

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khide, Biskra
Faculté des Sciences Exactes
Département de Mathématiques



Mémoire présenté pour obtenir le diplôme de
Master en “**Mathématiques Appliquées**”

Option : Statistique

Par Melle **Loucif Afaf**

Titre du mémoire
Mesures d'Association Définies par les Copules

Devant le Jury :

Prof. Abdelhakim Necir	U. Biskra	Président
Prof. Djamel Meraghni	U. Biskra	Rapporteur
Dr. Fairouz Afroun	U. Biskra	Examinatrice

Soutenu Publiquement le 03/06/2025

Dédicace

Je dédie ce mémoire avec toute ma gratitude :

- À ma famille, pour leur amour inestimable, leur soutien sans faille et leurs prières constantes. C'est grâce à vous que j'ai pu avancer avec force et sérénité.
- À mes enseignants, pour leur encadrement, leur patience et les précieuses connaissances qu'ils m'ont transmises tout au long de ce parcours.
- À mes amies, pour leur présence, leurs encouragements et les beaux moments partagés, qui ont rendu cette aventure encore plus spéciale.

Merci à chacun d'entre vous. Ce travail est aussi le vôtre.

Afaf Loucif

Remerciements

Tout d'abord, je remercie Dieu Le Tout-Pluissant de m'avoir aidée dans la réalisation de ce modeste travail.

Un grand merci à mes très chers parents pour m'avoir soutenue et encouragée durant toutes mes années d'études.

J'ai l'immense honneur de remercier mon encadrant, **Pr. Meraghni Djamel**, pour ses précieux conseils et pour les efforts qu'il m'a fournis durant la période de travail.

J'exprime également ma profonde gratitude aux membres du jury, **Pr. Necir Abdelhakim** et **Dr. Afroun Fairouz**, pour le temps consacré à l'évaluation de mon mémoire.

Enfin, je voudrais aussi remercier tous ceux et toutes celles, parmi mes enseignants et le personnel administratif, qui m'ont donné des conseils et des orientations appréciables.

Afaf Loucif

Abréviations et notations

$c\text{-à-d}$	c'est à dire.
v.a	variable aléatoire.
C	Copule.
c	Densité de la copule.
$Cov(X, Y)$	Covariance de X et Y
$E[X]$	Espérance de X
F	Fonction de répartition.
f	Densité de probabilité.
F^{-1}	Fonction quantile de F .
\mathcal{F}	Tribu.
M	Copule min.
Ω	Espace probabilisable.
(Ω, \mathcal{F})	Ensemble fondamental.
$(\Omega, \mathcal{F}, \mathbb{P})$	Espace de probabilité.
$p_{i,j}$	Loi de probabilité conjointe.

P_{X_i/X_j}	loi de probabilité conditionnelle.
r	Coefficient de corrélation linéaire.
W	Copule max.
(X_1, \dots, X_p)	échantillon de taille p de X .
α, β	Strictement croissante
μ	$E(X)$
Π	Copule produit.
ρ	Rho de Spearman.
τ	Tau de Kendall.
ϕ	Fonction génératrice de la copule archimédienne.
ω	événement élémentaire.

Table des matières

Dédicace	i
Remerciements	i
Notations et symbols	ii
Table des matières	iv
Liste des figures	vi
Liste des tableaux	vii
Introduction	1
1 Vecteurs Aléatoires	4
1.1 Généralités	4
1.1.1 Définition	4
1.1.2 Types	5
1.2 Lois de probabilité	5
1.2.1 Loi conjointe	6
1.2.2 Lois Marginales	6
1.2.3 Lois conditionnelles	7
1.2.4 Indépendance	8
1.3 Paramètres statistiques	8

1.3.1	Moyenne	8
1.3.2	Matrice de covariance	8
1.4	Vecteur aléatoire gaussien	11
2	Copules	13
2.1	Généralités	13
2.2	Définition	13
2.3	Exemple de copules	14
2.3.1	Copule produit	14
2.3.2	Copules min et max	15
2.4	Propriétés d'une copule	15
2.5	Théorème de Sklar	16
2.6	Familles de copules	18
2.6.1	Copules Archimédiennes	19
2.6.2	Copules elliptiques	21
3	Mesures d'association	25
3.1	Définition	25
3.2	Mesure de concordance	25
3.2.1	Concordance et discordance	25
3.2.2	Fonction de concordance	26
3.3	Structure de dépendance	29
3.3.1	Phi de Pearson	29
3.3.2	Tau de Kendall	30
3.3.3	Rho de Spearman	31

Table des figures

2.1 Visualisation de la Copule de Clayton	20
2.2 Visualisation de la Copule de Gumbel	20
2.3 Visualisation de la Copule de Frank	21
2.4 Visualisation de la Copule de Joe	21
2.5 Nuage de points et densité d'une copule gaussienne	23
2.6 Nuage de points et densité d'une copule de Student	23

Liste des tableaux

2.1 Principales copules Archimédiennes et leurs générateurs	19
3.1 Tableau générique	29
3.2 Principales mesures d'association	33

Introduction

L'étude de dépendance entre variables aléatoires constitue une problématique centrale en statistique, en économie, en finance ainsi que dans de nombreuses disciplines appliquées telles que l'hydrologie, la biologie... Pour quantifier cette dépendance, on utilise différentes mesures d'association, telles que le coefficient de corrélation de Pearson, la corrélation de Spearman ou encore le tau de Kendall. Ces mesures permettent d'évaluer dans quelle mesure les variations d'une variable sont liées à celles d'une autre. Toutefois, ces outils classiques présentent certaines limites, notamment lorsqu'il s'agit de modéliser les dépendances non linéaires ou asymétriques. C'est dans ce contexte que la théorie des copules s'avère particulièrement puissante et efficace. Introduites par Sklar en 1959, les copules permettent de dissocier la structure de dépendance entre les variables de leurs distributions marginales. Elles offrent ainsi un cadre théorique adéquat pour définir et analyser les mesures d'association, indépendamment des distributions marginales. En reliant les fonctions de répartition conjointes à leurs marginales à travers une copule, cette approche permet une modélisation plus flexible et précise des relations de dépendance. Ce travail est subdivisé en trois parties :

- **Chapitre 1 : Vecteurs aléatoires.** Ce chapitre, constitue un rappel des définitions et propriétés fondamentales des vecteurs aléatoires, à savoir le type de vecteurs, leur loi de probabilité et les paramètres associés.
- **Chapitre 2 : Copules.** Dans ce chapitre, on aborde la définition générale et spécifique des copules, accompagnée de plusieurs exemples illustratifs. On y présente les principales propriétés des copules, en mettant l'accent sur le théorème de Sklar, considéré comme la base de la théorie des copules. Par la suite, on mentionne quelques-unes des familles de copules les plus connues.
- **Chapitre 3 : Mesures d'association.** Ce dernier chapitre est consacré aux principales mesures

d'association, tel le tau de Kendall et le rho de Spearman. On commence par définir les notions de mesures de concordance et de discordance, puis on s'intéresse aux formes des différentes mesures en termes des copules.

Chapitre 1

Vecteurs Aléatoires

Dans ce chapitre, on présente les définitions et propriétés fondamentales des vecteurs aléatoires nécessaires à la suite du travail.

1.1 Généralités

Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité. Un vecteur aléatoire X est une application mesurable de $(\Omega, \mathcal{F}, \mathbf{P})$ dans \mathbb{R}^p muni de sa tribu borélienne. Si on munit \mathbb{R}^n de sa base canonique, alors on identifiera X au p-uplet de variables aléatoires (v.a) X_1, X_2, \dots, X_p . On écrit alors $X = (X_1, X_2, \dots, X_p)$. (voir [15])

1.1.1 Définition

Soient X_1, X_2, \dots, X_p ($p \geq 1$) des v.a définies sur l'espace probabilisable (Ω, \mathcal{F}) . Le p-uplet $X = (X_1, X_2, \dots, X_p)'$ est appelé vecteur aléatoire réel sur (Ω, \mathcal{F}) . Ce vecteur prend ses valeurs dans \mathbb{R}^p .

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R}^p \\ \omega &\mapsto (X_1(\omega), \dots, X_p(\omega)). \end{aligned}$$

On l'appelle aussi v.a multidimensionnelle ou multivariée.

