

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université A. MIRA-BEJAIA



Faculté des Sciences Exactes
Département de Mathématiques

Mémoire

Présenté par

BOUZEKRI Ikram

Pour l'obtention du diplôme de Masters

Filière : Mathématiques

Option : Probabilités Statistiques et Applications

Thème

Inférence Statistique des méthodes d'estimation par l'approche des copules

Dimanche le : 25/06/2023

Devant le Jury composé de :

Nom et Prénom

Grade

M. S. OUAZINE

M.C.B

Univ. de Béjaia

Président

M. O. SACI

M.C.B

Univ. de Béjaia

Rapporteur

M. M. BOURAINE

M.A.A

Univ. de Béjaia

Examineur

Année Universitaire : 2022/2023

- *Remerciements* -

Avant tout, je remercie Dieu le tout puissante de m'avoir donné santé, volonté, force et courage d'entamer et de continuer ce travail jusqu'à la fin.

Je tien à exprimer de tout cœur mon sincère remerciement à mon encadrant **Mr.SACI Oualid** pour avoir bien voulu me accompagner tout au long de la préparation de ce travail, pour les sages conseils et recommandations qui me est permis de réaliser ce modeste travail

Qu'elle trouve ici l'expression de mon profonde gratitude

Je remercie vivement monsieur **BOURAIN Mohand** d'avoir accepté de présider le jury et monsieur **OUAZINE Soufiane** d'examiner ce travail.

Un spécial remerciement à **BENAKLEF Nesrine**, qui est toujours à mes côtés

Mon sincère remerciement vont également à tous les enseignants du département d'Informatique et du département Mathématique de l'université **ABDERRAHMANE MIRA Bejaia**.

Ainsi qu'à ceux et celles qui par, leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé nos réflexions.

- *Dédicaces* -

Je dédie ce modeste travail :

A Mes chers parents

Vous représentez pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi

Aucune dédicace ne saurait être assez éloquente pour exprimer ce que vous méritez pour tous les sacrifices que vous n'avez cessé de me donner depuis ma naissance, Je vous dédie ce travail en témoignage de mon profond amour

Puisse Dieu, le tout puissant, vous préserver et vous accorder santé, longue vie, bonheur et santé

Mon cher frère **FAROUK** Mes chères sœurs **SOUHILA,ILHAM** et **MALAK**

En témoignage de l'attachement, de l'amour et de l'affection que je porte pour vous.

Les familles **BOUZEKRI, HAMMA**

une spéciale dédicace à ma sœur **KERKADENE LYDIA**, et à **IDIR FATEH** qui sont toujours à mes côtés

Mes très chers amis

Pour notre profonde amitié et agréables moments que nous avons vécus ensemble.

IKRAM

Table des Matières

Table des Matières	I
Liste des Figures	III
Liste des Tableaux	IV
Introduction générale	1
1 Préliminaires	3
1.1 La fonction de répartition	3
1.2 La fonction de répartition jointe	4
1.3 Définition des lois marginales	4
1.4 Notion d'indépendance	4
1.5 Fonction de répartition empirique	5
1.5.1 La fonction de répartition empirique univariée	5
1.5.2 La fonction de répartition empirique bidimensionnelle	6
2 Les copules	7
2.1 Copule bivariée	7
2.1.1 Théorème de Sklar	8
2.1.2 Propriétés de copules bivariées	11
2.1.3 Densité des copules bivariées	13
2.2 Copule multivariée	13
2.2.1 Théorème de Sklar	14
2.2.2 Propriétés de copules multivariées	14
2.2.3 Densité de copules multivariées	15
2.3 Les familles usuelles des copules	15
2.3.1 Les copules elliptiques	15
2.3.2 Les copules archimédiennes	18
2.3.3 Les copules Archimédiennes généralisées et hiérarchiques	21

2.3.4	Les copule de valeurs extrêmes	22
2.3.5	La copule Marchal-Olkin	24
2.3.6	Les copules empiriques	24
2.4	Mesures d'association bidimensionnelle	25
2.4.1	Les mesures de dépendance	25
3	Inférence statistique des copules	33
3.1	Méthodes d'estimations des copules	33
3.1.1	Estimation paramétrique des copules	33
3.1.2	Estimation semi-paramétrique des copules	35
3.1.3	Estimation non paramétrique	37
3.1.4	La fonction empirique	37
3.2	Les tests d'adéquation	38
3.2.1	Test du Chi-deux	38
3.2.2	Test de Cramér-von Mises	40
4	Simulation aléatoire des copules	42
4.1	Méthode de simulation	42
4.1.1	Méthode des distributions	42
4.1.2	Méthodes des distributions conditionnelles	43
4.1.3	Méthode analytique	44
4.2	Simulation	45
4.2.1	Simulation des copules elliptiques	45
4.2.2	Simulation des copules archimédiennes	47
4.3	Estimation des paramètres des copules	51
4.3.1	Estimation des paramètres des copules archimédiennes	51
4.3.2	Estimation des paramètres des copules elliptiques	52
4.4	Exemple d'application des copules pour la simulation des entrées dépendantes	54
	Bibliographie	60

Liste des Figures

2.1	Densités de copules Clayton (a) $\alpha = 2$, Gumbel (b) $\alpha = 2$ et Frank (c) $\alpha = 5.7369$	21
2.2	Copules Archimédiennes « fully nested » (à gauche) et « partially nested » (à droite).	22
2.3	Figures présentant en (a) la concordance et en (b) la discordance entre les deux points $A = (x_1, y_1)$ et $B = (x_1, y_2)$	26
4.1	Simulation de la copule gaussienne pour des différentes valeurs de ρ . . .	45
4.2	Figures présentant en (a) la densité de probabilité et en (b) la fonction distribution cumulative de la copule gaussienne.	46
4.3	Simulation de la copule de student pour des différentes valeurs de ρ . . .	47
4.4	Simulation de la copule de Frank pour des échantillons différents. . . .	48
4.5	Simulation de la copule de Gumbel pour des échantillons différents. . .	49
4.6	Simulation de la copule de Clayton pour des échantillons de tailles différentes et des valeurs de α différentes.	50
4.7	Histogramme représente les fréquences des données	56
4.8	Comparaison entre la cdf empirique et l'estimation cumulative de cdf lissé par noyau pour X_1	57
4.9	histogramme représente les réalisations de chaque entrée.	58

Liste des Tableaux

2.1	Quelque copules de valeurs extrêmes.	23
2.2	Exemples de mesures d'association bivariées	32
3.1	Tableau de contingence donnant les effectifs théoriques.	39
4.1	Estimation des paramètres de la copule gumble.	51
4.2	Estimation des paramètres de la copule clayton.	52
4.3	Estimation des paramètres de la copule frank.	52
4.4	Estimation des paramètres de la copule gaussienne.	53
4.5	Estimation des paramètres de la copule student.	53
4.6	Les données de deux action X_1 et X_2	55

Introduction générale

Dans la théorie des probabilités, lorsque les variables aléatoires sont indépendantes, la fonction de répartition jointe s'écrit comme le produit de ses marginales .

Dans la pratique, les variables aléatoires ne sont pas toujours indépendantes, comme par exemple les risques financiers, donc on ne peut pas négliger cette relation de dépendance. Pour cette raison il faut déterminer la loi jointe de ces variables. Les copules sont des fonctions qui permettent de coupler les lois marginales des variables afin d'obtenir la loi jointe, mesurer et modéliser complètement la dépendance.

Le terme copule vient du mot latin "copula", qui signifie au sens figuré, lien, alliance, liaison ou union. Cette notion apparaît sous d'autres appellations dans certains travaux de Fréchet[Fré51], Féron [Fér56] et Dall'Aglio [Dal56]. Mais elle est utilisée pour la première fois grâce à Sklar [Skl59] dans des études sur les espaces de probabilité métriques, ce dernier a élaboré un théorème fondamental dans la théorie des copules. Ce théorème permet la liaison entre la densité conjointe d'un vecteur aléatoire et ses densités marginales. La théorie des copules trouve des applications dans de nombreux domaines comme l'hydrologie, les sciences actuarielles et en finance. Par souci de simplicité et du fait que la théorie multivariée est une extension du cas bivariée on se limite à la théorie bivariée des copules.

L'objectif de ce travail consiste à introduire les différentes familles des copules et leurs propriétés, ainsi, de présenter les différentes méthodes d'estimation. Fournir ensuite des tests d'adéquation pour le choix de la bonne copule. Puis donner les différentes méthodes de simulation.

Ce mémoire est composé de quatre chapitres, le premier chapitre est un rappel mathématique. On y retrouve les principaux théorèmes qui serviront de base aux chapitres suivants.

Le deuxième chapitre est consacré à la théorie des copules dans le cas bivarié et multivarié. Il débute par une définition des copules et leurs propriétés fondamentales, notamment le théorème de Sklar. Ensuite on va donner les familles des copules les plus utilisées et aussi les copules associées à une copule. En fin, on va discuter sur la

structure de dépendance tel que le coefficient de corrélation linéaire, le tau de Kendall et le rho de Spearman.

Le troisième chapitre contient la partie d'estimation des paramètres de ces copules. L'objectif principal de la première section est de fournir les différentes méthodes d'estimation des copules à savoir l'approche paramétrique, semi-paramétrique et non-paramétrique. Où la première approche consiste à estimer le paramètre de la copule par l'une des méthodes suivantes :

1. Maximum de vraisemblance (MLE).
2. Inférence des marginales (IFM).

Pour la seconde approche on utilise la méthode du Maximum de vraisemblance canonique (CML) et la méthode des moments.

Quant à la troisième approche on utilise la fonction empirique. Tandis que, dans la deuxième section, on propose une méthode de sélection de la meilleure copule basée sur les tests d'adéquation à savoir : la statistique de chi-deux et le test de Cramér-von Mises.

Dans le quatrième chapitre, on s'intéresse aux différentes méthodes de simulation des copules. On va simuler quelques copules avec ses méthodes, et on estime les paramètres de quelques copules en calculant leur biais moyen et leur variance. Après on va donner un exemple d'application sur des données réelles. L'application est faite sur **MATLAB**.

Ce travail s'achève par une conclusion générale.

Chapitre 1

Préliminaires

1.1 La fonction de répartition

Soit X une variable aléatoire (v.a) à valeur dans \mathbb{R} . La fonction de répartition de X notée par F est :

$$F : \mathbb{R} \mapsto [0, 1]$$
$$x \mapsto F(x) = \mathbb{P}(X \leq x)$$

Remarques :

– F est une fonction (fct) continue et croissante avec :

$$\lim_{x \rightarrow -\infty} F(x) = 0;$$
$$\lim_{x \rightarrow +\infty} F(x) = 1.$$

– **la fonction inverse** de F noté F^{-1} (si elle existe) est :

$$F^{-1}(t) = \inf \{x \in \mathbb{R} : F(x) \geq t\} : \forall t \in [0, 1] \quad (1.1)$$

– Si F est continue alors x est une variable aléatoire uniforme sur l'intervalle $[0, 1]$. On note cet intervalle par \mathbf{I} .

Proposition 1 Soit X une v.a de fonction de répartition F . (voir [FM53])

- ★ Si U est uniforme dans $[0, 1]$ alors $F^{-1}(U) \xrightarrow{d} F$.
- ★ Si F est continue, alors $F(X) \xrightarrow{d} U_{[0,1]}$.

1.2 La fonction de répartition jointe

Soit (X, Y) un vecteur aléatoire à valeur dans \mathbb{R}^2 . La loi jointe du couple (X, Y) notée H est :

$$H_{(X, Y)} = \mathbb{P}(X \leq x, Y \leq y). \quad (1.2)$$

On peut noter H par $F_{(X, Y)}$.

1.3 Définition des lois marginales

On appelle lois marginales de X et Y les fonctions de répartition $F_{(X)}$ et $F_{(Y)}$ respectivement telles que :

$$\begin{cases} F_{(X)} = \lim_{y \rightarrow +\infty} H(x, y). \\ F_{(Y)} = \lim_{x \rightarrow +\infty} H(x, y). \end{cases}$$

1.4 Notion d'indépendance

Définition 1 Deux variables aléatoires X et Y sont indépendantes si pour tout intervalle A et B dans \mathbb{R} on a :

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Proposition 2 Soit X et Y deux v.a indépendantes [Mau85]

– La fonction de répartition vérifie

$$F_{X, Y}(x, y) = F_X(x)F_Y(y), \forall (x, y) \in \mathbb{R}^2.$$

– La transformée de Laplace vérifie pour tout (a, b)

$$L_{X, Y}(a, b) = L_X(a)L_Y(b) \quad \text{avec} \quad L_{X, Y}(a, b) = E[e^{aX+bY}].$$

– Dans le cas discret

$$P(X = x, Y = y) = P(X = x)P(Y = y), \forall (x, y) \in \mathbb{R}^2.$$

– Dans le cas continu, la densité du couple (X, Y) vérifie

$$f_{X, Y}(x, y) = f_X(x)f_Y(y), \forall (x, y) \in \mathbb{R}^2$$

Définition 2 Les variables aléatoires X_1, \dots, X_n sont **mutuellement indépendantes** si pour tous les intervalles A_1, \dots, A_n de \mathbb{R} on a :

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i).$$

Si les variables aléatoires X_1, \dots, X_n sont mutuellement indépendantes, alors elles sont indépendantes deux à deux .

Attention La réciproque est fausse !

1.5 Fonction de répartition empirique

1.5.1 La fonction de répartition empirique univariée

Soit X_1, \dots, X_n un échantillon aléatoire d'une loi F . La fonction de répartition empirique associée à cet échantillon est alors définie, en tout $x \in \mathbb{R}$, par :

$$F_n = \frac{1}{n} \sum_{i=1}^n 1(X_i \leq x), \quad (1.3)$$

où, pour chaque $i \in 1, \dots, n$, la fonction indicatrice est définie par :

$$1(X_i \leq x) = \begin{cases} 1 & \text{si } X_i \leq x, \\ 0 & \text{sinon.} \end{cases}$$

Remarque 1

- $F_n(x)$ correspond à la proportion des observations qui sont inférieures ou égales à x , pour un $x \in \mathbb{R}$ fixé.
- La variable aléatoire $nF_n(x)$ est donc une variable binomiale de paramètres n et $F(x)$.

