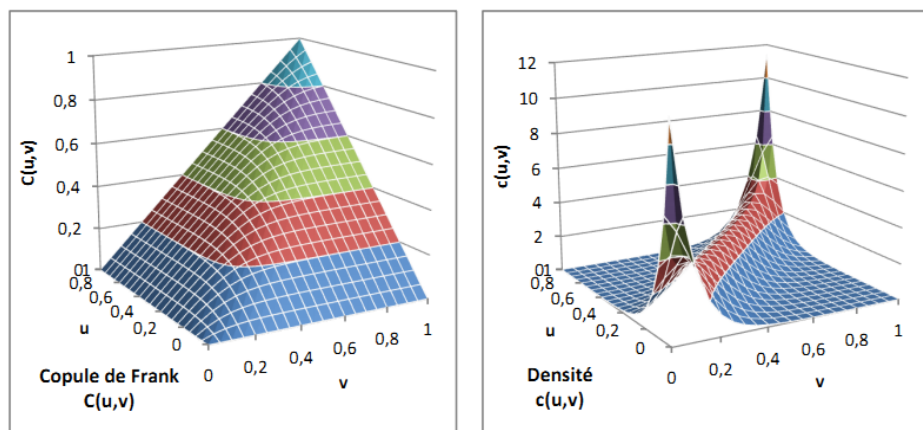
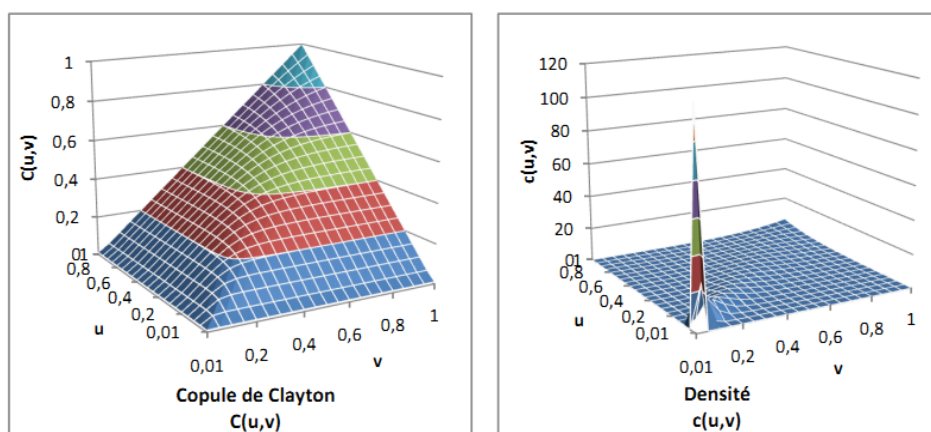
$$\mathbb{C}(\mathbf{u}) = \exp\left\{-\left[\sum_{i=1}^d (-\ln u_i)^\theta\right]^{\frac{1}{\theta}}\right\}.$$

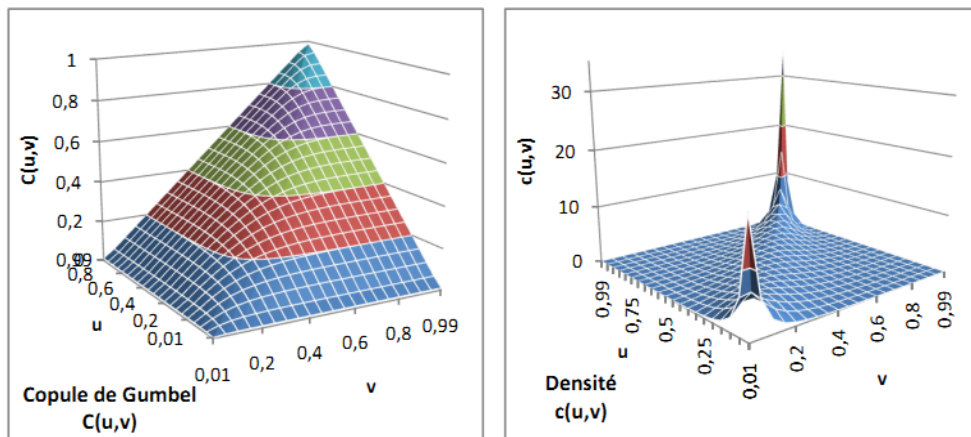
Le tableau suivant regroupe les caractéristiques des familles les plus connues. On rappelle que le paramètre θ mesure le degré de dépendance entre les deux variables ou plusieurs variables aléatoires. Plus il est élevé plus la dépendance est forte et une valeur positive de θ indique la dépendance positive

Copules	$\Phi(\mathbf{u})$	$\mathbb{C}(\mathbf{u})$
$\Pi(\mathbf{u})$	$-\ln \mathbf{u}$	$\prod_{i=1}^d u_i$
Gumbel	$(-\ln \mathbf{u})^\theta, \theta \geq 0$	$\exp\left\{-\left[\sum_{i=1}^d (-\ln u_i)^\theta\right]^{\frac{1}{\theta}}\right\}$
Frank	$-\ln \frac{\exp(-\theta \mathbf{u}) - 1}{\exp(-\theta) - 1}, \theta \in \mathbb{R} \setminus \{0\}$	$-\frac{1}{\theta} \ln \left(1 + \frac{\prod_{i=1}^d [\exp(-\theta u_i) - 1]}{(\exp(-\theta) - 1)^{d-1}}\right)$
Clayton	$u^{-\theta} - 1, \theta \geq 0$	$\left(\sum_{i=1}^d u_i^{-\theta} - d + 1\right)^{-\frac{1}{\theta}}$

Table 1.2 – Exemples de copules Archimédiennes.

Voici la représentation graphique de la fonction de densité et de la fonction de répartition :

FIG. 1.13 – La représentation de la copule de Frank ($\tau = 0,7$, $\theta = 11,4$)FIG. 1.14 – La representation de la copule de Clayton ($\tau = 0,7$, $\theta = 4.67$)

FIG. 1.15 – Représentation de la copule de Gumbel ($\tau = 0,7$, $\theta = 3.33$)

Chapitre 2

Mesures d'association bidimensionnelle

2.1 Introduction

Sans doute la relation de dépendance entre les variables aléatoires joue un rôle important dans plusieurs champs mathématiques. C'est une pratique largement répandue dans les milieux financiers et académiques. Une variété très vaste de ce concept a été étudiées par de nombreux auteurs pour proposer des définitions et des propriétés utiles pour avec des applications. Par exemple (le modèle d'évaluation des actifs financiers) utilisant la corrélation comme mesure de dépendance entre plusieurs instruments financiers et développant une théorie essentiellement basée sur la normalité multivariée des rendements afin de construire un portefeuille optimal bien que la corrélation linéaire soit l'un des concepts les plus présents, elle est aussi l'un des mal utilisée. En effet, la corrélation est seulement une mesure parmi d'autres (tau de Kendall, rho de Spearman,...) de la dépendance entre plusieurs variables.

Le coefficient de corrélation linéaire est la mesure canonique de la dépendance dans le cadre d'un modèle multivarié Gaussien et plus généralement pour les distributions elliptiques. Donc utiliser le coefficient de corrélation pour mesurer la dépendance en dehors de ce cadre peut être un générateur d'erreurs fréquentes. Le coefficient de corrélation linéaire est une mesure de dépendance facile à calculer il est très utile pour les familles de distributions elliptiques.

D'après Embrechts et al. (1999), la corrélation est naturellement utilisée

pour décrire la structure de dépendance lorsque la distribution jointe des indices financiers étudiés appartient à la famille des distributions elliptiques. Mais celui-ci perd de son intérêt si le modèle est différent.

1. Cette mesure de dépendance comporte plusieurs défauts en fait. Elle mesure que l'association linéaire entre deux variables aléatoires, de plus, cette mesure dépend du choix des marges.
2. Elle peut être près de 0 même en dépendance forte.
3. Le coefficient de corrélation linéaire n'est pas invariant lorsqu'on applique une transformation strictement croissante. $T : \mathbb{R} \longrightarrow \mathbb{R}$, c'est-à-dire :

$$\rho_l(T(X), T(Y)) \neq \rho_l(X, Y)$$

4. La corrélation linéaire ρ_l n'est pas définie si la variance de X et Y est infinie.

Dans ce chapitre nous rappelons un ensemble de notions de bases concernant la mesure d'association entre deux variables aléatoires, et le rôle que joue la copule. Comme nous allons le voir, la majorité des indices usuels de dépendance s'exprime uniquement en fonction de la copule, traduisant l'indépendance d'un quelconque indice de dépendance par rapport aux lois des marges.

Nous examinons les mesures d'association les plus connues tel que le coefficient de corrélation linéaire, rho de Spearman et le tau de Kendall, leurs propriétés, l'avantage et application. Premièrement nous concentrons sur la mesure d'association bivariée, puis après avoir présenté quelques notions de dépendance bidimensionnelles, on se penche sur la dépendance entre les composantes d'un vecteurs aléatoires (**Chapitre trois**)

2.2 Les mesures de dépendance

2.2.1 La corrélation linéaire de (X, Y)

Soient X et Y deux variables aléatoires. La covariance de X et Y est le nombre

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y).$$

Définition 2.2.1 Soient X et Y deux variables aléatoires ayant des variances finies. Le coefficient de corrélation entre X et Y se calcule par la formule suivante :

$$\rho(X, Y) = \frac{E(X - \bar{X}) E(Y - \bar{Y})}{\sqrt{\text{var}(X)} \sqrt{\text{var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}},$$

avec $\text{var}X > 0$, $\text{var}Y > 0$.

La définition du coefficient de corrélation linéaire est donc subordonnée à l'existence de des variances de X et Y .

Dans le cadre d'une dépendance linéaire parfaite, $Y = aX + b$, ($a \neq 0$, $b \in \mathbb{R}$), le coefficient de corrélation est égal à $+1$ ou -1 selon le signe de a .

D'autre part, ce coefficient de corrélation reste invariant par des transformations linéaire strictement croissantes des variables aléatoires. En effet :

$$\rho(aX + b, cY + d) = \text{sign}(ac) \rho(X, Y).$$

Lorsque $\rho(X, Y) = 0$, on dit que les deux variables aléatoires X et Y sont non corrélées. Attention, ne pas être corrélés ne signifie pas généralement être indépendants. Mais l'indépendance implique toujours l'absence de corrélation.

Exemple 2.2.1 prenons $X \sim \mathcal{N}(0, 1)$ et $Y = X^2$, on sait alors que $Y \sim \chi_1^2$ une loi du χ^2 à un degré de liberté. Ainsi les propriétés classiques d'une loi du χ^2 implique que $E[Y] = 1$ et $\text{var}[Y] = 2$. On alors :

$$\begin{aligned} \rho(X, Y) &= \text{corr}(X, Y) = \frac{\text{Cov}[X, Y]}{(\text{Var}[X] \text{Var}[Y])^{\frac{1}{2}}} = \frac{\mathbb{E}[(X - 0)(X^2 - 1)]}{(1 \times 2)^{\frac{1}{2}}} \\ &= \frac{1}{\sqrt{2}} \mathbb{E}[X^3 - X] = \frac{1}{\sqrt{2}} (\mathbb{E}[X^3] - \mathbb{E}[X]) = 0 - 0 = 0 \end{aligned}$$

Car X suit une loi normale centrée donc symétrique par rapport à 0. Pourtant il est clair que les deux variables X et $Y = X^2$ ne sont pas indépendantes !

Il faut donc toujours garder en mémoire que le coefficient de corrélation ne mesure que le **degré d'association linéaire** entre deux variables.

Remarque 2.2.1 Comme nous l'avons souligné précédemment, la corrélation linéaire est une mesure adaptée aux distributions elliptiques. En dehors de cet univers, nous avons recours à d'autres indicateurs de dépendance. Donc l'idée est de généraliser la notion de corrélation.

La section suivante présente des coefficients de corrélation non linéaire et non paramétrique au lieu de calculer le coefficient de corrélation linéaire entre deux variables aléatoires X et Y , nous pouvons utiliser leurs rangs. Par exemple pour un échantillon aléatoire de loi \mathbb{F} , nous prenons chacune des paires d'observations :

$$(X_1, Y_1), \dots, (X_n, Y_n).$$

Ensuite nous les classons par rangs et finalement nous normalisons leurs rangs pour obtenir les paires de rangs normalisées :

$$\left(\frac{R_1}{n}, \frac{S_1}{n}\right), \dots, \left(\frac{R_n}{n}, \frac{S_n}{n}\right).$$

On peut construire une mesure d'association indépendante des marges qui ne souffre pas nombreux problèmes identifiés pour le coefficient de corrélation linéaire. On remarque de plus que chacun des rangs normalisé est uniforme sur $\left(\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\right)$.

Il existe deux mesures de corrélation basées sur les rangs, le tau de Kendall et le rho de Spearman. Ces mesures présentent l'avantage de ne pas être influencées pour les lois marginales des observations.

2.2.2 Mesures d'association non paramétriques bivariées

Le tau de Kendall et le rho de Spearman sont deux mesures de concordance bien connues en statistiques. Elles donnent une mesure de la corrélation entre les rangs des observations, à la différence du coefficient de corrélation linéaire qui lui apprécie la corrélation entre les valeurs des observations. Elles offrent, par ailleurs, l'avantage de s'exprimer simplement en fonction de la copule associée au couple de variables aléatoires.

Ces notions figurent dans plusieurs travaux comme Cadoux et al. [2006], Embrechts et al. (2002), Demarta et McNeil (2004) et Lindskog et al. (2003). La définition de ces deux coefficients est, profondément liée à la notion de concordance (discordance) est essentielle à la compréhension de la définition du tau de Kendall et rho de Spearman.

De façon intuitive, une paire (X, Y) de variables aléatoires est concordante Si une grande valeur de X est souvent associée à une grande valeur de Y , et si

une petite valeur de X à de fortes chances d'être liée à une petite valeur de Y . En termes mathématiques, la nature des paires (x_i, y_j) et (x_i, y_j) est dite :

1. Concordante si l'un des deux cas suivants est rencontré :

i) $x_i > x_j$ et $y_i > y_j$.

ii) $x_i < x_j$ et $y_i < y_j$.

C'est-à-dire si $(x_i - x_j)(y_i - y_j) > 0$.

2. Discordante si l'un des deux cas suivantes est rencontré :

i) $x_i > x_j$ et $y_i < y_j$.

ii) $x_i < x_j$ et $y_i > y_j$.

C'est-à-dire si $(x_i - x_j)(y_i - y_j) < 0$. Noter que la probabilité d'avoir $(X_i - X_j)(Y_i - Y_j) = 0$ est nulle lorsque les lois des variables aléatoires sont continues

2.2.3 Les mesures de concordance

Nous pouvons interpréter la copule \mathbb{C} du vecteur $\mathbf{X} = (X_1, X_2)$ comme une reparamétrisation ou une normalisation de la répartition jointe \mathbb{F} après avoir éliminé les effets des marges F_1 et F_2 . Elle caractérise la structure de dépendance des composantes de \mathbf{X} .

De plus, c'est une statistique exhaustive celle-ci, et plusieurs mesures de dépendance peuvent être exprimées à partir de \mathbb{C} . Parmi ces statistiques, nous pouvons considérer de façon générale les mesures de concordance. Avant de les définir, nous définissons d'abord l'ordre de concordance.

Définition 2.2.2 Soient \mathbb{C}_1 et \mathbb{C}_2 deux copules bivariées. On dit que \mathbb{C}_1 est plus petite que \mathbb{C}_2 et on note $\mathbb{C}_1 \prec \mathbb{C}_2$, si et seulement si :

$$\mathbb{C}_1(u_1, u_2) \leq \mathbb{C}_2(u_1, u_2), \text{ pour tout } (u_1, u_2) \in [0, 1]^2.$$

L'ordre \prec est appelé l'ordre de concordance. Considérons, par exemple, la copule cubique définie par :

$$\mathbb{C}_{cub}(u_1, u_2, \theta) = \prod_{i=1}^2 u_i + \theta \prod_{i=1}^2 [u_i(u_i - 1)(2u_i - 1)], \text{ avec } \theta \in [-1, 2].$$

Nous avons :

$$\mathbb{C}_{cub} \left(\frac{3}{4}, \frac{3}{4}; 1 \right) = 0,5712 \geq \Pi \left(\frac{3}{4}, \frac{3}{4} \right) = \frac{3}{4} \times \frac{3}{4} = 0,5625.$$

Mais

$$\mathbb{C}_{cub} \left(\frac{3}{4}, \frac{1}{4}; 1 \right) = 0,1787 \leq \Pi \left(\frac{3}{4}, \frac{1}{4} \right) = \frac{3}{4} \times \frac{1}{4} = 0,1875.$$

Pour toute copule \mathbb{C} et d'après l'inégalité de Fréchet-Hoeffding, nous avons toujours la relation

$$W \prec \mathbb{C} \prec M.$$

Nous pouvons aussi vérifier que

$$W \prec \Pi \prec M.$$

Nous en déduisons qu'une structure de dépendance positive est une fonction copule \mathbb{C} qui vérifie l'inégalité suivante :

$$\Pi \prec \mathbb{C} \prec M.$$

De même, une structure de dépendance négative est une fonction copule \mathbb{C} qui vérifie l'inégalité suivante :

$$W \prec \mathbb{C} \prec \Pi.$$

Cette relation d'ordre est partielle car on ne peut pas comparer toutes les copules entre elles. Il existe donc des fonctions copules qui n'ont pas de structure de dépendance positive ni de structure négative.

Définition 2.2.3 *Une mesure numérique K d'association entre deux variables aléatoires continues X_1 et X_2 , dont la copule est \mathbb{C} , est une mesure de concordance si elle satisfait les propriétés suivantes :*

1. K est définie pour toute paire (X_1, X_2) de variables aléatoires continues
2. $-1 = K(X, -X) \leq K(\mathbb{C}) \leq K(X, X) = 1$;
3. $K(X_1, X_2) = K(X_2, X_1)$;
4. Si X_1 et X_2 sont indépendantes, alors $K(X_1, X_2) = 0$;
5. $K(-X_1, X_2) = K(X_1, -X_2) = -K(X_1, X_2)$;
6. Si $\mathbb{C}_1 \prec \mathbb{C}_2$ alors $K(\mathbb{C}_1) \leq K(\mathbb{C}_2)$;

7. Si $(X_{1,n}, X_{2,n})$ est une suite de variables aléatoires continues dont la copule est \mathbb{C}_n et si \mathbb{C}_n converge vers \mathbb{C} alors $\lim K(\mathbb{C}_n) = K(\mathbb{C})$.

A partir de la propriété (6) de la définition (2.2.2) on constate que l'ordre de concordance implique l'ordre sur K .

Lemme 2.2.1 *Le coefficient de corrélation ne définit pas une mesure de concordance d'où on a le théorème suivant :*

Théorème 2.2.1 *Soit K une mesure de concordance, entre deux variables aléatoires X et Y est*

1. Si Y est une fonction de X croissante alors :

$$K(X, Y) = 1 \Leftrightarrow \mathbb{C} = M.$$

2. Si Y est une fonction de X décroissante, alors :

$$K(X, Y) = -1 \Leftrightarrow \mathbb{C} = W.$$

3. Si α et β sont deux fonctions strictement monotones, alors :

$$k(\alpha(X), \beta(Y)) = K(X, Y).$$

Parmi les mesures de concordance, trois mesures très célèbres jouent un rôle important en statistique non paramétrique le tau de Kendall, le rhô de Spearman et l'indice de Gini. Ces mesures s'expriment en fonction de copule.

Lemme 2.2.2 *Soient (X, Y) et (X', Y') deux couples indépendantes de variables aléatoires de fonctions de répartition jointes, respectives \mathbb{F} et \mathbb{F}'*

1. Si le couple (X', Y') est fixé en un point (x, y) , alors :

$$\mathbb{P}(X < x, Y < y) = \mathbb{F}(x, y).$$

2. Si (X', Y') varie dans l'ensemble du plan nous avons :

$$\mathbb{P}(X < X', Y < Y') = \int_{\mathbb{R}^2} \mathbb{F} d\mathbb{F}'.$$

3. Si $(X, Y) \sim \mathbb{C}$ et si $(X', Y') \sim \mathbb{C}'$, alors :

$$\mathbb{P}(X < X', Y < Y') = \int_{\mathbb{R}^2} \mathbb{P}(X < x, Y < y) d\mathbb{F}'(x, y) = \int_{I^2} \mathbb{C} d\mathbb{C}'.$$

\mathbb{F}' est la distribution jointe du couple (X', Y')

Lemme 2.2.3 Soient $\mathbf{U} = (U_1, \dots, U_d)$ et $\mathbf{V} = (V_1, \dots, V_d)$ deux vecteurs de variables aléatoires continues et indépendantes de copule \mathbb{C} et \mathbb{C}' respectivement nous avons pour toute copules \mathbb{C}_1 et \mathbb{C}_2

$$\begin{aligned} \int_{[0,1]^d} \mathbb{C}_1(\mathbf{u}) d\mathbb{C}_2(\mathbf{u}) &= \int_{[0,1]^d} \mathbb{P}(\mathbf{U}_1 < \mathbf{u}) d\mathbb{C}_2(\mathbf{u}) = \mathbb{P}(\mathbf{U}_1 < \mathbf{U}_2) = \mathbb{P}(\mathbf{U}_2 > \mathbf{U}_1) \\ &= \int_{[0,1]^d} \mathbb{P}(\mathbf{U}_2 > \mathbf{u}) d\mathbb{C}_1(\mathbf{u}) = \int_{[0,1]^d} \overline{\mathbb{C}_2}(\mathbf{u}) d\mathbb{C}_1(\mathbf{u}). \end{aligned}$$

Exemple 2.2.2 Si $\mathbf{U}_\Pi \sim \Pi$, avec Π est la copule produit et si $\mathbf{U}_\mathbb{C} \sim \mathbb{C}$ on a :

$$\int_{[0,1]^d} \overline{\mathbb{C}}(\mathbf{u}) d\mathbf{u} = P[\mathbf{U}_\mathbb{C} > \mathbf{U}_\Pi] = P[\mathbf{U}_\Pi < \mathbf{U}_\mathbb{C}] = \int_{[0,1]^d} \Pi(\mathbf{u}) d\mathbb{C}(\mathbf{u})$$

Théorème 2.2.2 Soient (X_1, Y_1) et (X_2, Y_2) deux vecteurs indépendantes de variables aléatoires continues, de fonctions de répartition jointes, respectives, H_1 et H_2 . Soient F et G les marges associées, respectivement, à X_1, X_2 et à Y_1, Y_2 . Soient \mathbb{C}_1 et \mathbb{C}_2 les copules associées, respectivement, à H_1 et à H_2 données par $H_1(x, y) = \mathbb{C}_1(F(x), G(y))$ et $H_2(x, y) = \mathbb{C}_2(F(x), G(y))$. Si Q est la mesure de concordance et de discordance de (X_1, Y_1) et (X_2, Y_2) ; c'est à dire ;

$$Q = \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\},$$

alors:

$$Q(\mathbb{C}_1, \mathbb{C}_2) = 4 \int_{[0,1]^2} \mathbb{C}_2(u, v) d\mathbb{C}_1(u, v) - 1. \quad (2.1)$$

Preuve. Puisque toutes les variables aléatoires sont continues, alors :

$$\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\} = 1 - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) \geq 0\}.$$

Ainsi

$$Q = 2\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) \geq 0\} - 1.$$

Étudiant le premier terme

$$\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} = \mathbb{P}\{X_1 < X_2, Y_1 < Y_2\} + \mathbb{P}\{X_1 > X_2, Y_1 > Y_2\}.$$

Ces quantités peuvent être calculées par intégration :

$$\begin{aligned} \mathbb{P}\{X_1 > X_2, Y_1 > Y_2\} &= \mathbb{P}\{X_2 < X_1, Y_2 < Y_1\} \\ &= \int_{\mathbb{R}^2} \mathbb{P}\{X_2 < x, Y_2 < y\} d\mathbb{C}_1(F(x), G(y)) \\ &= \int_{\mathbb{R}^2} \mathbb{C}_2(F(x), G(y)) d\mathbb{C}_1(F(x), G(y)). \end{aligned}$$

Par changement de variables $u = F(x)$ et $v = G(y)$, on obtient alors :

$$\mathbb{P}\{X_1 < X_2, Y_1 < Y_2\} = \int_{\mathbb{R}^2} \mathbb{C}_2(F(x), G(y)) d(C_1 F(x), G(y))$$

D'une façon similaire :

$$\begin{aligned} \mathbb{P}\{X_1 > X_2, Y_1 > Y_2\} &= \int_{\mathbb{R}^2} \mathbb{P}\{X_2 > x, Y_2 > y\} d\mathbb{C}_1(F(x), G(y)) \\ &= \int_{\mathbb{R}^2} \{1 - F(x) - G(y) + C_2(F(x), G(y))\} dC_1(u, v) \\ &= \int_{[0,1]^2} \{1 - u - v + C_2(u, v)\} dC_1(u, v). \end{aligned}$$

Or, \mathbb{C}_1 est la fonction de répartition jointe du vecteur (U, V) où U et V sont des variables aléatoires uniformes, donc $E(U) = E(V) = \frac{1}{2}$ et par conséquent

$$\mathbb{P}\{X_1 < X_2, Y_1 < Y_2\} = 1 - \frac{1}{2} - \frac{1}{2} + \int_{I^2} \mathbb{C}_2(u, v) d\mathbb{C}_1(u, v).$$

Enfin,

$$\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} = 2 \int_{[0,1]^2} \mathbb{C}_2(u, v) d\mathbb{C}_1(u, v).$$

En regroupant ces résultats on en déduit que :

$$Q = Q(\mathbb{C}_1, \mathbb{C}_2) = 4 \int_{I^2} C_2(u, v) dC_1(u, v) - 1$$

■

Corollaire 2.2.1 Soient $\mathbb{C}_1, \mathbb{C}_2$ et Q données dans le théorème (2.2.2)

1. Q est symétrique, c'est-à-dire $Q(\mathbb{C}_1, \mathbb{C}_2) = Q(\mathbb{C}_2, \mathbb{C}_1)$
2. Q conserve l'ordre, c'est-à-dire si $\mathbb{C}_1 \prec \mathbb{C}'_1$ et $\mathbb{C}_2 \prec \mathbb{C}'_2$ pour tout (u, v) de I^2 alors $Q(\mathbb{C}_1, \mathbb{C}_2) \leq Q(\mathbb{C}'_1, \mathbb{C}'_2)$,
3. Q est invariante par rapport à la copule de survie, c'est-à-dire $Q(\mathbb{C}_1, \mathbb{C}_2) = Q(\tilde{\mathbb{C}}_1, \tilde{\mathbb{C}}_2)$.