Remarque 1.1.1

1. Les v.a X_1, X_2, \dots, X_p sont appelées composantes du vecteur X .

2. Dans le cas particulier $p = 2$, on parle de couple aléatoire (v.a bidimensionnelle ou bivariée) :

$$X = (X_1, X_2).$$

1.1.2 Types

Selon la nature des composantes X_1, X_2, \dots, X_p , on distingue deux types de vecteurs aléatoires : discret et continu.

Vecteur aléatoire discret

Le vecteur aléatoire X est dit discret si chacune de ses composantes est une v.a discrète. Ses valeurs sont généralement entières. On rencontre ce type de vecteurs dans les problèmes de dénombrement.

Exemple 1.1.1 *On s'intéresse aux nombres d'étudiants absents aux séances de cours et de travaux dirigés (TD) pendant le mois de mars. On désigne par X_1 le nombre d'absents au cours et par X_2 le nombre d'absents au TD. On obtient un couple aléatoire discret $X = (X_1, X_2)$.*

Vecteur aléatoire continu

Dans le cas où ces composantes sont des v.a continues, on dira que X est un vecteur aléatoire continu. Ses valeurs ne sont pas nécessairement entières. On rencontre ce type de vecteurs dans les sujets de mesure.

Exemple 1.1.2 *On s'intéresse aux mesures de taille et de poids d'un groupe d'élèves. Ces mesures sont généralement proportionnelles, c-à-d (c'est à dire) plus la taille augmente, plus le poids augmente en général, bien que cette relation ne soit pas toujours linéaire. On désigne par X_1 la taille et par X_2 le poids. On obtient un couple aléatoire continu $X = (X_1, X_2)$.*

1.2 Lois de probabilité

1.2.1 Loi conjointe

La loi d'un vecteur aléatoire X est donné par sa fonction de répartition F définie sur \mathbb{R}^p par :

$$F(x_1, \dots, x_p) := P(X_1 \leq x_1, \dots, X_p \leq x_p). \quad (1.1)$$

On l'appelle aussi fonction de répartition conjointe. (voir [7], page 48)

– Dans le cas où le vecteur X est discret, on définit, sur \mathbb{R}^p , la fonction de masse de probabilité conjointe par

$$P(x_1, \dots, x_p) = P(X_1 = x_1, \dots, X_p = x_p).$$

Elle est positive telle que $\sum \dots \sum P(x_1, \dots, x_p) = 1$.

– Dans le cas où le vecteur X est continu, la fonction F , définie par (1.1), est dérivable (presque partout) sur \mathbb{R}^p . On définit, sur \mathbb{R}^p , la fonction de densité de probabilité conjointe par :

$$f(x_1, \dots, x_p) = \frac{\partial^p F(x_1, \dots, x_p)}{\partial x_1 \dots \partial x_p}.$$

Elle vérifie les deux assertions suivantes :

f est positive.

f est intégrable avec $\int_{\mathbb{R}^p} f(x_1, \dots, x_p) dx_1 \dots dx_p = 1$.

Remarque 1.2.1 *On exprime la fonction de répartition en termes de la densité de probabilité.*

$$F(x_1, \dots, x_p) = \begin{cases} \sum^{x_1} \dots \sum^{x_p} p(u_1, \dots, u_p) & \text{si } X \text{ est discret} \\ \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} f(u_1, \dots, u_p) du_1 \dots du_p & \text{si } X \text{ est continu} \end{cases}.$$

1.2.2 Lois Marginales

La fonction de répartition marginale de la composante X_i , $i = 1, \dots, p$, est définie, sur \mathbb{R} , par :

$$F_{X_i}(x_i) = F(\infty, \dots, x_i, \dots, \infty)$$

On définit la fonction de masse marginale de la composante X_i , $i = 1, \dots, p$, par :

$$p_{X_i}(x_i) = \underbrace{\sum \dots \sum}_{(p-1) \text{ fois}} p(x_1, \dots, x_p); \quad x_i \in \mathbb{R}.$$

De la même façon on la densité marginale de la composante X_i , $i = 1, \dots, p$, par :

$$f_{X_i}(x_i) = \underbrace{\int \dots \int}_{(p-1) \text{ fois}} f(x_1, \dots, x_p) dx_1 \dots dx_p; \quad x_i \in \mathbb{R}.$$

(voir [10], page 108)

1.2.3 Lois conditionnelles

Elles sont définies en termes des fonctions de masse de densité conjointe et marginales. Dans le cas d'un couple aléatoire discret (X_1, X_2) , tel que les événements $\{X_1 = x_1\}$ et $\{X_2 = x_2\}$ sont de probabilités non nulles on définit alors deux familles de lois conditionnelles selon que l'on connaît la valeur de X_1 ou de X_2 . La fonction de masse conditionnelle de X_1 par rapport à X_2 , telle que $p_{X_2}(x_2) > 0$, est :

cas discret :

$$p(x_1/x_2) = P(X_1 = x_1 / X_2 = x_2) = \frac{p(x_1, x_2)}{p_{X_2}(x_2)}.$$

Pour un couple continu (X_1, X_2) , on définit la densité conditionnelle de X_1 par rapport à X_2 par :

cas continu :

$$f(x_1/x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)},$$

à condition bien sûr que $f_{X_2}(x_2) > 0$ (voir [10], page 109).

Remarque 1.2.2 *Les fonctions de masse et de densité conditionnelles de X_2 par rapport à X_1 sont définie de la même manière (en inversant les rôles).*

1.2.4 Indépendance

Les v.a X_1, X_2, \dots, X_p sont dites indépendantes si, pour tout $(x_1, \dots, x_p) \in \mathbb{R}^p$ on a

$$F(x_1, \dots, x_p) = F_{X_1}(x_1) \dots F_{X_p}(x_p).$$

Ce qui se traduit par

$$\begin{aligned} p(x_1, \dots, x_p) &= p_{X_1}(x_1) \dots p_{X_p}(x_p) \quad \text{si } X \text{ est discret,} \\ f(x_1, \dots, x_p) &= f_{X_1}(x_1) \dots f_{X_p}(x_p) \quad \text{si } X \text{ est continu.} \end{aligned}$$

(voir [15], page 87)

Remarque 1.2.3 *Dans le cas d'indépendance, les lois conditionnelles sont égales aux lois marginales.*

1.3 Paramètres statistiques

1.3.1 Moyenne

La moyenne ou espérance du vecteur aléatoire X est définie par :

$$\mu = E[X] := (\mu_1, \dots, \mu_p)'$$

où $\mu_k := E[X_k]$ désigne l'espérance de la composante X_k , $k = 1, \dots, p$. C'est le vecteur moyen (vecteur des moyennes). (voir [15])

1.3.2 Matrice de covariance

On appelle matrice de covariance du vecteur X , la matrice carré \sum de taille p de terme général

$$\sigma_{ij} = Cov(X_i, X_j), \quad i, j = 1, \dots, p,$$

où

$$\text{Cov}(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))] = E(X_i X_j) - E(X_i)E(X_j),$$

désigne la covariance des variables X_i et X_j . Pour $i = j$, le terme diagonal σ_{ii} , qu'on note σ_i^2 , est égal à la variance de la composante X_i .