Propriétés de La fonction répartition empirique univariée

$$\mathbb{E} \{F_n(x)\} = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \{1_{]-\infty, x]}(X_i)\} = \frac{1}{n} \sum_{i=1}^n Pr(X_i \leq x) = F(x),$$

et

$$\begin{aligned} \text{Var} \{F_n(x)\} &= \frac{1}{n^2} \sum_{i=1}^n \text{Var} \{1_{]-\infty, x]}(X_i)\} \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Pr}(X_i \leq x) \{1 - \text{Pr}(X_i \leq x)\} \\ &= \frac{F(x) \{1 - F(x)\}}{n}. \end{aligned}$$

Puisque F_n est une estimation sans biais de F et que $\text{Var} F_n \rightarrow 0$ quand $n \rightarrow \infty$, il est clair que $F_n \rightarrow F$ en probabilité. Autrement dit, F_n est une estimation convergente de F . De fait, la convergence est uniforme.

1.5.2 La fonction de répartition empirique bidimensionnelle

Soit un échantillon aléatoire $(X_1, Y_1), \dots, (X_n, Y_n)$ tiré d'une loi bivariée dont la fonction de répartition est H . On obtient une estimation de H par la fonction de répartition empirique bidimensionnelle définie en tous $(x, y) \in \mathbb{R}^2$, par :

$$H_n(x, y) = \frac{1}{n} \sum_{i=1}^n 1(X_i \leq x, Y_i \leq y). \quad (1.4)$$

Lorsque les variables x et y sont fixées, $nH_n(x, y)$ est une variable binomiale de paramètres n et $H(x, y)$.

Propriétés de La fonction répartition empirique bidimensionnelle

$$\mathbb{E} \{H_n(x, y)\} = H(x, y),$$

et

$$\text{Var} \{H_n(x, y)\} = \frac{H(x, y) \{1 - H(x, y)\}}{n}.$$

Pour $(x, y) \in \mathbb{R}^2$ fixé et $\sigma_{x,y}^2 = H(x, y) \{1 - H(x, y)\}$, le théorème de la centrale limite permet aussi d'affirmer que quand $n \rightarrow \infty$, on a :

$$\mathbb{H}_n(x, y) = \sqrt{n} \{H_n(x, y) - H(x, y)\} \rightsquigarrow \mathcal{N}(0, \sigma_{x,y}^2). \quad (1.5)$$

Chapitre 2

Les copules

Les copules permettent d'étudier la dépendance entre plusieurs variables aléatoires, avec l'idée que cette dépendance ne doit pas contenir d'information provenant des lois marginales des variables elles-mêmes.

2.1 Copule bivariée

Dans toute la suite I désigne l'intervalle $[0, 1]$.

Définition 3 Une copule 2-dimensionnelle est une fonction C de $I^2 = [0, 1]^2$ dans I ayant les propriétés suivantes :

1. La copule C est attachée, c'est à dire :

$$C(u, 0) = C(0, v) = 0, \forall (u, v) \in I^2. \quad (2.1)$$

2. Les marges sont uniformes, c'est à dire :

$$C(u, 1) = u \text{ et } C(1, v) = v, \forall (u, v) \in I^2. \quad (2.2)$$

3. La copule C est 2-croissante c'est à dire $\forall (u_1, u_2), (v_1, v_2) \in I^2$ telle que $u_1 \leq u_2$ et $v_1 \leq v_2$.

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

Cette définition signifie que C est une distribution de marginale uniforme.

Exemples

- **Copule produit** : La copule produit, notée par Π , est définie par :

$$C(u, v) = \Pi(u, v) = uv, \forall (u, v) \in I^2. \quad (2.4)$$

On observe que cette copule vérifie les propriétés $\boxed{(2.1)}$ et $\boxed{(2.2)}$, telle que :

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

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

Pour la propriété $\boxed{(2.3)}$, on a :

$$\begin{aligned} C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) &= \Pi(u_2, v_2) - \Pi(u_2, v_1) - \Pi(u_1, v_2) + \Pi(u_1, v_1) \\ &= u_2v_2 - u_2v_1 - u_1v_2 + u_1v_1 \geq 0, \end{aligned}$$

où

$$\text{si } u_2 \geq u_1 \Rightarrow \begin{cases} u_2v_2 & \geq u_1v_2, \\ u_2v_1 & \geq u_1v_1. \end{cases}$$

Alors

$$u_2v_2 - u_2v_1 \geq u_1v_2 - u_1v_1 \Rightarrow u_2v_2 - u_2v_1 - u_1v_2 + u_1v_1 \geq 0.$$

– **Copule min** : La copule min, notée par M, est définie par :

$$M(u, v) = \min(u, v), \forall (u, v) \in I^2. \quad (2.5)$$

– **Copule max** : La copule max, notée par W, est définie par :

$$W(u, v) = \max(u + v - 1, 0), \forall (u, v) \in I^2. \quad (2.6)$$

2.1.1 Théorème de Sklar

Ce théorème représente la base de la théorie des copules, car il valide l'existence de la copule C qui permet de relier les lois marginales de deux v.a pour obtenir la distribution jointe. Ce théorème a été développé par **Sklar (1959)** [Sk159] et porte son nom. La relation entre la distribution jointe, les fonctions de distribution marginales et la copule C est montré dans le théorème suivant.

Théorème 1 Soient X, Y deux v.a de fonction de répartition jointe H et des marginales F_X et F_Y , alors il existe une copule C telle que pour tout $(x, y) \in \mathbb{R}^2$:

$$H(x, y) = C(F_X(x), F_Y(y)). \quad (2.7)$$

Si F_X et F_Y sont continues, alors C est unique.

Inversement, si C est une copule et F_X, F_Y sont des fonctions de répartitions, alors la fonction H définie dans (2.7) est une fonction de répartition jointe dont les marginales sont F_X et F_Y [Nel07],[Vos08].

Preuve. A l'aide de (1.2) et la proposition 1, nous pouvons montrer (2.7) comme suit :

$$\begin{aligned} H(x, y) &= \mathbb{P}(x \leq X, Y \leq y) \\ &= \mathbb{P}(F_X^{-1}(U_1) \leq x, F_Y^{-1}(U_2) \leq y) \\ &= \mathbb{P}((U_1 \leq F(x), U_2 \leq F(y))) \\ &= C(F(x), F(y)). \end{aligned}$$

Ce qui nous donne la relation entre la loi jointe, les lois marginales et les copules. A partir du **théorème de Sklar** on peut extraire une définition de copules dépendant de la notion de variables aléatoires comme suit :

Soient F_X, F_Y les fonctions de répartition marginales de deux v.a X et Y respectivement.

Corollaire 1 Soit H une fonction de répartition 2-dimensionnelle de fonctions marginales $F_{(X)}$ et $F_{(Y)}$: Alors la copule C associée à H est donnée par :

$$C(u, v) = H(F_X^{-1}(u), F_Y^{-1}(v)) \text{ pour tout } (u, v) \in I^2 \quad (2.8)$$

Preuve : En appliquant le **théorème 1**, en posant $u = F_X(x)$ et $v = F_Y(y)$, on trouve que :

$$\begin{aligned} C(u, v) &= \mathbb{P}(U \leq u, V \leq v) \\ &= \mathbb{P}(F_X(X) \leq u, F_Y(Y) \leq v) \\ &= \mathbb{P}(X \leq F_X^{-1}(u), Y \leq F_Y^{-1}(v)) \\ &= H(F_X^{-1}(u), F_Y^{-1}(v)). \end{aligned}$$

Exemple 1

Soit (X, Y) un couple de v.a dont la fonction jointe H est définie par :

$$H(x, y) = \exp \left\{ -(x^{-\theta} + y^{-\theta})^{-\frac{1}{\theta}} \right\}, \forall (x, y) \in]0, \infty[\times]0, \infty[, \theta > 0,$$

et admet pour marginales

$$F_X(x) = \exp(-x), F_Y(y) = \exp(-y),$$

nous avons

$$F_X^{-1}(u) = -\ln u, F_Y^{-1}(v) = -\ln v, \forall (u, v) \in I^2,$$

d'après le **corollaire 1** la copule associée est exprimée comme suit :

$$C(u, v) = \exp \left\{ (-\ln u^{-\theta}) + (-\ln v^{-\theta}) \right\}^{-\frac{1}{\theta}}, \forall \theta \geq 1.$$

Exemple 2

Soit H la fonction jointe de vecteur aléatoire (X, Y) est définie par :

$$H_\theta(x, y) = \begin{cases} 1 - \exp(-x) - \exp(-y) + \exp(-(x + y + \theta xy)), & x \geq 0, y \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Avec $\theta \in [0, 1]$ et admet pour marginales :

$$F_X(x) = \exp(-(1 - x)), F_Y(y) = \exp(-(1 - y)),$$

nous avons :

$$F_X^{-1}(u) = -\log(1 - u), F_Y^{-1}(v) = -\log(1 - v),$$

la copule correspondante est donc :

$$C_\theta(u, v) = u + v - 1 + (1 - u)(1 - v)e^{-\theta \ln(1-u) \ln(1-v)}.$$

2.1.1.1 Les bornes de Fréchet-Hoeffding

Toute copule bivariée C est bornée par deux copules. Ces bornes sont déterminées dans le théorèmes suivant :

Théorème 2 Soit C une copule, alors :

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq M(u, v) = \min(u, v), \forall (u, v) \in I, \quad (2.9)$$

Preuve.

On a :

$$\begin{cases} C(u, v) \leq C(u, 1) & = u \\ C(u, v) \leq C(1, v) & = v \end{cases}$$

Ceci implique que :

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

D'autre part :

$$\forall (u, v) \in I^2 : C(u, v) \geq 0,$$

et de $\boxed{(2.3)}$ on pose : $u_1 = u, u_2 = 1, v_1 = v$ et $v_2 = 1$ on a :

$$\begin{aligned} C(1, 1) - C(1, v) - C(u, 1) + C(u, v) &\geq 0 \\ \implies 1 - v - u + C(u, v) &\geq 0 \\ \implies C(u, v) &\geq u + v - 1 \end{aligned}$$

donc :

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

alors :

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

telle que W et M représentent les bornes inférieures et supérieures de **Fréchet-Hoeffding** (respectivement).

Comme une conséquence du théorème de Sklar, si X, Y sont des variables aléatoires dont la fonction de répartition jointe est H et des marginales F_X et F_Y , respectivement, alors :

$$\max(F_X(x) + F_Y(y) - 1, 0) \leq H(x, y) \leq \min(F_X(x), F_Y(y)), \forall x, y \in \mathbb{R}. \quad (2.10)$$

Car M et W sont des copules, les bornes ci-dessus sont des fonctions de répartition jointes et sont appelées les bornes de **Fréchet-Hoeffding** pour une fonction de répartition jointe H et des marginales F_X et F_Y .

2.1.2 Propriétés de copules bivariées

Soient X, Y deux v.a continues, de fonction de répartition jointe H et dont les marginales sont F_X et F_Y , et soit C la copule associée :

1. **Symétrie** : On dit que $X;Y$ sont échangeable si et seulement si $F_X = F_Y$ et $C(u, v) = C(v, u)$ pour tout $(u, v) \in I^2$. Si $C(u, v) = C(v, u)$ pour tout $(u, v) \in I^2$, on dit que C est **symétrique**.
2. **Théorème d'invariance** : Ce théorème est essentiel à la théorie des copules, de façon particulière dans les notions de mesure d'association.

Théorème 3 Soient deux variables aléatoires continues X et Y des marges $F_{(X)}$ et $F_{(Y)}$ et de copule $C_{X,Y}$. Si α et β sont deux fonctions strictement croissantes, alors :

$$C_{\alpha(X),\beta(Y)} = C_{X,Y}. \quad (2.11)$$

La copule $C_{X,Y}$ est invariante par transformations strictement croissantes des variables aléatoires.

Preuve. On peut démontrer ce théorème facilement à l'aide des lois de probabilité, comme suit :

Soient $F_{X_1}, F_{Y_1}, F_{X_2}, F_{Y_2}$ les fonctions de répartition de $X, Y, \alpha(X)$ et $\beta(Y)$, respectivement. Comme α et β sont strictement croissantes, alors :

$$F_{X_2} = \mathbb{P}(\alpha(X) \leq x) = \mathbb{P}(X \leq \alpha^{-1}(x)) = F_{X_1}(\alpha^{-1}(x)),$$

aussi

$$F_{Y_2} = \mathbb{P}(\beta(Y) \leq y) = \mathbb{P}(Y \leq \beta^{-1}(y)) = F_{Y_1}(\beta^{-1}(y)).$$

Alors :

$$\begin{aligned} C_{\alpha(X),\beta(Y)}(F_{X_2}(x), F_{Y_2}(y)) &= \mathbb{P}(\alpha(X) \leq x, \beta(Y) \leq y) \\ &= \mathbb{P}(X \leq \alpha^{-1}(x), Y \leq \beta^{-1}(y)) \\ &= C_{X,Y}(F_{X_1}(\alpha^{-1}(x)), F_{Y_1}(\beta^{-1}(y))) \\ &= C_{X,Y}(F_{X_2}(x), F_{Y_2}(y)). \end{aligned}$$

3. **Ordre** : Soit C_1, C_2 deux copules. On dit que C_1 est plus petite que C_2 ou C_2 est plus grande que C_1 et on note $C_1 < C_2$ si :

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

Comme exemple, on a la copule $W = \max(u + v - 1, 0)$ est la plus petite copule et $M = \min(u, v)$ est la plus grande copule.

4. **Dérivées partielles** : Les dérivées partielles de $C(u, v)$ existent presque sûrement pour tout $(u, v) \in I^2$, et vérifient :

$$0 \leq \frac{\partial C(u, v)}{\partial u} \leq 1 \quad \text{et} \quad 0 \leq \frac{\partial C(u, v)}{\partial v} \leq 1.$$

2.1.3 Densité des copules bivariées

Soient X, Y deux variables aléatoires continues. Soit h la fonction de densité jointe de H , f_X et f_Y sont les fonctions de densités marginales de X, Y respectivement.

Définition 4 La densité $c(F_X(x), F_Y(y))$ associée à la copule $C(F_X(x), F_Y(y))$ est définie par :

$$\begin{aligned} c(F_X(x), F_Y(y)) &= \frac{\partial^2 C(F_X(x), F_Y(y))}{\partial F_X(x) \partial F_Y(y)} \\ &= \frac{h(x, y)}{f_X(x) f_Y(y)}. \end{aligned}$$

D'après le **théorème de Sklar** (Théorème 1), on donne la représentation canonique suivante :

$$h(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y) \tag{2.12}$$

2.2 Copule multivariée

Cette section développe les résultats de la section précédente au cas multidimensionnel ($d > 2$). Donc de la même manière, nous présentons les copules multivariées : définitions, théorèmes, propriétés. Nous notons $I^d = [0, 1]^d$.