Remarque 2.2.2 Pour les copules fréquemment utilisées Π , M , et W On peut facilement évaluer la mesure de concordance Q dans le cas bivarié. Rappelons que le support de M est la diagonale $u = v$, alors que la masse de probabilité de W se situe sur la diagonale secondaire $u = 1 - v$. Ainsi, on a

Pour toute fonction g intégrable sur le domaine I^2 , on a :

$$\int_{I^2} g(u, v) dM(u, v) = \int_{I^2} g(u, u) du.$$

$$\begin{aligned} Q(M, M) &= 4 \int_0^1 \int_0^1 \min(u, v) dM(u, v) - 1 = 4 \int_0^1 \int_0^1 \min(u, u) dM(u, u) \\ &= 4 \int_0^1 u du - 1 = 1; \end{aligned}$$

$$Q(M, \Pi) = 4 \int_0^1 \int_0^1 uv dM(u, v) - 1 = 4 \int_0^1 v^2 dv - 1 = \frac{1}{3}$$

$$Q(M, W) = 4 \int_0^1 \int_0^1 \max(u + v - 1, 0) dM(u, v) - 1 = 4 \int_{\frac{1}{2}}^1 (2u - 1) dv - 1 = 0,$$

et

$$Q(M, W) = 4 \int_0^1 \int_0^1 \max(u + v - 1, 0) dM(u, v) - 1 = 4 \int_{\frac{1}{2}}^1 (2u - 1) dv - 1 = 0.$$

De même comme le support de W est l'ensemble $D_W = \{(u, v) \in I^2 / u = 1 - v\}$, alors

$$\int_{I^2} g(u, v) dW(u, v) = \int_{I^2} g(u, 1 - u) du,$$

ce qui donne

$$Q(W, \Pi) = 4 \int_0^1 \int_0^1 uv dW(u, v) - 1 = 4 \int_0^1 (1-v) v dv - 1 = -\frac{1}{3}$$

$$Q(W, W) = 4 \int_0^1 \int_0^1 \max(u+v-1, 0) dW(u, v) - 1 = 4 \int_0^1 0 dv - 1 = -1;$$

Finalement comme $d\Pi = dudv$ alors :

$$Q(\Pi, \Pi) = 4 \int_0^1 \int_0^1 uv d\Pi = 4 \int_0^1 \int_0^1 uv dudv - 1 = 0.$$

Pour une copule quelconque \mathbb{C} on a :

$$\begin{aligned} W &< \mathbb{C} < M \\ \implies Q(W, M) &\leq Q(\mathbb{C}, M) \leq Q(M, M) \\ \implies 0 &\leq Q(\mathbb{C}, M) \leq 1 \end{aligned}$$

De la même manière, on établit que :

$$-1 \leq Q(\mathbb{C}, W) \leq 1 \quad \text{et} \quad -\frac{1}{3} \leq Q(\mathbb{C}, \Pi) \leq \frac{1}{3}.$$

2.2.4 Tau de Kendall

Le tau de Kendall joue le même rôle pour les fonctions de dépendance que le coefficient de corrélation linéaire pour la distribution Gaussienne. Le tau de Kendall est une mesure d'association non paramétrique entre deux variables aléatoires basées sur la notion de concordance.

Définition 2.2.4 Soient (X_1, Y_1) et (X_2, Y_2) deux couples de variables aléatoires indépendantes et identiquement distribuées. Le tau de Kendall du vecteur aléatoire (X_1, Y_1) s'écrit alors :

$$\begin{aligned} \tau(X_1, Y_1) &= \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\} \\ &= 2\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - 1. \end{aligned} \tag{2.2}$$

Le tau de Kendall n'est autre que la différence entre la probabilité de concordance et celle de discordance que l'on peut définir, aussi d'une manière empirique. Si on dispose d'un échantillon d'observation de taille n de données bivariées $\{(x_k, y_k)\}_{k=1}^n$. Définissons les nombres N_C et N_d comme étant respectivement le nombre de paires concordantes et discordantes dans cet échantillon. On définit la version empirique du tau de Kendall par :

$$\tau_n = \frac{N_C - N_d}{N_C + N_d} = \frac{(N_C - N_d)}{(C_2^n)},$$

avec

$$C_2^n = \frac{n(n-1)}{2}.$$

Finalement, on peut aussi donner cette version empirique en terme de densité de la copule associée :

$$\tau_n = \frac{2n}{n-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{p=1}^{i-1} \sum_{q=1}^{j-1} \left[c_n \left(\frac{i}{n}, \frac{j}{n} \right) c_n \left(\frac{p}{n}, \frac{q}{n} \right) - c_n \left(\frac{i}{n}, \frac{q}{n} \right) c_n \left(\frac{p}{n}, \frac{j}{n} \right) \right].$$

Soit (X, Y) une copule de variables aléatoires continues de copule \mathbb{C} compte tenu de (2.1) et de (2.2) la version populaire du tau de Kendall pour X et Y est donnée par :

$$\tau(X, Y) = Q(\mathbb{C}, \mathbb{C}) = 4 \int_{[0,1]^2} \mathbb{C}(u, v) d\mathbb{C}(u, v) - 1 = 4E\{\mathbb{C}(U, V)\} - 1, \quad (2.3)$$

où (U, V) sont de loi uniforme sur $[0, 1]$. L'expression de τ donnée dans la définition (2.3) représente la valeur théorique du tau de Kendall

Théorème 2.2.3 Soient \mathbb{C}_1 et \mathbb{C}_2 deux copules tel que le produit $\left(\frac{\partial}{\partial u}\mathbb{C}_1\right) \left(\frac{\partial}{\partial v}\mathbb{C}_2\right)$ est intégrable sur $[0, 1]^2$: Alors :

$$\int_{I^2} \mathbb{C}_1(u, v) d\mathbb{C}_2(u, v) = \frac{1}{2} - \int_{I^2} \frac{\partial}{\partial u}\mathbb{C}_1(u, v) \frac{\partial}{\partial v}\mathbb{C}_2(u, v) dudv. \quad (2.4)$$

Preuve. Lorsque les copules sont absolument continues on peut établir l'intégration par partie. Dans ce cas :

$$\int_{I^2} \mathbb{C}_1(u, v) d\mathbb{C}_2(u, v) = \int_0^1 \int_0^1 \mathbb{C}_1(u, v) \frac{\partial^2 \mathbb{C}_2(u, v)}{\partial u \partial v} dudv.$$

On a :

$$\begin{aligned} \int_0^1 \mathbb{C}_1(u, v) \frac{\partial^2 \mathbb{C}_2(u, v)}{\partial u \partial v} du &= \mathbb{C}_1(u, v) \frac{\partial \mathbb{C}_2(u, v)}{\partial v} \Big|_{u=0}^{u=1} - \int_0^1 \frac{\partial \mathbb{C}_1(u, v)}{\partial u} \frac{\partial \mathbb{C}_2(u, v)}{\partial v} du \\ &= v - \int_0^1 \frac{\partial \mathbb{C}_1(u, v)}{\partial u} \frac{\partial \mathbb{C}_2(u, v)}{\partial v} du. \end{aligned}$$

Car $\frac{\partial \mathbb{C}_2(1, v)}{\partial v} = 1$, et $\mathbb{C}_1(1, v) = v$. Nous intégrons une autre fois par rapport à v de 0 à 1, on obtient le résultat. ■

Nous venons de présenter quelques exemples sur la mesure d'association du tau de Kendall qui sont utiles dans la pratique.

Exemple 2.2.3 Soient X, Y deux variables aléatoires indépendantes ; alors leur copule est nécessairement définie par $\Pi(u, v) = uv$. Le calcul du tau de Kendall même alors on a :

$$\begin{aligned} \tau_{X, Y} &= \tau_{\Pi} = 4 \int_{I^2} uv d\Pi - 1 \\ &= 4 \int_0^1 \int_0^1 uv du dv - 1 = 4 \left(\int_0^1 u du \right)^2 - 1 = 0. \end{aligned}$$

$$\text{Car } d\Pi(u, v) = \frac{\partial^2 \Pi(u, v)}{\partial u \partial v} = 1 du dv.$$

Exemple 2.2.4 Pour la borne inférieure de Fréchet-Hoeffding le tau de Kendall vaut :

$$\tau_W = 4 \int_0^1 W(\mathbf{u}, 1 - \mathbf{u}) dW(u, v) - 1 = -1,$$

car toute la masse de probabilité de $dW(u, v)$ se trouve sur la droite $v = 1 - u$.

Alors que la borne supérieure de Fréchet-Hoeffding amène :

$$\tau_M = 4 \int_0^1 M(u, v) du - 1 = 4 \int_0^1 u du - 1 = 4 \left(\frac{1}{2} \right) - 1 = 1.$$

Exemple 2.2.5 Si on regarde la famille de copules Farlie-Gumbel-Morgenstern définie par :

$$\mathbb{C}_\theta(u, v) = uv + \theta uv(1-u)(1-v),$$

un calcul aboutit à :

$$\tau_{X,Y} = \frac{2\theta}{9},$$

puisque \mathbb{C}_θ est absolument continue, nous avons

$$d\mathbb{C}_\theta(u, v) = \frac{\partial^2 \mathbb{C}(u, v)}{\partial u \partial v} dudv = [1 + \theta(1-2u)(1-2v)] dudv.$$

et

$$\int_{[0,1]^2} \mathbb{C}_\theta(u, v) d\mathbb{C}_\theta(u, v) = \frac{1}{4} + \frac{\theta}{18}.$$

Pour (X, Y) de copule \mathbb{C}_θ , avec $\theta \in [-1, 1]$. Une conséquence de ce résultat est l'obtention du domaine de variation du tau de Kendall pour deux variables dont la copule est dans la famille de Farlie-Gumbel-Morgenstern,

$$\tau \in \left[\frac{-2}{9}, \frac{+2}{9} \right].$$

Exemple 2.2.6 Des familles paramétriques de copules basées sur les trois copules Π, W, M ont été largement étudiées. On peut citer à titre d'exemples et on se reportant à Nelsen (2005), et Joe (1997) pour une liste très complète.

-**Famille de fréchet** $\forall (\alpha, \beta) \in [0, 1]$, et $\alpha + \beta \leq 1$,

$$\mathbb{C}_{\alpha,\beta}(u, v) = \alpha M(u, v) + (1 - \alpha - \beta) \Pi(u, v) + \beta W(u, v),$$

et

$$d\mathbb{C}_{\alpha,\beta} = \alpha dM(u, v) + (1 - \alpha - \beta) d\Pi + \beta dW.$$

Utilisons les résultats des exemples de fonctions de concordances

On a :

$$\tau_{\alpha,\beta} = 4 \int_{I^2} \mathbb{C}_{\alpha,\beta}(u, v) d\mathbb{C}_{\alpha,\beta}(u, v) - 1 = \frac{(\alpha - \beta)(\alpha + \beta + 2)}{3}.$$

Théorème 2.2.4 Soient X et Y deux variables aléatoires d'une copule Archimédienne, Le Tau de Kendall de X et Y s'obtient simplement à l'aide du générateur de la copule, selon la formule :

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(\mathbf{u})}{\varphi'(\mathbf{u})} d\mathbf{u}.$$

-Copule de Clayton : Soit C_θ la copule de clayton, alors pour tout $\theta \geq -1$

$$\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{t^{\theta+1} - t}{\theta}.$$

Lorsque $\theta = 0$ et $\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = t \ln t$. Donc :

$$\tau_\theta = \frac{\theta}{\theta + 2}.$$

-Copule de Gumbel : Soit \mathbb{C}_θ la copule de Gumbel, alors pour $\theta \geq 1$

$$\frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = \frac{t \ln t}{\theta}.$$

et

$$\tau_\theta = \frac{\theta - 1}{\theta}.$$

-Copule Marchal-Olkin : La copule est définie par :

$$\mathbb{C}_{\alpha,\beta}(u, v) = \min(u^{1-\alpha}v, uv^{1-\beta}) = \begin{cases} u^{1-\alpha}v, u^\alpha \geq v^\beta \\ uv^{1-\beta}, u^\alpha \leq v^\beta \end{cases}, 0 < \alpha, \beta < 1,$$

qui est dérivable par tout sauf en $u^\alpha = v^\beta$. Alors :

$$\frac{\partial}{\partial u} \mathbb{C}_{\alpha,\beta}(u, v) \frac{\partial}{\partial v} \mathbb{C}_{\alpha,\beta}(u, v) = \frac{1}{4} \left(1 - \frac{\alpha\beta}{\alpha - \alpha\beta + \beta} \right).$$

D'après (2.4) nous obtenons :

$$\begin{aligned} \tau_{\alpha,\beta} &= 4 \int_{I^2} \mathbb{C}_{\alpha,\beta}(u, v) d\mathbb{C}_{\alpha,\beta}(u, v) - 1 \\ &= 4 \left(\frac{1}{2} - \int_{I^2} \frac{\partial}{\partial u} \mathbb{C}_{\alpha,\beta}(u, v) \frac{\partial}{\partial v} \mathbb{C}_{\alpha,\beta}(u, v) dudv \right) - 1 \\ &= \frac{\alpha\beta}{\alpha - \alpha\beta + \beta}. \end{aligned}$$

2.2.5 Le rhô de Spearman

Pour avoir une mesure de corrélation linéaire qui soit indépendante par transformations monotones des variables aléatoires, on définit le rhô de Spearman ρ_s qui est simplement la corrélation linéaire de Pearson entre $F(X)$ et $G(Y)$ où F et G sont les fonctions de répartition de X et de Y respectivement. C'est-à-dire que :

$$\rho_s(X, Y) = \text{Corr}(F(X), G(Y)),$$

où $F(X)$ et $G(Y)$ sont uniformes sur $[0, 1]$. Cette mesure a été développée par Charles Spearman (1904). Le coefficient de corrélation de Spearman est basé sur l'étude de la corrélation des rangs.

La mesure de la dépendance au sens de Spearman entre deux variables aléatoires continues X et Y qui sera notée $\rho_s(X, Y)$ a le même champ d'application que la statistique de Kendall $\tau(X, Y)$. Autre définition du rhô de Spearman dite version populaire est la suivante :

Définition 2.2.5 Soient $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ trois couples de variables aléatoires indépendantes et identiquement distribuées. Le rhô de Spearman, noté ρ_s , est défini par :

$$\rho_s = 3 \{ \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3)] > 0 - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_3) < 0] \},$$

notez que si la distribution jointe de (X_2, Y_3) est $F(x)G(y)$. Donc la copule associée au vecteur (X_2, Y_3) est la copule Π .

Remarque 2.2.3 Les propriétés du rhô de Spearman sont les mêmes que celle du tau de Kendall :

1. On sait que pour toute copule \mathbb{C} , on a : $W(u, v) \leq \mathbb{C}(u, v) \leq M(u, v)$, le résultat précédent assure que : $-\frac{1}{3} = Q(W, \Pi) \leq Q(\mathbb{C}, \Pi) \leq Q(M, \Pi) = \frac{1}{3}$, ce qui justifie la multiplication de Q par la facteur 3 permettant que : $-1 \leq \rho_s \leq 1$. En d'autres termes : $\rho(W) \leq \rho(\mathbb{C}) \leq \rho(M)$.

2. Si $X \sim F$ et $Y \sim G$, soit $U = F(X)$ et $V = G(Y)$, on a :

$$\begin{aligned} \rho_s(X, Y) &= 12 \int_{[0,1]^2} \mathbb{C}(u, v) \, dudv - 3 & (2.5) \\ &= \frac{E(UV) - \frac{1}{4}}{\frac{1}{12}} = \frac{\text{cov}(U, V)}{\sqrt{\text{var}(U)}\sqrt{\text{var}(V)}} \\ &= \rho(F(X), G(Y)). \end{aligned}$$

3. Donc le rhô de Spearman pour une paire (X, Y) est équivalent au coefficient de corrélation usuel de Pearson de la paire (U, V)

4. On peut considérer que ρ_s est une mesure de distance moyenne entre la copule \mathbb{C} et la copule d'indépendance car :

$$\rho_s(\mathbb{C}) = 12 \int_{[0,1]^2} \{\mathbb{C}(u, v) - uv\} \, dudv = \frac{\int_{[0,1]^2} [\mathbb{C}(u, v) - \Pi(u, v)] \, dudv}{\int_{[0,1]^2} [M(u, v) - \Pi(u, v)] \, dudv}$$

Puisque :

$$\begin{aligned} \int_0^1 \int_0^1 \left[\min(u, v) - \int_0^1 \Pi(u, v) \right] \, dudv &= \int_0^1 \int_0^u v \, dv \, du + \int_0^1 \int_0^v u \, dv \, du - \frac{1}{4} \\ &= \int_0^1 \frac{u^2}{2} \, du + \int_0^1 \frac{v^2}{2} \, dv - \frac{1}{4} = \frac{1}{12}. \end{aligned}$$

5. Le tau de Kendall et le rhô de Spearman ne dépend pas du comportement des lois marginales, mais uniquement de la structure de dépendance.

6. Pour toutes fonctions croissantes f et g , on a $\tau(f(X), g(Y)) = \tau(X, Y)$ et $\rho(f(X), g(Y)) = \rho(X, Y)$.

7. Soit un échantillon de taille n de données bivariées $\{(x_k, y_k)\}_{k=1}^n$. On définit la version empirique du rhô de Spearman en terme de densité de la copule associée par :

$$\rho_n = \frac{12}{n^2 - 1} \sum_{i=1}^n \sum_{j=1}^n \left[c_n \left(\frac{i}{n}, \frac{j}{n} \right) - \frac{i}{n} \frac{j}{n} \right].$$

8. Supposons une certaine famille paramétrique $\{\mathbb{C}_\theta, \theta \in \Theta\}$, de copules dont les membres sont ordonnées par θ , c'est-à-dire pour tout $u \in [0, 1]^2$ on a $C_\theta(\mathbf{u}, \mathbf{v}) \leq C_{\theta'}(\mathbf{u}, \mathbf{v})$ lorsque $\theta < \theta'$. On dit alors que θ est un paramètre de dépendance. Dans ce cas, on montre que ρ_s est monotone croissant en fonction de θ . En effet :

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

Théorème 2.2.5 Les coefficients ρ_S de Spearman et τ de Kendall vérifient les propriétés ci-dessous :

1. **Symétrie** : $\delta(X, Y) = \delta(Y, X)$
2. **Indépendance** : Si X et Y sont indépendantes alors : $\delta(X, Y) = 0$
3. **Normalisation** : $-1 \leq \delta(X, Y) \leq 1$.
4. **Cas extrêmes** :
 - $\delta(X, Y) = 1$ Si et seulement si M est la copule associée à (X, Y)
 - $\delta(X, Y) = -1$ Si et seulement si W est la copule associée à (X, Y)
5. **Invariance par transformation monotone** :
 - Si f est strictement croissante alors $\delta(f(X), Y) = \delta(X, Y)$
 - Si f est strictement décroissante alors $\delta(f(X), Y) = -\delta(X, Y)$

Nous allons maintenant présenter quelques exemples de mesures d'associations du rhô de Spearman

a) **Copule de Marshall-Olkin** La copule est définie par :

$$\mathbb{C}_{\alpha, \beta}(u, v) = \min(u^{1-\alpha}v, uv^{1-\beta}) = \begin{cases} u^{1-\alpha}v, & u^\alpha \geq v^\beta \\ uv^{1-\beta}, & u^\alpha \leq v^\beta \end{cases}, \quad (2.6)$$

où $0 < \alpha, \beta < 1$. Puisque nous avons $0 \leq v \leq u^{\frac{\alpha}{\beta}} \leq 1$, $0 \leq u \leq v^{\frac{\beta}{\alpha}} \leq 1$.
Intégrant la quantité (2.6)

$$\begin{aligned}
\int_0^1 \int_0^1 \mathbb{C}_{\alpha, \beta}(u, v) &= \int_0^1 \int_0^1 \min(u^{1-\alpha}v, uv^{1-\beta}) dudv \\
&= \int_0^1 \int_{v^\beta < u^\alpha} u^{1-\alpha}v dudv + \int_0^1 \int_{v^\beta > u^\alpha} uv^{1-\beta} dudv \\
&= \int_0^1 \left(u^{1-\alpha} \int_0^{u^{\frac{\alpha}{\beta}}} v dv \right) du + \int_0^1 \left(v^{1-\beta} \int_0^{v^{\frac{\beta}{\alpha}}} u du \right) dv
\end{aligned}$$

Nous avons :

$$\begin{aligned}
\int_0^1 \int_0^1 \min(\mathbf{u}^{1-\alpha}\mathbf{v}, \mathbf{u}\mathbf{v}^{1-\beta}) d\mathbf{u}d\mathbf{v} &= \frac{1}{2} \int_0^1 \mathbf{u}^{1-\alpha} \mathbf{u}^{\frac{2\alpha}{\beta}} d\mathbf{u} + \frac{1}{2} \int_0^1 \mathbf{v}^{1-\beta} \mathbf{v}^{\frac{2\beta}{\alpha}} d\mathbf{v} \\
&= \frac{1}{2} \int_0^1 u^{1-\alpha+\frac{2\alpha}{\beta}} du + \int_0^1 v^{1-\beta+\frac{2\beta}{\alpha}} dv \\
&= \frac{1}{2} \frac{\beta + \alpha}{2\alpha + 2\beta - \alpha\beta}.
\end{aligned}$$