Remarque 1.3.1

1. La matrice \sum est une matrice symétrique semi définie positive. On écrit alors

$$\sum = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_2 & \sigma_1^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{pmatrix}. \tag{1.2}$$

En effet, pour $i, j = 1, \dots, p$, on a

$$\sigma_{ij} = \text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = E(X_j X_i) - E(X_j)E(X_i) = \sigma_{ji}.$$

2. La covariance est bilinéaire.

En effet, ona, pour $a, b, c, d \in \mathbb{R}$,

$$\begin{aligned} \text{Cov}(aX_i + bX_j, cX_i + dX_j) &= a\text{Cov}(X_i, cX_i + dX_j) + b\text{Cov}(X_j, cX_i + dX_j) \\ &= ac\text{Cov}(X_i, X_i) + ad\text{Cov}(X_i, X_j) \\ &\quad + bc\text{Cov}(X_j, X_i) + bd\text{Cov}(X_j, X_j). \end{aligned}$$

3. Si X_i et X_j sont indépendantes, alors elles sont non corrélées. La matrice \sum est donc diagonale.

En effet, ona

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = E(X_i)E(X_j) - E(X_i)E(X_j) = 0.$$

4. Pour $a, b \in \mathbb{R}$, on a

$$\text{Var}(aX_i + bX_j) = a^2\text{Var}(X_i) + 2ab\text{Cov}(X_i, X_j) + b^2\text{Var}(X_j). \quad (1.3)$$

En particulier, pour $a = b = 1$, on a

$$\text{Var}(X_i + X_j) = \text{Var}(X_i) + 2\text{Cov}(X_i, X_j) + \text{Var}(X_j).$$

Proposition 1.3.1 (Inégalité de Cauchy-Schwarz) Pour $i, j = 1, \dots, p$, on a

$$|\sigma_{ij}| = |\text{Cov}(X_i, X_j)| \leq \sqrt{\text{Var}(X_i)}\sqrt{\text{Var}(X_j)} = \sigma_i\sigma_j.$$

Preuve. On pose $\mathcal{P}(\lambda) = \text{Var}(\lambda X_i + X_j) \geq 0$, $\lambda \in \mathbb{R}$. D'après le résultat (1.3), on a

$$\mathcal{P}(\lambda) = \lambda^2\text{Var}(X_i) + 2\lambda\text{Cov}(X_i, X_j) + \text{Var}(X_j) \geq 0,$$

où $\mathcal{P}(\lambda)$ est un polynôme de degré 2 par rapport à λ . Son discriminant est négatif, c-à-d

$$(\text{Cov}(X_i, X_j))^2 - \text{Var}(X_i)\text{Var}(X_j) \leq 0.$$

On en déduit le résultat. ■

Matrice de corrélation

La matrice de corrélation du vecteur aléatoire X est une matrice carrée d'ordre p , symétrique définie positive :

$$\mathbf{R} = \begin{pmatrix} 1 & \cdots & \cdots & \rho_{1p} \\ \vdots & \ddots & \cdot & \cdot \\ \cdot & \cdot & \ddots & \cdot \\ \vdots & \cdot & \cdot & \cdot \\ \rho_{1p} & \cdot & \cdot & 1 \end{pmatrix},$$

de le terme général

$$r_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}, \quad i, j = 1, \dots, p,$$

appelé coefficient de corrélation linéaire entre les composantes X_i et X_j .

Remarque 1.3.2

1. On a $-1 \leq \rho_{ij} \leq 1$, $i, j = 1, \dots, p$.
2. Le terme diagonal est égal à l'unité : $\rho_{ii} = 1$, $i = 1, \dots, p$.
3. Si X_i et X_j sont indépendantes, alors la matrice \mathbf{R} est égale à la matrice identité. (voir [15])

1.4 Vecteur aléatoire gaussien

Un vecteur aléatoire $X = (X_1, X_2, \dots, X_p)'$ est dit gaussien si et seulement si toute combinaison linéaire de ses composantes est une v.a gaussienne. Autrement dit, pour toute suite réelle (a_1, a_2, \dots, a_p) la v.a

$$Z = a_1 X_1 + a_2 X_2 + \dots + a_p X_p.$$

est gaussienne.

Densité de probabilité

La densité de probabilité du vecteur X est définie sur \mathbb{R}^p par :

$$f(x) = \frac{1}{(2\pi)^{p/2} (\det \Sigma)^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right\},$$

où $\mu \in \mathbb{R}^p$ et Σ une matrice définie positive ($\det \Sigma \neq 0$) de la forme [1.2]. On dit que X est de distribution multinormale de paramètres μ et Σ qui représentent le vecteur moyen et la matrice de covariance. On écrit $X \sim \mathcal{N}_p(\mu, \Sigma)$.

Remarque 1.4.1

1. Le cas particulier où $\mu = 0$ et $\Sigma = I_p$ correspond au vecteur multinormal standard Z de densité

$$f(z) = \frac{1}{(2\pi)^{p/2}} \exp \left\{ -\frac{1}{2} z' z \right\}.$$

Le passage du cas général au cas standard se fait via le changement de variable

$$Z = \Sigma^{-1/2} (X - \mu),$$

où $\Sigma^{-1/2}$ est l'inverse de la matrice racine $\Sigma^{1/2}$. Pour des détails sur la définition de cette dernière se référer à (voir [15], page 88)

2. Les lois marginales et conditionnelles sont normales. (voir [3], page 94)

Proposition 1.4.1 Les composantes d'un vecteur multinormal sont indépendantes si et seulement si elles sont non corrélées (matrice de covariance diagonale).

Cas particulier : couple aléatoire gaussien

Dans ce cas, on a

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}, \quad \det \Sigma = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 \quad \text{et} \quad \Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}.$$

En exprimant la covariance σ_{12} en termes du coefficient de corrélation linéaire r , on obtient

$$\det \Sigma = (1 - r^2) \sigma_1^2 \sigma_2^2 \quad \text{et} \quad \Sigma^{-1} = \frac{1}{(1 - r^2) \sigma_1^2 \sigma_2^2} \begin{pmatrix} \sigma_2^2 & -r \sigma_1 \sigma_2 \\ -r \sigma_1 \sigma_2 & \sigma_1^2 \end{pmatrix}.$$

La densité s'écrit alors, pour $(x, y) \in \mathbb{R}^2$:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)\sigma_1^2\sigma_2^2} \left[(x - \mu_1)^2 - 2r(x - \mu_1)(y - \mu_2) + (y - \mu_2)^2 \right] \right\}. \quad (1.4)$$

Chapitre 2

Copules

Dans ce chapitre on va présenter les fonctions copules, leurs définitions et certaines de leurs caractéristiques ainsi que quelques formes usuelles.

2.1 Généralités

Une copule d -dimensionnelle ($d \geq 2$) est une fonction : $C : [0, 1]^d \rightarrow [0, 1]$ telle qu'il existe un espace probabiliste $(\Omega, \mathcal{F}, \mathbb{P})$ contenant un vecteur aléatoire (U_1, \dots, U_d) tel que $U_k \sim U[0, 1]$ pour tout $k = 1, \dots, d$

$$C(u_1, \dots, u_d) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d), \quad u_1, \dots, u_d \in [0, 1].$$

Par la suite, on ne va s'intéresser qu'aux copules bidimensionnelles correspondant au cas $d = 2$.

2.2 Définition

Une copule bidimensionnelle est une fonction C de $\mathbb{I}^2 = [0, 1]^2$ dans \mathbb{I} telle que :

$$C(u_1, u_2) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2), \quad u_1, u_2 \in [0, 1].$$

Elle a les propriétés suivantes :

i) C est attachée, c-à-d

$$C(u, 0) = C(0, v) = 0, \forall u, v \in \mathbb{I}. \quad (2.1)$$

ii) Ses marges sont uniformes, c-à-d

$$C(u, 1) = u \text{ et } C(1, v) = v, \forall u, v \in \mathbb{I}. \quad (2.2)$$

iii) C est 2-croissante : $\forall u_1, v_1, u_2, v_2 \in \mathbb{I}$, telle que $u_1 \leq u_2$ et $v_1 \leq v_2$

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \quad (2.3)$$

2.3 Exemple de copules

On présente ici quelques modèles de copules les plus utilisées.

2.3.1 Copule produit

La copule produit, appelée aussi copule d'indépendance, notée Π et définie par :

$$\Pi(u, v) = uv, \forall u, v \in \mathbb{I}.$$

On observe qu'elle vérifie les propriétés [2.1](#) et [2.2](#). En effet, on a :

$$\Pi(0, v) = 0 \times v = 0 \text{ et } \Pi(u, 0) = u \times 0 = 0.$$

et

$$\Pi(1, v) = 1v = v \text{ et } \Pi(u, 1) = u.$$

Pour la propriété [2.3](#), on a :

$$\Pi(u_2, v_2) - \Pi(u_2, v_1) - \Pi(u_1, v_2) + \Pi(u_1, v_1) = u_2v_2 - u_2v_1 - u_1v_2 + u_1v_1 \geq 0,$$

où

$$u_2 \geq u_1 \implies \begin{cases} u_2 v_2 \geq u_1 v_2 \\ u_2 v_1 \geq u_1 v_1 \end{cases}$$

alors

$$u_2 v_2 - u_2 v_1 \geq u_1 v_2 - u_1 v_1 \implies u_2 v_2 - u_2 v_1 - u_1 v_2 + u_1 v_1 \geq 0$$

2.3.2 Copules min et max

La copule min, noté M , est définie par :

$$M(u, v) = \min(u, v), \forall u, v \in I.$$

La copule max, noté W , est définie par :

$$W(u, v) = \max(u + v - 1, 0) \forall u, v \in I.$$

Pour toute copule C on a la relation :

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

Pour cela, M et W sont appelées bornes supérieure et inférieure de Fréchet respectivement.