Définition 5 Une copule d -dimensionnelle est une fonction C de I^d dans I ayant les propriétés suivantes :

1. $\forall u \in I^d$

$$C(u) = 0 \quad \text{pour au moins une coordonnée de } u = 0, \tag{2.13}$$

2. pour toutes les coordonnées égales 1 sauf u_i , alors :

$$C(u) = u_i. \quad (2.14)$$

3. C est d -croissante, c'est-à-dire $\forall u = (u_1, \dots, u_d), v = (v_1, \dots, v_d)$ dans I^d tels que $u_i \leq v_i$, pour $i = 1, \dots, d$ on a :

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} \times C(x_{1i_1}, \dots, x_{di_d}) \leq 0, \quad (2.15)$$

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

2.2.1 Théorème de Sklar

Théorème 4 Soit H une fonction de répartition d -dimensionnelle de fonctions de répartition marginales F_1, \dots, F_d . Alors il existe une d -copule C telle que pour tout $x \in \mathbb{R}^d$,

$$H(x) = H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \quad (2.16)$$

Si les fonctions F_1, \dots, F_d sont continues, alors C est unique.

Inversement, si C est une d -copule et si F_1, \dots, F_d sont des fonctions de répartition univariées, alors la fonction H définie dans (2.16) est une fonction de répartition d -dimensionnelle de marginales F_1, \dots, F_d .

Corollaire 2 Soit H une fonction de répartition d -dimensionnelle de fonctions de répartition marginales F_1, \dots, F_d . Alors, la copule C associée à H est donnée par :

$$C(u_1, \dots, u_d) = H(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \text{ pour tout } u \in I^d. \quad (2.17)$$

2.2.2 Propriétés de copules multivariées

1. **Continuité uniforme** : Une copule C est uniformément continue sur son domaine. En particulier, pour tout u, v dans I^d , nous avons :

$$|C(u) - C(v)| \leq \sum_{i=1}^n |v_i - u_i| \quad (2.18)$$

2. **Invariance** : Soient (X_1, \dots, X_d) un vecteur de variables aléatoires continues, de fonction de répartition F associée à une copule C et $(\alpha_1, \alpha_2, \dots, \alpha_d)$ une suite de fonctions strictement croissantes. Alors, la fonction de répartition jointe du vecteur aléatoire $(\alpha_1(X_1), \dots, \alpha_d(X_d))$ est aussi associée à la même copule C .

$$C_{\alpha_1(X_1), \dots, \alpha_d(X_d)} = C_{X_1, \dots, X_d}(u)$$

3. **Dérivées partielles** : Soit C une copule. Les dérivées partielles de C existent presque sûrement, pour tout $i = 1, \dots, d$ et pour tout $u \in I^d$, et on a :

$$0 \leq \frac{\partial C(u)}{\partial u_i} \leq 1. \quad (2.19)$$

2.2.3 Densité de copules multivariées

La densité c associée à la copule C est définie par :

$$\begin{aligned} c(F_1(x_1), \dots, F_d(x_d)) &= \frac{\partial^d C(F_1(x_1), \dots, F_d(x_d))}{\partial F_1(x_1), \dots, \partial F_d(x_d)} \\ &= \frac{h(x_1, \dots, x_d)}{f_1(x_1), \dots, f_d(x_d)} \\ &= \frac{h(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{\prod_{i=1}^d f_i(F_i^{-1}(u_i))} \end{aligned} \tag{2.20}$$

telle que h est la densité de H et f_i est la densité de F_i .

Si la fonction de répartition multivariée H est absolument continue et en utilisant le théorème de Sklar, on peut présenter la fonction de densité en fonction de la densité c et ses fonctions de répartition marginales par :

$$h(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i). \tag{2.21}$$

2.3 Les familles usuelles des copules

Il existe une grande variété de copules, elles se divisent en plusieurs principales catégories : les copules elliptiques, archimax, Archimédiennes, copules des valeurs extrêmes et les copules empiriques.

2.3.1 Les copules elliptiques

Les copules elliptiques sont définies à partir des lois de distribution elliptique qui sont symétriques et donc faciles à simuler. Les deux classes les plus utilisées des copules elliptiques sont la copule normale et la copule de Student.

Définition 6 *Un vecteur aléatoire $X = (X_1, X_2)$ est de distribution elliptique s'il admet la représentation suivante :*

$$X = \mu + RAU.$$

Où :

- $\mu = (\mu_1, \mu_2) \in \mathbb{R}^2$.
- U est un vecteur aléatoire uniforme sur la sphère unité de \mathbb{R}^2 .
- R est un vecteur aléatoire uniforme sur la sphère unité de U .
- A est une matrice de dimension $n \times n$ telle que $\Sigma = AA^t$ est non singulière.

Définition 7 La fonction de densité d'une distribution elliptique (si elle existe) est donnée par :

$$f(x) = |\Sigma|^{-\frac{1}{2}} g((X - \mu)^t (X - \mu)),$$

où g est une fonction définie de \mathbb{R}^+ dans \mathbb{R} , dite génératrice de densité, uniquement déterminée par la fonction de distribution de R .

Exemple : Soit la fonction suivante :

$$g(t) = k \exp(-t/2)$$

génère la fonction de distribution normale multivariée et :

$$g(t) = k(1 + t/m)^{(n+m)/2}$$

génère la fonction de distribution de student multivariée, avec k une constante de normalisation et $m \in \mathbb{N}$.

Définition 8 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de distribution elliptique. La fonction de caractéristique $\phi_X(t), t \in \mathbb{R}$ est donnée par :

$$\begin{aligned} \phi_X(t) &= \mathbb{E} \left(\exp(it^T \mathbf{X}) \right) \\ &= \mathbb{E} \left(\exp(it^T \mu + \mathbf{R} \mathbf{A} U) \right) \\ &= \mathbb{E} \left(\exp(it^T \mu) g(t^T \Sigma t) \right) \end{aligned}$$

Notons par $X \sim E_n(\mu, \Sigma, g)$ la classe de la distribution elliptique de vecteur moyen, de matrice de covariance $\Sigma = (\sigma_{ij})$ et de générateur caractéristique g .

Définition 9 On appelle copule elliptique toute copule de la forme suivante :

$$C_\theta(u, v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \left(-\frac{s^2 - 2\theta st + s^2}{2(1-\theta^2)} \right) ds dt, \forall (u, v) \in I^2.$$

Où $\theta \in [-1, 1]$ et ϕ^{-1} est l'inverse de distribution Gaussienne univariée.

2.3.1.1 La copule gaussienne

La copule gaussienne de deux variables aléatoires X, Y est définie par :

$$C_r^\phi(u, v) = \phi_r(\phi^{-1}(u), \phi^{-1}(v)), \quad (2.22)$$

telle que ϕ_r est la fonction de répartition jointe de la loi normale bivariée, de coefficient de corrélation linéaire $r \in [-1, 1]$ et ϕ^{-1} est la fonction quantile de la loi normale standard, par suite :

$$\phi_r(\phi^{-1}(u), \phi^{-1}(v)) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-r_{X,Y}^2}} \exp\left(\frac{2r_{X,Y}st - s^2 - t^2}{1-r_{X,Y}^2}\right) dsdt.$$

Cette copule est paramétrée par le coefficient de corrélation linéaire r .

Remarque 2

$-C_r^\phi(u, v) \rightarrow W(u, v)$ quand $r \rightarrow -1$ et $-C_r^\phi(u, v) \rightarrow M(u, v)$ quand $r \rightarrow 1$.

2.3.1.2 La copule student

La copule de Student est une copule paramétrique, paramétrée par le coefficient de corrélation linéaire r et le degré de liberté v . Cette copule est définie par :

$$C_{r,v}^t(u, v) = t_{r,v}(t_v^{-1}(u), t_v^{-1}(v)) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi\sqrt{1-r^2}} \left(1 + \frac{s^2 + t^2 - 2rst}{v(1-r^2)}\right)^{-\frac{v+2}{2}} dsdt.$$

La densité est définie par :

$$c_{r,v}^t(u, v) = r^{\frac{-1}{2}} \frac{\Gamma\left(\frac{v+2}{2}\right) \Gamma\left(\frac{v}{2}\right)}{\Gamma\left(\frac{v+1}{2}\right)^2} \left(\frac{\left(\frac{1 + \varsigma_1^2 + \varsigma_2^2 - 2r\varsigma_1\varsigma_2}{v(1-r^2)}\right)^{-\left(\frac{v+2}{2}\right)}}{\prod_{j=1}^2 (1 + \varsigma_j^2)^{-\left(\frac{v+2}{2}\right)}} \right),$$

où $\varsigma_1 = t_v^{-1}(u)$, $\varsigma_2 = t_v^{-1}(v)$.

Remarque :

– Si le degré de liberté $v \rightarrow \infty$, alors la copule de student converge vers la copule Gaussienne [MS06].

2.3.2 Les copules archimédiennes

Cette famille de copule a été nommée par Ling (1965) [Lin65], mais elle a été recon- nue par Schweizer et Sklar (1961) [Sch61] dans l'étude de t-norme. Avant d'être intro- duit dans la finance et aussi dans d'autres domaines [GM86a]. Ces copules trouvent une large gamme d'applications pour plusieurs raisons :

1. La facilité avec laquelle elles peuvent être construites .
2. La grande variété de familles de copules qui appartiennent à cette famille .
3. Les nombreuses propriétés possédées par les membres de cette famille.

Mais avant de définir cette famille de copules, on a besoin de présenter quelque défi- nitions, et propriétés utiles pour définir les copules archimédiennes.

Définition 10 *La copule Archimédienne, noté $C^A(u, v)$, est définie par :*

$$C^A(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)), \quad (2.23)$$

telle que $\varphi \mapsto \mathbb{R}^{+}$, est une fonction continue, décroissante et convexe, où $\varphi(1) = 0$, la fonction φ est dite générateur, et φ^{-1} est le pseudo-inverse de φ définie comme suit :*

$$\varphi^{-1}(u) = \begin{cases} \varphi^{-1}(u) & \text{si } 0 \leq u \leq \varphi(0) \\ 0 & \text{si } \varphi(0) \leq u \leq +\infty. \end{cases} \quad (2.24)$$

Propriétés de cette famille : la famille de copules Archimidiennes a les propriétés suivantes :

a. **La symétrie :**

$$C^A(u, v) = C^A(v, u), \forall (u, v) \in I^2. \quad (2.25)$$

b. **L'associativité :**

$$C^A(C^A(u, v), z) = C^A(u, C^A(v, z)), \forall (u, v, z) \in I^3. \quad (2.26)$$

c. **Contour convexe :**

$$(u, v) \in I^2 : \varphi(u) + \varphi(v) = \varphi(k), k > 0. \quad (2.27)$$

d. **La densité :**

$$c^A(u, v) = \frac{-\varphi''(C(u, v))\varphi'(u)\varphi'(v)}{(\varphi'(C(u, v)))^3}. \quad (2.28)$$

e. **Si c est une constante strictement positive, alors $c\varphi$ est un générateur de la copule C^A .**

Exemple :

1. On prend la fonction $\varphi(t) = -\ln(t)$ comme générateur et le pseudo inverse $\varphi^{-1}(t) = \exp(-t)$ et on construit la copule C en utilisant [\(2.23\)](#) comme suit :

$$C^A(u, v) = \exp(-[(-\ln(u)) + (-\ln(v))]) = uv = \prod(u, v),$$

qui nous donne la copule d'indépendance \prod alors on déduit que la copule \prod est une copule Archimédienne.

2. Soit $\varphi(t) = 1 - t$ pour $t \in [0, 1]$. Donc $\varphi^{-1}(t) = 1 - t$ pour $t \in [0, 1]$ et 0 pour $t > 1$, alors $\varphi^{-1}(t) = \max(1 - t, 0)$. D'après [\(2.23\)](#), on a :

$$C^A(u, v) = \max(u + v - 1, 0) = W(u, v),$$

donc la copule $W(u, v)$ est Archimédienne.

Théorème 5 : Soit C une copule Archimédienne générée par φ . Soit $K_c(t)$ la C -mesure de l'ensemble $(u, v) \in I^2$ $C(u, v) \leq t$, alors :

$$K_c(t) = t - \frac{\varphi(t)}{\varphi^{-1}(t)}, \forall t \in I. \tag{2.29}$$

Corollaire 3 : Soient U_1 et U_2 deux v.a uniformes sur $[0, 1]$ dont la fonction de répartition jointe est la copule Archimédienne C générée par φ , alors la fonction K_c donnée par [\(2.29\)](#) est la fonction de répartition de la v.a $C(U_1, U_2)$.

Exemples de copules Archimédiennes

Les copules les plus utilisés de cette famille sont les copule de Gumbel , de Clayton et de Frank, définies dans ce qui suit .

1. **Copule de Gumbel (1960) :** C'est une copule asymétrique. Elle est définie par :

$$C_\alpha^{Gum}(u, v) = \exp -(-\ln u)^\alpha + (-\ln v)^\alpha \frac{1}{\alpha}, \tag{2.30}$$

et dont le générateur est :

$$\varphi_\alpha(t) = (-\ln t)^\alpha,$$

où le paramètre de dépendance $\alpha \in [1, \infty[$ [\[Gum60\]](#),[\[Hut90\]](#) et [\[Vos08\]](#). Nous avons :

- Si $\alpha \rightarrow 1$ alors $C_\alpha^{Gum} \rightarrow \Pi$.
- Si $\alpha \rightarrow \infty$ alors $C_\alpha^{Gum} \rightarrow M$.

2. **Copule de Clayton (1978)** : La copule de Clayton s'appelle aussi la copule de cook et Johnson [CJ81], et d'abord étudiée par Kimeldorf et sampson [KS75]. Elle est définie par :

$$C_\alpha^{cla}(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{\frac{-1}{\alpha}}, \quad (2.31)$$

où le générateur et le pseudo inverse sont :

$$\varphi_\alpha(t) = \frac{1}{\alpha}(t^{-\alpha} - 1).$$

$$\varphi_\alpha^{-1} = (t + 1)^{\frac{-1}{\alpha}},$$

avec le paramètre de dépendance $\alpha \in [-1, 0[\cup]0, \infty[$.