Finalement

$$\rho_s = 6 \frac{\beta + \alpha}{2\alpha + 2\beta - \alpha\beta} - 3 = \frac{3\alpha\beta}{2\alpha + 2\beta - \alpha\beta}.$$

b) **Copule de Farlie-Gembel-Morgenstern** : Cette copule est définie par :

$$\mathbb{C}_\alpha(u, v) = uv [1 + \alpha(1-u)(1-v)], \quad \alpha \in [-1, 1].$$

Alors :

$$\begin{aligned}
 \rho_s(X, Y) &= \rho_s(F(X), G(Y)) = 12 \int_{[0,1]^2} \mathbb{C}(u, v) \, dudv - 3 \\
 &= 12 \int_{[0,1]^2} (uv + \theta uv(1-u)(1-v)) \, dudv - 3 \\
 &= 12\theta \int_{[0,1]^2} uv(1-u)(1-v) \, dudv \\
 &= 12\theta \int_0^1 \left(u(1-u) \int_0^1 v(1-v) \, dv \right) du \\
 &= 2\theta \left[\frac{1}{2} - \frac{1}{3} \right] = 2\theta \times \frac{1}{6} = \frac{\theta}{3}.
 \end{aligned}$$

Donc

$$\rho_s = \frac{\theta}{3} \Leftrightarrow -\frac{1}{3} \leq \rho_s \leq \frac{1}{3}.$$

c) **Famille de Fréchet**

$$\forall \alpha > 0, \forall \beta > 0, \alpha + \beta \leq 1. \mathbb{C}_{\alpha, \beta} = \alpha M + (1 - \alpha - \beta) \Pi + \beta W.$$

On a

$$\begin{aligned}
 Q(\mathbb{C}_{\alpha, \beta}, \Pi) &= \alpha Q(M, \Pi) + (1 - \alpha - \beta) Q(\Pi, \Pi) + \beta Q(W, \Pi) \\
 &= \alpha \left(\frac{1}{3} \right) + (1 - \alpha - \beta) \times 0 + \beta \left(-\frac{1}{3} \right) = \frac{\alpha - \beta}{3},
 \end{aligned}$$

par conséquent :

$$\rho_s(\mathbb{C}_{\alpha, \beta}) = 3Q(\mathbb{C}_{\alpha, \beta}, \Pi) = \alpha - \beta.$$

d) **Copule Normale** : Si (X, Y) est un couple de loi normale bivarieé de coefficient de corrélation ρ alors :

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

e) **Copule de Clayton** : une procédure d'intégration numérique sur le carée $[0, 1]^2$ montre que :

$$\rho_s = 12I(\theta) - 3, \text{ où } I(\theta) = \int_0^1 \int_0^1 (u^{-\theta} + v^{-\theta}) \, dudv.$$

f) Famille de Ali-Mikhail-Haq

$$C(u, v, \theta) = \frac{uv}{[1 - \theta(1-u)(1-v)]},$$

où $-1 \leq \theta \leq +1$. Le tau de Kendall associé vaut :

$$\tau = \frac{3\theta - 2}{3\theta} - \frac{2(1-\theta)^2}{3\theta^2} \log(1-\theta).$$

Le rhô de Spearman est donné par :

$$\rho = -\frac{12(1+\theta)}{\theta^2} \int_0^1 \frac{\log(1-x)}{x} dx - \frac{3(12+\theta)}{\theta} - \frac{24(1-\theta)}{\theta^2} \log(1-\theta).$$

Les résultats concernant le rhô de Spearman dans le cas bidimensionnel peuvent être étendus dans le cas multidimensionnel de la manière suivante

$$\rho_s = \frac{1}{(d+1)^{-1} - 2^{-d}} \int u_1 \dots u_d dC(u_1, \dots, u_d) - 2^{-d}.$$

Nous énoncerons au **troisième chapitre** les mesures d'association multivariée du rhô de Spearman.

2.2.6 Liens entre le τ de Kendall et le ρ de Spearman

Généralement, les valeurs τ et ρ pour une copule \mathbb{C} sont différentes. Certaines relations lient les deux mesures de concordance.

Théorème 2.2.6 (Daniels, 1950) *Si X et Y deux variables aléatoires continues, alors :*

$$-1 \leq 3\tau - 2\rho_s \leq 1.$$

Théorème 2.2.7 (Durbin et Stuart, 1951) *Si X et Y deux variables aléatoires continues, alors :*

$$\frac{1+\rho_s}{2} \geq \left(\frac{1+\tau}{2}\right)^2 \quad \text{et} \quad \frac{1-\rho_s}{2} \geq \left(\frac{1-\tau}{2}\right)^2.$$

En combinant ces deux dernières inégalités, on obtient :

$$\frac{3\tau - 1}{2} \leq \rho_s \leq \frac{1 + 2\tau - \tau^2}{2} \quad \text{si } \tau \geq 0,$$

et

$$\frac{-1 + 2\tau + \tau^2}{2} \leq \rho_s \leq \frac{3\tau + 1}{2} \quad \text{si } \tau \leq 0.$$

Une autre mesure de concordance qui s'exprime à l'aide de la copule, c'est l'indice de Gini. Elle est définie par :

$$\gamma = \int_{[0,1]^2} (|u + v - 1| - |u - v|) d\mathbb{C}(u, v).$$

Comme la différentielle $d\mathbb{C}$ intervient dans l'expression de la mesure γ , il est possible de la calculer via l'expression équivalente suivante (voir Nelsen, 1999, page 174)

$$\gamma = 4 \int_{[0,1]^2} [\mathbb{C}(u, u) + \mathbb{C}(u, 1 - u) - u] du. \tag{2.7}$$

Pour certaines copules, nous disposons d'expressions analytiques. Par exemple, nous avons :

Copules	Tau de Kendall	Rhô de Spearman
Indépendance	0	0
La borne supérieure	1	1
La borne inférieure	-1	-1
Gaussienne	$\frac{2}{\pi} \arcsin(\rho)$	$\rho_s = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right)$
Student	$\frac{2}{\pi} \arcsin(\rho)$	$\rho_s = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right)$
Frank	$\tau(\theta) = 1 - \frac{4}{\theta} + \frac{4}{\theta^2} \int_0^\theta \left(\frac{t}{e^t - 1}\right) dt$	$\rho_s(\theta) = 1 - \frac{12}{\theta^2} \int_0^\theta \left(\frac{t}{e^t - 1}\right) dt + \frac{24}{\theta^3} \int_0^\theta \left(\frac{t^2}{e^t - 1}\right) dt$
Clayton	$\tau(\theta) = \frac{\theta}{\theta + 2}$	$\rho_s(\theta) = 12 \int_0^1 \int_0^1 (u^{-\theta} + v^{-\theta}) dudv - 3$
Gumbel	$\tau(\theta) = \frac{\theta - 1}{\theta}$	

Tableau 2.1 Exemples de mesures d'association bivariées

Chapitre 3

Mesures d'association multivariées

3.1 Introduction

Le problème de mesures d'association entre les variables aléatoires est largement étudié en littérature. Le développement de la théorie de copules a eu un grand impact dans l'étude de la mesure non paramétrique dans le cas des variables aléatoires continues, plusieurs mesures d'association bivariées à savoir le tau de Kendall, rho de Spearman, et l'indice de Gini ont été étudiés en termes de copules. Ces mesures, vérifiant un ensemble d'axiomes de scarini sont dites mesures de concordance.

Ce chapitre constitue un aperçu sur les mesures d'association multivariées basées sur les copules multidimensionnelles. C'est à dire mesure d'association entre d -variables aléatoires du vecteur $\mathbf{X} = (X_1, \dots, X_d)$, où $d \geq 2$, certaines de ces mesures connues en statistiques économique, la propriété importante pour les variables aléatoires continues est que les mesures sont invariantes par transformation continue cette propriété connue sous le nom (invariance par échel). Notez aussi que toutes les mesures présentent l'avantage demeurent inchangés sous l'hypothèse d'une transformation strictement croissante des variables aléatoires sauf le coefficient de corrélation lineaire. L'analyse des données multivariées continues a une grande importance dans une multitude domaine, citons à titre d'exemple, le cas de l'analyse multivariée de la finance (rendement des actifs).

Dans ce chapitre, nous montrons que les mesures de dépendance bivariée de

Spearman et de Kendall peuvent s'étendent à toute famille finie de variables aléatoires X_1, \dots, X_d . Le rhô de Spearman est une mesure non paramétrique de dépendance basées sur les rangs des observations. Ainsi, plutôt que d'utiliser les résultats bruts, on classe ces mêmes résultats selon leur rang. Il ya plusieurs méthodes pour faire une extention à cette mesure pour les fonctions multivariées X_1, \dots, X_d , pour plus de details sur ce sujet, on pourra se référer, entre d'autres, à Kendall (1970), Joe (1990), Nelsen (1996) et Wolff (1980)

3.2 Extentions multivariées du rhô de Spearman

Définition 3.2.1 (*Rhô de Spearman*), Soit C une copule d -dimensionnelle du vecteur aléatoire (X_1, \dots, X_d) $d \geq 2$ de loi marginales uniforme sur $[0,1]$, de loi jointe F telque $F(x_1, \dots, x_d) = C(u_1, \dots, u_d)$. Alors, il existe plusieurs adaptations du coefficient de Spearman dans le cadre multivariée. Donc les trois versions associées à un couple (X, Y) peuvent s'étendre en général à un vecteur aléatoire (X_1, \dots, X_d) de la manière suivante :

$$\rho_1(C) = h(d) \left\{ 2^d \int_{I^d} C(u_1, \dots, u_d) du_1 \dots du_d - 1 \right\} \quad (3.1)$$

$$\rho_2(C) = h(d) \left\{ 2^d \int_{I^d} \Pi(u_1, \dots, u_d) dC(u_1 \dots u_d) - 1 \right\} \quad (3.2)$$

$$\rho_3(C) = h(2) \left\{ 2^2 \sum_{k < l} \frac{1}{\binom{d}{2}} \int_{I^2} C_{kl}(u_k, u_l) du_k du_l - 1 \right\} \quad (3.3)$$

Avec

$$h(d) = \frac{d+1}{2^d - (d+1)}$$

où C_{kl} est la copule marginale bivariée de (X_k, X_l) qui correspond à la k^{ieme} et la l^{ieme} marge, ces trois mesures se coincident dans le cas où $d = 2$

Exemple 3.2.1 Soient x, y, z trois variables aléatoires de loi uniforme sur $[0, 1]^3$ et

de copule C alors les coefficients de Spearman à l'ordre 3 sont définient par :

$$\rho_1(C) = 8 \left(\int_{I^3} C(u, v, w) dudvdw - \frac{1}{8} \right) \quad (3.4)$$

$$\rho_2(C) = 8 \left(\int_{I^3} uvwdC(u, v, w) - \frac{1}{8} \right) \quad (3.5)$$

$$\rho_3(C) = h(2) \left\{ 2^2 \sum_{1 \leq k < l \leq 3} \binom{d}{2}^{-1} \int_{I^2} c_{kl}(u_k, u_l) du_k du_l - 1 \right\} \quad (3.6)$$

Remarque 3.2.1 La moyenne de ρ_1 et ρ_2 conduit à une autre mesure de concordance multidimensionnelle aussi discutée par Taylor (2006), Dolati et Ubeda Flores (2006). ρ_1 à été analysé par Joe (1997), Nelsen (1999), et Wolff (1980), tandis que ρ_2 à été introduit par Joe (1997), et Nelsen (1999), ρ_3 présenté par Kendall (1970), correspond à une version pondérée du coefficient de Spearman en dimation 2. Les calculs explicites pour ρ_1 et ρ_2 sont assez délicat. ρ_3 est plus souvent utilisé compte tenu de la simplicité de son calcul parce que peut s'ecrire sous la forme suivante :

$$\rho_3 = \frac{2}{d(d-1)} \sum_{k < l} \rho_1(C_{kl})$$

Exemple 3.2.2 Soit l'échantillon de variables aléatoires (X_1, X_2, X_3) . Alors :

$$\begin{aligned} \rho_3 &= h(2) \left\{ 2^2 \sum \frac{1}{C_2^3} \int_{I^2} C_{kl}(u_k, u_l) du_k du_l - 1 \right\} \\ &= 4 \sum_{\mathbf{k} < \mathbf{l}} \int_{I^2} C_{kl}(u_k, u_l) du_k dv_l - 3 \\ &= 4 \left\{ \int_{I^2} C_{12}(u_1, u_2) du_1 du_2 + \int_{I^2} C_{13}(u_1, u_3) du_1 du_3 + \int_{I^2} C_{23}(u_2, u_3) du_2 du_3 \right\} \\ &= \frac{1}{3} \left\{ \underbrace{12 \int_{I^2} C_{12}(u_1, u_2) du_1 du_2 - 3}_{\text{}} + \underbrace{12 \int_{I^2} C_{13}(u_1, u_3) du_1 du_3 - 3}_{\text{}} \right. \\ &\quad \left. + \underbrace{12 \int_{I^2} C_{23}(u_2, u_3) du_2 du_3 - 3}_{\text{}} \right\} \\ &= \frac{1}{3} \sum_{1 \leq k < l \leq 3} h(2) \left\{ 2^2 \int_{I^2} C_{kl}(u_k, u_l) du_k du_l - 1 \right\} = \frac{1}{3} \sum_{1 \leq k < l \leq 3} \rho_1(C_{kl}) \end{aligned}$$

On sait que ρ dans le cas bivariée ne dépend que de la copule de (X, Y) . En effet, soit deux variables aléatoires X et Y de distributions marginales F_X et F_Y continues, et avec une copule unique C . Dès lors, ρ s'écrit comme suit :

$$\rho_s = 12 \int_0^1 \int_0^1 uv dC(u, v) - 3, \quad (3.7)$$

que l'on peut réécrire :

$$\begin{aligned} \rho_s &= \frac{\text{cov}(U_1, U_2)}{\sqrt{\text{var}U_1}\sqrt{\text{var}U_2}} = \frac{\int_0^1 \int_0^1 uv dC(u, v) - \frac{1}{4}}{\frac{1}{\sqrt{12}} \frac{1}{\sqrt{12}}} \\ &= \frac{\int_{I^2} uv dC(u, v) - \int_{I^2} uv d\Pi(u, v)}{\int_{I^2} uv dM(u, v) - \int_{I^2} uv d\Pi(u, v)} \\ &= \frac{\int_{I^2} C(u, v) dudv - \int_{I^2} \Pi(u, v) dudv}{\int_{I^2} M(u, v) dudv - \int_{I^2} \Pi(u, v) dudv}. \end{aligned} \quad (3.8)$$

Car :

$$\begin{aligned} \int_{I^2} M(u, v) dudv &= \int_{I^2} \min(u, v) dudv \\ &= \int_0^1 \left[\int_0^1 \left(u \times \mathbf{1}_{[0, v]}^{(u)} + v \times \mathbf{1}_{[v, 1]}^{(u)} \right) du \right] dv = \frac{1}{3} \end{aligned}$$

et :

$$\int_{I^2} uv dudv = \int_0^1 u du \int_0^1 v du = \frac{1}{4},$$

donc le ρ_s peut s'interpréter comme mesure moyenne de la distance normalisée entre C et Π . En suivant les mêmes idées, il est possible d'étendre cette définition en dimension $d > 2$ pour le rhô de Spearman comme suit :

$$\begin{aligned} \rho_s(C) &= h(d) \left\{ 2^d \int_{I^d} C(u_1, \dots, u_d) du_1 du_2 \dots du_d - 1 \right\} \\ &= \frac{\int_{I^d} C(u) du - \int_{I^d} \Pi(u) du}{\int_{I^d} M(u) du - \int_{I^d} \Pi(u) du}. \end{aligned} \quad (3.9)$$

Car :

$$\begin{aligned} \int_{I^d} M(u) du &= E \{ \min(U_1, \dots, U_d) \} \\ &= E(X) = \int_0^1 u^d du = \frac{1}{d+1} \text{ et } \int_{I^d} \Pi(u) du = \frac{1}{2^d}. \end{aligned}$$

M est la fonction de distribution jointe du vecteur $\mathbf{U}=(U, \dots, U)$

3.2.1 La relation existante entre $\rho_1(C)$ et $\rho_2(C)$

Nous avons vu que si $d = 2$, on a :

$$\rho_1(C) = \rho_2(C) = \rho_3(C),$$

étudions le cas $d > 2$. si $d > 2$ la valeurs des ρ_i pour $i = 1, 2, 3$, sont différente en générale, il existe une relation intéressante entre les deux versions ρ_1 et ρ_2 , pour cela considérons le vecteur aléatoire $X = (X_1, \dots, X_d)$, et l'indice $I \subseteq \{1, \dots, d\}$, avec $2 \leq k \leq d$, et $|I| = \text{Card}I$ notons par $\rho_{1,I}$ la $|I|$ -dimensionnelle de la mesure ρ_1 , correspond à la valeur des X_i , $i \in I$, $I = \{i_1, \dots, i_d\}$, avec $2 \leq k \leq d$ et $|I| = k$ alors :

$$\rho_1(C_I) = h(k) \left\{ 2^k \int_{[0,1]^k} C_{i_1 \dots i_k}(u_{i_1}, \dots, u_{i_k}) du - 1 \right\}.$$

La relation est donnée par :

$$\rho_2(C) = \rho_{2,\{1, \dots, d\}} = \sum_{k=2}^d (-1)^k \frac{h(d) 2^d}{h(k) 2^k} \sum_{\substack{I \subseteq \{1, \dots, d\} \\ |I|=k}} \rho_1(C_I). \quad (3.10)$$

Dans le cas où C est radialement symétrique on a :

$$\rho_{2,\{1, \dots, d\}} = \rho_{1,\{1, \dots, d\}},$$

et les deux expressions ρ_1 , ρ_2 s'exprime en fonction de $\rho_1(C_I)$ comme suit :

1. Si d est impair nous avons :

$$\rho_{2,\{1, \dots, d\}} = \rho_{1,\{1, \dots, d\}} = \sum_{k=2}^{d-1} (-1)^k \frac{2^{d-1} h(d)}{2^k h(k)} \sum \rho_1(C_I)$$

2. Si d est pair : la relation (3.10) est vérifiée si $\sum_{k=2}^{d-1} (-1)^k \frac{2^d h(d)}{2^k h(k)} \sum \rho_{1(C_I)} = 0$

Preuve. L'expression (3.10) peut s'écrire sous la forme :

$$\begin{aligned} \rho_{2,\{1, \dots, d\}} &= \sum_{k=2}^{d-1} (-1)^k \frac{2^d h(d)}{2^k h(k)} \sum_{|I|=k} \rho_1(C_I) + (-1)^d \frac{2^d h(d)}{2^d h(d)} \sum_{|I|=d} \rho_1(C_I) \\ &= \sum_{k=2}^{d-1} (-1)^k \frac{2^d h(d)}{2^k h(k)} \sum_{|I|=k} \rho_1(C_I) + (-1)^d \rho_{1,\{1, \dots, d\}}. \end{aligned}$$

Premier cas si d est impair alors :

$$\rho_{2,\{1,\dots,d\}} = \sum_{k=2}^{d-1} (-1)^k \frac{2^d h(d)}{2^k h(k)} \sum \rho_1(C_I) - \rho_{1,\{1,\dots,d\}}$$

Ceci equivalent à :

$$2\rho_{2,\{1,\dots,d\}} = \sum_{k=2}^{d-1} (-1)^k \frac{2^d h(d)}{2^k h(k)} \sum_{|I|=k} \rho_1(C_I)$$

Alors :

$$\rho_{2,\{1,\dots,d\}} = \rho_{1,\{1,\dots,d\}} = \sum_{k=2}^{d-1} (-1)^k \frac{2^{d-1} h(d)}{2^k h(k)} \sum_{|I|=k} \rho_1(C_I)$$

Deuxième cas : Si d est pair on a :

$$\rho_{2,\{1,\dots,d\}} = \rho_{1,\{1,\dots,d\}} + \sum_{k=2}^{d-1} (-1)^k \frac{2^d h(d)}{2^k h(k)} \sum_{|I|=k} \rho_1(C_I).$$

L'égalité est vérifiée si $\sum_{k=2}^{d-1} (-1)^k \frac{2^d h(d)}{2^k h(k)} \sum_{|I|=k} \rho_1(C_I) = 0$. ■

Exemple 3.2.3 Prenons maintenant l'exemple suivant et étudions la relation qui existe entre ρ_1 et ρ_3 . Nous présenterons ici la copule C_0 définie par :

$$C_0 = \sum \binom{d}{2}^{-1} C_{kl}(u_k, u_l) \cdot \prod_{j \neq l, k} u_j,$$

où C_{kl} désigne la copule marginale qui correspond à la $k^{\text{ème}}$ et la $l^{\text{ème}}$ marge d'une copule quelconque C dans ce cas :

$$\begin{aligned}
& h(2) \left\{ 2^d \int_{[0,1]^d} C_0(u) du - 1 \right\} \tag{3.11} \\
&= h(2) \left\{ 2^d \int_{[0,1]^d} \sum_{1 \leq k < l \leq d} \binom{d}{2}^{-1} C_{kl}(u_k, u_l) \cdot \prod_{j \neq l, k} u_j du - 1 \right\} \\
&= h(2) \left\{ 2^d \sum_{1 \leq k < l \leq d} \binom{d}{2}^{-1} \int_{[0,1]^d} C_{kl}(u_k, u_l) \cdot \prod_{j \neq l, k} u_j du - 1 \right\} \\
&= h(2) \left\{ 2^d \sum_{1 \leq k < l \leq d} \binom{d}{2}^{-1} \int_{[0,1]^2} C_{kl}(u_k, u_l) du_k du_l \cdot \int_{[0,1]^{d-2}} \prod_{j \neq l, j \neq k} u_j du - 1 \right\} \\
&= h(2) \left\{ 2^d \sum_{1 \leq k < l \leq d} \binom{d}{2}^{-1} \int_{[0,1]^2} C_{kl}(u_k, u_l) du_k du_l \times \frac{1}{2^{d-2}} - 1 \right\} \\
&= h(2) \left\{ 2^2 \sum_{1 \leq k < l \leq d} \binom{d}{2}^{-1} \int_{[0,1]^2} C_{kl}(u_k, u_l) du_k du_l - 1 \right\} \\
&\stackrel{Def}{=} \rho_3(C).
\end{aligned}$$

D'autre part on a :

$$\rho_1(C_0) \stackrel{Def}{=} h(d) \left\{ 2^d \int_{[0,1]^d} C_0(u) du - 1 \right\}, \tag{3.12}$$

De l'équation (3.11) et (3.12) on en déduit qu'il existe donc une relation entre ρ_1 et ρ_3 donnée par :

$$\rho_3(C_0) = \frac{3\rho_1(C_0)}{h(d)}.$$

Si $d = 2$, alors $\rho_1 = \rho_3$. Notez aussi que $\rho_1(C_0) \leq h(d) \setminus 3$

Remarque 3.2.2 ρ_3 est une mesure d'association multidimensionnelle déterminée seulement par les copules bivariées par exemple la copule FGM, de dimension trois donnée par :

$$C_\lambda(u, v, w) = uvw + \lambda u(1-u)v(1-v)w(1-w), \quad |\lambda| \leq 1,$$

cette copule possède des marges bivariées indépendantes et $\rho_3 = 0$, par contre $\rho_1 = 8\lambda \cdot 6^{-3} \neq 0$ si $\lambda \neq 0$.