2.4 Propriétés d'une copule

Dans cette partie on va présenter quelques propriétés des copules (en plus de celles vues ci-dessus) telles que la continuité, la symétrie, l'ordre, la croissance et l'invariance. (voir [II], page 9)

a. Continuité : les copules sont des fonctions continues. Plus précisément, elles vérifient la condition de Lipschitz. $\forall u_1, u_2, v_1, v_2 \in \mathbb{I}$ on a

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

b. Symétrie : on dit que C est symétrique si

$$\forall (u, v) \in \mathbb{I}^2, C(u, v) = C(v, u).$$

c. Ordre : on dit que C_1 est plus petite que C_2 si

$$\forall (u, v) \in \mathbb{I}^2 : C_1(u, v) \leq C_2(u, v).$$

2.5 Théorème de Sklar

Ce théorème représente la base de la théorie de copules, car il valide l'existence de la copule C qui permet de relier les lois marginales F et G de deux v.a X et Y pour obtenir la distribution conjointe H . Ce théorème a été développé par Sklar en 1959 et porte son nom [16] (voir, par exemple, [13], page 17).

Théorème 2.5.1 *Si H est la fonction répartition de (X, Y) alors il existe une copule C de dimension 2 telle que*

$$\forall (x, y) \in \mathbb{R}^2 : H(x, y) = C(F(x), G(y)). \quad (2.4)$$

Pour la démonstration, on a besoin de la définition suivante.

Définition 2.5.1 (Fonction quantile) *Soit F une fonction répartition univarié. Le quantile de F (l'inverse généralisé) est une fonction, notée F^{-1} de domaine de définition \mathbb{I} telle que*

$$F^{-1}(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\}.$$

Preuve. Soit C la fonction répartition du vecteur (U, V) où $U = F(X)$ et $V = G(Y)$. A l'aide de la définition [2.5.1](#) et l'équation [2.5.1](#) :

$$\begin{aligned} H(x, y) &= \mathbb{P}[X \leq x, Y \leq y] \\ &= \mathbb{P}[F^{-1}(U) \leq x, G^{-1}(V) \leq y] \\ &= \mathbb{P}[U \leq F(x), V \leq G(y)] \\ &= C(F(x), G(y)). \end{aligned}$$

Ce qu'il fallait démontrer (CQFD). ■

On déduit un des résultats essentiels de la théorie des copules, connu sous le nom de théorème d'invariance (par transformations strictement croissantes). Pour plus de détails,

Théorème 2.5.2 (Invariance) *Soient X et Y deux v.a continues de marginales F et G et de copule $C_{X,Y}$, Si α et β sont deux fonctions strictement croissantes, alors*

$$C_{\alpha(X),\beta(Y)} = C_{X,Y}.$$

En d'autres termes, la copule $C_{X,Y}$ est invariante par transformations strictement croissantes des v.a X et Y .

Preuve. Soient α et β deux fonctions strictement croissantes et X et Y deux v.a continues ayant pour copule $C_{X,Y}(u, v) = H_1(x, y)$ et pour fonctions de répartition marginales respectivement F_1 et G_1 . Notons les fonctions de distribution univariées de $\alpha(X)$ et $\beta(Y)$ respectivement F_2 et G_2 et la loi conjointe H_2 . Puisque α et β sont strictement croissantes et que X et Y sont des v.a continues,

pour tout $(u, v) \in \mathbb{I}^2$ on a : (voir [13], page 25)

$$\begin{aligned}
 C_{\alpha(X),\beta(Y)}(u, v) &= C_{\alpha(X)\beta(Y)}(F_2(x), G_2(y)) \\
 &= H_2(x, y) \\
 &= \mathbb{P}(\alpha(X) \leq x, \beta(Y) \leq y) \\
 &= \mathbb{P}(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)) \\
 &= H_1(\alpha^{-1}(x), \beta^{-1}(y)) \\
 &= C_{X,Y}(F_1(\alpha^{-1}(x)), G_1(\beta^{-1}(y))) \\
 &= C_{X,Y}(F_2(x), G_2(y)) \\
 &= C_{X,Y}(u, v).
 \end{aligned}$$

Ce nous qui donne la formule de théorème 2.5.2. ■

Remarque 2.5.1 (Densité d'une copule) Soient X, Y deux v.a continues de densité conjointe, h et de densités marginales f et g respectivement. La densité $c(F(x), G(y))$ associée à la copule $C(F(x), G(y))$ est définie par :

$$c(F(x), G(y)) = \frac{\partial^2 C(F(x), G(y))}{\partial F(x) \partial G(y)} = \frac{h(x, y)}{f(x)g(y)}.$$

D'après 2.5.1 on donne la représentation canonique suivante :

$$h(x, y) = c(F(x), G(y))f(x)g(y).$$

2.6 Familles de copules

Il existe de nombreuses familles de copules qui diffèrent les unes des autres. En général, une famille a plusieurs paramètres qui se rapportent à la force et la forme de la dépendance. Certaines familles de copules sont décrites ci-dessous.

2.6.1 Copules Archimédiennes

Définition 2.6.1 On appelle copule Archimédienne la copule C définie par :

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), u, v \in \mathbb{I}, \quad (2.5)$$

où $\phi : \mathbb{I} \rightarrow [0, +\infty[$ une fonction continue, strictement décroissante vérifiant $\phi(1) = 0$, appelée générateur de la copule. (voir [17], page 43)

Remarque 2.6.1

1. On définit l'inverse de ϕ par $\phi^{[-1]}$ tel que

$$\phi^{[-1]}(t) := \begin{cases} \phi^{-1}(t) & \text{si } 0 \leq t \leq \phi(0). \\ 0 & \text{si } \phi(0) \leq t \leq \infty. \end{cases}$$

Il est clair que si $\phi(0) = \infty$, alors $\phi^{[-1]} = \phi^{-1}$.

2. ϕ est au moins deux fois continument dérivable telle que $\phi'(u) < 0$ et $\phi''(u) > 0$ pour tout $u \in \mathbb{I}$.

Le tableau 2.1 présente quelques familles Archimédiennes classiques.

Copule	$C_\theta(u, v)$	$\phi_\theta(t)$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(t^{-\theta} - 1)/\theta; \theta \in [-1, \infty[-\{0\}]$
Gumbel	$\exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\}$	$(-\ln t)^\theta; \theta \geq 1$
Franck	$-\frac{1}{\theta} \ln(1 + (e^{-\theta u} - 1)(e^{-\theta v} - 1)/(e^{-\theta} - 1))$	$-\ln((e^{-\theta t} - 1)/(e^{-\theta} - 1)); \theta \in \mathbb{R}$
Joe	$1 - \left[(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right]^{1/\theta}$	$-\ln(1 - (1-t)^\theta); \theta \geq 1$

TAB. 2.1 – Principales copules Archimédiennes et leurs générateurs

On note les cas particuliers suivants (voir, par exemple, [6], [9] et [18]).

- **Copule de Clayton** : $C_{-1} = W$, $C_0 = \Pi$ et $C_\infty = M$.
- **Copule de Gumbel-Hougaard** : $C_1 = \Pi$ et $C_\infty = M$.
- **Copule de Franck** : $C_{-\infty} = W$, $C_0 = \Pi$ et $C_\infty = M$.