- Si le paramètre $\alpha \rightarrow 0$ alors les marginales sont indépendantes.
- Si $\alpha \rightarrow \infty$ alors on obtient la copule de borne supérieur de Fréchet-Hoeffding M .

3. **Copule de Frank (1979)** : Cette copule est symétrique de paramètre de dépendance $\alpha \in]-\infty, 0[\cup]0, +\infty[$ [Gen87]. Elle est définie par :

$$C_\alpha^{Fr}(u, v) = \frac{-1}{\alpha} \ln \left(1 + \frac{(e^{-u\alpha} - 1)(e^{-v\alpha} - 1)}{e^{-\alpha} - 1} \right), \quad (2.32)$$

de générateur et de densité :

$$\varphi_\alpha(t) = -\ln \left[\frac{\exp(-\alpha t) - 1}{\exp(-\alpha) - 1} \right].$$

$$c^{Fr}(u, v) = \frac{(\alpha - 1) \ln \alpha^{u+v}}{(\alpha + 1 + (\alpha^u - 1)(\alpha^v - 1))^2}.$$

- Si $\alpha \rightarrow 0$ alors $C_\alpha^{Fr} \rightarrow \Pi$.
- Si $\alpha \rightarrow +\infty$ alors $C_\alpha^{Fr} \rightarrow M$.
- Si $\alpha \rightarrow -\infty$ alors $C_\alpha^{Fr} \rightarrow W$.

La figure (2.1) illustre les densités de copules Archimédiennes : Clayton, Gumbel et Frank.

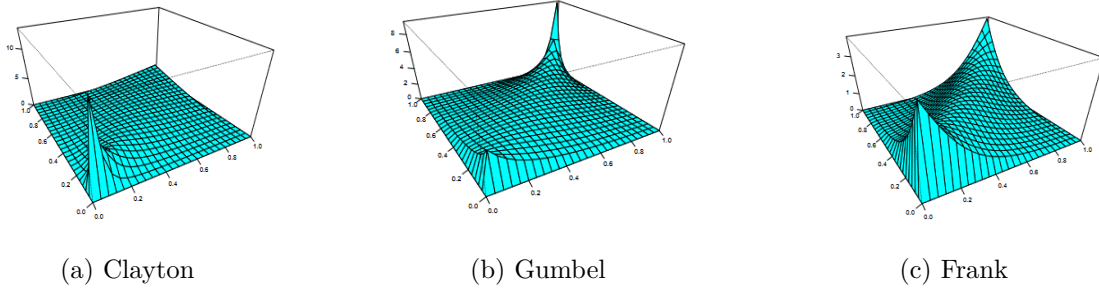


FIGURE 2.1 – Densités de copules Clayton (a) $\alpha = 2$, Gumbel (b) $\alpha = 2$ et Frank (c) $\alpha = 5.7369$.

2.3.3 Les copules Archimédiennes généralisées et hiérarchiques

Considérons le cas où $C(u_1, \dots, u_d)$ peut s'écrire :

$$\phi_1^{-1}[\phi_1[\phi_2^{-1}(\phi_2[\dots\phi_{d-1}^{-1}[\phi_{d-1}(u_1) + \phi_{d-1}(u_2)] + \dots + \phi_2(u_{d-1})]] + \phi_1(u_d)],$$

où les ϕ_i sont des générateurs de copules. C est une copule si $\phi_i \circ \phi_{i-1}^{-1}$ est l'inverse d'une transformée de Laplace. Cette copule est parfois appelée « fully nested Archimedean » (FNA). On commence par coupler deux composantes (ici U_1 et U_2), puis on couple cette paire avec U_3 , etc. On dimension $d = 5$, elle donne :

$$\phi_1^{-1}[\phi_1(\phi_2^{-1}[\phi_2(\phi_3^{-1}[\phi_3(\phi_4^{-1}[\phi_4(u_1) + \phi_4(u_2)]) + \phi_3(u_3)]) + \phi_2(u_4))] + \phi_1(u_5)].$$

On peut aussi envisager une construction plus hiérarchique, avec les copules dite « partially nested Archimedean » (PNA), en considérant un couplage complet pour (U_1, U_2, U_3) , un autre pour (U_4, U_5) , et en couplant les deux vecteurs,

$$\phi_4^{-1}[\phi_4(\phi_1^{-1}[\phi_1(\phi_2^{-1}[\phi_2(u_1) + \phi_2(u_2)]) + \phi_1(u_3)]) + \phi_4(\phi_3^{-1}[\phi_3(u_4) + \phi_3(u_5)])].$$

La condition pour avoir effectivement une copule est que $\phi_2 \circ \phi_1^{-1}$ soit l'inverse d'une transformée de Laplace, mais aussi $\phi_4 \circ \phi_1^{-1}$ ainsi que $\phi_4 \circ \phi_3^{-1}$. Ces deux constructions sont visualisées sur la figure (2.2).

Enfin, on peut envisager une construction réellement hiérarchique, ne visant pas uniquement à faire du couplage (deux à deux),

$$\phi_3^{-1}[\phi_3(\phi_1^{-1}[\phi_1(u_1) + \phi_1(u_2) + \phi_1(u_3)]) + \phi_3(\phi_2^{-1}[\phi_2(u_4) + \phi_2(u_5)])].$$

Dans ce cas, on a effectivement construit une copule si $\phi_3 \circ \phi_1^{-1}$ ainsi que $\phi_3 \circ \phi_2^{-1}$ sont des inverses de transformées de Laplace. On pourrait aussi envisager.

$$\phi_3^{-1}[\phi_3(\phi_1^{-1}([\phi_1(u_1) + \phi_1(u_2)] + \phi_3(u_3) + \phi_3(\phi_2^{-1}[\phi_2(u_4) + \phi_2(u_5)]))].$$



FIGURE 2.2 – Copules Archimédiennes « fully nested » (à gauche) et « partially nested » (à droite).

2.3.4 Les copule de valeurs extrêmes

Avant de citer les copules de valeurs extrêmes, on va d'abord présenter les lois de ces variables aléatoires. Pour cela, considérons (X_1, \dots, X_n) une suite de variables aléatoires (i.i.d). Soit $M_n = \max(X_1, \dots, X_n)$.

Théorème 6 : (Fisher-Tippett, 1928) *S'il existe deux suite $c_n > 0$ et $d_n \in \mathbb{R}$, telle que $\frac{M_n - d_n}{c_n}$ converge en distribution non dégénérée [KM97], alors :*

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = G_\alpha(x).$$

$G_\alpha(x)$ prend une des trois lois suivantes :

Fréchet	$\Phi_\alpha(x) = \exp(-x^{-\alpha})$, pour $x > 0$, $\alpha > 0$.
Weibull	$\Psi_\alpha(x) = \exp(-(-x)^\alpha)$, pour $x \leq 0$, $\alpha < 0$.
Gumbel	$\Lambda(x) = \exp(-e^{-x})$, pour $x \in \mathbb{R}$.

Définition 11

Une copule C^* est une copule de valeurs extrêmes bivariées s'il existe une copule C telle que :

$$C^*(u, v) = \lim_{n \rightarrow \infty} C^n(u^{\frac{1}{n}}, v^{\frac{1}{n}}).$$

Loi de valeurs extrêmes bivariées

Soit $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ couples de variables aléatoires (i.i.d) de distribution commune H . Alors il existe des $a_n, c_n > 0$ et $b_n, d_n \in \mathbb{R}$ telle que :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{M_{X,n} - b_n}{a_n} \leq x, \frac{M_{Y,n} - d_n}{c_n} = F_{Y_\theta}(x, y) \right),$$

telle que F_Y est une distribution non dégénérée si et seulement si les lois marginales de F_Y sont des lois de valeurs extrêmes univariées [HF06].

Théorème 7 : *Pour toute copule de valeurs extrêmes bivariées C^* il existe une fonction convexe A définie de $[0, 1]$ dans $[\frac{1}{2}, 1]$ telle que :*

$$C^*(u, v) = \exp \left[\left(-\ln(u) + \ln(v) A \left(\frac{\ln u}{\ln(u) + \ln(v)} \right) \right) \right]. \quad (2.33)$$

De plus A vérifie :

$$\max(t, 1 - t) < A(t) < t, \forall t \in [0, 1]. \quad (2.34)$$

A est dite générateur ou fonction de dépendance de Pickands [AG05].

Famille paramétrique des copules de valeurs extrêmes bivariées : Il existe essentiellement deux grandes familles de modèles paramétriques usuels de copules de valeurs extrêmes bivariées : le modèle mixte ou de Tawn (1988) [Taw88] et le modèle logistique ou de Gumbel (1960)[Gum60], [Nel91]. Les autres modèles proviennent généralement d'une extension symétrique de ces modèles. Les modèles Galambos [Gal75] et Husler-Reiss [MFM08] et le modèle Marchal-Olkin [MO67] et [CRRL00], cette copule dépend de deux paramètres.

Les copules des différentes familles sont présentées dans le tableau suivant :

<i>Copule</i>	C_θ^*	θ
Indépendance	uv	/
Gumbel A	$\exp \left\{ -(\tilde{u}^\theta + \ln \tilde{v}^\theta)^{\frac{1}{\theta}} \right\}, \theta \geq 1$	$[1, +\infty[$
Gumbel B	$uv \exp \left\{ \theta \frac{\tilde{u}\tilde{v}}{\tilde{u}+\tilde{v}} \right\}, \theta \geq 0$	I
Galambos	$uv \exp \left\{ -(\tilde{u}^{-\theta} + \tilde{v}^{-\theta})^{\frac{1}{\theta}} \right\}$	$[0, +\infty[$
Marchal-Olkin	$u^{1-\theta_1} v^{1-\theta_2} \min(u^{\theta_1}, v^{\theta_2})$	I
Logistique de Joe	$\exp \left\{ (\tilde{u}^{\theta_1} + \tilde{v}^{\theta_1} - V(\tilde{u}, \tilde{v}, \theta))^{\frac{1}{\theta_1}} \right\}$	$[1, +\infty[^2$

TABLE 2.1 – Quelques copules de valeurs extrêmes.

Avec $\tilde{u} = -\ln u$, $\tilde{v} = -\ln v$ et Φ est la fonction de répartition de la loi normale centrée réduite.

Remarque : La copule de Gumbel appartient à la famille de copules Archimédiennes et à la famille de copules de valeurs extrêmes. Cette famille de copules est définie par Capéraà et al(2000)[CFG00] qui s'appelle la famille de copules archimax(voir aussi [GR89]).

2.3.5 La copule Marchal-Olkin

La copule Marchal-Olkin est définie par :

$$C_{\alpha,\beta}(u, v) = \min(u^{1-\alpha}v, uv^{1-\beta}) = \begin{cases} u^{1-\alpha}v, & u^\alpha \geq v^\beta, \\ uv^{1-\beta}, & u^\alpha < v^\beta. \end{cases} \quad (2.35)$$

avec $0 < \alpha, \beta < 1$.

2.3.6 Les copules empiriques

Les copules empiriques notées par \hat{C}_n ont été introduite par Deheuvels [Deh79]. Elles sont construites en fonction des marginales empiriques F_{X_n} et F_{Y_n} et la distribution empirique jointe H_n comme suit :

$$\hat{C}_n(u, v) = H_n(F_{X_n}^{-1}(u), F_{Y_n}^{-1}(v)), \quad (2.36)$$

avec :

$$F_{X_n}(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} \quad \text{et} \quad F_{Y_n}(y) = \frac{1}{n} \sum_{i=1}^n 1_{\{Y_i \leq y\}}.$$

La copule empirique est définie sur l'ensemble $L = \left\{ \left(\frac{j}{n}, \frac{k}{n} \right), \text{ avec } j, k = 1, \dots, n \right\}$.

Elle peut aussi s'écrire comme suit :

$$\hat{C}_n \left(\frac{j}{n}, \frac{k}{n} \right) = \frac{1}{n} \sum_{i=1}^n 1_{(X_i \leq x_{(j)}, Y_i \leq y_{(k)})}, \quad (2.37)$$

où $x_{(j)}$ et $y_{(k)}$ sont les statistiques de rang de l'échantillon (X_1, X_2, \dots, X_n) et (Y_1, Y_2, \dots, Y_n) respectivement.

2.4 Mesures d'association bidimensionnelle

2.4.1 Les mesures de dépendance

2.4.1.1 Coefficient de corrélation linéaire

Soient X et Y deux variables aléatoires ayant des variances finies. Le coefficient de corrélation linéaire des variables X et Y est donné par :

$$\rho = r(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Avec $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ est la covariance entre X et Y ;
 $Var(X)$, $Var(Y)$ correspondent aux variances respectives des variables X, Y .
La définition du coefficient de corrélation linéaire est donc subordonnée à l'existence 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 .

Exemple : Prenons $X \rightsquigarrow \mathcal{N}(0, 1)$ et $Y = X^2$ on sait alors que $Y \rightsquigarrow \mathcal{X}_1^2$ une loi du \mathcal{X}^2 à un degré de liberté. Ainsi les propriétés classiques d'une loi du \mathcal{X}^2 implique que $E[Y] = 1$ et $var[Y] = 2$. On a alors :

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

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.

2.4.1.2 Les mesures de concordance

Soient $(x_1, y_1), (x_2, y_2)$ deux observations d'un couple de variables aléatoires continues (X, Y) , elles sont dites :

concordantes si :

$$(x_1 - x_2)(y_1 - y_2) > 0,$$

discordantes si :

$$(x_1 - x_2)(y_1 - y_2) < 0.$$

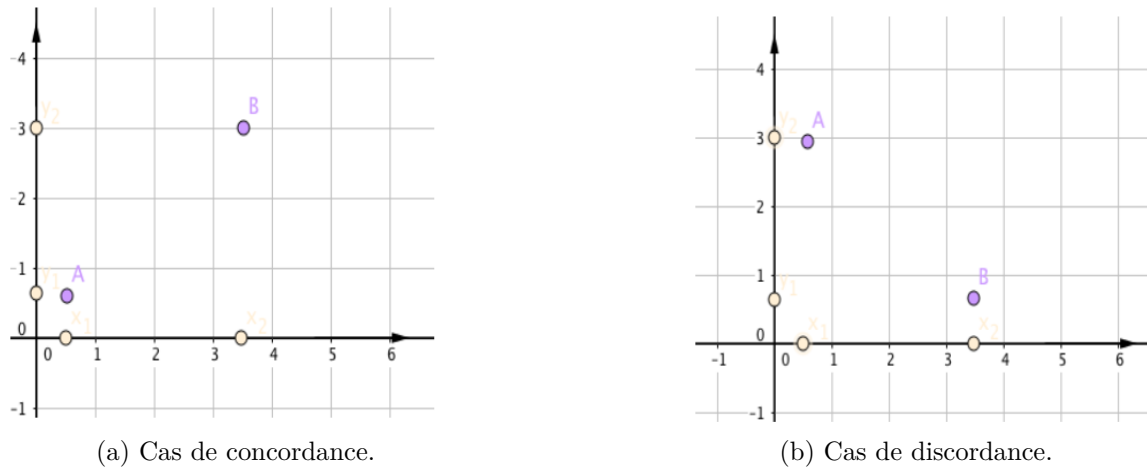


FIGURE 2.3 – Figures présentant en (a) la concordance et en (b) la discordance entre les deux points $A = (x_1, y_1)$ et $B = (x_2, y_2)$.