Preuve. Par définition :

$$\begin{aligned}
\rho_3 &= h(2) \left\{ 2^2 \sum_{1 \leq k < l \leq 3} \binom{3}{2}^{-1} \int_{[0,1]^2} C_{kl}(u_k, u_l) du_k du_l - 1 \right\} \\
&= 3 \left\{ 2^2 \sum_{1 \leq k < l \leq 3} \frac{1}{3} \int_{[0,1]^2} C_{kl}(u_k, u_l) du_k du_l - 1 \right\} \\
&= 4 \sum_{1 \leq k < l \leq 3} \int_{[0,1]^2} C_{kl}(u_k, u_l) du_k du_l - 3 \\
&= 4 \int_{[0,1]^2} C_{12}(u_1, u_2) du_1 du_2 + 4 \int_{[0,1]^2} C_{13}(u_1, u_3) du_1 du_3 \\
&\quad + 4 \int_{[0,1]^2} C_{23}(u_2, u_3) du_2 du_3 - 3 \\
&= 4 \left\{ \int_{[0,1]^2} C(u_1, u_2, 1) du_1 du_2 + \int_{[0,1]^2} C(u_1, 1, u_3) du_1 du_3 \right. \\
&\quad \left. + \int_{[0,1]^2} C(1, u_2, u_3) du_2 du_3 - 3 \right\} \\
&= 4 \left[\frac{1}{4} + \frac{1}{4} + \frac{1}{4} \right] - 3 = 0
\end{aligned}$$

Maintenant nous calculerons ρ_1 :

$$\begin{aligned}
\rho_1(C) &= h(3) \left\{ 2^3 \int_{[0,1]^3} C(u, v, w) dudvdw - 1 \right\} \\
&= 8 \int_0^1 \int_0^1 \int_0^1 uvwdudvdw + 8\lambda \int_0^1 \int_0^1 \int_0^1 uvwdudvdw - 1 \\
&= +8\lambda \left[\frac{1}{2} - \frac{1}{3} \right]^3 = \frac{8\lambda}{6^3} \neq 0 \text{ si } \lambda \neq 0.
\end{aligned}$$

C'est-à-dire dans ce cas $\rho_1 \neq \rho_3$. ■

3.3 Estimation non paramétrique de ρ_i par la copule empirique

Nous proposons des estimateurs de ρ_i , ($i = 1, 2, 3$) dans le cas multivariée. Dans cette section, nous abordons le problème d'estimation non paramétriques des ρ_i , ($i = 1, 2, 3$), dans ce cas nous supposons à voir un échantillon (X_1, \dots, X_d) ,

où $X_j = (X_{1j}, \dots, X_{dj})$, d -dimension $j = 1, \dots, n$. C'est une suite de variables aléatoires à valeurs dans \mathbb{R}^d , $(X_{11}, \dots, X_{d1}), \dots, (X_{1n}, \dots, X_{dn})$, et définissons les vecteurs de rangs associées à chaque observation par $(R_{11}, \dots, R_{d1}), \dots, (R_{1n}, \dots, R_{dn})$, où R_{ij} est le rang de X_{ij} parmi X_{i1}, \dots, X_{id} , soit F la fonction de répartition jointe du vecteur $X = (X_1, \dots, X_d)$, dont les marges F_i de copules C sont complètement inconnues. Alors les distributions F_i sont estimées par les fonctions marginales empiriques univariées $\hat{F}_{i,n}(x)$ tel que

$$\hat{F}_{i,n}(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_{ij} \leq x\}} \quad \text{pour } i = 1, \dots, d, \text{ et } x \in \mathbb{R}. \quad (3.13)$$

L'estimateur non paramétriques des ρ_i se base sur les rangs non pas sur les observations originales. Nous introduisons la copule empirique définie par :

$$\begin{aligned} \hat{C}_n(\mathbf{u}) &= \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{\hat{U}_{1,n} \leq u_1, \hat{U}_{2,n} \leq u_2, \dots, \hat{U}_{d,n} \leq u_d\}} \\ &= \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d \mathbf{1}_{\{\hat{U}_{ij} \leq u_i\}} \quad \text{pour tout } \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d \end{aligned} \quad (3.14)$$

$\hat{U}_{ij,n} = \frac{1}{n} R_{ij} = \hat{F}_{i,n}(X_{ij})$, et $\hat{U}_{j,n} = (\hat{U}_{1j,n}, \dots, \hat{U}_{dj,n})$ La distribution C est inconnue, mais on peut la remplacer par son estimateur \hat{C}_n (dans les formules des ρ_i avec $i = 1, 2, 3$) on obtient les valeurs de $\hat{\rho}_{i,n}$:

$$\begin{aligned} \hat{\rho}_{1,n}(\hat{C}_n) &= h(d) \left\{ 2^d \int_{I^d} \hat{C}_n(u) du - 1 \right\} \\ &= h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \prod_{i=1}^d (1 - \hat{U}_{ij,n}) - 1 \right\}. \end{aligned} \quad (3.15)$$

$$\begin{aligned} \hat{\rho}_{2,n}(\hat{C}_n) &= h(d) \left\{ 2^d \int_{I^d} \Pi d \hat{C}_n(u) du - 1 \right\} \\ &= h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \prod_{i=1}^d \hat{U}_{ij,n} - 1 \right\}. \end{aligned} \quad (3.16)$$

$$\begin{aligned} \hat{\rho}_{3,n}(\hat{C}_{kl}) &= 12 \sum_{1 \leq k < l \leq d} \frac{1}{\binom{d}{2}} \int \hat{C}_{kl}(u_k, u_l) du_k \cdot du_l - 3 \\ &= \frac{12}{n} \frac{1}{\binom{d}{2}} \sum_{1 \leq k < l \leq d} \sum_{j=1}^n (1 - \hat{U}_{kj,n}) (1 - \hat{U}_{lj,n}) - 3 \end{aligned} \quad (3.17)$$

Preuve. En utilisant les équations (3.1) et (3.2) on remplace les copules C , et \bar{C} par leur estimateur \hat{C}_n , $\hat{\bar{C}}_n$ respectivement et d'après la définition de la fonction indicatrice $\mathbf{1}_{\{A\}}$. Alors nous avons :

$$\begin{aligned} \rho_1(\hat{C}_n) &= h(d) \left\{ 2^d \int_{I^d} \hat{C}_n(\mathbf{u}) du - 1 \right\} = h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \int_{I^d} \prod_{i=1}^d \mathbf{1}_{\{\hat{U}_{ij} < u_i\}} du - 1 \right\} \\ &= h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \int_{I^d} [\mathbf{1}_{\{\hat{U}_{1j} < u_1\}} \times \dots \times \mathbf{1}_{\{\hat{U}_{dj} < u_d\}}] du_d \dots du_1 - 1 \right\} \\ &= h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \int_{\hat{U}_{1j}}^1 du_1 \times \int_{\hat{U}_{2j}}^1 du_2 \times \dots \times \int_{\hat{U}_{dj}}^1 du_d - 1 \right\} \\ &= h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \prod_{i=1}^d (1 - \hat{U}_{ij,n}) \right\}, \end{aligned}$$

Pour la deuxième version on a :

$$\rho_2(\hat{C}_n) = h(d) \left\{ 2^d \int_{I^d} \Pi(u) d\hat{C}_n(\mathbf{u}) - 1 \right\} = h(d) \left\{ 2^d \int_{I^d} \hat{\bar{C}}_n(\mathbf{u}) du - 1 \right\}.$$

Avec :

$$\hat{\bar{C}}_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d \mathbf{1}_{\{\hat{U}_{ij} > u_i\}}.$$

Nous avons :

$$\begin{aligned} \rho_2(\hat{C}_n) &= h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \int_{I^d} \prod_{i=1}^d \mathbf{1}_{\{\hat{U}_{ij} > u_i\}} du - 1 \right\} \\ &= h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \prod_{i=1}^d \int_0^{\hat{U}_{ij}} 1 du - 1 \right\} = h(d) \left\{ \frac{2^d}{n} \sum_{j=1}^n \prod_{i=1}^d \hat{U}_{ij,n} - 1 \right\} \end{aligned}$$

Pour la troisième version on à :

$$\begin{aligned} \rho_3(\hat{C}_{kl}) &= 12 \sum_{1 \leq k < l \leq d} \frac{1}{\binom{d}{2}} \int_{[0,1]^2} \hat{C}_{kl,n}(u_k, u_l) du_k du_l - 3 \\ &= \frac{24}{n} \frac{1}{d(d-1)} \sum_{1 \leq k < l \leq d} \int_{[0,1]^2} \sum_{j=1}^n \mathbf{1}_{\{\hat{U}_{kj} \leq u_k, \hat{U}_{lj} \leq u_l\}} du_k du_l - 3 \\ &= \frac{24}{n} \frac{1}{d(d-1)} \sum_{1 \leq k < l \leq d} \sum_{j=1}^n \int_{\hat{U}_{kj}}^1 \mathbf{1}_{\{\hat{U}_{kj,n} \leq u_k\}} du_k \times \int_{\hat{U}_{lj}}^1 \mathbf{1}_{\{\hat{U}_{lj,n} \leq u_l\}} du_l - 3. \end{aligned}$$

On en deduit que :

$$\rho_3 \left(\hat{C}_{kl} \right) = \frac{24}{n} \frac{1}{d(d-1)} \sum_{1 \leq k < l} \sum_{j=1}^n \left(1 - \hat{U}_{kj} \right) \left(1 - \hat{U}_{lj} \right) - 3.$$

■

Il est intéressant de donner un cas special ($d = 2$). Donc la mesure $\hat{\rho}_{1,n}$ définit par

$$\hat{\rho}_{1,n} = \frac{12}{n} \sum_{j=1}^n \left(1 - \hat{U}_{1j,n} \right) \left(1 - \hat{U}_{2j,n} \right), \quad (3.18)$$

est un estimateur de $\rho = \int_{[0,1]^2} C(u,v) dudv - 3$, il existe une légère différence entre la version classique $\hat{\rho}_{s,n}$ (Coefficient de rang) et la version estimée $\hat{\rho}_{1,n}$ avec :

$$\hat{\rho}_{s,n} = 1 - \frac{6n}{n^2 - 1} \sum_{j=1}^n \left(\hat{U}_{1j,n} - \hat{U}_{2j,n} \right)^2.$$

On montrera ici que ces deux statistiques ont la même loi asymptotique.

Remarque 3.3.1 L'équation (3.18) peut aussi s'écrire de la façon suivante :

$$\hat{\rho}_{1,n} = \frac{12}{n} \left[\sum_{j=1}^n \hat{U}_{1j,n} \hat{U}_{2j,n} - 1 \right] - 3.$$

Nous pouvons montrer que $\hat{\rho}_{1,n} - \hat{\rho}_{s,n} \leq 0$, et $\lim_{n \rightarrow \infty} \sqrt{n} (\hat{\rho}_{1,n} - \hat{\rho}_{s,n}) = 0$

Preuve. Nous avons :

$$\begin{aligned} \hat{\rho}_{1,n} - \hat{\rho}_{s,n} &= \frac{12}{n} \left[\sum_{j=1}^n \hat{U}_{1j,n} \hat{U}_{2j,n} - 1 \right] - 3 - \left[1 - \frac{6n}{n^2 - 1} \sum_{j=1}^n \left(\hat{U}_{1j,n} - \hat{U}_{2j,n} \right)^2 \right] \\ &= \frac{12}{n^3} \sum_{j=1}^n p_j q_j - \frac{12}{n} - 3 - \frac{12}{n(n^2 - 1)} \sum_{j=1}^n p_j q_j + \frac{3(n+1)^2}{n^2 - 1}, \end{aligned}$$

avec $p_j = n \hat{U}_{1j,n}$ et $q_j = n \hat{U}_{2j,n}$ (p_j représente le rang des X_{1j} parmi X_{11}, \dots, X_{1n}), (q_j représente le rang des X_{2j} parmi X_{21}, \dots, X_{2n}), On trouve donc :

$$\hat{\rho}_{1,n} - \hat{\rho}_{s,n} = \left(\frac{12}{n^3} - \frac{12}{n^3 - n} \right) \sum_{j=1}^n p_j q_j + \frac{6(-n+2)}{n(n-1)} \leq 0,$$

et

$$\lim_{n \rightarrow \infty} \sqrt{n} (\hat{\rho}_{1,n} - \hat{\rho}_{s,n}) = 0, \text{ pour tout } n \in \mathbb{N}.$$

Par conséquent dans le cas bidimensionnel. $\hat{\rho}_{1,n}$ et $\hat{\rho}_{s,n}$ la loi asymptotique de $\hat{\rho}_{1,n}$ coïncide donc avec celle de $\hat{\rho}_{s,n}$. ■

Si $(X_{11}, X_{21}), \dots, (X_{1n}, X_{2n})$ est une suite de variables aléatoires i. i. d. Alors $\sqrt{n}(\hat{\rho}_{1,n} - \rho_s)$ est asymptotiquement Gaussien centré, sa variance asymptotique a été étudiée par. Rüschendorf (1976), Genest et Rémillard (2004) et définie par la relation :

$$Var_{asym}(\hat{\rho}_{1,n}) = 144 \int_{I^2} \int_{I^2} E \{ \mathbb{G}_C(u, v) \mathbb{G}_C(s, t) \} dudvdsdt.$$

3.4 Convergence en loi du processus empirique de la copule

Avant d'étudier la normalité asymptotique du processus $\sqrt{n} \{ \hat{\rho}_{i,n} - \rho_i \}$, il faut étudier tout d'abord le comportement asymptotique du processus gaussien :

$$\sqrt{n} \{ \hat{C}_n(\mathbf{u}) - C(\mathbf{u}) \},$$

la dérivation des lois limites $\sqrt{n} \{ \hat{\rho}_{i,n}(\hat{C}_n) - \rho_i(C) \}$, exige deux théorèmes essentiels cités après qui concerne le comportement asymptotique du processus empirique multivarié de la copule.

$$\mathbb{G}_n(\mathbf{u}) = \sqrt{n} \{ \hat{C}_n(\mathbf{u}) - C(\mathbf{u}) \}, \quad \mathbf{u} \in [0, 1]^d,$$

qui a été développé dans plusieurs travaux et dans divers espaces. Par exemple la convergence faible aussi dite convergence en loi du processus \mathbb{G}_n vers un processus Gaussien a fait l'objet d'étude de plusieurs auteurs.

Rüschendorf (1976) ou bien Gänßler et Stute (1987) ont prouvé que \mathbb{G}_n converge faiblement dans l'espace $\mathbb{D}([0, 1]^d)$ (l'espace des fonctions continues à droite admetant une limite à gauche en chaque point de $[0, 1]^d$).

Vandert Vaart et Wellner, (1996) ont montré que la convergence faible a lieu dans $\ell^\infty([a, b]^d)$ lorsque $0 < a < b < 1$. Enfin, Fermanian et al. (2004) ont établi la convergence faible dans l'espace $\ell^\infty([0, 1]^d)$ moyennant la delta-méthode et sous réserve que les dérivées partielles d'ordre 1 de la copule C existent et soient continuent.

Tous ces auteurs, cités précédemment ont montré que le processus Gaussien limite s'écrit comme combinaisons de ponts browniens, et ont établi le théorème suivant :

Théorème 3.4.1 *Soit F une fonction de répartition d -dimensionnelle associée à une copule C . On suppose que pour tout $i = 1, \dots, d$ les dérivées partielles d'ordre 1 existent et sont continues. Alors, le processus empirique de la copule*

$$\mathbb{G}_n(\mathbf{u}) = \sqrt{n} \left\{ \hat{C}_n(\mathbf{u}) - C(\mathbf{u}) \right\} \overset{C.f.aible}{\rightsquigarrow} \mathbb{G}_C(\mathbf{u}), \quad (3.19)$$

dans l'espace $\ell^\infty([0, 1]^d)$ de plus le processus $\mathbb{G}_C(\mathbf{u})$ admet la représentation suivante :

$$\mathbb{G}_C(\mathbf{u}) = \mathbb{B}_C(\mathbf{u}) - \sum_{i=1}^d D_i C(\mathbf{u}) \mathbb{B}_C(u^{(i)}), \quad \text{pour } \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d, \quad (3.20)$$

où $u^{(i)} = (1, \dots, 1, u_i, 1, \dots, 1)$ et $\mathbb{B}_C(\mathbf{u}) = (\mathbb{B} \circ C)(\mathbf{u})$ est un processus Gaussien centré (Pont Brownien) de dimension d de fonction de covariance :

$$E \{ \mathbb{B}_C(\mathbf{u}) \mathbb{B}_C(\mathbf{v}) \} = C(\mathbf{u} \wedge \mathbf{v}) - C(\mathbf{u}) C(\mathbf{v}) \quad \text{pour tout } \mathbf{u}, \mathbf{v} \in [0, 1]^d. \quad (3.21)$$

Remarque 3.4.1 *La convergence de ce processus permet d'étudier la normalité asymptotique des estimateurs du rhô de Spearman*

Remarque 3.4.2 *L'équation (3.20) est équivalente à :*

$$\begin{aligned} \mathbb{G}_C(u_1, \dots, u_d) &= \mathbb{B}_C(u_1, \dots, u_d) - \partial_1 C(u_1, \dots, u_d) \mathbb{B}_C(u_1, 1, \dots, 1) \\ &\quad - \partial_2 C(u_1, \dots, u_d) \mathbb{B}_C(1, u_2, \dots, 1) \\ &\quad - \dots - \partial_d C(u_1, \dots, u_d) \mathbb{B}_C(1, \dots, 1, u_d), \end{aligned}$$

où \mathbb{B}_C un pont Brownien défini sur $[0, 1]^d$ de fonction de covariance

$$E \{ \mathbb{B}_C(\mathbf{u}) \mathbb{B}_C(\mathbf{v}) \} = C(u_1 \wedge v_1, \dots, u_d \wedge v_d) - C(u_1, \dots, u_d) C(v_1, \dots, v_d),$$

pour tout $0 \leq u_1, \dots, v_d \leq 1$.

Remarque 3.4.3 *Si $d = 2$ l'expression (3.20) devient :*

$$\mathbb{G}_C(u, v) = \mathbb{B}_C(u, v) - \partial_1 C(u, v) \mathbb{B}_C(u, 1) - \partial_2 C(u, v) \mathbb{B}_C(1, v),$$

et la fonction de covariance est donnée par :

$$E(\mathbb{B}_C(u, v) \mathbb{B}_C(u', v')) = C(u \wedge u', v \wedge v') - C(u, v) C(u', v'), \quad (3.22)$$

pour tout $(u, v, u', v') \in [0, 1]^4$.

Le théorème (3.3.1) est étroitement lié au deuxième théorème suivant dans lesquels nous avons besoins d'établir la normalité asymptotique du processus $\sqrt{n} \{ \hat{\rho}_{2,n} - \rho_2 \}$.

Théorème 3.4.2 *Soit F une fonction de distribution d -dimensionnelle associée à la copule C , supposons que la distribution jointe de \mathbf{U} et C , soit la fonction de survie $\bar{C}(\mathbf{u}) = \mathbb{P}[\mathbf{U} > \mathbf{u}]$. Considérons l'estimateur :*

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d \mathbf{1}_{\{\hat{U}_{ij} > u_i\}} \quad \text{pour } \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d.$$

Sous ces hypothèses et l'utilisation des notations du théorème (3.3.1) nous avons :

$$\bar{\mathbb{G}}_n(\mathbf{u}) = \sqrt{n} \left\{ \hat{C}_n(\mathbf{u}) - \bar{C}(\mathbf{u}) \right\} \overset{C.\text{faible}}{\rightsquigarrow} \mathbb{G}_{\bar{C}}(\mathbf{u}) \text{ dans l'espace } \ell^\infty \left([0, 1]^d \right),$$

et

$$\mathbb{G}_{\bar{C}}(\mathbf{u}) = \mathbb{B}_{\bar{C}}(\mathbf{u}) + \sum_{i=1}^d D_i \bar{C}(\mathbf{u}) \mathbb{B}_{\bar{C}}(u_{(i)}),$$

où $\mathbb{B}_{\bar{C}}(u_{(i)}) = -\mathbb{B}_{\bar{C}}(u^{(i)})$ avec $u_{(i)}$ le vecteur dont la i^{iem} composante vaut u_i et la $(d-1)$ composantes égale à zéro, $u^{(i)} = (0, 0, \dots, u_i, 0, \dots, 0)$. Le processus $\mathbb{B}_{\bar{C}}(\mathbf{u})$ est un processus Gaussien centré dans $[0, 1]^d$ de fonction de covariance

$$E \{ \mathbb{B}_{\bar{C}}(\mathbf{u}) \mathbb{B}_{\bar{C}}(\mathbf{v}) \} = \bar{C}(\mathbf{u} \vee \mathbf{v}) - \bar{C}(\mathbf{u}) \bar{C}(\mathbf{v}).$$

Preuve. Considérons l'estimateur : \bar{F}

$$\hat{C}_n^*(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d \mathbf{1}_{\{U_{ij} > u_i\}} \quad \text{pour tout } \mathbf{u} \in [0, 1]^d.$$

On note par F_i , $i = 1, \dots, d$,. Les fonctions de distributions marginales supposées connues. Soit $U_{ij} = F_i(X_{ij})$, Neahous (1971), Bickel et Wichura (1971) ont montré que le processus empirique de la copule dans le cas où les marges sont connues convergent vers le pont Brownien, $\mathbb{B}_{\bar{C}}(\mathbf{u})$ dans l'espace $\ell^\infty \left([0, 1]^d \right)$. D'autre part le processus empirique de la copule $\bar{\mathbb{G}}_n(\mathbf{u})$ peut se décomposer de

la manière suivante :

$$\begin{aligned}
\bar{\mathbb{G}}_n(\mathbf{u}) &= \sqrt{n} \left[\widehat{F}_n \left\{ \widehat{F}_{1,n}^{-1}(u_1), \dots, \widehat{F}_{d,n}^{-1}(u_d) \right\} - \bar{C}(\mathbf{u}) \right] \\
&= \sqrt{n} \left[\left(\widehat{F}_n - \bar{F} \right) \left\{ \widehat{F}_{1,n}^{-1}(u_1), \dots, \widehat{F}_{d,n}^{-1}(u_d) \right\} \right. \\
&\quad \left. - \left(\widehat{F}_n - \bar{F} \right) \left\{ F_1^{-1}(u_1), \dots, F_d^{-1}(u_d) \right\} \right] \\
&\quad + \sqrt{n} \left[\bar{F} \left\{ \widehat{F}_{1,n}^{-1}(u_1), \dots, \widehat{F}_{d,n}^{-1}(u_d) \right\} \right. \\
&\quad \left. - \bar{F} \left\{ F_1^{-1}(u_1), \dots, F_d^{-1}(u_d) \right\} \right] \\
&\quad + \sqrt{n} \left(\widehat{F}_n - \bar{F} \right) \left\{ F_1^{-1}(u_1), \dots, F_d^{-1}(u_d) \right\}
\end{aligned}$$

où $\widehat{F}_n(x) = \frac{1}{n} \sum_{j=1}^d \prod_{i=1}^d \mathbf{1}_{\{X_j > x_i\}}$, et avec $\bar{F}(x) = \mathbb{P}(X > x)$

Noter que $\sqrt{n} \left[\widehat{C}_n^*(\mathbf{u}) - \bar{C}(\mathbf{u}) \right] \overset{C.f}{\rightsquigarrow} B_{\bar{C}}(\mathbf{u})$ dans $\ell^\infty([0, 1]^d)$ par contre

$$\sqrt{n} \left[\bar{F} \left\{ \widehat{F}_{1,n}^{-1}(u_1), \dots, \widehat{F}_{d,n}^{-1}(u_d) \right\} - \bar{C}(\mathbf{u}) \right] \overset{C.f}{\rightsquigarrow} - \sum_{i=1}^d D_i \bar{C}(\mathbf{u}) \mathbb{B}_{\bar{C}}(u^{(i)})$$

Celle ci découle par la representation du processus empirique uniforme (Bahadur 1966) et par applications de formule de Delta methode fonctionnelle (Vandert Vart et Wellener1996 : page394), la convergence faible a lieu dans l'espace $\ell^\infty([0, 1]^d)$ le reste converge vers zero en probabilité. en raison de la convergence faible du $\sqrt{n} \left(\widehat{F}_n - \bar{F} \right)$. dans l'espace $\ell^\infty([-\infty, \infty]^d)$ ■

3.5 La normalité asymptotique des estimateurs

L'objet de cette section est d'établir la normalité asymptotique de l'estimateur $\hat{\rho}_i, i = 1, 2, 3$. On présente ici un théorème de la normalité asymptotique On étudie deux applications concrètes de ce théorème. On souhaite cependant ici construire un intervalle de confiance de niveau $1 - 2\alpha$ pour ρ_i . C'est-à-dire pour σ^2 fixée,

$$\sqrt{n} \left\{ \hat{\rho}_{i,n} \left(\hat{C}_n \right) - \rho_i(C) \right\} \overset{C.f}{\rightsquigarrow} Z_i \sim \mathcal{N}(0, \sigma_i^2)$$

On pourrait directement construire un intervalle de confiance asymptotique de niveau asymptotique $1 - 2\alpha$ pour ρ_i . Deuxième cas, les bornes de l'intervalle de

confiance font intervenir la quantité $\sigma_i(\theta)$, qui dépend d'un paramètre inconnu θ . Nous allons résoudre ce problème par la Delta méthode.