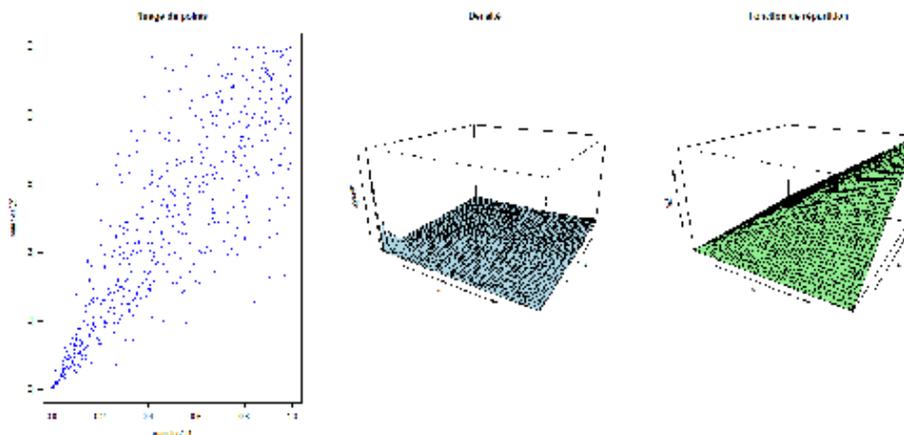


FIG. 2.1 – Visualisation de la Copule de Clayton

– **Copule de Joe** : $C_1 = \Pi$ et $C_\infty = M$.

La figure 2.1 illustre la copule de Clayton : un nuage de points montrant une forte dépendance dans la queue inférieure, ce qui indique un lien significatif entre les variables lorsque leurs valeurs sont faibles. Une densité où la concentration est maximale dans le coin inférieur gauche. La fonction de répartition cumulative montre une dépendance croissante pour les petites valeurs

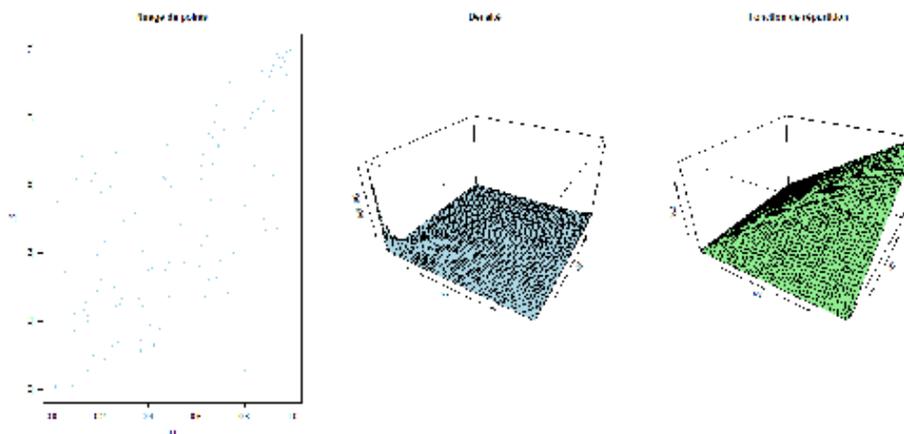


FIG. 2.2 – Visualisation de la Copule de Gumbel

La figure 2.2 illustre la copule de Gumbel : le nuage de points montre une forte dépendance dans la queue supérieure, la densité met en évidence une concentration dans le coin supérieur droit et la fonction de répartition montre une croissance plus marquée pour les grandes valeurs.

La figure 2.3 illustre la copule de Frank : le nuage de points montre une dépendance positive modérée et symétrique entre les deux composantes, sans concentration dans les coins. Ceci indique une absence de dépendance dans les queues. La densité est maximale au centre, impliquant une dépendance moyenne

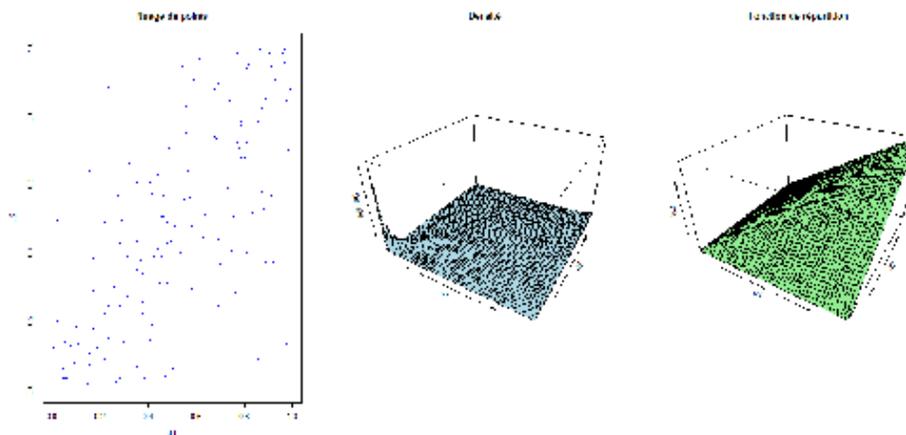


FIG. 2.3 – Visualisation de la Copule de Frank

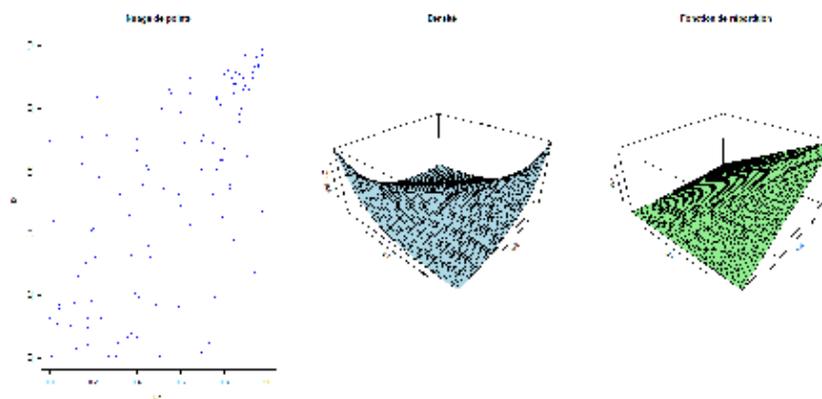


FIG. 2.4 – Visualisation de la Copule de Joe

autour des valeurs centrales. La fonction de répartition est régulière, traduisant une liaison monotone et symétrique.

La figure 2.4 illustre la copule de Joe : le nuage de points illustre la dépendance positive entre U_1 et U_2 , avec une concentration dans la diagonale. La densité représente la densité de probabilité de la copule, mettant en évidence une forte concentration aux coins. Fonction de répartition Montre comment la probabilité cumulée évolue selon les valeurs de U_1 et U_2 .

2.6.2 Copules elliptiques

Une copule est dite elliptique si elle s'écrit, pour $0 \leq u, v \leq 1$, sous la forme :

$$\begin{aligned}
 C_r(u, v) &= H_r(F^{-1}(u), G^{-1}(v)) \\
 &= \frac{1}{\sqrt{1-r^2}} \int_{-\infty}^{F^{-1}(u)} \int_{-\infty}^{G^{-1}(v)} \varphi\left(\frac{x^2 - 2rxy + y^2}{\sqrt{1-r^2}}\right) dx dy,
 \end{aligned}
 \tag{2.6}$$

où H_r est la fonction de répartition conjointe de X et Y , F^{-1} et G^{-1} leurs fonctions quantiles respectives et r leur coefficient de corrélation. φ étant une fonction réelle de deux variables. Les deux classes les plus utilisées de copule elliptique sont les copules gaussiennes et les copules de Student.

Copule Gaussienne

Cette copule est utilisée lorsque la mesure de dépendance est exprimée le coefficient de corrélation r . Elle est paramétrée par ce dernier. (voir [8], page 15)

Définition 2.6.2 *La copule Gaussienne de deux v.a X, Y est définie par :*

$$C_r(u, v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(-\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right) dx dy.$$

Il suffit d'appliquer les résultats [1.4] de la section [1.4] du chapitre 1 dans (2.6).

On a les cas particuliers suivants :

$$C_r(u, v) = \begin{cases} W(u, v), & r = -1, \\ \Pi(u, v), & r = 0, \\ M(u, v), & r = 1. \end{cases}$$

La figure [2.5] illustre les propriétés de la copule Gaussienne : le nuage de points montre une dépendance modérée et symétrique entre les composantes. La densité présente une surface régulière et symétrique, caractéristique de la copule normale.