Définition 12

La fonction de concordance entre les deux vecteurs aléatoires (X_1, Y_1) et (X_2, Y_2) est définie par :

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

2.4.1.3 Le 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 13

Soient (X_1, Y_1) et (X_2, Y_2) deux vecteurs aléatoires continus i.i.d de fonction de répartition conjointe H et de fonctions marginales F_X (pour X_1, X_2) et F_Y (pour Y_1, Y_2). Le tau de Kendall noté par τ est définie comme suit :

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

Le tau de Kendall possède les propriétés suivantes :

- Si X et Y sont comonotones alors $\tau = 1$ (la concordance est la plus solide).
- Si X et Y sont antimonotones alors $\tau = -1$ (la discordance est la plus solide).

Soient a et b deux fonctions strictement croissantes, alors :

$$\tau(a(X), b(Y)) = \tau(X, Y).$$

Proposition : Soit (X, Y) un couple de variables aléatoires continues de copule C . Le tau de Kendall $\tau(X, Y)$ a pour expression :

$$\begin{aligned} \tau_{(X,Y)} &= 4 \iint_{I^2} C(u, v)c(u, v)dudv - 1. \\ &= 4\mathbb{E}((C(U, V))) - 1. \end{aligned} \tag{2.38}$$

Preuve :

On note $U = F_X(X)$, $U' = F_X(X')$, $V = F_Y(Y)$ et $V' = F_Y(Y')$ dont les lois sont uniformes sur $[0, 1]$. On a :

$$\begin{aligned} \tau(X, Y) &= \mathbb{P} \{ (X - X')(Y - Y') \geq 0 \} - \mathbb{P} \{ (X - X')(Y - Y') < 0 \} \\ &= 2\mathbb{P} \{ (X - X')(Y - Y') \geq 0 \} - 1 \\ &= 2\mathbb{P} \{ (X - X') \geq 0, (Y - Y') \geq 0 \} + 2\mathbb{P} \{ (X - X') < 0, (Y - Y') < 0 \} - 1 \\ &= 4\mathbb{P} \{ X \geq X', Y \geq Y' \} - 1 \\ &= 4\mathbb{P} \{ F_X(X) \geq F_X(X'), F_Y(Y) \geq F_Y(Y') \} - 1 \\ &= 4\mathbb{P} \{ U \geq U', V \geq V' \} - 1 \\ &= 4 \int_0^1 \int_0^u \int_0^1 \int_0^v c(u, v)c(u', v')dudvdv' - 1 \\ &= 4 \int_0^1 \int_0^1 C(u, v)c(u, v)dudv - 1 \\ &= 4\mathbb{E}(C(U, V)) - 1. \end{aligned}$$

Si on dispose d'un échantillon d'observations de taille n de données bivariées (x_n, y_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 :

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

2.4.1.4 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 $F_X(X)$ et $F_Y(Y)$ où F_X et F_Y sont les fonctions de marginales de X et de Y respectivement. C'est-à-dire que :

$$\rho_s = \text{Corr}(F_X(X), F_Y(Y)),$$

où $F_X(X)$ et $F_Y(Y)$ sont uniformes sur $[0, 1]$. 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 15

Soient (X_1, Y_1) , (X_2, Y_2) et (X_3, Y_3) trois vecteurs aléatoires indépendants de même loi H . Le coefficient de corrélation de Spearman est défini par :

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

Théorème 8 : 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 $\delta(f(X), Y) = -\delta(X, Y)$.

Proposition : Si a et b sont deux fonctions strictement croissantes, alors :

$$\rho_s(a(X), b(Y)) = \rho_s(X, Y).$$

Preuve : On a

$$\begin{aligned} \rho_s(X, Y) &= 12 \int_{-\infty}^{+\infty} [H(x, y) - F_X(x)F_Y(y)] dx dy \\ &= \frac{1}{\sqrt{\text{Var}(a(X)\text{Var}(b(Y)))}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [H(x, y) - F_X(x)F_Y(y)] dx dy, \end{aligned}$$

Le rho de Spearman peut aussi être exprimé en fonction de la copule comme le montre la proposition suivante.

Proposition : Si C désigne la copule des variables X et Y , le rho de Spearman $\rho_s(X, Y)$ a pour expression :

$$\begin{aligned} \rho_s(X, Y) &= 12 \int_0^1 \int_0^1 C(u, v) dudv - 3 \\ &= 12 \int_0^1 \int_0^1 uv dC(u, v) - 3. \end{aligned}$$

Preuve : Soient $U = F_X(X)$ et $V = F_Y(Y)$, deux variables qui suivent la loi uniforme sur $[0, 1]$. Par conséquent leurs espérances et leurs variances sont respectivement égales à :

$$\mathbb{E}(U) = \mathbb{E}(V) = \int_0^1 u du = 1/2.$$

et

$$\text{Var}(U) = \text{Var}(V) = \int_0^1 u^2 du - \left(\int_0^1 u du \right)^2 = 1/3 - 1/4 = 1/12.$$

D'autre part on a :

$$\begin{aligned}
 \rho_s(X, Y) &= \rho_s(a(X), b(Y)) \\
 &= \frac{1}{\sqrt{\text{Var}(a(X)\text{Var}(b(Y)))}} \int_0^1 \int_0^1 (C(u, v) - uv) dudv \\
 &= 12 \int_0^1 \int_0^1 C(u, v) dudv - 3.
 \end{aligned}$$

Aussi :

$$\begin{aligned}
 \rho_s(X, Y) &= \rho(U, V) \\
 &= \frac{\text{Cov}(U, V)}{\sigma_U \sigma_V} \\
 &= 12 \text{Cov}(U, V) \\
 &= 12 \int_0^1 \int_0^1 uv dC(u, v) - 3.
 \end{aligned}$$

Remarque : Si l'on dispose d'un échantillon d'observations de taille n de (X, Y) , $(x_i, y_i)_{1 \leq i \leq n}$, on peut construire un estimateur empirique du rho de Spearman. Celui-ci est donné par la formule suivante :

$$\rho_n(X, Y) = \frac{\sum_{i=1}^n (r_{x_i} - \bar{r}_x)(r_{y_i} - \bar{r}_y)}{\sqrt{\sum_{i=1}^n (r_{x_i} - \bar{r}_x)^2} \sqrt{\sum_{i=1}^n (r_{y_i} - \bar{r}_y)^2}},$$

où r_{x_i} , \bar{r}_x (resp. r_{y_i} , \bar{r}_y) désigne le rang de x_i et la moyenne $\bar{r}_x = \frac{1}{n} \sum_{i=1}^n r_{x_i}$ (resp. le rang de y_i et la moyenne $\bar{r}_y = \frac{1}{n} \sum_{i=1}^n r_{y_i}$).

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

Généralement, les valeurs τ et ρ_s pour une copule C sont différentes. Certaines relations lient les deux mesures de concordance.

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

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

Théorème (Durbin et Stuart, 1951) : Si X et Y sont 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 :

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

et

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

Démonstration : Soient (X_1, Y_1) , (X_2, Y_2) et (X_3, Y_3) trois vecteurs aléatoires continues i.i.d de fonction de répartition conjointe H . Si on note p la probabilité que deux de ces vecteurs soient concordants avec la troisième, on a :

$$\begin{aligned} p &= \mathbb{P}[(X_2, Y_2) \text{ et } (X_3, Y_3) \text{ sont concordants avec } (X_1, Y_1)] \\ &= \int \int_{\mathbb{R}^2} \mathbb{P}[(X_2, Y_2) \text{ et } (X_3, Y_3) \text{ sont concordants avec } (x, y)] dH(x, y) \\ &= \int \int_{\mathbb{R}^2} \mathbb{P}[(X_2 - x)(Y_2 - y) > 0] \mathbb{P}[(X_3 - x)(Y_3 - y) > 0] dH(x, y) \\ &= \int \int_{\mathbb{R}^2} (\mathbb{P}[(X_2 - x)(Y_2 - y) > 0])^2 dH(x, y) \\ &\geq \left[\int \int_{\mathbb{R}^2} (\mathbb{P}[(X_2 - x)(Y_2 - y) > 0]) dH(x, y) \right]^2 \\ &= \left[\int \int_{\mathbb{R}^2} (\mathbb{P}[(X_2 - x)(Y_2 - y) > 0] + \mathbb{P}[(X_2 - x)(Y_2 - y) < 0]) dH(x, y) \right]^2 \\ &= [\mathbb{P}[(X_2 - X_1)(Y_2 - Y_1) > 0]]^2 \\ &= \left(\frac{1 + \tau}{2} \right)^2. \end{aligned}$$

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

<i>Copules</i>	Tau kendall	Rh� de Spearman
Ind�pendance	0	0
La borne sup�rieure	1	1
La borne inf�rieure	-1	-1
Gaussienne	$\frac{2}{\pi} \arcsin(\rho)$	$\rho_s = \frac{6}{\pi} \arcsin(\frac{\rho}{2})$
Student	$\frac{2}{\pi} \arcsin(\rho)$	$\rho_s = \frac{6}{\pi} \arcsin(\frac{\rho}{2})$
Frank	$\tau(\alpha) = 1 - \frac{4}{\alpha}$ $+ \frac{4}{\alpha^2} \int_0^\alpha \left(\frac{t}{e^t-1}\right) dt$	$\rho_s(\alpha) = 1 - \frac{12}{\alpha^2} \int_0^\alpha \left(\frac{t}{e^t-1}\right) dt$ $+ \frac{24}{\alpha^3} \int \left(\frac{t^2}{e^t-1}\right) dt$
Clayton	$\tau(\alpha) = \frac{\alpha}{\alpha+2}$	$\rho_s = 12 \int_0^1 \int_0^1 (u^{-\alpha} + v^{-\alpha}) dudv - 3$
Gumbel	$\tau(\alpha) = \frac{\alpha-1}{\alpha}$	/

TABLE 2.2 – Exemples de mesures d'association bivari es

———— Chapitre 3 ————

Inférence statistique des copules

La fonction de copule, comme ses paramètres, reste inconnue en pratique et doit être estimée. Dans ce chapitre, nous présentons plusieurs méthodes d'estimation et deux tests de vraisemblance pour choisir une copule appropriée.

3.1 Méthodes d'estimations des copules

3.1.1 Estimation paramétrique des copules

Lorsqu'il s'agit d'une estimation paramétrique d'une copule, on impose un modèle paramétrique pour la copule et pour les distributions marginales. Pour estimer ces paramètres il existe différentes méthodes. Dans la suite, nous nous intéressons qu'aux deux méthodes d'estimation les plus usuelles, la méthode du maximum de vraisemblance (MLE), et la méthode des fonctions d'inférence des marginales (IFM).

1. la méthode du maximum de vraisemblance

Soit C_θ une copule paramétrique multivariée de paramètre θ . On estime sous les deux hypothèses suivantes :

$$H_0 : C \in C_0$$

telle que $C_0 = \{C_\theta : \theta \in \Theta\}$, où Θ est un sous ensemble de \mathbb{R}^p pour tout entier $p \geq 1$, on a

$$H'_0 : F_1 \in \mathcal{F}_1, \dots, F_d \in \mathcal{F}_d$$

telle que F_j pour $j = 1, \dots, d$ sont les lois marginales de X_1, \dots, X_d est $\mathcal{F}_j = \{\mathcal{F}_{\gamma_j} : \gamma_j \in \Gamma_j\}$, où Γ_j est un sous ensemble de \mathbb{R} et en maximisant la fonction log-vraisemblance $L(\theta)$ définie par :

$$L(\theta) = \sum_{j=1}^d \ln f(x_j),$$

telle que $f(x_1, \dots, x_d)$ est la fonction de densité jointe de X_1, \dots, X_d donnée par :

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{j=1}^d f_j(x_j),$$

on peut voir que la fonction $L(\theta)$ de l'échantillon $(x_1^i, \dots, x_d^i)_{1 \leq i \leq n}$ peut s'écrire comme suit :

$$\begin{aligned} L(\theta) &= \sum_{i=1}^n \ln(c_\theta(F_1(x_{i1}), \dots, F_d(x_{id}))) \prod_{j=1}^d f_j(x_{ij}) \\ &= \sum_{i=1}^n \ln(c_\theta(F_1(x_{i1}), \dots, F_d(x_{id}))) + \sum_{i=1}^n \sum_{j=1}^d \ln f_j(x_{ij}). \end{aligned}$$

Où c_θ représente la densité de la copule C_θ , alors l'estimateur de θ , notée par $\hat{\theta}_{MLE}$ est :

$$\hat{\theta}_{MLE} = \arg \max L(\theta).$$

Cet estimateur est consistant et vérifie la propriété de normalité asymptotique

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \longrightarrow \mathcal{N}(0, I^{-1}(\theta)),$$

telle que $I(\theta)$ est la matrice d'information de **Fisher**, cette matrice est estimée par l'inverse de la matrice **Hessienne** de la fonction de vraisemblance.

2. Méthode des fonctions d'inférence des marginales

La méthode d'estimation MLE estime conjointement les paramètres des distributions marginales et ceux de la fonction de copule, donc elle exige des calculs intensifs parce que l'estimation des paramètres des lois marginales et des paramètres de la copule se fait d'une façon simultanée ce qui engendre des temps de calcul très longs dans le cas d'une grande dimension. Joe et Xu [JX96] ont proposé la méthode IFM (Inference Functions for Margins) et ont estimé les paramètres séparément.