Définition 3.5.1 (*Normalité asymptotique*) .Une suite (\hat{g}_n) d'estimateurs de $g(\theta_0)$ est dite asymptotiquement normale, à vitesse \sqrt{n} et de variance asymptotique σ_g^2 , lorsque

$$\sqrt{n}(\hat{g}_n - g(\theta_0)) \rightarrow \mathcal{N}(0, \sigma_g^2)$$

Proposition 3.5.1 (*Propriétés de la convergence en loi*) .Si on a la convergence en loi $Y_n \rightarrow Y$ et si ϕ est une fonction continue, alors on a encore $\phi(Y_n) \rightarrow \phi(Y)$. Si la suite de variables aléatoires (Y_n) est asymptotiquement normale, telle qu'il existe y et σ_y^2 telque

$$\sqrt{n}(Y_n - Y) \rightarrow \mathcal{N}(0, \sigma_y^2)$$

et si ϕ est de classe C^1 , alors $(\phi(Y_n))$ est également asymptotique, normale

$$\sqrt{n}(\phi(Y_n) - \phi(Y)) \rightarrow \mathcal{N}(0, \phi'(Y)^2 \sigma_y^2)$$

Théorème 3.5.1 Soit $\hat{\rho}_{i,n}$, $i = 1, 2, 3$ l'estimateur définie par les formules précédentes Soit F la fonction de distribution jointe de copule C dont les marges F_i sous les hypothèses du théorèmes (3.3.1) et (3.3.2). Alors les processus :

$$\sqrt{n} \left\{ \hat{\rho}_{i,n}(\hat{C}_n) - \rho_i(C) \right\} \stackrel{C.f.}{\rightsquigarrow} Z_i \sim \mathcal{N}(0, \sigma_i^2) \quad \text{pour tout } i = 1, 2, 3$$

et les variances asymptotiques sont données par :

$$\sigma_1^2 = 2^{2d} h(d)^2 \int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(\mathbf{u}) \mathbb{G}_C(\mathbf{v}) \} d\mathbf{u} d\mathbf{v}, \quad (3.23)$$

$$\sigma_2^2 = 2^{2d} h(d)^2 \int_{I^d} \int_{I^d} E \{ \mathbb{G}_{\bar{C}}(\mathbf{u}) \mathbb{G}_{\bar{C}}(\mathbf{v}) \} d\mathbf{u} d\mathbf{v}, \quad (3.24)$$

$$\sigma_3^2 = 144 \sum_{\substack{k < l \\ s < t}} \binom{d}{2}^{-2} \int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(s,t)}) \} d\mathbf{u} d\mathbf{v} \quad (3.25)$$

avec $u^{(k,l)} = (1, \dots, 1, u_k, 1, \dots, 1, u_l, 1, \dots, 1)$.

Preuve. Posons dans la suite

$$S_{n,1} = \sqrt{n} \{ \hat{\rho}_{1,n} - \rho_1 \}$$

$$S_{n,2} = \sqrt{n} \{ \hat{\rho}_{2,n} - \rho_2 \}$$

$$S_{n,3} = \sqrt{n} \{ \hat{\rho}_{3,n} - \rho_3 \}$$

Étudions chaque terme séparément :

L'étude du terme $S_{n,1}$ Rappelons les données du théorème (3.3.1) et (3.3.2) et le théorème de l'image continue ((1.3.6) de Vander Vaart, (1996)), parceque l'opérateur integral est une application lineaire continue de $\ell^\infty([0, 1]^d)$ dans \mathbb{R} on a :

$$Z_1 = \lim_{n \rightarrow \infty} S_{n,1} = h(d) 2^d \int_{I^d} \lim_{n \rightarrow \infty} \mathbb{G}_n(\mathbf{u}) d\mathbf{u} = h(d) 2^d \int_{I^d} \mathbb{G}_C(\mathbf{u}) d\mathbf{u}.$$

Calculons $E(Z_1)$, et σ_1^2

$$E(Z_1) = h(d) 2^d E \int_{I^d} \mathbb{G}_C(u) d\mathbf{u} = h(d) 2^d \int_{I^d} E(\mathbb{G}_C(\mathbf{u})) d\mathbf{u} = 0.$$

Car le processus Gaussien limite $G_C(\mathbf{u})$ s'ecrit comme une combinaison de ponts Brownien C'est-à-dire que :

$$E(\mathbb{G}_C(\mathbf{u})) = 0.$$

Pour la variance σ_1^2 nous trouvons :

$$\begin{aligned} \sigma_1^2 &= \text{Var}(Z_1) = E(Z_1^2) = 2^{2d} h(d)^2 E \left[\left(\int_{I^d} \mathbb{G}_C(\mathbf{u}) d\mathbf{u} \right)^2 \right] \\ &= 2^{2d} h(d)^2 E \left[\int_{I^d} \mathbb{G}_C(\mathbf{u}) d\mathbf{u} \int_{I^d} \mathbb{G}_C(\mathbf{u}) d\mathbf{u} \right]. \end{aligned}$$

Selon le théorème de Fubini, on peut ecire cette expression sous la forme :

$$\begin{aligned} \sigma_1^2 &= 2^{2d} h(d)^2 E \left\{ \int_{I^d} \int_{I^d} (\mathbb{G}_C(\mathbf{u}) \mathbb{G}_C(\mathbf{v})) d\mathbf{u} d\mathbf{v} \right\} \\ &= 2^{2d} h(d)^2 \int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(u) \mathbb{G}_C(\mathbf{v}) \} d\mathbf{u} d\mathbf{v}. \end{aligned}$$

L'étude du terme $S_{n,2}$

Les résultats de la convergence en loi pour les autres s'obtient de façon similaire on a :

$$\begin{aligned} Z_2 &= \lim_{n \rightarrow \infty} S_{n,2} = h(d) 2^d \lim_{n \rightarrow \infty} \int_{I^d} \bar{\mathbb{G}}_n(\mathbf{u}) d\mathbf{u} \\ &= \lim_{n \rightarrow \infty} S_{n,2} = h(d) 2^d \int_{I^d} \lim_{n \rightarrow \infty} \bar{\mathbb{G}}_n(\mathbf{u}) d\mathbf{u} = h(d) 2^d \int_{I^d} \mathbb{G}_{\bar{C}}(\mathbf{u}) d\mathbf{u}. \end{aligned}$$

$$\sigma_2^2 = \text{Var}(Z_2) = E(Z_2^2) = 2^{2d} h(d)^2 \int_{I^d} \int_{I^d} E\{G_{\bar{C}}(\mathbf{u}) G_{\bar{C}}(\mathbf{v})\} d\mathbf{u} d\mathbf{v}.$$

L'étude du terme $S_{n,3}$

On a si $1 < i < j < n$, alors :

$$C_{ij}(u_i, u_j) = C(1, \dots, 1, u_i, 1, \dots, 1, u_j, 1, \dots, 1) = C(u^{(i,j)}) \quad (3.26)$$

Donc :

$$\begin{aligned} S_{n,3} &= 12 \sum_{1 \leq k < l \leq d} \binom{d}{2}^{-1} \int_{[0,1]^2} \sqrt{n} [\hat{C}_{kl,n}(u_k, u_l) - C_{kl}(u_k, u_l)] du_k du_l \\ &= \frac{24}{d(d-1)} \sum_{1 \leq k < l \leq d} \int_{I^d} \sqrt{n} \{ \hat{C}_n(u^{(k,l)}) - C(u^{(k,l)}) \} d\mathbf{u} d\mathbf{v}. \end{aligned}$$

D'après le théorème (1) et de la continuité du processus $\sqrt{n} \{ \hat{C}_n(\mathbf{u}) - C(\mathbf{u}) \}$ on a :

$$\sqrt{n} \{ \hat{\rho}_{3,n} - \rho_3 \} \xrightarrow{C.f} Z_3,$$

où

$$Z_3 = \frac{24}{d(d-1)} \sum_{1 \leq k < l} \int_{I^d} \mathbb{G}_C(u^{(k,l)}) du.$$

Calculons la variance asymptotique de Z_3 :

$$\sigma_3^2 = \text{Var}(Z_3) = E(Z_3^2) = E \left\{ 12 \sum_{1 \leq k < l} \binom{d}{2}^{-1} \int_{I^d} \mathbb{G}_C(u^{(k,l)}) du \right\}^2,$$

et par application du théorème de Fubini :

$$\begin{aligned} \sigma_3^2 &= 144 \sum_{1 \leq k < l} \sum_{1 \leq k < l} \binom{d}{2}^{-2} E \left\{ \int_{I^d} \mathbb{G}_C(u^{(k,l)}) du \right\}^2 \\ &= 144 \sum_{\substack{1 \leq k < l \\ 1 \leq s < t}} \binom{d}{2}^{-2} E \left\{ \int_{I^d} \mathbb{G}_C(u^{(k,l)}) du \int_{I^d} \mathbb{G}_C(v^{(s,t)}) dv \right\}. \end{aligned}$$

Finalement :

$$\sigma_3^2 = 144 \sum_{\substack{k < l \\ s < t}} \binom{d}{2}^{-2} \int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(s,t)}) \} dudv.$$

■

Remarque 3.5.1 *Le processus $\mathbb{G}_C(u^{(k,l)})$ donnée en equation (3.25) s'écrit d'une autre manière sous la forme :*

$$\mathbb{G}_C(u^{(k,l)}) = \mathbb{B}_C(u^{(k,l)}) - D_k C(u^{(k,l)}) \mathbb{B}_C(u^{(k)}) - D_l C(u^{(k,l)}) \mathbb{B}_C(u^{(l)})$$

Nous avons :

$$\begin{aligned} \mathbb{G}_C(u^{(k,l)}) &= \mathbb{G}_C(1, \dots, 1, u_k, 1, \dots, 1, u_l, 1, \dots, 1) \\ &= \mathbb{B}_C(u^{(k,l)}) - \sum_{i=1}^d D_i C(u^{(k,l)}) \mathbb{B}_C\left[(u^{(k,l)})^{(i)}\right] \\ &= \mathbb{B}_C(u^{(k,l)}) - [\partial_1 C(u^{(k,l)}) \mathbb{B}_C(\underline{1}, 1, \dots, 1) \\ &\quad + \partial_2 C(u^{(k,l)}) \mathbb{B}_C(1, \underline{1}, \dots, 1) + \dots + \partial_k C(u^{(k,l)}) \mathbb{B}_C(1, 1, \dots, \underline{u}_k, 1, \dots, 1) \\ &\quad + \dots + \dots + \partial_l C(u^{(k,l)}) \mathbb{B}_C(1, 1, \dots, \underline{u}_l, \dots, 1)] \\ &= \mathbb{B}_C(u^{(k,l)}) - \partial_k C(u^{(k,l)}) \mathbb{B}_C(u^{(k)}) - \partial_l C(u^{(k,l)}) \mathbb{B}_C(u^{(l)}). \end{aligned}$$

Puisque $\mathbb{B}_C(1, 1, \dots, 1) = 0$. On peut formuler σ_3^2 d'une autre manière :

$$\begin{aligned} \sigma_3^2 &= \left[144 \sum_{k < l} \binom{d}{2}^{-2} \int_{I^2} \int_{I^2} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(k,l)}) \} d(u_k, u_l) d(v_k, v_l) \right. \\ &\quad \left. + \sum_{k < l, r < s, \{k,l\} \neq \{r,s\}} \binom{d}{2}^{-2} \int_{I^2} \int_{I^2} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(r,s)}) \} d(u_k, u_l) d(v_r, v_s) \right]. \end{aligned} \quad (3.27)$$

Nous définissons par la suite quelques processus Gaussiens qui interviennent dans l'étude du processus empirique de la copule multivarié.

3.5.1 Processus Gaussiens

Définition 3.5.2 *Soit C une copule.*

- *Le processus $\{\mathbb{W}_C(\mathbf{u}) : \mathbf{u} \in [0, 1]^d\}$ est appelé **processus de Wiener** à d -paramètres est un processus Gaussien centré de fonction de covariance*

$$E \{ \mathbb{W}_C(\mathbf{u}) \mathbb{W}_C(\mathbf{v}) \} = C(\mathbf{u} \wedge \mathbf{v}). \text{ pour tout } \mathbf{u} \in [0, 1]^d.$$

- *Plusieurs processus Gaussiens ont été construits à partir de $\{\mathbb{W}_C(\mathbf{u}) : \mathbf{u} \in [0, 1]^d\}$. Par exemple le processus dit **pont Brownien***
- ***Le pont Brownian** multivariée définie en fonction de \mathbb{W}_C sur $[0, 1]^d$ par :*

$$\mathbb{B}_C(\mathbf{u}) = \mathbb{W}_C(\mathbf{u}) - C(\mathbf{u}) \mathbb{W}_C(1, \dots, 1),$$

vérifie :

$$E \{ \mathbb{B}_C(\mathbf{u}) \} = 0, \quad \text{et} \quad E \{ \mathbb{B}_C(\mathbf{u}) \mathbb{B}_C(\mathbf{v}) \} = C(\mathbf{u} \wedge \mathbf{v}) C(\mathbf{u}) C(\mathbf{v})$$

Remarque 3.5.2 (Les variables aléatoires sont indépendantes)

Lorsque $C(\mathbf{u}) = \prod_{i=1}^d u_i$ nous avons pour tout $\mathbf{u}, \mathbf{v} \in [0, 1]^d$:

1. $E \{ \mathbb{W}_C(\mathbf{u}) \} = 0$ et $E \{ \mathbb{W}_C(\mathbf{u}) \mathbb{W}_C(\mathbf{v}) \} = \prod_{i=1}^{i=d} \min(u_i, v_i)$.
2. $E \{ \mathbb{B}_C(\mathbf{u}) \} = 0$ et $E \{ \mathbb{B}_C(\mathbf{u}) \mathbb{B}_C(\mathbf{v}) \} = \prod_{i=1}^{i=d} \min(u_i, v_i) - \prod_{i=1}^{i=d} (u_i v_i)$.
3. Le processus \mathbb{G}_C définie par :

$$\begin{aligned} \mathbb{G}_C(\mathbf{u}) &= \mathbb{B}_C(\mathbf{u}) - \sum_{i=1}^{i=d} \frac{\partial_i \Pi(\mathbf{u})}{\partial u_i} \mathbb{B}_C(\mathbf{1}, \dots, 1, u_i, 1, \dots, 1) \\ &= \mathbb{B}_C(\mathbf{u}) - \sum_{i=1}^{i=d} \left\{ \prod_{j \neq i}^d u_j \right\} \mathbb{B}_C(\mathbf{1}, u_i, \mathbf{1}), \quad \mathbf{1} = \{1, \dots, 1\}, \end{aligned}$$

est un processus Gaussien centré de fonction de covariance

$$E \{ \mathbb{G}_C(\mathbf{u}) \mathbb{G}_C(\mathbf{v}) \} = \prod_{i=1}^{i=d} \{ (u_i \wedge v_i) - u_i v_i \}, \quad \text{pour } u, v \in [0, 1]^d.$$

Proposition 3.5.2 Soit C une copule radialement symétrique sous les hypothèses et l'utilisation des notation du théorème (3.3.1) et (3.3.2) le processus $\mathbb{G}_C(\mathbf{u})$ et $\mathbb{G}_{\bar{C}}(\mathbf{1} - \mathbf{u})$ sont également distribuées. Par contre les variances asymptotiques σ_1^2 et σ_2^2 données dans les formules (3.23) et (3.24) sont égales

3.6 Exemples (Variance asymptotique)

3.6.1 Exemple de la copule d'indépendance

Considérons le vecteur aléatoire $X = (X_1, \dots, X_d)$, où X_1, \dots, X_d sont stochastiquement indépendantes leurs copule est $C(u) = \prod_{i=1}^{i=d} u_i = u_1 \times u_2 \times \dots \times u_d$.

Proposition 3.6.1 Soit C la copule produit $\Pi(u) = \prod_{i=1}^{i=d} u_i$. Alors les variances asymptotiques dans le théorème (3.4.1) sont données par :

$$\sigma_1^2 = \sigma_2^2 = -\frac{(d+1)^2 \left\{ 3 + d - 3 \left(\frac{4}{3} \right)^d \right\}}{3(1+d-2^d)^2} \quad \text{et} \quad \sigma_3^2 = \binom{d}{2}^{-2}.$$

Preuve. On a la copule produit Π est radialement symétrique. D'après la proposition (3.4.2) on a :

$$\sigma_1^2 = \sigma_2^2 \quad \text{et} \quad D_i C(\mathbf{u}) = \frac{\partial_i \Pi(\mathbf{u})}{\partial u_i} = \prod_{k \neq i}^d u_k.$$

C'est la copule k -dimensionnelle de la copule $\prod u_i = u_1 \dots u_d$ on a :

$$\sigma_1^2 = \sigma_2^2 = \frac{2^{2d} (d+1)^2}{[2^d - (d+1)]^2} \int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(\mathbf{u}) \mathbb{G}_C(\mathbf{v}) \} d\mathbf{u} d\mathbf{v},$$

avec $\mathbf{u} = (u_1, \dots, u_d)$, et $\mathbf{v} = (v_1, \dots, v_d)$

Pour être précis, on doit calculer le produit $\mathbb{G}_C(\mathbf{u}) \mathbb{G}_C(\mathbf{v})$. Compte tenu de la relation (3.19) et (3.20) nous avons :

$$\begin{aligned} \mathbb{G}_C(\mathbf{u}) \mathbb{G}_C(\mathbf{v}) &= \left[\mathbb{B}_C(\mathbf{u}) - \sum_{i=1}^d D_i C(\mathbf{u}) \mathbb{B}_C(u^{(i)}) \right] \left[\mathbb{B}_C(\mathbf{v}) - \sum_{i=1}^d D_j C(\mathbf{v}) \mathbb{B}_C(v^{(j)}) \right] \\ &= \mathbb{B}_C(\mathbf{u}) \mathbb{B}_C(\mathbf{v}) - \sum_{j=1}^d D_j C(\mathbf{v}) \mathbb{B}_C(\mathbf{u}) \mathbb{B}_C(v^{(j)}) \\ &\quad - \sum_{i=1}^d D_i C(\mathbf{u}) \mathbb{B}_C(\mathbf{v}) \mathbb{B}_C(u^{(i)}) \\ &\quad + \sum_{i=1}^d D_i C(\mathbf{u}) \mathbb{B}_C(u^{(i)}) \sum_{j=1}^d D_j C(\mathbf{v}) \mathbb{B}_C(v^{(j)}). \end{aligned}$$

Première étape :

$$\begin{aligned}
& \int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(\mathbf{u}) \mathbb{G}_C(\mathbf{v}) \} d\mathbf{u} d\mathbf{v} \\
= & \int_{I^d} \int_{I^d} E \{ \mathbb{B}_C(\mathbf{u}) \mathbb{B}_C(\mathbf{v}) \} d\mathbf{u} d\mathbf{v} \\
& - \sum_{j=1}^d \int_{I^d} \int_{I^d} E \{ \mathbb{B}_C(\mathbf{u}) D_j C(\mathbf{v}) \mathbb{B}_C(v^{(j)}) \} d\mathbf{u} d\mathbf{v} \\
& - \sum_{i=1}^d \int_{I^d} \int_{I^d} E \{ \mathbb{B}_C(\mathbf{v}) D_i C(\mathbf{u}) \mathbb{B}_C(u^{(i)}) \} d\mathbf{u} d\mathbf{v} \\
& + \sum_{i=1}^d \sum_{j=1}^d \int_{I^d} \int_{I^d} E \{ D_i C(\mathbf{u}) \mathbb{B}_C(u^{(i)}) D_j C(\mathbf{v}) \mathbb{B}_C(v^{(j)}) \}
\end{aligned}$$

a) **D'abord traitant le cas** $i \neq j$, dans ce qui suit nous supposons que $i < j$, et utilisant l'indépendance des variables aléatoires nous pouvons écrire :

$$\begin{aligned}
& E \{ D_i C(\mathbf{u}) \mathbb{B}_C(u^{(i)}) D_j C(\mathbf{v}) \mathbb{B}_C(v^{(j)}) \} \\
= & E \{ \Pi u_k \Pi v_{k'} \mathbb{B}_C(u^{(i)}) \mathbb{B}_C(v^{(j)}) \}
\end{aligned} \tag{3.28}$$

et

$$\begin{aligned}
& E \{ \Pi u_k \Pi v_{k'} \} = \left(\frac{1}{2^{d-1}} \right)^2, \quad k \neq i, \quad j \neq k'. \\
E \{ \mathbb{B}_C(u^{(i)}) \mathbb{B}_C(v^{(j)}) \} &= \Pi(u^{(i)} \wedge v^{(j)}) - \Pi(u^{(i)}) \Pi(v^{(j)}), \\
\min(u^{(i)}, v^{(j)}) &= (1, \dots, 1, u_i, 1, \dots, 1, v_j, 1, \dots, 1), \\
\Pi(u^{(i)} \wedge v^{(j)}) &= u_i v_j = \Pi(u^{(i)}) \Pi(v^{(j)}).
\end{aligned} \tag{3.29}$$

Donc Si $i \neq j$ et après calculs, nous obtenons les relations suivante :

$$E \{ \mathbb{B}_C(u^{(i)}) \mathbb{B}_C(v^{(j)}) \} = 0.$$

Et enfin

$$\int_{I^d} \int_{I^d} E \{ D_i C(\mathbf{u}) \mathbb{B}_C(u^{(i)}) D_j C(\mathbf{v}) \mathbb{B}_C(v^{(j)}) \} d\mathbf{u} d\mathbf{v} = 0. \tag{3.30}$$

b) Prenant le cas $i = j$, Soit :

$$\begin{aligned}
I_1 &= \int_{I^d} \int_{I^d} E \{ DiC(\mathbf{u}) \mathbb{B}_C(u^{(i)}) DiC(\mathbf{v}) \mathbb{B}_C(v^{(i)}) \} d\mathbf{u}d\mathbf{v} \\
&= \int_{I^d} \int_{I^d} E \{ \Pi(u_k) \Pi(v_k) \mathbb{B}_C(u^{(i)}) \mathbb{B}_C(v^{(i)}) \} d\mathbf{u}d\mathbf{v}, \quad k \neq i \\
&= \frac{1}{2^{d-1}} \frac{1}{2^{d-1}} \int_{I^d} E \{ \mathbb{B}_C(u^{(i)}) \mathbb{B}_C(v^{(i)}) \} d\mathbf{u}d\mathbf{v} \\
&= \frac{1}{2^{2d-2}} \int_{I^2} [\min\{u_i, v_i\} - u_i v_i] d\mathbf{u}d\mathbf{v},
\end{aligned}$$

$$\text{avec } u^{(i)} \wedge v^{(i)} = (1, \dots, 1, u_i \wedge v_i, 1, \dots, 1)$$