Copule de Student

La copule Student est extraite de la distribution de Student. (voir [4], page 15)

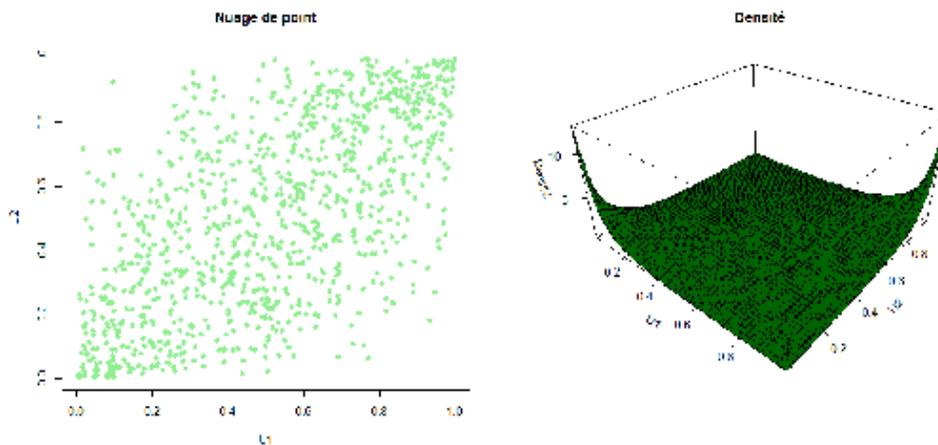


FIG. 2.5 – Nuage de points et densité d’une copule gaussienne

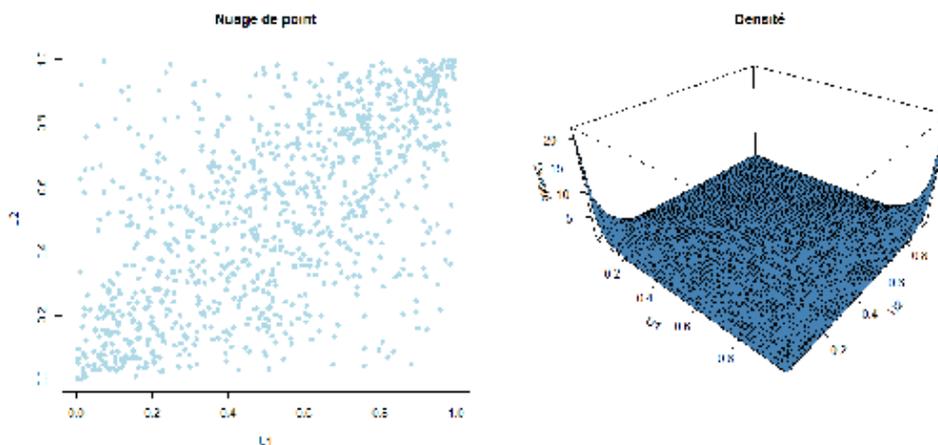


FIG. 2.6 – Nuage de points et densité d’une copule de Student

Définition 2.6.3 La copule de Student (*T* copule) est une copule paramétrique, paramétrée par le coefficient de corrélation linéaire r et le degré de liberté k ($k > 3$) Cette copule est définie par :

$$C_{r,k}(u, v) = T_{r,k}(T_k^{-1}(u), T_k^{-1}(v))$$

tell que $T_{r,k}$ la distribution de Student bivariée alors la copule Student bivariée est :

$$C_{r,k}(u, v) = \int_{-\infty}^{T_k^{-1}(u)} \int_{-\infty}^{T_k^{-1}(v)} \frac{1}{2\pi\sqrt{1-r^2}} \left(1 + \frac{x^2 - 2rxy + y^2}{k(1-r^2)} \right)^{\frac{2-k}{2}} dx dy$$

La figure [2.6](#) illustre les caractéristiques de la copule T (copule de Student) : le nuage de points montre une forte dépendance dans les queues, ce qui indique une corrélation extrême entre les variables. la

densité met en évidence une concentration dans les coins, typique de la copule T .

Chapitre 3

Mesures d'association

*D*ans ce chapitre, on présente les définitions des mesures d'association (concordance-discordance), leurs caractéristiques ainsi que leur intérêt et le lien qu'elles ont avec les copules.

3.1 Définition

Les mesures d'association sont des outils d'analyse de données qui mesurent la force d'une relation entre deux variables. A titre d'exemple, on peut citer le coefficient de corrélation de Pearson introduit dans la section [1.3](#) du premier chapitre. Il mesure le degré de la relation linéaire entre deux v.a.

3.2 Mesure de concordance

3.2.1 Concordance et discordance

Soient (x_i, y_i) et (x_j, y_j) , pour $i, j \in \{1, \dots, n\}$, deux ensembles d'observations d'un couple de v.a (X, Y) .

Définition 3.2.1 *On dit que (x_i, y_i) et (x_j, y_j) sont concordantes si et seulement si*

$$(x_i - x_j)(y_i - y_j) > 0,$$

c-à-d : $(x_i < x_j$ et $y_i < y_j)$ ou $(x_i > x_j$ et $y_i > y_j)$. Sinon, on dira qu'elles sont discordantes.

Propriété 3.2.1 Une mesure numérique d'association $\kappa_{X,Y}$ entre deux v.a continues X et Y dont la copule est C est une mesure de concordance si et seulement si elle satisfait les propriétés qui suivent : (voir [4])

1. $-1 \leq \kappa_{X,Y} \leq 1$, $\kappa_{X,Y} = 1$ et $\kappa_{X,-X} = -1$.
2. $\kappa_{X,Y} = \kappa_{Y,X}$.
3. $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$.
4. Si X et Y sont indépendantes, alors $\kappa_{X,Y} = 0$.
5. Si (X_n, Y_n) est une suite de couples aléatoires continus convergeant vers (X, Y) , alors

$$\lim_{n \rightarrow \infty} \kappa_{X_n, Y_n} = \kappa_{X, Y}.$$

6. Si $\alpha(X)$ et $\beta(Y)$ sont des fonctions strictement croissantes, alors

$$\kappa_{\alpha(X), \beta(Y)} = \kappa_{X, Y}.$$

7. Si les copules respectives de (X_1, Y_1) et (X_2, Y_2) sont telles que $C_1 < C_2$, alors $\kappa_{X_1, Y_1} \leq \kappa_{X_2, Y_2}$.

3.2.2 Fonction de concordance

Définition 3.2.2 La fonction de concordance Q entre deux couples (X_1, Y_1) et (X_2, Y_2) est égale à la différence entre la probabilité de concordance et la probabilité de discordance, c-à-d :

$$Q := \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\}. \quad (3.1)$$

Théorème 3.2.1 Soient (X_1, Y_1) et (X_2, Y_2) deux vecteurs aléatoires continus indépendants de fonctions de répartition H_1 et H_2 de copules C_1 et C_2 respectivement. Soient F et G les marges respectives associées à X_1, X_2 et à Y_1, Y_2 , c-à-d

$$H_1(x, y) = C_1(F(x), G(y)) \text{ et } H_2(x, y) = C_2(F(x), G(y)).$$

Alors, la fonction de concordance Q s'exprime en termes des copules selon la relation suivantes :

$$Q = Q(C_1, C_2) = 4 \int \int_{\mathbb{I}^2} C_2(u, v) dC_1(u, v) - 1. \quad (3.2)$$

Preuve. D'après la probabilité de l'événement contraire et puisque toutes les v.a sont, dans ce cas, continues, alors on a

$$\begin{aligned} \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\} &= 1 - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) \geq 0\} \\ &= 1 - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\}. \end{aligned}$$

Ainsi

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

En étudiant le premier terme de [3.1](#), on a

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

en integrant ces quantités

$$\begin{aligned} \int \mathbb{P}(X_2 < X_1, Y_2 < Y_1) &= \int \int_{\mathbb{R}^2} \mathbb{P}(X_2 < x, Y_2 < y) dC_1(F(x), G(y)) \\ &= \int \int_{\mathbb{R}^2} C_2(F(x), G(y)) dC_1(F(x), G(y)) \end{aligned}$$