Cette méthode repose sur le fait que la représentation en copule permet de séparer les paramètres spécifiques des distributions marginales de ceux de la structure de dépendance, cette méthode est composée de deux étapes :

- Estimer les paramètres de distribution marginales $\gamma_1, \dots, \gamma_d$ qui sont définie par :

$$\hat{\gamma}_j = \arg \max L_j(\gamma_j). \tag{3.1}$$

Telle que :

$$L_j(\gamma_j) = \sum_{i=1}^n \ln f_j(x_{ij}).$$

Où f_j est la densité de F_j .

- l'estimation du paramètre θ de la copule. En utilisant les estimateurs $\hat{\gamma}_j$ obtenues par (3.1). Nous avons alors :

$$\hat{\theta}_{IFM} = \arg \max L(\theta).$$

Où

$$L(\theta) = \sum_{i=1}^n \ln c_\theta(F_{\hat{\gamma}_1(x_{i1})}, \dots, F_{\hat{\gamma}_d(x_{i1})}).$$

L'estimateur $\hat{\theta}_{IFM}$ vérifie aussi la propriété de normalité asymptotique.

$$\sqrt{n}(\hat{\theta}_{IFM} - \theta) \longrightarrow \mathcal{N}(0, V^{-1}(\theta)).$$

Avec $V(\theta)$ est la matrice d'information de **Godambe**, définie par :

$$V(\theta) = D^{-1}M(D^{-1})^t.$$

Où

$$D = \mathbb{E}\left[\frac{\partial}{\partial \theta}(g(\theta)^t)\right], M = \mathbb{E}[g(\theta)^t g(\theta)], \text{ et } g(\theta) = \left(\frac{\partial}{\partial \gamma_1} L_1, \dots, \frac{\partial}{\partial \gamma_d} L_d\right).$$

3.1.2 Estimation semi-paramétrique des copules

Ce type d'estimation est utilisé dans le cas où les paramètres des distributions marginales sont inconnus contrairement à ceux des copules associées qui eux sont connues. On a deux méthodes : méthode du maximum de vraisemblance canonique (CML) et la méthode des moments.

1. la méthode de maximum de vraisemblance canonique(pseudo-maximum de vraisemblance)

Cette méthode a été proposé dans le cas où les marginales F_{X_1}, \dots, F_{X_d} associées aux X_1, \dots, X_d sont inconnues, elle est compose de deux étapes :

- On remplace les marginales F_{X_1}, \dots, F_{X_d} par leurs estimations naturelles (estimateur empirique), définies par :

$$\hat{F}_{X_j}(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_{ij} \leq x\}},$$

avec $1_{\{X_{ij} \leq x\}}$ est la fonction indicatrice .

- Estimer le paramètre de la copule paramétrique par la méthode MLE

$$\hat{\theta}_{CML} = \arg \max L(\theta),$$

telle que :

$$L(\theta) = \sum_{i=1}^n \ln c_{\theta}(F_{x_1}(x_{i1}), \dots, F_{x_d}(x_{id}))$$

Remarque

Nous avons pour tout $i \in 1, \dots, n$ et $j \in 1, \dots, d$, $R_{ij} = nF_j(\hat{x}_{ij})$, où R_{ij} est le rang de X_{ij} . Donc l'estimateur tiré de cette méthode est basé sur les rangs des observations.

Exemple : (Dans le cas bivariée)

Soit X, Y deux variables aléatoires dont les marginales sont F_X et F_Y respectivement. Soit C_{θ} la copule associée à X et Y de paramètre θ . La pseudo log-vraisemblance est définie par :

$$L(\theta) = \sum_{i=1}^n \ln c_{\theta}(F_{X_n}(x_i), F_{Y_n}(y_i)),$$

telle que :

$$F_{X_n}(x_i) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_1 \leq x\}} \text{ et } F_{Y_n}(y_i) = \frac{1}{n} \sum_{i=1}^n 1_{\{Y_1 \leq y\}}.$$

Comme $nF_{X_n} = R_i$ et $nF_{Y_n} = S_i$, où R_i et S_i sont les rangs de X_i et Y_i dans leurs échantillons respectivement, donc $L(\theta)$ devient :

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

2. La méthode des moments

Cette méthode consiste à estimer les paramètre $\gamma_j, j = 1, \dots, d$ des lois marginales et le paramètre θ de la copule par la méthode des moments, i.e,

- Résoudre le système des d'équations à d inconnus.

$$\left\{ \begin{array}{l} \bar{X}_t = f(\gamma_1, \dots, \gamma_d) \\ S_t^2 = g(\gamma_1, \dots, \gamma_d) \\ \mu_{3,t} = h(\gamma_1, \dots, \gamma_d) \\ \cdot \\ \cdot \\ \cdot \end{array} \right.$$

Où d désigne la dimension de θ , f, g et h sont les expressions des moments (ordinaires) d'ordre 1, 2 et 3 en fonction des paramètres γ_j . Répéter cette étape pour toutes les marginales.

- Inverser le tau de kendall ou le rho de Spearman pour obtenir le paramètre θ de la copule.

3.1.3 Estimation non paramétrique

Si le modèle paramétrique de la copule est mal spécifié, l'approche paramétrique et semi-paramétrique conduisent généralement à un mauvais ajustement des données. Dans ce cas, les méthodes d'estimation non paramétriques peuvent être une bonne alternative.

3.1.4 La fonction empirique

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

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq x\}},$$

pour (x_1, \dots, x_n) un échantillon de taille n de la loi F .

Si on se donne $(x_1^i, \dots, x_d^i)_{1 \leq i \leq n}$ un n-échantillon du vecteur X (de dimension d), on peut généraliser l'expression de la fonction de répartition empirique en posant :

$$F_n(x_1, \dots, x_d) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_{i1} \leq x_1, \dots, x_{id} \leq x_d\}}.$$

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

Propriétés asymptotiques

- F_n est un estimateur sans biais.

$$\mathbb{E}(F_n(x)) = F(x).$$

- Par la loi forte des grands nombres

$$F_n(x) \xrightarrow[p.s.]{} F(x).$$

- F_n est asymptotiquement normale de variance $F(x)(1 - F(x))$.

$$\sqrt{n}(F_n(x) - F(x)) \longrightarrow \mathcal{N}(0, F(x)(1 - F(x))).$$

- F_n est un estimateur uniformément consistant de F , c'est à dire :

$$\|F_n - F\|_\infty \xrightarrow[p.s.]{} 0.$$

3.2 Les tests d'adéquation

Les tests d'adéquation permettent d'identifier la copule qui s'ajuste le mieux aux données. On va exposer dans cette partie, les différents tests d'adéquation possibles pour une distribution bivariée.

3.2.1 Test du Chi-deux

La sélection de la meilleure copule doit s'appuyer sur un test statistique. Le test du chi-deux est utilisé dans le cadre de l'ajustement d'une distribution paramétrique à une distribution empirique qui constitue à cet égard un outil intéressant pour le choix d'une copule. Le processus de sélection d'une copule est décrit ci-après.

1. Estimation du paramètre de la copule :

On estime le paramètre de la copule bivariée à partir de la méthode CML afin de s'affranchir d'une éventuelle erreur de spécification des lois marginales. On note $C_{\hat{\theta}}$ la copule paramétrique obtenue.

2. Calcul de la copule empirique :

On définit un treillis d'ordre n adapté au nombre d'observations et aux contraintes de temps de calcul, puis on calcule la copule empirique bivariée sur ce treillis en utilisant la formule $\boxed{(2.37)}$.

3. Choix d'un nombre d'intervalles pour découper les uniformes univariées :

On se donne un nombre k afin de découper chacune des uniformes en k intervalles et de construire un tableau de contingence bivarié des effectifs. k et n doivent être choisis de manière à obtenir aisément les effectifs empiriques à partir de la copule empirique. Sans à priori sur la distribution empirique bivariée, on découpe chaque uniforme sur des intervalles identiques. On note $]u_0, u_1],]u_1, u_2], \dots,]u_{k-1}, u_k]$ et $]v_0, v_1],]v_1, v_2], \dots,]v_{k-1}, v_k]$ les k intervalles pour découper U et V respectivement :

4. Calcul des effectifs et construction du tableau de contingence bivarié :

L'objet de cette étape est de calculer les effectifs empiriques et théoriques sur chacun des k^2 intervalles bidimensionnels $]u_{i-1}, u_i] \times]v_{j-1}, v_j]$ avec $i, j = 1, \dots, k$. Pour ce faire, il convient de calculer les probabilités bivariées $p_{i,j}$ d'appartenance à un intervalle. On a :

$$p_{i,j} = C(u_i, v_j) - C(u_{i-1}, v_j) - C(u_i, v_{j-1}) + C(u_{i-1}, v_{j-1}).$$

Les effectifs $f_{i,j}$ sont calculés en multipliant les probabilités bivariées par le nombre total d'observation. On obtient alors un tableau de contingence bivarié donnant les effectifs empiriques issus de la copule empirique \hat{C} et les effectifs théoriques issus de la copule paramétrique $C_{\hat{\theta}}$. Il a la forme suivante :

$U \setminus V$	$]v_0, v_1]$	$]v_1, v_2]$...	$]v_{k-1}, v_k]$
$]u_0, u_1]$	$f_{1,1}$	$f_{1,2}$...	$f_{1,k}$
$]u_1, u_2]$	$f_{2,1}$	$f_{2,2}$...	$f_{2,k}$
...
$]u_{k-1}, u_k]$	$f_{k,1}$	$f_{k,2}$...	$f_{k,k}$

TABLE 3.1 – Tableau de contingence donnant les effectifs théoriques.

5. Application du critère de Cochran et regroupement en classes :

Dans la pratique, les effectifs théoriques de certaines cases peuvent être très faibles voir proches de zéro. Il faut alors procéder à un regroupement des k^2 intervalles initiaux en R classes permettant de respecter le critère de Cochran qui recommande d'avoir des effectifs théoriques au moins supérieurs à 1% du nombre total d'observations dans chaque classe, et supérieurs à 5% du nombre total d'observations dans au moins 80% des classes. Ce regroupement est aussi à effectuer sur le tableau bivarié des effectifs empiriques.

6. Calcul de la statistique bivariée du Chi-deux et sélection de la copule :

Après avoir défini le regroupement en R classes vérifiant le critère de Cochran, on calcule la statistique bivariée du Chi-deux observée comme suit :

$$\chi_{obs}^2 = \sum_{r=1}^R \frac{(O_r - E_r)^2}{E_r}, \quad (3.2)$$

où O_r et E_r représentent respectivement les effectifs théoriques et empiriques et attendus dans chaque classe. Cette statistique suit une loi de Chi-deux à $(R - v - 1)$ degrés de liberté où v est le nombre de paramètres estimés de la copule. On se donne ensuite un seuil critique pour la zone de rejet et on calcule la p-valeur $\hat{\alpha} = \mathbb{P}(\chi_{R-v-1}^2 > \chi_{obs}^2)$. On rejette l'adéquation de la copule paramétrique si $\hat{\alpha} < \alpha$. Il faut préciser pour les praticiens que $\hat{\alpha}$ se situe en fait dans l'intervalle suivant :

$$\mathbb{P}(\chi_{R-1}^2 > \chi_{obs}^2) \leq \hat{\alpha} \leq \mathbb{P}(\chi_{R-v-1}^2 > \chi_{obs}^2). \quad (3.3)$$

Cet encadrement étant de faible amplitude lorsque v est petit vis à vis de R , le seuil critique sera le plus souvent en dehors de cet encadrement et on conclura sans ambiguïté (Besson et Partrat [PBG05]). Parmi les copules acceptées, on sélectionne alors celle qui maximise la p-valeur.

3.2.2 Test de Cramér-von Mises

On suppose qu'on cherche à tester $C \in \mathcal{C}$ (l'hypothèse nulle), où \mathcal{C} est une famille de copules (l'hypothèse alternative étant $C \notin \mathcal{C}$). Si la famille \mathcal{C} est une famille théorique, Fermanian ou Genest et Remillard ont suggéré d'utiliser le test de Cramér-von Mises, avec comme statistique :

$$T = n \int \int_{I^2} (\hat{C}(u, v) - C_{\hat{\theta}}(u, v))^2 d\hat{C}(u, v). \quad (3.4)$$

Une approche un peu plus simple est d'utiliser la fonction de Kendall définie par : $K(t) = \mathbb{P}(C(u, v) \leq t)$. Si on suppose que cette fonction appartient à une famille paramétrique, alors la statistique prétendante peut s'écrire :

$$T = n \int \int_{I^2} (\hat{K}(t) - K_{\hat{\theta}}(t))^2 dK_{\hat{\theta}}(t).$$

Comme suggéré par Genest et Rivest. On rejette l'hypothèse nulle si la valeur de cet indicateur T est supérieure au seuil critique α .

Chapitre 4

Simulation aléatoire des copules

Simuler une copule bivariée C signifie simuler les arguments u et v de cette fonction, tirés d'un couple aléatoire (U_1, U_2) de loi uniforme sur $[0, 1]$. Ceci permet alors de déterminer la simulation d'un couple aléatoire (X, Y) dont la structure de dépendance est définie par C et de distribution jointe H dont les marges sont F_X et F_Y . Pour cela on utilise la transformation $\boxed{(2.8)}$ du **corollaire 1** du chapitre 2.

Toute la difficulté est de simuler des nombres aléatoires issus de C . Par exemple, la simulation de la copule indépendance est relativement simple : on simule des réalisations de loi uniforme sur $[0, 1]$ de manière indépendante. Mais pour la majorité des copules, cela s'avère plus complexe. Nous présentons dans cette section trois méthodes de simulation des copules qui sont décrites dans [\[PLA17\]](#) ou [\[DMK01\]](#).

4.1 Méthode de simulation

4.1.1 Méthode des distributions

On suppose que l'on se trouve dans une situation où la loi jointe du couple (X, Y) est plus facile à simuler directement que la copule C , c'est par exemple le cas de la copule gaussienne. Un vecteur gaussien de dimension 2 est aisé à simuler, alors que la copule gaussienne n'est pas simple à simuler directement.

Simuler des réalisations (u, v) du couple aléatoire (U_1, U_2) de distribution C revient à :

1. Simuler des réalisations (x, y) du couple aléatoire (X, Y) de distribution H .
2. Appliquer la transformation $(u, v) = (F_X(x), F_Y(y))$.