On conclue que

$$\begin{aligned}
&\sum_{i=1}^d \int_{I^d} \int_{I^d} E \{ DiC(u) \mathbb{B}_C(u^{(i)}) DiC(v) \mathbb{B}_C(v^{(i)}) \} dudv \\
&= \frac{1}{2^{2d-2}} \sum_{i=1}^d \left[\int_{I^2} u_i \wedge v_i dudv - \int_{I^2} u_i v_i dudv \right] \\
&= \frac{d}{3 \times 2^{2d-2}} - \frac{d}{2^{2d}} = d \left[\frac{2^{-2d+2}}{3} - 2^{-2d} \right] = d \left[\frac{2^{-2d+2}}{3} - 2^{-2d} \right] \quad (3.31)
\end{aligned}$$

En suivant le même raisonnement pour I_2 :

$$I_2 = \sum_{i=1}^d \int_{I^d} \int_{I^d} E \{ \mathbb{B}_C(u) DiC(v) \mathbb{B}_C(v^{(i)}) \} dudv,$$

on a :

$$\begin{aligned}
I_2 &= \sum_{i=1}^d \int_{I^d} \int_{I^d} E \{ \mathbb{B}_C(\mathbf{u}) DiC(\mathbf{v}) \mathbb{B}_C(v^{(i)}) \} d\mathbf{u}d\mathbf{v} \\
&= \sum_{i=1}^d \int_{I^d} \int_{I^d} E \{ \mathbb{B}_C(\mathbf{v}) DiC(\mathbf{u}) \mathbb{B}_C(u^{(i)}) \} d\mathbf{u}d\mathbf{v} \\
&= \sum_{i=1}^d \int_{I^d} \int_{I^d} E \{ \Pi(u_k) \mathbb{B}_C(\mathbf{v}) \mathbb{B}_C(u^{(i)}) \} d\mathbf{u}d\mathbf{v}, \text{ pour tout } k \neq i \\
&= \sum \int_{I^{d-1}} E \{ \Pi(u_k) \} \int_{I^{d+1}} E \{ \mathbb{B}_C(\mathbf{v}) \mathbb{B}_C(u^{(i)}) \} dudv \\
&= \frac{1}{2^{d-1}} \sum_{i=1}^d \int_{I^{d+1}} [\Pi(u \wedge v^{(i)}) - \Pi(u) C(v^{(i)})] dudv, \\
&= \frac{1}{2^{d-1}} \sum_{i=1}^d \left[\int_{I^{d-1}} u_1 \times \dots \times u_{d-1} du \int_{I^2} u_i \wedge v_i du_i dv_i - \int_{I^{d+1}} (u_1 \times \dots \times u_d \times v_i) dudv \right] \\
&= \frac{d}{2^{2d-2}} \frac{1}{3} - \frac{d}{2^{2d-1}} \frac{1}{2} = d \left[\frac{2^{-2d+2}}{3} - 2^{-2d} \right] \tag{3.32}
\end{aligned}$$

Maintenant on va calculer I_3

$$\begin{aligned}
I_3 &= \int_{I^d} \int_{I^d} E \{ \mathbb{B}_C(\mathbf{u}) \mathbb{B}_C(\mathbf{v}) \} d\mathbf{u}d\mathbf{v} = \int_{I^d} \int_{I^d} [C(\mathbf{u} \wedge \mathbf{v}) - C(\mathbf{u}) C(\mathbf{v})] d\mathbf{u}d\mathbf{v} \\
&= \int_{I^d} \int_{I^d} [\Pi(\mathbf{u} \wedge \mathbf{v}) - \Pi(\mathbf{u}) \Pi(\mathbf{v})] d\mathbf{u}d\mathbf{v} \\
&= \int_{I^d} \int_{I^d} (u_1 \wedge v_1) \times \dots \times (u_d \wedge v_d) d\mathbf{u}d\mathbf{v} \\
&\quad - \int_{I^d} \int_{I^d} (u_1 \times \dots \times u_d) \times (v_1 \times \dots \times v_d) d\mathbf{u}d\mathbf{v} \\
&= \int_{I^2} (u_1 \wedge v_1) du_1 dv_1 \times \dots \times \int_{I^2} (u_d \wedge v_d) du_d dv_d \\
&\quad - \int_{I^d} \Pi(\mathbf{u}) d\mathbf{u} \int_{I^d} \Pi(\mathbf{v}) d\mathbf{v} \\
&= \frac{1}{3} \times \dots \times \frac{1}{3} - \frac{1}{2^d} \times \frac{1}{2^d} = \frac{1}{3^d} - \frac{1}{2^{2d}}. \tag{3.33}
\end{aligned}$$

En regroupant les equations (3.31), (3.32), et (3.33), nous obtenons

$$\begin{aligned}
\sigma_1^2 &= \sigma_2^2 \\
&= 2^{2d} h^2(d) \left\{ \frac{1}{3^d} - \frac{1}{2^{2d}} - 2d \left[\frac{2^{2-2d}}{3} - 2^{-2d} \right] + d \left[\frac{2^{2-2d}}{3} - 2^{-2d} \right] \right\} \\
&= \frac{2^{2d} [d+1]^2}{[2^d - (d+1)]^2} \left\{ 3^{-d} - 2^{-2d} - d \frac{2^{2-2d}}{3} + d 2^{-2d} \right\} \\
&= \frac{[d+1]^2}{[2^d - (d+1)]^2} \left\{ 3^{-d} 2^{2d} - 1 - \frac{4d}{3} + d \right\}.
\end{aligned}$$

On déduit que :

$$\sigma_1^2 = \sigma_2^2 = \frac{[d+1]^2}{[2^d - (d+1)]^2} \left\{ \left(\frac{4}{3} \right)^d - \frac{d+3}{3} \right\}.$$

Calculons σ_3^2

$$\sigma_3^2 = 144 \sum_{\substack{k < l \\ s < t}} \binom{d}{2}^{-2} \int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(s,t)}) \} d\mathbf{u} d\mathbf{v}.$$

1-Montrons tout d'abord que pour $k, l, r, s \in \{1, \dots, d\}$, $k < l$, $s < t$ et $\{k, l\} \cap \{s, t\} = \emptyset$, nous avons :

$$\int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(s,t)}) \} d\mathbf{u} d\mathbf{v} = 0,$$

On a :

$$\begin{aligned}
\mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(s,t)}) &= [\mathbb{B}_C(u^{(k,l)}) - u_l \mathbb{B}_C(u^{(k)}) - u_k \mathbb{B}_C(u^{(l)})] \\
&\quad \times [\mathbb{B}_C(v^{(s,t)}) - v_t \mathbb{B}_C(v^{(s)}) - v_s \mathbb{B}_C(v^{(t)})] \\
&= \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(s,t)}) - v_t \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(s)}) \\
&\quad - v_s \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(t)}) - u_l \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(s,t)}) \\
&\quad + u_l v_t \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(s)}) + u_l v_s \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(t)}) \\
&\quad - u_k \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(s,t)}) + u_k v_t \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(s)}) \\
&\quad + u_k v_s \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(t)}).
\end{aligned}$$

$$E \{ \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(s,t)}) \} = \Pi(u^{(k,l)} \wedge v^{(s,t)}) - \Pi(u^{(k,l)}) \Pi(v^{(s,t)}).$$

$$\min(u^{(k,l)}, v^{(s,t)}) = (1, \dots, 1, u_k, 1, \dots, 1, u_l, 1, \dots, 1, v_s, 1, \dots, 1, v_t, 1, \dots, 1).$$

Donc si $\{k, l\} \cap \{s, t\} = \phi$, alors :

$$\Pi(u^{(k,l)} \wedge v^{(s,t)}) = \Pi(u^{(k,l)}) \Pi(v^{(s,t)}).$$

C'est-à-dire :

$$E \{ \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(s,t)}) \} = 0.$$

Nous suivons la même procédure présentée dans la calcul précédent :

$$\begin{aligned} E \{ \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(s)}) \} &= E \{ \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(s,t)}) \} \\ &= E \{ \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(s,t)}) \} \\ &= E \{ \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(t)}) \} \\ &= \Pi(u^{(k,l)} \wedge v^{(s)}) - \Pi(u^{(k,l)}) \Pi(v^{(s)}) = 0. \end{aligned}$$

Car pour tout $k \neq l \neq s \neq t$,

$$\min(u^{(k,l)} \wedge v^{(s)}) = u_k u_l u_s = \Pi(u^{(k,l)}) \Pi(v^{(s)}).$$

Ainsi :

$$\begin{aligned} E \{ \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(s)}) \} &= E \{ \mathbb{B}_C(u^{(l)}) \mathbb{B}(v^{(t)}) \} = \{ \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(t)}) \} \\ &= E \{ \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(s)}) \} = 0. \end{aligned}$$

Finalement on a montré que dans le cas : $\{k, l\} \cap \{s, t\} = \phi$

on a bien

$$\int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(s,t)}) \} d\mathbf{u} d\mathbf{v} = 0.$$

Cette dernière expression est égale à zéro, même si nous considérons

le cas où $\{k, l\} \cap \{s, t\}$ ont exactement un élément en commun.

2-Considérons le cas où $s = k, t = l$

Pour tout $k \in \{1, \dots, d\}$ et $t \in \{1, \dots, d\}$, nous avons, d'après la relation (3.27)

$$\begin{aligned} \sigma_3^2 &= 144 \left(\sum_{k < l} \frac{1}{\binom{d}{2}} \right)^2 \int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(k,l)}) \} d\mathbf{u} d\mathbf{v} \\ &= 144 \left(\sum_{k < l} \frac{1}{\binom{d}{2}} \right)^2 \int_{I^2} \int_{I^2} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(k,l)}) \} d(u_k, u_l) d(v_k, v_l). \end{aligned}$$

En prenant le produit $\mathbb{P}_r = \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(s,t)})$. Les calculs nous donnent avec $s = k$, et $t = l$

$$\begin{aligned} \mathbb{P}_r &= \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(k,l)}) - v_l \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(k)}) \\ &\quad - v_k \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(l)}) - u_l \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(k,l)}) \\ &\quad + u_l v_l \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(k)}) + u_l v_k \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(l)}) \\ &\quad - u_k \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(k,l)}) + u_l v_l \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(k)}) \\ &\quad + u_k v_k \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(l)}). \end{aligned}$$

Ici les intégrales sont définies par :

$$\begin{aligned} 1^\circ) J_1 &= \int_{I^d} \int_{I^d} E \{ \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(k,l)}) \} d\mathbf{u} d\mathbf{v} \\ &= \int_{I^d} \int_{I^d} \Pi(u^{(k,l)}) \Pi(v^{(k,l)}) d\mathbf{u} d\mathbf{v} \\ &= \int_{I^2} \int_{I^2} (u_k \wedge v_k) (u_l \wedge v_l) d(u_k, u_l) d(v_k, v_l) \\ &\quad - \int_{I^2} \Pi(u^{(k,l)}) d(u_k, u_l) \int_{I^2} \Pi(v^{(k,l)}) d(v_k, v_l) \\ &= \int_{I^2} \left[\left(\int_{I^2} (u_k \wedge v_k) d(u_k, v_k) \right) (u_l \wedge v_l) \right] d(u_l, v_l) - \frac{1}{16} \\ &= \frac{1}{3} \int_I (u_l \wedge v_l) d(u_l, v_l) - \frac{1}{16} = \frac{1}{9} - \frac{1}{16} = \frac{7}{144}. \end{aligned} \quad (3.34)$$

$$\begin{aligned} 2^\circ) J_2 &= \int_{I^2} \int_{I^2} E \{ v_l \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(k)}) \} du dv \\ &= \int_{I^2} \int_{I^2} E(v_l) E \{ \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(k)}) \} du dv \\ &= \frac{1}{2} \int_0^1 \int_0^1 \int_0^1 [\Pi(u \wedge v^{(k)}) - \Pi(u^{(k,l)}) \Pi(v^{(k)})] d(u_k, u_l) dv_l \\ &= \frac{1}{2} \int_0^1 \int_0^1 \int_0^1 [\Pi(u \wedge v^{(k)}) - \Pi(u^{(k,l)}) \Pi(v^{(k)})] d(u_k, u_l) dv_l \\ &= \frac{1}{2} \int_0^1 \int_{I^2} [(u_k \wedge u_l) u_k du_k du_l] v_l dv_l - \frac{1}{8} \\ &= \frac{1}{2} \times \frac{1}{3} \int_0^1 v_l dv_l - \frac{1}{8} - \frac{1}{16} = \frac{1}{48}. \end{aligned} \quad (3.35)$$

Même procédure, que décrite ci-dessus :

$$\begin{aligned}
3^\circ) J_3 &= \int \int E \{ v_k \mathbb{B}_C(u^{(k,l)}) \mathbb{B}_C(v^{(l)}) \} d\mathbf{u} d\mathbf{v} \\
&= \int \int E \{ u_l \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(k,l)}) \} d\mathbf{u} d\mathbf{v} \\
&= \int \int E \{ u_k \mathbb{B}_C(u^l) \mathbb{B}_C(v^{(k,l)}) \} d\mathbf{u} d\mathbf{v} = \frac{1}{48}. \quad (3.36)
\end{aligned}$$

$$\begin{aligned}
4^\circ) J_4 &= \int_{I^2} \int_{I^2} E \{ \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(k)}) \} d\mathbf{u} d\mathbf{v} \\
&= \int_{I^2} \int_{I^2} E \{ \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(l)}) \} d\mathbf{u} d\mathbf{v} = 0. \quad (3.37)
\end{aligned}$$

$$\begin{aligned}
5^\circ) J_5 &= \int_{I^2} \int_{I^2} E \{ \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(l)}) \} d\mathbf{u} d\mathbf{v} \\
&= \int_{I^2} \int_{I^2} E \{ \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(k)}) \} d\mathbf{u} d\mathbf{v} = 0 \quad (3.38)
\end{aligned}$$

Puisque :

$$\begin{aligned}
E \{ \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(k)}) \} &= \Pi(u^{(l)} \wedge v^{(k)}) - \Pi(u^{(l)}) \Pi(v^{(k)}) \\
&= u_l u_k - u_l u_k = 0, \quad k \neq l
\end{aligned}$$

Finalement :

$$\begin{aligned}
6^\circ) J_6 &= \int_{I^2} \int_{I^2} E \{ u_l v_l \mathbb{B}_C(u^{(k)}) \mathbb{B}_C(v^{(k)}) \} d\mathbf{u} d\mathbf{v} \\
&= \int_{I^2} \int_{I^2} E \{ u_k v_k \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(l)}) \} d\mathbf{u} d\mathbf{v} \\
&= \int_{I^2} \int_{I^2} E(u_k v_k) E \{ \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(l)}) \} d\mathbf{u} d\mathbf{v} \\
&= \frac{1}{4} \int_{I^2} E \{ \mathbb{B}_C(u^{(l)}) \mathbb{B}_C(v^{(l)}) \} d\mathbf{u} d\mathbf{v} \\
&= \frac{1}{4} \int_{I^2} [C(u^{(l)} \wedge v^{(l)}) - C(u^{(l)}) C(v^{(l)})] du_l dv_l \\
&= \frac{1}{4} \left[\int_{I^2} (u_l \wedge v_l) du_l dv_l - \int_{I^2} u_l v_l du_l dv_l \right] \\
&= \frac{1}{4} \left[\frac{1}{3} - \frac{1}{4} \right] = \frac{1}{48}. \quad (3.39)
\end{aligned}$$

En additionnant les équations, (3.34), (3.35), (3.36), (3.37), (3.38), et (3.39) On obtient :

$$\int_{I^2} \int_{I^2} E \{ G_C(u^{(k,l)}) G_C(v^{(k,l)}) \} dudv = \frac{7}{144} - \frac{4 \times 1}{48} + 2 \times \frac{1}{48} = \frac{1}{144}.$$

Par conséquent, nous avons :

$$\int_{I^d} \int_{I^d} E \{ \mathbb{G}_C(u^{(k,l)}) \mathbb{G}_C(v^{(s,t)}) \} d\mathbf{u}d\mathbf{v} = \frac{1}{144}$$

Il s'ensuit que

$$\sigma_3^2 = 144 \times \left(\sum_{k < l} \right)^2 \binom{d}{2}^{-2} \times \frac{1}{144} = \left(\sum_{k < l} \binom{d}{2}^{-1} \right)^2.$$

Est finalement la variance asymptotique σ_3^2 est donnée par la formule :

$$\sigma_3^2 = \binom{d}{2}^{-2}, \quad \sigma_3 = \binom{d}{2}^{-1}.$$

■

Dans cette section, nous donnons des exemples sur les copules paramétriques dans le cas où la variance σ^2 dépend d'un ou de plusieurs paramètres de dépendance.

3.7 Stabilisation de la variance par transformation :

Un cas usuel est celui pour lequel on dispose de résultats de normalité asymptotique pour la suite de statistique T_n :

$$\sqrt{n} \{T_n - \theta\} \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \mathcal{N}(0, \sigma^2(\theta)).$$

Les intervalles de confiance, au niveau (asymptotique) $1 - 2\alpha$ sont alors donnée par

$$\left(T_n - Z_\alpha \frac{\sigma(\theta)}{\sqrt{n}}, T_n + Z_\alpha \frac{\sigma(\theta)}{\sqrt{n}} \right).$$

Z_α est le quantile de loi $\mathcal{N}(0, 1)$. En pratique, ce résultat n'est pas exploitable, si les écarts-type $\sigma(\theta)$ dépend de θ , qui l'inconnue du problème. Une première

solution est de remplacer cette quantité par un estimateur convergent, auquel cas, on conserve le niveau asymptotique de $1 - 2\alpha$, la seconde est de transformer notre problème en un problème où la variance de la loi limite ne dépend plus de θ .

Ici, on se concentre sur une autre méthode (Delta méthode) qui consiste à utiliser une transformation de notre suite statistique qui soit de variance constante. Cette démarche conduit souvent à de meilleures approximations. Le problème est de rechercher la fonction ϕ . Si on considère notre problème transformé par ϕ , le paramètre à estimer est $\phi(\theta)$ et une séquence d'estimateurs naturels est $\phi(T_n)$. Si impose à la transformation ϕ recherchée d'être différentiable, la Delta méthode assure que :

$$\sqrt{n} \{ \phi(T_n) - \phi(\theta) \} \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \mathcal{N} \left(0, \dot{\phi}(\theta)^2 \sigma^2(\theta) \right).$$

En choisissant ϕ telle que $\dot{\phi}(\theta) \sigma(\theta) = 1$, C'est à dire :

$$\phi(\theta) = \int \frac{1}{\sigma(\theta)} d\theta$$

On obtient ensuite un intervalle de confiance de niveau asymptotique $1 - 2\alpha$ pour $\phi(\theta)$. qui ne dépend pas de θ et que l'on peut ainsi déterminer. ϕ étant bijective, on retrouve un intervalle de confiance pour θ par ϕ^{-1} .

En résumé : la transformation delta méthode transforme la convergence en loi du processus $\sqrt{n} \{T_n - \theta\}$ en une convergence en loi de $\sqrt{n} \{ \phi(T_n) - \phi(\theta) \}$. Dans le deuxième exemple, nous nous intéressons à une copule particulière, soit la copule FGM

3.7.1 La copule FGM

La copule est définie par :

$$C_\lambda(u, v) = uv + \lambda uv(1-u)(1-v), \quad \text{pour tout } |\lambda| \leq 1$$

Qui est largement utilisé en statistiques une liste d'applications et de références est donnée dans Nelsen (2006, P77). Rappelons que pour $d = 2$ toutes les variances asymptotiques figurent dans le théorème (3.4.1) sont identiques ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2$) par conséquent un calcul nous donne :

$$\sigma^2 = 1 - \frac{11}{45}\lambda^2, \quad |\lambda| \leq 1 \text{ et } \rho = \frac{\lambda}{3}$$

Preuve. Calcul de ρ_s

On a :

$$\begin{aligned} \rho_s &= 12 \int_0^1 \int_0^1 (C(u, v) - uv) dudv = 12\lambda \int_0^1 \int_0^1 uv(1-u)(1-v) dudv \\ &= 12 \int_0^1 \left[\int_0^1 u(1-u)(1-v) du \right] dv = 12 \int_0^1 \left[\int_0^1 u(1-u)(1-v) dv \right] du \\ &= 2\lambda \int_0^1 v(1-v) dv = 2\lambda \left[\frac{v^2}{2} - \frac{v^3}{3} \right] = \frac{\lambda}{3} \end{aligned}$$

Calcul de la variance σ^2 cas ($d = 2$)

$$\sigma^2 = 144 \int_{I^2} \int_{I^2} E \{ \mathbb{G}_C(u, v) \mathbb{G}_C(u', v') \} dudvd'ud'v'$$

De la même manière, que les, calculs précédent on trouve aussi :

$$E \{ \mathbb{G}_C(u, v) \mathbb{G}_C(u', v') \} = C(u \wedge u', v \wedge v') - C(u, v) C(u', v')$$

$$\begin{aligned} & - \frac{\partial}{\partial u'} C(u', v') [C(u \wedge u', v) - C(u, v) u'] - \frac{\partial}{\partial u} C(u, v) [C(u \wedge u', v') - C(u', v') u] \\ & - \frac{\partial}{\partial v'} C(u', v') [C(u, v \wedge v') - C(u, v) v'] - \frac{\partial}{\partial v} C(u, v) [C(u', v \wedge v') - C(u', v') v] \\ & + \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial u'} C(u', v') [u \wedge u' - uu'] + \frac{\partial}{\partial v} C(u, v) \frac{\partial}{\partial v'} C(u', v') [v \wedge v' - vv'] \\ & + \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v'} C(u', v') [C(u, v') - uv'] + \frac{\partial}{\partial v} C(u, v) \frac{\partial}{\partial u'} C(u', v') [C(u', v) - u'v] \end{aligned}$$

Après calcul de l'intégrale $\int_{I^2} \int_{I^2} E \{ \mathbb{G}_C(u, v) \mathbb{G}_C(u', v') \} dudvd'ud'v'$ on trouve la variance σ^2 qui s'exprime en fonction de ρ , sous la forme :

$$\sigma^2(\rho) = 1 - \frac{11}{5}\rho^2.$$

■

Alors σ^2 dépend d'un paramètre ρ . Autrement dit si $\hat{\rho}_i$ est un estimateur de ρ alors on a $\sqrt{n} \{ \hat{\rho}_{i,n} - \rho \}$ converge en loi vers $\mathcal{N}(0, \sigma^2)$. Par application de la méthode delta on aura :

$$\sqrt{n} \{ \phi(\hat{\rho}_{i,n}) - \phi(\rho) \} \rightsquigarrow \mathcal{N}(0, 1).$$

Ainsi, à partir de ce résultat on peut effectuer un intervalle de confiance sur ρ au niveau asymptotique $1 - 2\alpha$ par ϕ^{-1} car :

$$\phi(\rho) \in \left[\phi(\hat{\rho}_{i,n}) - z_\alpha \frac{1}{\sqrt{n}}, \phi(\hat{\rho}_{i,n}) + z_\alpha \frac{1}{\sqrt{n}} \right] \quad i = 1, 2, 3$$

et

$$\phi(\rho) = \int \frac{d\rho}{\sigma(\rho)} = \int \frac{d\rho}{\sqrt{1 - \frac{11}{5}\rho^2}} = \sqrt{\frac{5}{11}} \int \frac{d\rho}{\sqrt{1 - \rho^2}}$$

Or

$$\phi(\rho) = \sqrt{\frac{5}{11}} \arcsin \sqrt{\frac{11}{5}} \rho \quad \text{pour } |\rho| \leq \frac{5}{11}$$

On en déduit un intervalle de confiance de niveau asymptotique $1 - 2\alpha$ pour ρ (z_α désigne le quantile d'ordre $1 - 2\alpha$ de loi $\mathcal{N}(1, 0)$) :

$$\rho \in \left[\phi^{-1} \left(\phi(\hat{\rho}_{i,n}) - z_\alpha \frac{1}{\sqrt{n}} \right), \phi^{-1} \left(\phi(\hat{\rho}_{i,n}) + z_\alpha \frac{1}{\sqrt{n}} \right) \right].$$

3.7.2 La copule de Kotz-J. (Copule de deux paramètres)

La copule C est dite copule de Kotz-Johnson, si sa forme est donnée par

$$C(u, v) = uv + \lambda uv(1 - u)(1 - v)(1 + \theta uv),$$

pour tout $-1 \leq \theta \leq 1$. C'est la forme générale de la copule *Farlie-Gumbel-Morgenstern*. La mesure de Spearman pour cette copule est :

$$\rho = \frac{\lambda}{3} + \frac{\lambda\theta}{12}.$$

Et la variance en fonction de λ et θ est donnée par :

$$\sigma^2 = 1 + \frac{1}{25}\lambda\theta - \frac{11}{45}\lambda^2 - \frac{11}{90}\lambda^2\theta - \frac{53}{5040}\lambda^2\theta^2 + \frac{1}{450}\lambda^3\theta + \frac{1}{1800}$$

Deuxième partie

Distributions Continues bivariées

Chapitre 4

Définition et représentation de la loi normale asymétrique

Dans ce chapitre, nous allons à présent successivement proposer des rappels sur les notions de deux distributions et densités des variables et des couples de variables aléatoires continues. Nous présentons la distribution normale et la distribution normale asymétrique (Skew-Normal Distribution) introduite par O'Hagan et Leonard en (1976) par la suite par Azzalini (1985) dans les deux cas univariée et bivariée à travers leurs fonctions de densité de probabilités ou leurs fonctions de distributions, ainsi la distribution normale de dimension trois.