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

$$\mathbb{P}(X_1 > X_2, Y_1 > Y_2) = \int \int_{\mathbb{I}^2} C_2(u, v) dC_1(u, v).$$

D'une façon similaire, on a

$$\begin{aligned} \mathbb{P}(X_1 < X_2, Y_1 < Y_2) &= \int \int_{\mathbb{R}^2} \mathbb{P}(X_2 > x, Y_2 > y) dC_1(F(x), G(y)) \\ &= \int \int_{\mathbb{R}^2} (1 - F(x) - G(y) + C_2(F(x), G(y))) dC_1(F(x), G(y)) \\ &= \int \int_{\mathbb{I}^2} (1 - u - v + C_2(u, v)) dC_1(u, v), \end{aligned}$$

car U et V sont des v.a uniformes, donc $E(U) = E(V) = \frac{1}{2}$ et donc

$$\mathbb{P}(X_1 < X_2, Y_1 < Y_2) = 1 - \frac{1}{2} - \frac{1}{2} + \int \int_{\mathbb{I}^2} C_2(u, v) dC_1(u, v).$$

Par conséquent, on a

$$\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} = 2 \int \int_{\mathbb{I}^2} C_2(u, v) dC_1(u, v).$$

En regroupant ces résultats, on obtient le résultat **3.2**. ■

Corollaire 3.2.1 *Soient C_1 et C_2 deux copules et Q leur fonction de concordance. Alors Q possède les propriétés suivantes :*

1. Q est symétrique :

$$Q(C_1, C_2) = Q(C_2, C_1).$$

2. Q est non décroissante :

$$\text{si } C_1 < C'_1 \text{ et } C_2 < C'_2 \text{ pour tout } (u, v) \in \mathbb{I}^2, \text{ alors } Q(C_1, C_2) < Q(C'_1, C'_2).$$

Exemple :

Nous calculons la mesure Q pour les copules M , W et Π : Le support de M est la diagonale $u = v$ dans \mathbb{I}^2 , et car M a des marges uniforme, il s'ensuit que si g est une fonction intégrable dont le domaine est \mathbb{I}^2 , alors (voir **[13]**, page 160)

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

Par conséquent, nous avons :

- $Q(M, M) = 4 \int \int_{\mathbb{I}^2} \min(u, v) dM(u, v) - 1 = 4 \int_0^1 u du - 1 = 1.$
- $Q(M, \Pi) = 4 \int \int_{\mathbb{I}^2} uv dM(u, v) - 1 = 4 \int_0^1 u^2 du - 1 = 1/3.$
- $Q(M, W) = 4 \int \int_{\mathbb{I}^2} \max(u + v - 1, 0) dM(u, v) - 1 = 4 \int_0^1 (2u - 1) du - 1 = 0.$

X vs Y	1	0
1	a	b
0	c	d

TAB. 3.1 – Tableau générique

De la même façon, car le support de W est la diagonale $v = 1 - u$, nous avons :

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

et donc

- $Q(W, \Pi) = 4 \int \int_{\mathbb{I}^2} uv dW(u, v) - 1 = 4 \int_0^1 u(1 - u) du - 1 = -1/3$.
- $Q(W, W) = 4 \int \int_{\mathbb{I}^2} \max(u + v - 1, 0) dW(u, v) - 1 = 4 \int_0^1 0 du - 1 = -1$.

et pour la copule Π , nous avons $d\Pi(u, v) = dudv$, alors

- $Q(\Pi, \Pi) = 4 \int \int_{\mathbb{I}^2} uv d\Pi(u, v) - 1 = 4 \int_0^1 \int_0^1 uv dudv - 1 = 0$.

3.3 Structure de dépendance

Dans cette section, on présente les trois principales mesures d'association, à savoir le phi de Pearson, le rho de Spearman et le tau de Kendall.

3.3.1 Phi de Pearson

Le coefficient de corrélation phi de Pearson permet de mesurer l'intensité de la liaison entre deux variables binaires (codées 0 ou 1). Le calcul est réalisé à travers le coefficient de Pearson sur les variables binaires ou sur un tableau [3.1](#) composé de deux lignes et deux colonnes comme suit :

Sa formule est :

$$\phi := \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

Remarque 3.3.1 *On utilise souvent le codage 1 de la modalité qui nous intéresse et 0 pour la seconde. De plus, ce codage détermine le signe de ϕ , mais il n'a pas d'incidence sur la valeur absolue du coefficient. (voir [19](#)) Le coefficient phi de Pearson est similaire, dans son interprétation, au coefficient de corrélation linéaire de la section [1.3](#) du chapitre 1.*

- Il varie entre -1 et 1 .
- Plus il est proche de ces bornes plus la liaison est forte entre les deux variables. $\phi = 0$ indique une situation d'indépendance.

3.3.2 Tau de Kendall

Le tau de Kendall de la série exprimé en fonction des observations écrit sous la forme suivante :

Définition 3.3.1 (Version échantillon) Soit une série de n observations $\{(x_i, y_i)\}_{1 \leq i \leq n}$ d'un couple (X, Y) . Le tau de Kendall est défini par :

$$\tau := \frac{2}{n(n-1)}(n_c - n_d); \tau \in [-1, 1], \quad (3.3)$$

où $\begin{cases} n_c : \text{nombre de paires concordantes,} \\ n_d : \text{nombre de paires discordantes,} \\ n : \text{nombre total de paires.} \end{cases}$

Sachant que $n_c - n_d = \sum_{i=1}^{j-1} \sum_{j=2}^n \text{sign}\{(x_i - x_j)(y_i - y_j)\}$.

En posant $z_{ij} = (x_i - x_j)(y_i - y_j)$, on a $\text{sign}\{z_{ij}\} = \begin{cases} 1 & \text{si } z_{ij} > 0. \\ -1 & \text{si } z_{ij} < 0. \end{cases}$

Le tau de Kendall n'est autre que la différence entre la probabilité de concordance et celle de discordance.

Définition 3.3.2 (Version population) Soient (X_1, Y_2) et (X_2, Y_2) deux couples de vecteurs aléatoires. Le tau de Kendall est donné par :

$$\tau := \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Proposition 3.3.1 Soit (X, Y) un couple de v.a continues de copule C . Le tau de Kendall a comme expression :

$$\tau = \tau_c = 4 \int \int_{\mathbb{I}^2} C(u, v) dC(u, v) - 1. \quad (3.4)$$

En effet, on remplace C_1 et C_2 par C dans la formule (3.2) et on obtient le résultat.

Remarque 3.3.2

1. La double intégrale dans (3.4) n'est autre que $\mathbb{E}[C(u, v)]$. Par conséquent, on peut écrire :

$$\tau_c = 4\mathbb{E}[C(u, v)] - 1$$

2. Le tau de Kendall de X et Y d'une copule Archimédienne s'obtient simplement à l'aide du générateur de la copule, selon la formule ci-dessous (voir, par exemple, [1], page 70) :

$$\tau_c = 1 + 4 \int_{\mathbb{I}} \frac{\phi(u)}{\phi'(u)} du.$$

3.3.3 Rho de Spearman

Tout comme le tau de Kendall, le rho de Spearman est une mesure de dépendance basée sur la notion de concordance. Soient (X_1, Y_1) , (Y_2, Y_2) et (X_3, Y_3) trois couples aléatoires indépendants de même distribution H dont les distributions marginales sont F et G et dont la copule associée est C .