4.1.2 Méthodes des distributions conditionnelles

Cette méthode consiste à simuler directement les marges (U_1, U_2) de la copule. L'algorithme de cette méthode est le suivant :

1. Simuler deux variables aléatoires indépendantes U_1 et U_2 uniformes sur $[0, 1]$, soient u et v les nombres simulés.
2. Déterminer la distribution conditionnelle de la copule C de U_2 sachant $U_1 = u$, cette dernière se définit de la manière suivante :

$$\begin{aligned}
 C_{U_2|U_1}(u, v) &= \mathbb{P}(U_2 \leq v | U_1 = u) \\
 &= \lim_{\Delta U \rightarrow 0^+} \frac{\mathbb{P}(U_2 \leq v, u \leq U_1 \leq u + \Delta U)}{\mathbb{P}(u \leq U_1 \leq u + \Delta U)} \\
 &= \lim_{\Delta U \rightarrow 0^+} \frac{\mathbb{P}(U_1 \leq u + \Delta U, U_2 \leq v) - \mathbb{P}(U_2 \leq v, U_1 \leq u)}{\mathbb{P}(U_1 \leq u + \Delta U) - \mathbb{P}(U_1 \leq u)} \\
 &= \lim_{\Delta U \rightarrow 0^+} \frac{C(u + \Delta U, v) - C(u, v)}{C(u + \Delta U, 1) - C(u, 1)} \\
 &= \lim_{\Delta U \rightarrow 0^+} \frac{C(u + \Delta U, v) - C(u, v)}{\Delta U - 0} \\
 &= \frac{\partial C(u, v)}{\partial u}.
 \end{aligned}$$

3. Inverser la distribution conditionnelle :

$$C_{U_2|U_1}^{-1}(U, u) = \{v : C_{U_2|U_1}(u, v) = u\}.$$

4. Remplacer U par v pour obtenir l'expression de V .

$$v = C_{U_2|U_1}^{-1}(v, u).$$

On obtient alors (U_1, U_2) les marges de la copule.

4.1.2.1 Cas particulier des copules archimédiennes

Dans le cas particulier d'une copule archimédienne, la distribution conditionnelle s'exprime à l'aide du générateur de la copule. Cette méthode de simulation de copules

archimédiennes est introduite par Genest et Mackay (voir [GM86b]). Si C est une copule archimédienne de générateur φ , on a vu que C s'écrit sous la forme :

$$C_\varphi(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)), (u, v) \in I^2. \quad (4.1)$$

Par définition même de $C_{U_2|U_1}(u, v)$, on a :

$$C_{U_2|U_1}(u, v) = \frac{\partial}{\partial u} C_\varphi(u, v) = \frac{\varphi'(u)}{\varphi'[\varphi^{-1}\{\varphi(u) + \varphi(v)\}]}.$$

Ainsi, l'inverse de $C_{U_2|U_1}(u, v)$ est donné en tout $U \in [0, 1]$ par :

$$C_{U_2|U_1}^{-1}(U, u) = \varphi^{-1} \left[\varphi \left[(\varphi')^{-1} \left\{ \frac{\varphi'(u)}{U} \right\} \right] - \varphi(u) \right].$$

Pour simuler les marges (U_1, U_2) d'une copule archimédienne C_φ , on procède comme suit :

1. On génère deux variables aléatoires de façon indépendantes U_1, U_2 qui sont uniformes sur l'intervalle $[0, 1]$, soient u et v des nombres simulés.
2. On calcule :

$$C_{U_2|U_1}^{-1}(U, u) = \varphi^{-1} \left[\varphi \left[(\varphi')^{-1} \left\{ \frac{\varphi'(u)}{U} \right\} \right] - \varphi(u) \right].$$

3. Ainsi pour $U = v$, on calcule v à partir de l'expression :

$$v = C_{U_2|U_1}^{-1}(v, u) = \varphi^{-1} \left[\varphi \left[(\varphi')^{-1} \left\{ \frac{\varphi'(u)}{v} \right\} \right] - \varphi(u) \right].$$

Le couple (U_1, U_2) est alors de loi C_φ . Cette procédure fonctionne bien pour les copules de Clayton et de Frank, mais pour la copule de Gumbel, il n'y a pas une formule analytique pour $(\varphi')^{-1}$.

4.1.3 Méthode analytique

Une méthode analytique est une méthode de simulation spécifique à chaque copule. On choisit de présenter ici la méthode analytique de simulation d'une copule de Clayton proposée par Devroye (voir [Dev86]). L'expression de la copule de Clayton est :

$$C_\alpha^{cla}(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{\frac{-1}{\alpha}}. \quad (4.2)$$

Cette méthode revient à :

1. Simuler deux variables aléatoires S et T indépendantes et uniformes sur $[0, 1]$, soient s et t des nombres simulés à partir desquelles on détermine $x = -\ln(s)$ et $y = -\ln(t)$.
2. Simuler une variable aléatoire Z de loi Gamma de paramètres 1 et α : $Z \rightsquigarrow \Gamma(1, \alpha)$, soit z le nombre simulé.
3. Déterminer les réalisations de la copule de Clayton à partir des expressions suivantes :

$$u = \left(1 + \frac{x}{z}\right)^{-\alpha} \text{ et } v = \left(1 + \frac{y}{z}\right)^{-\alpha}.$$

4.2 Simulation

4.2.1 Simulation des copules elliptiques

4.2.1.1 Simulation de la copule gaussienne

Il est possible de simuler une paire de variables aléatoires (U_1, U_2) ayant pour distribution jointe la copule normale. La première étape consiste à simuler une paire de variables aléatoires (X, Y) ayant pour distribution la loi normale bivariée $\mathcal{N}_2(0, R)$, R étant la matrice de corrélation ayant ρ comme élément non diagonal, et bien sûr 1 sur la diagonale. En prenant ensuite $U_1 = \phi(X)$ et $U_2 = \phi(Y)$, alors (U_1, U_2) aura pour distribution la copule normale.

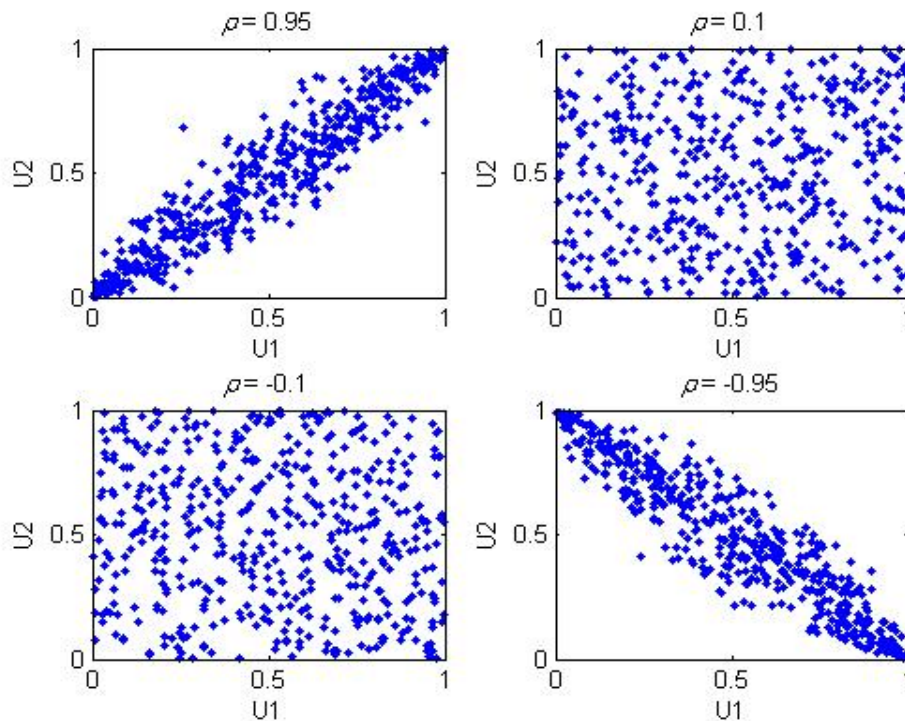
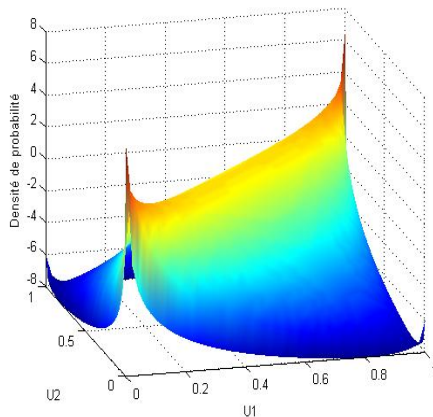


FIGURE 4.1 – Simulation de la copule gaussienne pour des différentes valeurs de ρ .

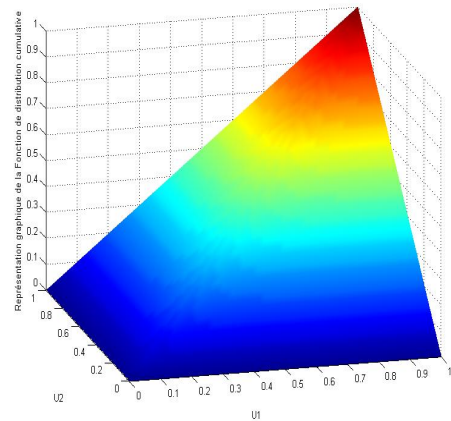
Remarques

- La dépendance entre les deux v.a.s est à droite si ρ est positif et elle est à gauche si ρ est négatif .
- Pour $\rho = 0.1$ et $\rho = -0.1$ il y a une faible dépendance entre les v.a.s (indépendance).
- Pour $\rho = -0.95$ il y a une fort dépendance entre les deux v.a.s (corrélacion négative).
- Pour $\rho = 0.95$ il y a une fort dépendance entre les deux v.a.s (corrélacion positive).

Ci-dessous les représentations graphiques de la densité de probabilité et la fonction distribution cumulative de la copule gaussienne pour $\rho = 0.95$.



(a) La densité de probabilité.

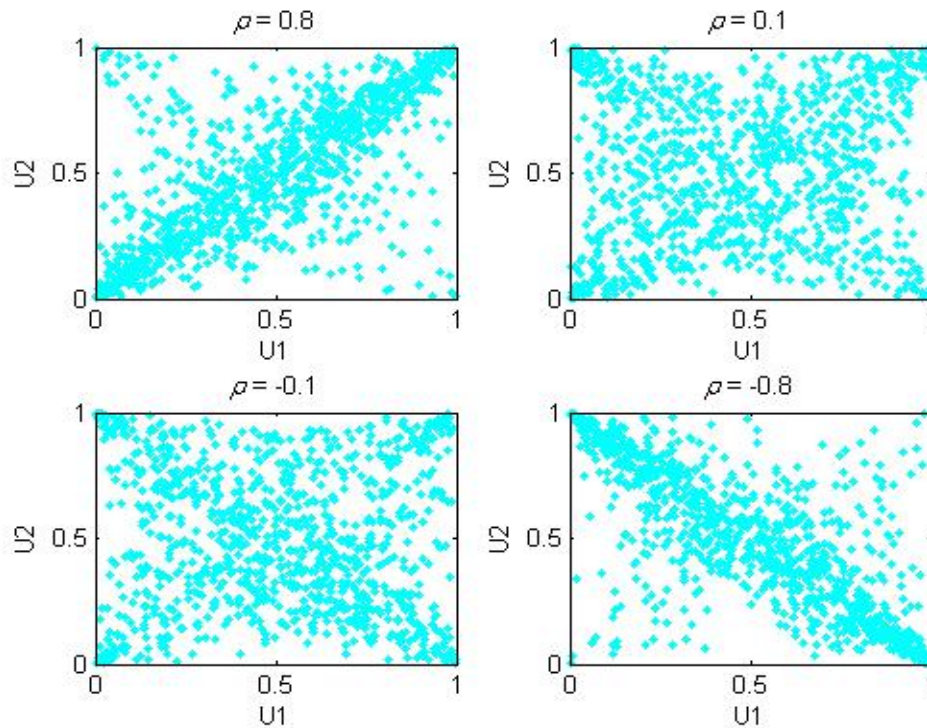


(b) La fonction distribution cumulative.

FIGURE 4.2 – Figures présentant en (a) la densité de probabilité et en (b) la fonction distribution cumulative de la copule gaussienne.

4.2.1.2 Simulation de la copule student

Une copule de student a des distributions marginales uniformes pour U_1 et U_2 , tout comme une copule gaussienne. Alors (U_1, U_2) aura pour distribution la copule de student.


 FIGURE 4.3 – Simulation de la copule de student pour des différentes valeurs de ρ .

Remarque

Les remarques sur la copule gaussienne sont maintenues pour la copule de student.

4.2.2 Simulation des copules archimédiennes

4.2.2.1 Copule frank

Dans le cas de la copule de Frank, la méthode des distributions conditionnelles peut être mise en œuvre simplement. L'expression de cette copule est donnée par :

$$C_{\alpha}^{Fr}(u, v) = \frac{-1}{\alpha} \ln \left(1 + \frac{(e^{-u\alpha} - 1)(e^{-v\alpha} - 1)}{e^{-\alpha} - 1} \right), \alpha \in \mathbb{R}^*.$$

On en déduit que :

$$C_{U_2|U_1}(u, v) = \frac{\partial C(u, v)}{\partial u} = \frac{e^{-\alpha u}(e^{-\alpha v} - 1)}{e^{-\alpha} - 1 + (e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}.$$

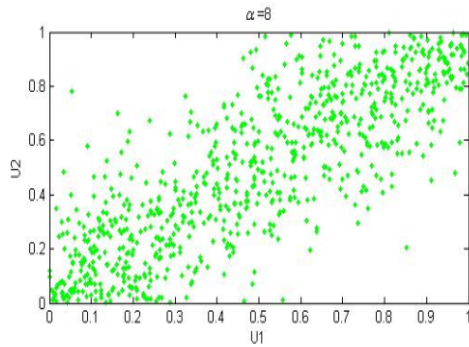
On inverse alors $C_{U_2|U_1}(u, v)$ en résolvant l'équation $C_{U_2|U_1}(u, v) = U$ en v , pour obtenir :

$$C_{U_2|U_1}^{-1}(U, u) = -\frac{1}{\alpha} \ln \left\{ 1 + \frac{U(e^{-\alpha} - 1)}{U + (1 - U)e^{-\alpha u}} \right\}.$$

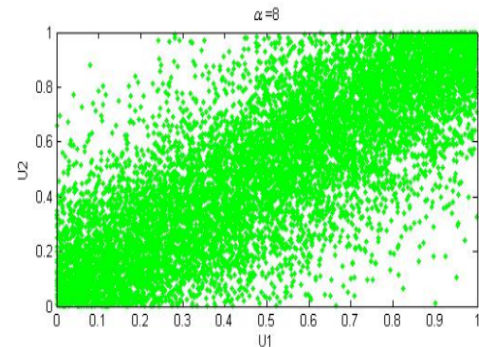
Donc :

$$v = C_{U_2|U_1}^{-1}(v, u) = -\frac{1}{\alpha} \ln \left\{ 1 + \frac{v(e^{-\alpha} - 1)}{v + (1 - v)e^{-\alpha u}} \right\}.$$

La simulation peut donc être mise en oeuvre simplement.