Le but de ce chapitre est d'introduire la notion de la loi normale asymétrique et plus particulièrement la fonction de densité et sa représentation en fonction des variables aléatoires normales.

4.1 La loi normale

Commençons par la définition de la loi normale unidimensionnelle

Définition 4.1.1 *Une variable aléatoire absolument continue X suit une loi normale centrée réduite si sa fonction de densité de probabilité ϕ est telle :*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

On note la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$. C'est à dire, si Z est une variable aléatoire $\mathcal{N}(0, 1)$,

$$\begin{aligned} \Phi(z) &= P(Z \leq z) \\ &= \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx. \end{aligned}$$

4.1.1 Distribution normale bidimensionnelle

Définition 4.1.2 *La loi normale bidimensionnelle standard (loi de 2 variables) à une fonction de distribution jointe de deux variables aléatoires X et Y si sa densité de copule (X, Y) est définie par :*

$$f(x, y) = \frac{1}{2\pi |\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} Z^t \Sigma^{-1} Z \right] = \frac{1}{2\pi \sqrt{1-\rho^2}} \exp \left[-\frac{1}{2} \frac{x^2 - 2\rho xy - y^2}{(1-\rho^2)} \right]$$

de matrice de covariance Σ telle que :

$$\begin{aligned} \Sigma &= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad |\Sigma| = \det \Sigma = 1 - \rho^2. \\ \Sigma^{-1} &= \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \end{aligned}$$

ρ le coefficient de corrélation simple entre X et Y

Définition 4.1.3 *On dit que les deux variables aléatoires X et Y suivent une loi normale bivariée standardisée avec corrélation ρ si est seulement si : $Y = \rho X + \sqrt{1-\rho^2}Z$, où Z et X sont deux variables aléatoires Gaussiennes standardisées indépendantes*

4.1.2 Distribution normale standard de dimension trois

Définition 4.1.4 *On dit qu'un vecteur (X_1, X_2, X_3) suit la loi normale standard si sa densité peut s'écrire sous forme de :*

$$f(x_1, x_2, x_3) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp \left\{ -\frac{1}{2} Q(x_1, x_2, x_3) \right\} \quad (4.1)$$

où Σ est la matrice de covariance donnée par :

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix}$$

La matrice Σ est symétrique avec :

$$|\Sigma| = \det \Sigma = 1 + 2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2$$

La matrice inverse est donnée par :

$$\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{pmatrix} 1 - \rho_{23}^2 & \rho_{12} - \rho_{13}\rho_{23} & \rho_{12}\rho_{23} - \rho_{13} \\ \rho_{12} - \rho_{13}\rho_{23} & 1 - \rho_{13}^2 & \rho_{23} - \rho_{12}\rho_{13} \\ \rho_{12}\rho_{23} - \rho_{13} & \rho_{23} - \rho_{12}\rho_{13} & 1 - \rho_{12}^2 \end{pmatrix}$$

La forme quadratique de $Q(x_1, x_2, x_3)$ donnée en équation (4.1) est :

$$\begin{aligned} Q(x_1, x_2, x_3) &= (x_1, x_2, x_3)' \Sigma^{-1} (x_1, x_2, x_3) \\ &= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{23}x_2x_3 + 2a_{13}x_1x_3 - 1 \end{aligned}$$

où a_{ij} sont les éléments de la matrice covariance inverse Σ^{-1}

4.1.3 La probabilité quadrant (Orthant probability)

1. La distribution normale bivariée standard prend $\sigma_1 = \sigma_2 = 1$ et $\mu_1 = \mu_2 = 0$. La probabilité quadrant dans ce cas particulier est alors donnée analytiquement par :

$$\begin{aligned} \mathbb{P}\{Z_1 > 0, Z_2 > 0\} &= P\{Z_1 < 0, Z_2 < 0\} \\ &= \int_0^\infty \int_0^\infty \frac{1}{2\pi |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}Z^T \Sigma^{-1} Z\right\} \\ &= \frac{1}{4} + \frac{\sin^{-1}(\rho_{12})}{2\pi}, \end{aligned}$$

avec $Z = (Z_1, Z_2)$ est un vecteur Gaussien

2. La distribution normale trivariée prend des variances unitaires et $\mu_1 = \mu_2 = \mu_3 = 0$. La probabilité quadrant dans ce cas particulier est alors

donnée analytiquement par

$$\begin{aligned} \mathbb{P}\{Z_1 > 0, Z_2 > 0, Z_3 > 0\} &= \mathbb{P}\{Z_1 < 0, Z_2 < 0, Z_3 < 0\} \\ &= \int_0^\infty \int_0^\infty \int_0^\infty \frac{1}{2\pi^{\frac{3}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} Z^T \Sigma^{-1} Z\right\} \\ &= \frac{1}{8} + \frac{1}{4\pi} (\sin^{-1} \rho_{12} + \sin^{-1} \rho_{13} + \sin^{-1} \rho_{23}) \end{aligned}$$

Avec $Z = (Z_1, Z_2, Z_3)$ est un vecteur Gaussien et $\rho_{ij} = \text{Corr}(Z_i, Z_j)$, pour tout $i, j \in \{1, 2, 3\}$

4.2 Loi normale asymétrique (Skew-Normal Distribution)

Dans cette section nous allons définir la distribution normale asymétrique univariée cette distribution sera définie par sa fonction de densité et trois représentations en termes de la distribution normale. Nous discuterons un ensemble de propriétés de la densité donnée en équation (4.2)

En théorie de probabilité et en statistique, la distribution normale asymétrique (Skew-Normal Distribution) est une distribution de probabilité continue qui généralise la distribution normale en introduisant une asymétrie nulle.

Enfin, avant d'aller plus loin dans l'étude, nous commençons par donner un bref rappel des définitions et théorèmes utiles et nécessaires et nous intéressons plus particulièrement au cas de la deuxième section qui vienne, de compléter l'étude de la distribution de Cauchy bivariée, c'est à dire nous introduisons une distribution dite loi normale asymétrique.

Définition 4.2.1 Soit $\lambda \in \mathbb{R}$ on dit que la variable aléatoire Z suit une loi normale asymétrique de paramètre λ (Coefficient d'asymétrie ou Skewness) si sa densité de probabilité est donnée par

$$f_\lambda(z) = 2\phi(z) \Phi(\lambda z). \quad (4.2)$$

Avec :

$$\mathbb{P}(Z \leq z) = \int_{-\infty}^z 2\phi(t) \Phi(\lambda t) dt = \int_{-\infty}^{\Phi(x)} 2\Phi(\lambda \Phi^{-1}(t)) dt$$

ϕ est la densité de probabilité de la loi normale réduite et Φ sa fonction de répartition.

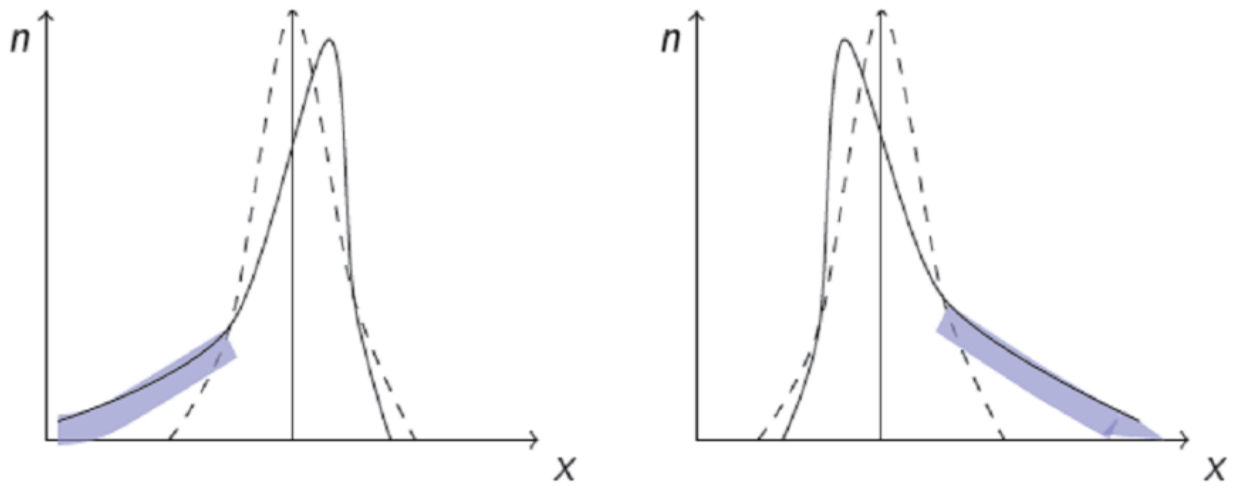


FIG. 4.1 – *Skewness a gauche*, ($\lambda < 0$) *Skewness a droite*, ($\lambda > 0$)

Si Z suit la loi (Skew-Normal Distribution) nous écrivons $Z \sim SN(\lambda)$. On peut vérifier qu'on retrouve une distribution normale quand $\lambda = 0$, et que la valeur absolue de l'asymétrie vers la droite si $\lambda > 0$ et asymétrie vers a gauche si $\lambda < 0$.

4.2.1 Représentation de la loi normale asymétrique univariée

Théorème 4.2.1 Soit U et V deux variables aléatoires indépendantes et soit :

$$Z = \frac{\lambda}{\sqrt{1+\lambda^2}}|U| + \frac{1}{\sqrt{1+\lambda^2}}V$$

Alors $Z \sim SN(\lambda)$

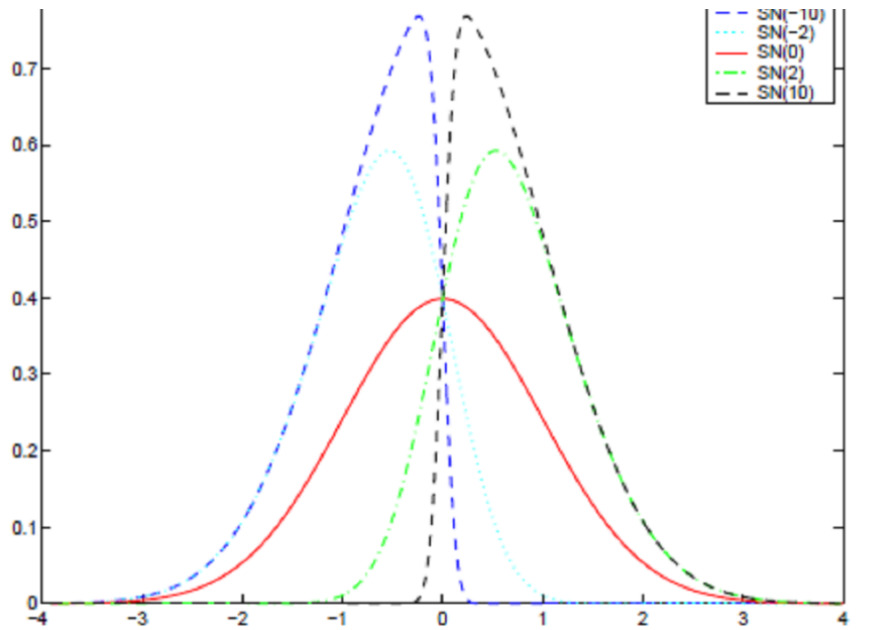


FIG. 4.2 – La densité de $SN(-10)$, $SN(-2)$, $SN(0)$, $SN(2)$ et $SN(10)$

Lemme 4.2.1 Soit X et Y deux variables aleatoires independantes suivant la loi normale standard et soit $\lambda \in \mathbb{R}$. Alors La loi conditionnelle de Y étant donnée $X < \lambda Y$ suit la loi $SN(\lambda)$

Preuve. Si X et Y sont deux variables aleatoires independantes de loi normale standard et $\lambda \in \mathbb{R}$, Alors :

$$\begin{aligned}
 \mathbb{P}(Y \leq t \mid X < \lambda Y) &= \frac{\mathbb{P}(Y \leq t, X < \lambda Y)}{\mathbb{P}(X - \lambda Y < 0)} \\
 &= \frac{\int_{-\infty}^t \Phi(\lambda y) \phi(y) dy}{\int_{-\infty}^0 \frac{1}{\sqrt{2\pi}\sqrt{1+\lambda^2}} e^{-\frac{u^2}{2(1+\lambda^2)}} du} \\
 &= \frac{\int_{-\infty}^t \Phi(\lambda y) \phi(y) dy}{\Phi(0)} \\
 &= 2 \int_{-\infty}^t \Phi(\lambda y) \phi(y) dy
 \end{aligned}$$

La différenciation par rapport à t , nous donne la fonction densité de la loi $SN(\lambda)$

$$\frac{d}{dt} \mathbb{P}(Y \leq t \setminus X < \lambda Y) = 2\Phi(\lambda t) \phi(t).$$

■

Théorème 4.2.2 Soient Y et W deux variables aléatoires indépendantes et soit

$\lambda \in \mathbb{R}$, définissons X tel que $X = \frac{\lambda Y - W}{\sqrt{1 + \lambda^2}}$. Alors :

i) (X, Y) admet une distribution normale standard bivariée avec

$$\text{coefficient de corrélation } \rho = \text{Corr}(X, Y) = \mathbb{E}(XY) = \frac{\lambda}{\left(\sqrt{1 + \lambda^2}\right)^{\frac{1}{2}}}.$$

ii) La distribution conditionnelle de Y sachant $X > 0$ est $SN(\lambda)$.

Preuve. Soit $(X, Y) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ et soit $\lambda \in \mathbb{R}$, nous définissons la variable aléatoire $X = \frac{\lambda Y - W}{\sqrt{1 + \lambda^2}}$. Alors X suit la loi normale standard est :

$$\begin{aligned} \text{Corr}(X, Y) &= \mathbb{E}(XY) \\ &= \mathbb{E}\left(\frac{(\lambda Y - W)Y}{\sqrt{1 + \lambda^2}}\right) \\ &= \frac{\lambda}{\sqrt{1 + \lambda^2}}. \end{aligned}$$

Appliquons le théorème (4.2.2) nous avons :

$$\begin{aligned} \mathbb{P}\{Y \leq y \setminus X > 0\} &= \mathbb{P}\left\{Y \leq y \setminus \frac{(\lambda Y - W)}{\sqrt{1 + \lambda^2}} > 0\right\} \\ &= \mathbb{P}\{Y \leq y \setminus W < \lambda Y\} \end{aligned}$$

ceci permet de dire que $Y \sim SN(\lambda)$ ■

Théorème 4.2.3 Si Y_0 et Y_1 deux variables aléatoires indépendantes de loi normale centré réduite. soit $\delta \in [-1, 1]$, alors :

$$Z = \delta |Y_0| + (1 - \delta^2)^{\frac{1}{2}} Y_1 \sim SN(\lambda(\delta)).$$

Les paramètres λ et δ , sont étroitement liés l'un à l'autre par les deux relations suivantes :

$$\lambda(\delta) = \frac{\delta}{\sqrt{1 - \delta^2}}, \text{ et } \delta(\lambda) = \frac{\lambda}{\sqrt{1 + \lambda^2}}, \text{ avec } \delta \in [-1, 1].$$

4.2.2 Représentation de la distribution normale asymétrique bivariée

Définition 4.2.2 *Soient*

$$\begin{cases} Z_{\lambda_1} = \lambda_1 |Y_0| + (1 - \lambda_1^2)^{\frac{1}{2}} Y_1 \\ Z_{\lambda_2} = \lambda_2 |Y_0| + (1 - \lambda_2^2)^{\frac{1}{2}} Y_2 \end{cases} ,$$

où le vecteur $(Y_1, Y_2) \sim N(0, 0, 1, 1, \rho^*)$, et $Y_0 \sim \mathcal{N}(0, 1)$, telque $Y_0 \perp Y_1, Y_0 \perp Y_2$. Alors le couple $(Z_{\lambda_1}, Z_{\lambda_2}) \sim BSN(\lambda_1, \lambda_2, \rho^*)$ et de fonction de densité normale asymétrique bivariée définie par : $h(x, y) = 2\phi(x, y; \rho^*) \Phi(\lambda_1 x + \lambda_2 y)$, où $\rho^* = \text{Corr}(Y_1, Y_2)$.

Chapitre 5

Généralisation de la distribution de Cauchy bivariée

Ce chapitre est consacré à l'étude de la distribution de Cauchy bivariée qui à été discuté dans la littérature par [Fang, Kotz, et Ng (1990)] Cette distribution déduite de la distribution Gaussienne trivariée avec une matrice de corrélation Σ c'est à dire une matrice symétrique définie positive dont les éléments diagonaux sont tous égaux à 1.

Alors nous obtenons des expressions explicites pour la fonction de distribution jointe et la fonction de densité. Ainsi nous montrons que ces expressions réduites dans un cas particulier sont des expressions correspondantes à la loi de Cauchy bivariée. Finalement, nous pouvons utiliser cette distribution pour déterminer ce qu'on appelle. La loi normale bivariée asymétrique (Orthant probability of Bivariate Skew-Normal Distribution).

5.1 Loi de Cauchy unidimensionnelle

Une variable aléatoire X admet une loi de Cauchy, si elle possède une densité de probabilité

$$f(t) = \frac{1}{\pi} \frac{\beta}{[\beta^2 + (t - \alpha)^2]}.$$

Si X suit la loi de Cauchy nous écrivons $X \sim C(\alpha, \beta)$. C'est une loi à deux paramètres : α et β ($\beta > 0$), où α est un paramètre de **position** et β un paramètre **d'échelle** (voir figure 5.1). Cette distribution est symétrique par

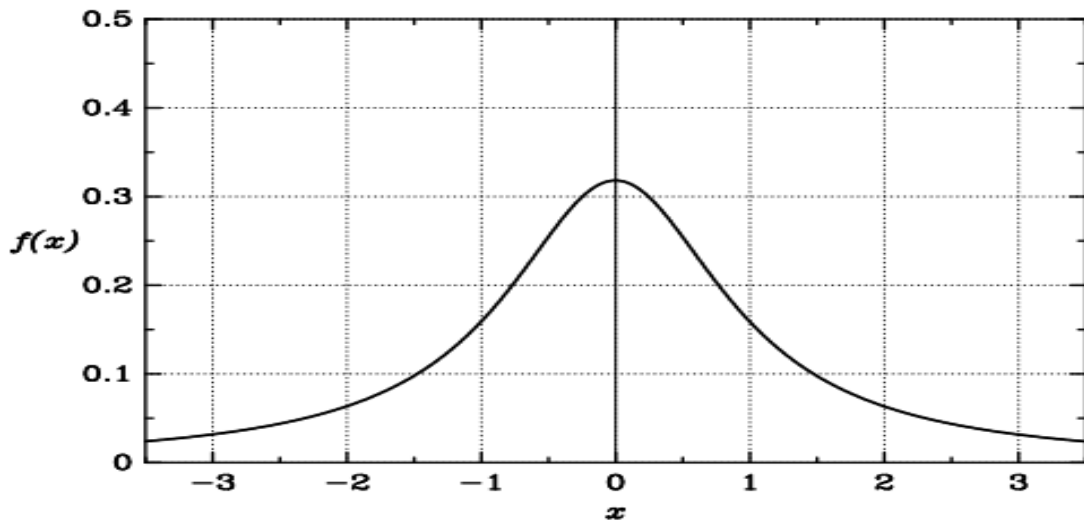


FIG. 5.1 – Densité de probabilité de la loi de Cauchy standard, à deux paramètres $\alpha = 0$ et $\beta = 1$

rapport à α , le paramètre β donnant une information sur l'étalement de la fonction f

5.1.1 Fonction de répartition

La fonction de distribution est donnée par :

$$F(t) = \frac{1}{\pi} \arctan\left(\frac{t - \alpha}{\beta}\right) + \frac{1}{2},$$

avec $t \in \mathbb{R}$.

5.1.2 Espérance et écart type

La loi de Cauchy n'admet ni espérance ni écart type. Et il en va de même pour tout moment d'ordre supérieur. En effet, $\frac{1}{\pi} \frac{\beta}{\beta^2 + (t - \alpha)^2}$ n'est pas intégrable au sens de Lebesgue car $\left| \frac{\beta}{\beta^2 + (t - \alpha)^2} \right| \sim \left| \frac{1}{t} \right|$ (à l'infini) d'où la divergence de l'intégrale : l'espérance n'existe pas. Cependant α , qui en est la **médiane** est souvent considéré comme la « moyenne » de la loi de Cauchy, car :

$$\lim_{R \rightarrow +\infty} \int_{-R}^{+R} \frac{\beta}{\pi} \frac{t}{[\beta^2 + (t - \alpha)^2]} dt = \alpha.$$

- L'inverse d'une variable aléatoire de la loi de Cauchy suit une loi de Cauchy.
- Le quotient de deux variables aléatoires réelles indépendantes suivant des lois **normales** standards suit une loi de Cauchy.

Définition 5.1.1 Soient U_1 et U_2 deux variables aléatoires indépendantes tel que $U_1 \sim \mathcal{N}(0, 1)$ et $U_2 \sim \mathcal{N}(0, 1)$. Alors la variable $Y = \frac{U_2}{|U_1|}$ suit la loi de Cauchy standard est on écrit $Y \sim C(0, 1)$ si sa densité de probabilité est définie par :

$$f(t) = \frac{1}{\pi(t^2 + 1)}, \quad t \in \mathbb{R}$$

Et la fonction de distribution est donnée par

$$F(t) = \mathbb{P}(Y \leq t) = \mathbb{P}\left(\frac{U_2}{|U_1|} \leq t\right) = \frac{1}{2} + \frac{1}{\pi} \operatorname{tg}^{-1}(t).$$

5.2 Loi de Cauchy bivariée

Soient U_1 et U_2 deux variables aléatoires indépendantes tel que $U_1 \sim \mathcal{N}(0, 1)$ et $U_2 \sim \mathcal{N}(0, 1)$. On sait que la variable aléatoire $\frac{U_2}{|U_1|}$ suit la loi de Cauchy standard. On continue par la même procédure avec U_1 comme une variable aléatoire normale standard et le couple (U_2, U_3) comme vecteur normale standard bivariée de coefficient de corrélation ρ alors le vecteur :

$$\left(\frac{U_2}{|U_1|}, \frac{U_3}{|U_1|}\right) \sim C\left(0, \Sigma\right) \quad \text{où } \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \quad (5.1)$$

Et sa fonction de densité est donnée par :

$$f(t_1, t_2, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}(1+t_1^2+t_2^2)^{\frac{3}{2}}}, \quad t = (t_1, t_2) \in \mathbb{R}^2. \quad (5.2)$$

5.3 Généralisation

Jamalizedeh et balakrishnan (2008) ont proposé une étude d'application sur la distribution de Cauchy bidimensionnelle, ils ont donné une définition de la fonction de distribution jointe du vecteur $(W_1, W_2) = \left(\frac{U_2}{|U_1|}, \frac{U_3}{|U_1|}\right)$, où $(U_1, U_2, U_3) \sim \mathcal{N}(0, 1, \Sigma)$.