Définition 3.3.3 (Version échantillon) Soit une série de n observations $\{(x_i, y_i)\}_{1 \leq i \leq n}$ d'un couple (X, Y) . On définit le rho de Spearman par :

$$\rho := 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n D_i^2,$$

où $D_i = R_{X_i} - R_{Y_i}$, R_{X_i} et R_{Y_i} les rangs des observations de X et Y respectivement. (voir [14])

Définition 3.3.4 (Version population) La version populaire du rho de Spearman est définie comme étant proportionnelle à la différence de la probabilité de concordance et celle de discordance des couples aléatoires (X_1, Y_1) et (X_2, Y_3) (voir [11]) :

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

La distribution de (X_1, Y_1) étant H et celle de (X_2, Y_3) étant Π (car les variables X_2 et Y_3 sont indépendantes) alors d'après ce qui précède on a le théorème suivant :

Théorème 3.3.1 Soient X et Y deux v.a continues de copule C . Le rho de Spearman de X et Y est exprimé par :

$$\rho = \rho_C = 12 \int \int_{\mathbb{I}^2} C(u, v) dudv - 3 \quad (3.5)$$

Preuve. On a

$$\begin{aligned} \rho &= 3Q(C, \Pi) = 3 \left(\int \int_{\mathbb{I}^2} \Pi(u, v) dC(u, v) - 1 \right) \\ &= 12 \int \int_{\mathbb{I}^2} uv dC(u, v) - 3. \end{aligned}$$

Une integration par parties nous donne le résultat (3.5). ■

Remarque 3.3.3 Le rho de Spearman des v.a X et Y est égal au coefficient de corrélation linéaires des marginales correspondantes $F(X)$ et $G(Y)$, c-à-d :

$$\rho = \rho(X, Y) = r(F(X), G(Y)).$$

Preuve. Puisque les variables $U = F(X)$ et $V = F(Y)$ sont uniformes sur \mathbb{I} alors $E(U) = E(V) = 1/2$ et $Var(U) = Var(V) = 1/12$. D'autre part, on a :

$$\int \int_{\mathbb{I}^2} uv dC(u, v) = \mathbf{E}(UV).$$

Donc,

$$\begin{aligned} \rho &= 12 \int \int_{\mathbb{I}^2} uv dC(u, v) - 3 \\ &= 12\mathbf{E}(UV) - 3 = \frac{\mathbf{E}(UV) - 1/4}{1/12} \\ &= \frac{\mathbf{E}(UV) - \mathbf{E}(U)\mathbf{E}(V)}{\sqrt{Var(U)Var(V)}} \\ &= \frac{\mathbf{Cov}(U, V)}{\sigma_U \sigma_V} = r(F(X), G(Y)). \end{aligned}$$

CQFD. ■

Le tableau 3.2 résume les trois principales mesures d'association et leur expressions en termes de copules.

	Mesure	Copule
tau de Kendall	$\tau := \frac{2}{n(n-1)} (n_c - n_d)$	$4 \int \int_{\mathbb{I}^2} C(u, v) dC(u, v) - 1$
rho de Spearman	$\rho := 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^n D_i^2$	$12 \int \int_{\mathbb{I}^2} uv dC(u, v) - 3$
phi de Pearson	$\phi := \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$	$12 \int \int_{\mathbb{I}^2} C(uv) dudv - 3$

TAB. 3.2 – Principales mesures d'association

Conclusion

Ce travail a porté sur l'étude des mesures d'association définies par les copules, un outil puissant permettant de modéliser la dépendance entre variables aléatoires au-delà des approches classiques.

Après un rappel des définitions et principales propriétés des vecteurs aléatoires, dans le premier chapitre, on a consacré le deuxième chapitre à l'étude des copules, en mettant en avant leur rôle essentiel dans la caractérisation de la structure de dépendance entre variables aléatoires. Quelques unes des familles de copules les plus couramment utilisées (Clayton, Gumbel, Frank,...) y sont présentées. Le troisième chapitre est dédié essentiellement à la présentation des trois mesures d'association les plus populaires, à savoir le coefficient de corrélation de Pearson, le rho de Spearman et le tau de Kendall. On a montré comment les copules permettent non seulement de redéfinir ces mesures, mais aussi d'améliorer leur performance dans des contextes non linéaires. Ce travail met en évidence la pertinence des copules dans l'analyse statistique moderne.

Enfin, il convient toutefois de signaler qu'il existe d'autres méthodes d'analyse de la dépendance qui n'ont pas été abordées dans ce mémoire et qui méritent d'être explorées pour faire l'objet d'un travail futur.

Bibliographie

- [1] Benelmir, I. (2018). Modélisation de la Dépendance par les Copules. Thèse de doctorat, Université Mohamed Khider, Biskra, Algérie.
- [2] Betteka, S. (2017). Les valeurs extrêmes bivariées. Thèse de doctorat, Université Mohamed Khider, Biskra, Algérie.
- [3] Bigot, J. (2014) Notes de cours de Probabilités. [https : //www.math.u – bordeaux.fr/~jbigot/Site/Enseignement_le/cours_proba_ISAE.pdf](https://www.math.u-bordeaux.fr/~jbigot/Site/Enseignement_le/cours_proba_ISAE.pdf).
- [4] Chine, A. (2017). Sur la statistique de copules. Thèse de doctorat, Université Mohamed Khider, Biskra, Algérie.
- [5] Christophe, J. (2006) Processus Gaussiens. Université de La Rochelle.
- [6] Cook, R.D., Johnson, M.E., (1981). A family of distributions for modelling nonelliptically symmetric multivariate data, Royal Statistical Society. 43, 210-218.
- [7] Dauxois, J.Y. (2011) Cours de probabilités. Université de Toulouse.
- [8] Femmam, K. (2023). Contribution on the estimation of the copulas parameters. Doctoral dissertation, Mohamed Khider University, Biskra, Algeria.
- [9] Genest, C. (1987). Frank's family of bivariate distributions. Biometrika, 74, 549-555.
- [10] Lecoutre, J.P. (2016). Statistique et probabilités : cours et exercices corrigés. Dunod.
- [11] Lehmann, E. L. (1966). Some concepts of dependence. The Annals of Mathematical Statistics, 1137-1153.
- [12] Lounas, F. (2011). Modélisation de la dépendance par les copules et applications. Mémoire de Master, Université Mouloud Mammeri, Tizi-Ouzou, Algérie.

- [13] Nelsen, R. B. (2006). An Introduction to copula. Springer Verlag, New York.
- [14] Roubi, A. (2012). Estimation des paramètres des lois de Pareto multivariées. Mémoire de magistère, Université Mohamed Khider, Biskra, Algérie.
- [15] Saporta, G. (2011). Probabilités, analyse des données et statistique. Technip, Paris.
- [16] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Univ. Paris. 8, 229-231.
- [17] Trivedi, P. K., & Zimmer, D. M. (2005). Copula modeling : An introduction for practitioners. Foundations and Trends®in Econometrics, 1(1), 1-111.
- [18] Vose, D. (2008). Risk analysis : a quantitative guide. John Wiley and Sons, England.
- [19] Yahia, D. (2021). Statistique Non Paramétrique. Support de cours de Master 2-Statistique. Département de Mathématiques, Université Mohamed Khider, Biskra.

ملخص

تتناول هذه المذكرة موضوع "مقاييس الارتباط المعرفة بواسطة الصلات (الروابط)" كأدوات إحصائية لدراسة الاقتران بين المتغيرات العشوائية. تتيح الروابط التعبير عن التوزيع المشترك بدلالة التوزيعات الهامشية. تم التطرق إلى مقاييس مختلفة مثل معاملات بيرسون، كندالوسبيرمان وتفسيرها باستخدام الروابط.

الكلمات المفتاحية: مقاييس الارتباط، توافق، الروابط، علاقة، معامل كيندال، معامل بيرسون، معامل سبيرمان، مبرهنة سكلار، أشعة عشوائية.

Abstract

This dissertation explores the use of copulas as statistical tools to model the dependence between random variables. Copulas allow the separation of the dependence structure from marginal distributions. Various association measures such as Pearson's phi, Kendall's tau and Spearman's rho are presented and interpreted through copulas.

Keywords: Association measures; Concordance; Copulas; Correlation; Kendall's Tau; Pearson's Phi; Random Vectors; Sklar's theorem; Spearman's Rho.

Résumé

Ce mémoire explore les copules comme outils statistiques pour modéliser la dépendance entre variables aléatoires. Les copules permettent de séparer la structure de dépendance des distributions marginales. Différentes mesures d'association telles le phi de Pearson, le tau de Kendall et le rho de Spearman sont présentées et interprétées via les copules.

Mots clés: Concordance; Corrélation; Copules; Mesures d'association; Phi de Pearson; Rho de Spearman; Tau de Kendall; Théorème de Sklar; Vecteurs aléatoires.