Définition 5.3.1 Soit le vecteur $(U_1, U_2, U_3) \sim \mathcal{N}(0, \Sigma)$ de matrice de covariance

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{23} & 1 \end{pmatrix}, \quad |\rho_{ij}| < 1.$$

Alors le vecteur $(W_1, W_2) = \left(\frac{U_2}{|U_1|}, \frac{U_3}{|U_1|} \right)$ admet une distribution dite la généralisation de la distribution de Cauchy bivariée.

Maintenant, avant d'établir l'expression de la distribution jointe d'un vecteur (W_1, W_2) on va utiliser comme point de départ le lemme suivant :

Lemme 5.3.1 Si le couple $(Y_1, Y_2) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ et $X \sim \mathcal{N}(0, 1)$ sont indépendantes, alors :

$$\begin{aligned} & \mathbb{P}\{Y_1 < \gamma_1 X, Y_2 < \gamma_2 X, X > 0\} \\ &= \frac{1}{4\pi} \left\{ \cos^{-1} \left(\frac{\rho + \gamma_1 \gamma_2}{\sqrt{1 + \gamma_1^2}} \right) + \text{tg}^{-1}(\gamma_1) + \text{tg}^{-1}(\gamma_2) \right\}. \end{aligned} \quad (5.3)$$

Preuve. Nous avons :

$$\begin{aligned} & \mathbb{P}(Y_1 < \gamma_1 X, Y_2 < \gamma_2 X, X > 0) \\ &= \mathbb{P}(Y_1 - \gamma_1 X < 0, Y_2 - \gamma_2 X < 0, X > 0) \\ &= \mathbb{P}\left(\frac{Y_1 - \gamma_1 X}{\sqrt{1 + \gamma_1^2}} < 0, \frac{Y_2 - \gamma_2 X}{\sqrt{1 + \gamma_2^2}} < 0, -X < 0 \right) \\ &= \mathbb{P}(V_1 < 0, V_2 < 0, V_3 < 0), \end{aligned}$$

où

$$(V_1, V_2, V_3) = \left(\frac{Y_1 - \gamma_1 X}{\sqrt{1 + \gamma_1^2}}, \frac{Y_2 - \gamma_2 X}{\sqrt{1 + \gamma_2^2}}, -X \right) \sim \mathcal{N}\left(0, 1, \sum_v\right).$$

Avec :

$$\sum_v = \begin{pmatrix} 1 & \frac{\rho + \gamma_1 \gamma_2}{\sqrt{1 + \gamma_1^2} \sqrt{1 + \gamma_2^2}} & \frac{\gamma_1}{\sqrt{1 + \gamma_1^2}} \\ \frac{\rho + \gamma_1 \gamma_2}{\sqrt{1 + \gamma_1^2} \sqrt{1 + \gamma_2^2}} & 1 & \frac{\gamma_2}{\sqrt{1 + \gamma_2^2}} \\ \frac{\gamma_1}{\sqrt{1 + \gamma_1^2}} & \frac{\gamma_2}{\sqrt{1 + \gamma_2^2}} & 1 \end{pmatrix}.$$

$$\begin{aligned} \text{Corr}(V_1, V_2) &= \rho_{12} = \frac{\rho + \gamma_1 \gamma_2}{\sqrt{1 + \gamma_1^2} \sqrt{1 + \gamma_2^2}}, \\ \text{Corr}(V_1, V_3) &= \rho_{13} = \frac{\gamma_1}{\sqrt{1 + \gamma_1^2}}, \\ \text{Corr}(V_2, V_3) &= \rho_{23} = \frac{\gamma_2}{\sqrt{1 + \gamma_2^2}}. \end{aligned}$$

Utilisant la formule de la probabilité quadrant (Orthant probability) étudiée en section (4.1) On à :

$$\begin{aligned} &\mathbb{P}(V_1 < 0, V_2 < 0, V_3 < 0) \\ &= \frac{1}{8} + \frac{1}{4\pi} \sin^{-1} \left(\frac{\rho + \gamma_1 \gamma_2}{\sqrt{1 + \gamma_1^2} \sqrt{1 + \gamma_2^2}} \right) + \sin^{-1} \left(\frac{\gamma_1}{\sqrt{1 + \gamma_1^2}} \right) + \sin^{-1} \left(\frac{\gamma_2}{\sqrt{1 + \gamma_2^2}} \right). \end{aligned}$$

Et par application des formules de fonctions inverses on obtient :

$$P(V_1 < 0, V_2 < 0, V_3 < 0) = \frac{1}{4\pi} \left\{ \cos^{-1} \left(-\frac{\rho + \gamma_1 \gamma_2}{\sqrt{1 + \gamma_1^2} \sqrt{1 + \gamma_2^2}} \right) + tg^{-1}(\gamma_1) + tg^{-1}(\gamma_2) \right\}.$$

Car :

$$\cos^{-1}(-x) = \frac{\pi}{2} + \sin^{-1}(x), \quad g^{-1}(x) = \sin^{-1} \left(\frac{x}{\sqrt{1 + x^2}} \right), \quad x \in [-1, 1].$$

■

Tous les résultats qui suivent s'appuient sur ce lemme

Théorème 5.3.1 Soit $F(t_1, t_2; \Sigma)$ la fonction de distribution jointe du vecteur $(W_1, W_2) = \left(\frac{U_2}{|U_1|}, \frac{U_3}{|U_1|} \right)$. Alors :

$$\begin{aligned} F(t_1, t_2; \Sigma) &= \frac{1}{4\pi} \left\{ \cos^{-1} \left(-\frac{\rho_{23} - \rho_{12}t_1 - \rho_{13}t_2 + t_1t_2}{\sqrt{1 + t_1^2 - 2\rho_{12}t_1} \sqrt{1 + t_2^2 - 2\rho_{13}t_2}} \right) \right. \\ &\quad + tg^{-1} \left(\frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}} \right) + tg^{-1} \left(\frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}} \right) \\ &\quad + \cos^{-1} \left(-\frac{\rho_{23} + \rho_{12}t_1 + \rho_{13}t_2 + t_1t_2}{\sqrt{1 + t_1^2 + 2\rho_{12}t_1} \sqrt{1 + t_2^2 + 2\rho_{13}t_2}} \right) \\ &\quad \left. + tg^{-1} \left(\frac{t_1 + \rho_{12}}{\sqrt{1 - \rho_{12}^2}} \right) + tg^{-1} \left(\frac{t_2 + \rho_{13}}{\sqrt{1 - \rho_{13}^2}} \right) \right\}. \end{aligned} \quad (5.4)$$

Où U_1, U_2, U_3 sont définis comme suit:

$$(U_1, U_2, U_3) \stackrel{d}{=} \left(X, \rho_{12}X + \sqrt{1 - \rho_{12}^2}Y_1, \rho_{13}X + \sqrt{1 - \rho_{13}^2}Y_2 \right),$$

tel que :

$$(Y_1, Y_2) \sim \mathcal{N}_2 \left(0, 0, 1, 1, \frac{\rho_{23} - \rho_{12}\rho_{13}}{\sqrt{1 - \rho_{12}^2}\sqrt{1 - \rho_{13}^2}} \right),$$

et

$$X \sim \mathcal{N}(0, 1), \quad X \perp Y_1, \quad X \perp Y_2.$$

Preuve. Nous avons :

$$\begin{aligned} F(t_1, t_2; \Sigma) &= \mathbb{P} \left\{ \frac{U_2}{|U_1|} \leq t_1, \frac{U_3}{|U_1|} \leq t_2 \right\} \\ &= \mathbb{P} \{ U_2 \leq t_1 U_1, U_3 \leq t_2 U_1, U_1 > 0 \} \\ &\quad + \mathbb{P} \{ U_2 \leq -t_1 U_1, U_3 \leq -t_2 U_1, U_1 < 0 \}. \end{aligned} \quad (5.5)$$

Calculons le premier terme de l'expression (5.5) nous obtenons :

$$\begin{aligned} &\mathbb{P} \{ U_2 \leq U_1 t_1, U_3 \leq U_1 t_2, U_1 > 0 \} \\ &= \mathbb{P} \left\{ \rho_{12}X + \sqrt{1 - \rho_{12}^2}Y_1 \leq X t_1, \rho_{13}X + \sqrt{1 - \rho_{13}^2}Y_2 \leq X t_2, X > 0 \right\} \\ &= \mathbb{P} \left\{ Y_1 \leq \frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}}X, Y_2 \leq \frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}}X, X > 0 \right\}. \end{aligned} \quad (5.6)$$

Maintenant, nous pouvons utiliser le lemme précédent avec

$$\gamma_1 = \frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \gamma_2 = \frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}}, \rho = \frac{\rho_{23} - \rho_{12}\rho_{13}}{\sqrt{1 - \rho_{12}^2}\sqrt{1 - \rho_{13}^2}}.$$

1°) **Le premier terme** de l'équation (5.5), égale à :

$$\begin{aligned}
& \mathbb{P}\{U_2 \leq t_1 U_1, U_3 \leq t_2 U_1, U_1 > 0\} \\
&= \frac{1}{4\pi} \left\{ \cos^{-1} \left(-\frac{\frac{\rho_{23} - \rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13}^2}} + \frac{(t_1 - \rho_{12})(t_2 - \rho_{13})}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{13}^2}}}{\sqrt{1 + \frac{(t_1 - \rho_{12})^2}{1-\rho_{12}^2}} \sqrt{1 + \frac{(t_2 - \rho_{13})^2}{1-\rho_{13}^2}}} \right) \right. \\
&\quad \left. + tg^{-1} \left(\frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}} \right) + tg^{-1} \left(\frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}} \right) \right\} \\
&= \frac{1}{4\pi} \left\{ \cos^{-1} \left(-\frac{\rho_{23} - \rho_{12}t_1 - \rho_{13}t_2 + t_1t_2}{\sqrt{1 + t_1^2 - 2\rho_{12}t_1} \sqrt{1 + t_2^2 - 2\rho_{13}t_2}} \right) \right. \\
&\quad \left. + tg^{-1} \left(\frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}} \right) + tg^{-1} \left(\frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}} \right) \right\}. \tag{5.7}
\end{aligned}$$

2°) Le deuxième terme du (5.5) nous donne :

$$\begin{aligned}
& \mathbb{P}\{U_2 \leq t_1(-U_1), U_3 \leq t_2(-U_1), -U_1 > 0\} \\
&= \mathbb{P}\left\{ \rho_{12}X + \sqrt{1 - \rho_{12}^2}Y_1 \leq X(-t_1), \rho_{13}X + \sqrt{1 - \rho_{13}^2}Y_2 \leq X(-t_2), -X > 0 \right\} \\
&= \mathbb{P}\left\{ Y_1 \leq \frac{t_1 + \rho_{12}}{\sqrt{1 - \rho_{12}^2}}(-X), Y_2 \leq \frac{t_2 + \rho_{13}}{\sqrt{1 - \rho_{13}^2}}(-X), -X > 0 \right\} \\
&= \frac{1}{4\pi} \left\{ \cos^{-1} \left(-\frac{\rho_{23} + \rho_{12}t_1 + \rho_{13}t_2 + t_1t_2}{\sqrt{1 + t_1^2 + 2\rho_{12}t_1} \sqrt{1 + t_2^2 + 2\rho_{13}t_2}} \right) \right. \\
&\quad \left. + tg^{-1} \left(\frac{t_1 + \rho_{12}}{\sqrt{1 - \rho_{12}^2}} \right) + tg^{-1} \left(\frac{t_2 + \rho_{13}}{\sqrt{1 - \rho_{13}^2}} \right) \right\}. \tag{5.8}
\end{aligned}$$

Avec :

$$\gamma_1 = \frac{t_1 + \rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \quad \gamma_2 = \frac{t_2 + \rho_{13}}{\sqrt{1 - \rho_{13}^2}}, \text{ prenons les équations (5.7), et (5.8) dans } \tag{5.5}$$

Le théorème est démontré. ■

Remarque 5.3.1 De la même façon nous pouvons aussi obtenir la fonction de distribution jointe du vecteur $(W_1^*, W_2^*) = \left(\frac{U_2}{U_1}, \frac{U_3}{U_1} \right)$. Soit F^* la fonction de distribution jointe du (W_1^*, W_2^*) . Alors F^* est définie par :

$$F^*(t_1, t_2; \Sigma) = \frac{1}{2\pi} \left\{ \cos^{-1} \left(-\frac{\rho_{23} - \rho_{12}t_1 - \rho_{13}t_2 + t_1t_2}{\sqrt{1+t_1^2 - 2\rho_{12}t_1}\sqrt{1+t_2^2 + 2\rho_{13}t_2}} \right) + \tan^{-1} \left(\frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}} \right) + \tan^{-1} \left(\frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}} \right) \right\}.$$

Preuve. Soit F^* la fonction de distribution du vecteur (W_1^*, W_2^*) on a par définition :

$$\begin{aligned} F^*(t_1, t_1; \Sigma) &= \mathbb{P} \left(\frac{U_2}{U_1} \leq t_1, \frac{U_3}{U_1} \leq t_2 \right) \\ &= \mathbb{P} \{U_2 \leq t_1 U_1, U_3 \leq t_2 U_1, U_1 > 0\} \\ &\quad + \mathbb{P} \{U_2 \geq t_1 U_1, U_3 \geq t_2 U_1, U_1 < 0\} \\ &= 2\mathbb{P} \{U_2 \leq t_1 U_1, U_3 \leq t_2 U_1, U_1 > 0\}. \end{aligned}$$

Puisque

$$\begin{aligned} &\mathbb{P} \{U_2 \geq t_1 U_1, U_3 \geq t_2 U_1, U_1 < 0\} \\ &= \mathbb{P} \{U_2 \geq t_1 U_1, U_3 \geq t_2 U_1, -U_1 > 0\} \\ &= \mathbb{P} \left\{ Y_1 \geq \frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}} X, Y_2 \geq \frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}} X, X < 0 \right\} \\ &= \mathbb{P} \{Y_1 \geq \gamma_1(X), Y_2 \geq \gamma_2(X), -X > 0\} \\ &= \mathbb{P} \{Y_1 - \gamma_1 X \geq 0, Y_2 - \gamma_2 X \geq 0, -X > 0\} \\ &= \mathbb{P} \left\{ \frac{Y_1 - \gamma_1 X}{\sqrt{\gamma_1^2 + 1}} \geq 0, \frac{Y_2 - \gamma_2 X}{\sqrt{\gamma_2^2 + 1}} \geq 0, -X > 0 \right\} \\ &= \mathbb{P} (\acute{V}_1 \geq 0, \acute{V}_2 \geq 0, \acute{V}_3 \geq 0) = \mathbb{P} (\acute{V}_1 < 0, \acute{V}_2 < 0, \acute{V}_3 < 0). \end{aligned}$$

$$\text{Avec } \acute{V}_1 = \frac{Y_1 - \gamma_1 X}{\sqrt{\gamma_1^2 + 1}}, \acute{V}_2 = \frac{Y_2 - \gamma_2 X}{\sqrt{\gamma_2^2 + 1}}, \acute{V}_3 = -X \text{ et } (\acute{V}_1, \acute{V}_2, \acute{V}_3) \sim \mathcal{N} \left(0, 1, \sum_{\acute{V}} \right)$$

Par applications du lemme (4.1.1) on conclut que :

$$F^*(t_1, t_2) = \frac{1}{2\pi} \left\{ \cos^{-1} \left(-\frac{\rho_{23} - \rho_{12}t_1 - \rho_{13}t_2 + t_1t_2}{\sqrt{1+t_1^2 - 2\rho_{12}t_1}\sqrt{1+t_2^2 + 2\rho_{13}t_2}} \right) + \tan^{-1} \left(\frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}} \right) + \tan^{-1} \left(\frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}} \right) \right\},$$

où

$$\gamma_1 = \frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \quad \gamma_2 = \frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}}.$$

■

Remarque 5.3.2 *Un cas spéciale lorsque $\rho_{12} = \rho_{13} = 0$ et $\rho_{23} = \rho^*$, alors le vecteur (W_1^*, W_2^*) Suit la loi de Cauchy bivariée standard définie dans (5.1) cette distribution à été discuté par exemple par [Fang, Kotz, et Ng (1990)]. Aussi on peut vérifiée que la distribution jointe donnée en équation (5.4) dans le théorème se réduire a une distribution G définie par :*

$$\begin{aligned} & G(t_1, t_2, \rho^*) \\ &= \frac{1}{2\pi} \left\{ \cos^{-1} \left(-\frac{\rho^* + t_1 t_2}{\sqrt{1 + t_1^2} \sqrt{1 + t_2^2}} \right) + \tan^{-1}(t_1) + \tan^{-1}(t_2) \right\} \quad (5.9) \end{aligned}$$

Remarque 5.3.3 *La différentiation de cette expression nous donne la fonction densité conjointe $f(t_1, t_1, \rho)$ définie en équation (5.2)*

Remarque 5.3.4 *Il est facile de montrer que :*

$$\begin{aligned} & F(t_1, t_1; \Sigma) \\ &= \frac{1}{2} \left\{ G \left(\frac{t_1 - \rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \frac{t_2 - \rho_{13}}{\sqrt{1 - \rho_{13}^2}}; \varrho \right) \right. \\ & \quad \left. + G \left(\frac{t_1 + \rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \frac{t_2 + \rho_{13}}{\sqrt{1 - \rho_{13}^2}}; \varrho \right) \right\}. \quad (5.10) \end{aligned}$$

où

$$\varrho = \rho_{23,1} = \frac{\rho_{23} - \rho_{12}\rho_{13}}{\sqrt{1 - \rho_{12}^2}\sqrt{1 - \rho_{13}^2}}$$

5.4 La probabilité quadrant de la loi normale bivariée asymétrique

Rappelons que si $(U_2, U_3) \sim \mathcal{N}(0, 0, 1, 1, \rho^*)$ on a :

$$\mathbb{P}\{U_2 < 0, U_3 < 0\} = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1}(\rho_{23}), \quad (5.11)$$

avec $\rho_{23} = \text{Corr}(U_2, U_3)$

A l'aide des formules de trigonométrie des fonctions circulaires suivantes :

$$\sin^{-1}(\rho_{23}) = \frac{\pi}{2} - \cos^{-1}(\rho_{23}) \text{ et } \cos^{-1}(-\rho_{23}) = \pi - \cos^{-1}(\rho_{23})$$

On retrouve facilement la formule de l'équation (5.11) d'une autre manière :

$$\begin{aligned} \mathbb{P}\{U_2 < 0, U_3 < 0\} &= \frac{1}{4} + \frac{1}{2\pi} \left\{ \frac{\pi}{2} - \cos^{-1}(\rho_{23}) \right\} \\ &= \frac{1}{2} - \frac{1}{2\pi} \left\{ \pi - \cos^{-1}(-\rho_{23}) \right\} \\ &= \frac{1}{2} - \frac{1}{2} + \frac{1}{2\pi} \cos^{-1}(-\rho_{23}) \\ &= \frac{1}{2\pi} \cos^{-1}(-\rho_{23}) \end{aligned} \quad (5.12)$$

De plus si $(Z_{\lambda_1}, Z_{\lambda_2}) \sim BSN(\lambda_1, \lambda_2, \rho_{12})$ voir la définition (4.2.2) on a la représentation stochastique.

$$Z_{\lambda_1} = \frac{\lambda_1}{\sqrt{1+\lambda_1^2}} |U_1| + \frac{1}{\sqrt{1+\lambda_1^2}} U_2 \text{ et } Z_{\lambda_2} = \frac{\lambda_2}{\sqrt{1+\lambda_2^2}} |U_1| + \frac{1}{\sqrt{1+\lambda_2^2}} U_3.$$

On utilisant le résultat précédent, nous obtenons la formule suivante dite formule de la probabilité quadrant de la loi normale asymétrique bivariée

$$\begin{aligned} &\mathbb{P}\{Z_{\lambda_1} < 0, Z_{\lambda_2} < 0\} \\ &= \mathbb{P}\left\{ \frac{U_2}{|U_1|} < -\lambda_1, \frac{U_3}{|U_1|} < -\lambda_2 \right\} = G(-\lambda_1, -\lambda_2; \rho^*) \\ &= \frac{1}{2\pi} \left\{ \cos^{-1} \left(-\frac{\rho^* + \lambda_1 \lambda_2}{\sqrt{1+\lambda_1^2} \sqrt{1+\lambda_2^2}} \right) - \tan^{-1}(\lambda_1) - \tan^{-1}(\lambda_2) \right\} \end{aligned} \quad (5.13)$$

Car $\tan^{-1}(-x) = -\tan^{-1}(x)$

Remarque 5.4.1 Si $\lambda_1 = \lambda_2 = 0$, alors la formule de probabilité (5.13) devient (5.12)

Bibliographie

- [1] Azzalini, A., Dalla Valle, A., (1996) . The multivariate Skew-Normal Distribution. *Biometrika* **83**, 715-726.
- [2] Deheuvels, P., (1979). La fonction de dépendance empirique et ses propriétés. *Acad. Roy. Belg. Bull. Cl. Sci.* **65** (5), 274–292.
- [3] Dolati, A., Úbeda-Flores, M., (2006). On measures of multivariate concordance. *J. Prob. Stat. Sc.* **4** (2) , 147-164.
- [4] Embrechts, P., F. Lindskog., A.J. McNeil., (2003). Modelling Dependence with Copulas and Applications to Risk Management, in : S.T. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, Chapter **8**, 329–384.
- [5] Fermanian, J.D., Radulovic´, D., Wegkamp, M., (2004). Weak convergence of empirical copula processes. *Bernoulli* **10** (5), 847–860.
- [6] Genest, C., B. R´emillard., (2004). Tests of Independence and Randomness Based on the Empirical Copula Process, *Test* **13** (2), 335–369.
- [7] Jamalizadeh, A. Balakrishnan, N., (2008) . ‘On a Generalization of Bivariate Cauchy Distribution, *Communications in Statistics-Theory and Methods*, **37** (4), 469-474
- [8] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994). *Continuous univariate Distribution. Vol. 1. 2nd Edition.* New York : John Wiley & Son
- [9] Kotz, S., Balakrishnan, N., Johnson, N. L. (2000) . *Continuous Multivariate Distributions, Volume 1, Models and Applications, 2nd Edition.* New York : John Wiley & Son.
- [10] Nelsen, R.B., (1999). *An introduction to Copulas.*, *Lecture Notes in Statistics*, 139, Springer Verlag, New-York.
- [11] Rodríguez-Lallena, J.A., Úbeda-Flores, M., (2004). A new class of bivariate copulas. *Statist. Prob. Lett.* **66** (3), 315–325.

-
- [12] Rüschendorf, L., (1976). Asymptotic normality of multivariate rank order statistics. *Ann. Statist.* **4**, 912–923.
- [13] Schmid, F., R. Schmidt., (2007). Multivariate conditional versions of Spearman's rho and related measures of tail dependence, *J. Multivar. Anal.* **98** (6), 1123–1140.
- [14] Schmid, F., R. Schmidt., (2007). Multivariate extensions of Spearman's rho and related Statistics, *Stat. Probab. Lett.* **77** (4), 407–416.
- [15] Stute, W., (1984) . The oscillation behavior of empirical process : the multivariate case. *Ann. Prob.* **12**, 361-379.
- [16] Taylor, M.D., (2007). Multivariate measures of concordance, *Ann. Inst. Stat. Math.* **59** (4), 789–806.
- [17] Tsukahara, H., (2005). Semiparametric estimation in copula models, *Canad. J. Statist.* **33** (3) 357–375.
- [18] Tsukahara, H., (2000) Copulas and some applications
- [19] Van Der Vaart, A.W., Wellner, J.A., (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- [20] Wolff, E.F., (1980). N-dimensional measures of dependence. *Stochastica* **4** (3), 175–188.