(a) Echantillon de 800 paires pour $\alpha = 8$



(b) Echantillon de 10000 paires pour $\alpha = 8$

FIGURE 4.4 – Simulation de la copule de Frank pour des échantillons différents.

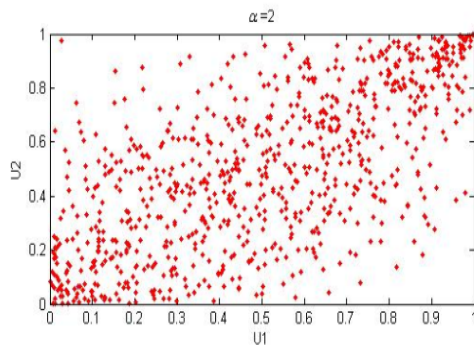
4.2.2.2 Copule Gumbel

L'expression de cette copule est :

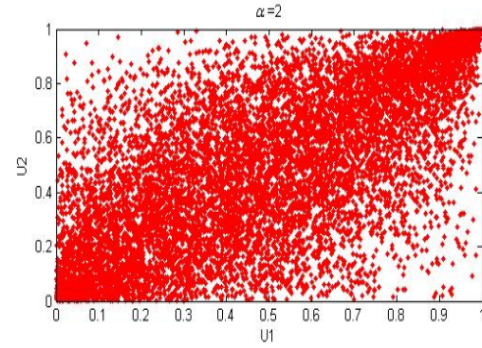
$$C_{\alpha}^{Gum}(u, v) = \exp \left[- \{ (-\ln u)^{\alpha} + (-\ln v)^{\alpha} \}^{\frac{1}{\alpha}} \right], \quad (4.3)$$

On en déduit que :

$$\begin{aligned} C_{U_2|U_1}(u, v) &= \frac{\partial \exp \left[- \{ (-\ln u)^{\alpha} + (-\ln v)^{\alpha} \}^{\frac{1}{\alpha}} \right]}{\partial u} \\ &= \frac{1}{u} \left[1 + \left(\frac{\ln v}{\ln u} \right)^{\alpha} \right]^{-1 - \frac{1}{\alpha}} \exp \left[- \{ (-\ln u)^{\alpha} + (-\ln v)^{\alpha} \}^{\frac{1}{\alpha}} \right]. \end{aligned}$$



(a) Echantillon de 800 paires pour $\alpha = 2$



(b) Echantillon de 10000 paires pour $\alpha = 2$

FIGURE 4.5 – Simulation de la copule de Gumbel pour des échantillons différents.

Remarque

- Dans les figures (4.4) et (4.5) on remarque que plus la taille de l'échantillon augmente plus la dépendance devient visible aux points $(0, 0)$ et $(1, 1)$.

4.2.2.3 Copule clayton

L'expression de la copule de Clayton bidimensionnelle s'écrit donc sous la forme :

$$C_{\alpha}^{cla}(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{\frac{-1}{\alpha}}, \quad (4.4)$$

où le générateur de la copule de Clayton est défini par :

$$\varphi_{\alpha}(t) = \frac{1}{\alpha}(t^{-\alpha} - 1).$$

Les expressions $\varphi_{\alpha}^{-1}(t)$, $\varphi'_{\alpha}(t)$ et $(\varphi'_{\alpha})^{-1}$ sont données par :

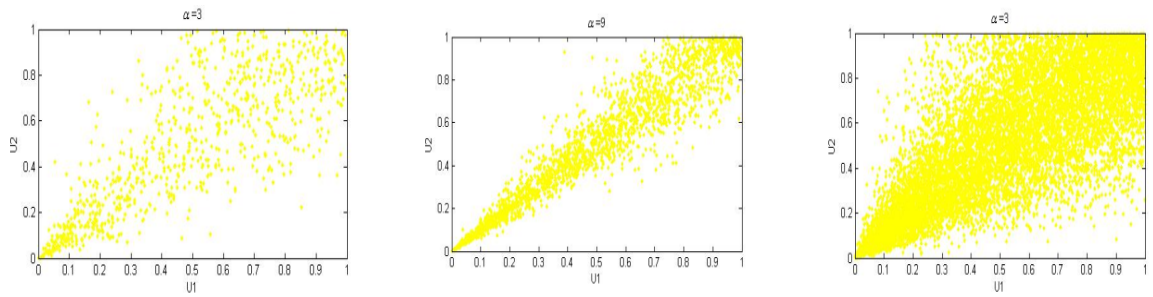
$$\varphi_{\alpha}^{-1}(t) = \left(\frac{1}{\alpha t + 1}\right)^{\frac{1}{\alpha}}, \quad \varphi'_{\alpha}(t) = -t^{-\alpha-1} \text{ et } (\varphi'_{\alpha})^{-1} = \left(-\frac{1}{t}\right)^{\frac{1}{\alpha+1}}.$$

- L'inverse $C_{U_2|U_1}^{-1}(U, u)$ est donnée par :

$$C_{U_2|U_1}^{-1}(U, u) = \left[u^{-\alpha} U^{-\frac{\alpha}{\alpha+1}} - u^{-\alpha} + 1 \right]^{-\frac{1}{\alpha}},$$

- pour $U = v$, on obtient :

$$v = \left[u^{-\alpha} v^{-\frac{\alpha}{\alpha+1}} - u^{-\alpha} + 1 \right]^{-\frac{1}{\alpha}}.$$



(a) Echantillon de 2000 paires pour $\alpha = 3$

(b) Echantillon de 2000 paires pour $\alpha = 9$

(c) Echantillon de 10000 paires pour $\alpha = 3$

FIGURE 4.6 – Simulation de la copule de Clayton pour des échantillons de tailles différentes et des valeurs de α différentes.

Remarques

- Plus la valeur de α augmente plus la dépendance devienne forte.
- Plus la taille de l'échantillon augmente plus la dépendance devient visible aux points $(0, 0)$ et $(1, 1)$.

4.3 Estimation des paramètres des copules

Dans cette partie, on va estimer les paramètres des copules. Pour cela, on va répéter l'expérience 100 fois pour un échantillonnage de 10000. Ensuite on va calculer le biais moyen et la variance du tau de Kendall et $\hat{\alpha}$ pour différentes valeurs de α . On prend comme exemple d'application les copules archimédiennes et elliptiques.

4.3.1 Estimation des paramètres des copules archimédiennes

Copule de Gumble

On rappelle que le tau de Kendall théorique de C_α^{Gum} est :

$$\tau = \frac{\alpha - 1}{\alpha}.$$

On obtient les résultats dans le tableau suivant :

Valeur de α	$\alpha = 2$		$\alpha = 4$		$\alpha = 5$	
Paramètres estimés	$\hat{\alpha}$	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\tau}$
Biais moyen	0.0034	0.0015	0.0144	$5.1564 * 10^{-4}$	0.0145	$9.1980 * 10^{-4}$
Variance	0.0052	$3.3087 * 10^{-4}$	0.0247	$9.5459 * 10^{-5}$	0.0421	$6.9861 * 10^{-5}$

TABLE 4.1 – Estimation des paramètres de la copule gumble.

Copule de Clayton

On rappelle que le tau de Kendall théorique de C_α^{cla} est :

$$\tau = \frac{\alpha}{\alpha + 2}.$$

On obtient les résultats dans le tableau suivant :

Valeur de α	$\alpha = 2$		$\alpha = 4$		$\alpha = 6$	
Paramètres estimés	$\hat{\alpha}$	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\tau}$
Biais moyen	0.0019	0.0012	0.0268	0.0036	0.0110	0.0011
Variance	0.0054	$3.4496 * 10^{-4}$	0.0169	$2.1478 * 10^{-4}$	0.0271	$1.0648 * 10^{-4}$

TABLE 4.2 – Estimation des paramètres de la copule clayton.

Copule de Frank

On rappelle que :

$$C_{\alpha}^{Fr}(u, v) = \frac{-1}{\alpha} \ln \left(1 + \frac{(e^{-u\alpha} - 1)(e^{-v\alpha} - 1)}{e^{-\alpha} - 1} \right), \alpha \in \mathbb{R}^*.$$

On obtient les résultantes dans le tableau suivant :

Valeur de α	$\alpha = 0.2$		$\alpha = 1$		$\alpha = 2$	
Paramètres estimés	$\hat{\alpha}$	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\tau}$	$\hat{\alpha}$	$\hat{\tau}$
Biais moyen	0.0228	0.0923	0.1157	0.1021	0.0244	0.1002
Variance	0.0251	$5.1384 * 10^{-4}$	$5.84422 * 10^{-4}$	$3.6691 * 10^{-4}$	0.0011	0.0035

TABLE 4.3 – Estimation des paramètres de la copule frank.

4.3.2 Estimation des paramètres des copules elliptiques

Copule gaussienne

On rappelle que le tau de kendall théorique de C_{ρ}^{ϕ} est :

$$\tau = \frac{2}{\pi} \arcsin(\rho)$$

On obtient les résultantes dans le tableau suivant :

Valeur de ρ	$\rho = -0.5$		$\rho = -0.9$		$\rho = 0.5$	
Paramètres estimés	$\hat{\rho}$	$\hat{\tau}$	$\hat{\rho}$	$\hat{\tau}$	$\hat{\rho}$	$\hat{\tau}$
Biais moyen	0.0074	0.0794	0.0013	0.0095	0.0196	0.0456
Variance	0.0995	$8.5190 * 10^{-5}$	0.00185	0.0154	0.0854	$9.5635 * 10^{-6}$

TABLE 4.4 – Estimation des paramètres de la copule gaussienne.

Copule de student

On rappelle que le tau de kendall théorique de $C_{r,v}^t$ est :

$$\tau = \frac{2}{\pi} \arcsin(\rho)$$

On obtient les résultantes dans le tableau suivant :

Valeur de ρ	$\rho = -0.5$		$\rho = -0.9$		$\rho = 0.5$	
Paramètres estimés	$\hat{\rho}$	$\hat{\tau}$	$\hat{\rho}$	$\hat{\tau}$	$\hat{\rho}$	$\hat{\tau}$
Biais moyen	0.0284	0.0322	0.0549	0.0334	0.0981	0.2857
Variance	0.0668	0.0467	0.0399	0.0681	0.0392	0.0290

TABLE 4.5 – Estimation des paramètres de la copule student.

Remarques

- On remarque que pour des différentes valeurs de α (pour les copules Archimédiennes) et ρ (pour les copules elliptiques) le biais moyen est très proche de 0, alors on a une bonne estimation des paramètres.
- la variance des paramètres de différentes copules est très petite pour différentes valeurs de α et ρ alors les estimateurs obtenus sont faiblement variants, donc ils sont efficaces.

4.4 Exemple d'application des copules pour la simulation des entrées dépendantes

Dans cette partie on va enrichir notre étude par une application sur des données réelles.

Considérons les données de deux actions X_1 et X_2 suivant :

X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
0.4478	0.0673	-1.1189	-1.2053	0.7050	0.7050	1.3072	-1.1355
0.9881	1.5816	0.5979	2.1019	-0.3877	-0.9825	-0.6575	-0.4761
0.8746	0.7562	0.0417	0.0552	1.8221	2.2286	-1.0556	-0.30561
0.7144	-0.6898	0.5230	0.0875	0.2937	-0.2738	-0.5842	-1.0150
0.0054	-1.3429	-1.5404	-2.5562	0.7817	-0.6284	-0.6263	0.0386
0.0823	-0.5475	-0.3229	0.0995	0.0903	0.0717	-1.5278	-2.1143
-0.7116	-0.8024	0.3807	0.9166	-0.6297	-1.6153	-1.4187	-1.2010
-1.2352	-1.0896	-0.4476	-1.0733	1.0013	-1.0996	-0.2085	-0.9540
1.7091	2.2787	-2.3527	-1.1213	1.4174	0.6817	-1.0944	-1.8561
0.2822	-0.5338	-1.4586	0.4026	0.2836	-0.6547	0.2233	0.7685
-1.0400	-0.3950	-1.4345	0.0233	-0.3086	-1.4369	0.4382	0.8561
-0.1889	-1.7312	-0.7666	-1.9470	0.4141	1.2592	-0.4607	0.7363
-2.2852	-2.3245	-1.6030	-1.2482	-0.3264	-0.4429	-1.6621	-0.4945

-0.8414	-0.1041	-0.5846	0.0364	-0.3375	0.1962	-0.4818	-0.6888
-1.8086	-2.4949	0.3195	-0.0563	-0.1486	1.1146	-1.0633	-0.7411
2.0228	1.7290	0.4073	0.4786	-1.2028	-0.4754	0.0334	-0.2944
1.1048	2.0519	2.2280	1.7168	-0.8395	-1.3725	-1.6621	-0.4945
-0.1681	-0.6563	-1.4459	-1.4534	0.0704	-0.0958	-2.8717	-1.9454
1.8103	2.1797	-1.4459	-1.4534	-0.6805	-0.9082	0.1288	1.3720
0.3939	0.1471	-1.7440	-2.8875	-2.0160	1.6442	1.6968	3.6961
1.1313	1.1645	-2.0160	1.6442	0.0899	0.8309	-2.8875	0.8669
-0.1681	-0.6563	-0.5842	-1.0150	1.3517	1.2818	-0.9519	-0.3585
-0.2085	-0.9540	1.1682	0.6902	-1.0944	-1.8561	0.2233	0.7685
-0.5338	0.1204	-0.3950	-1.9299	-2.2852	-2.3245	0.5230	0.0875
-1.4586	0.4026	-1.9898	0.3473	-0.9107	-1.5671	0.3170	3.2004

TABLE 4.6 – Les données de deux action X_1 et X_2 .

La représentation de ces données sous forme d'histogrammes est donnée comme suit :

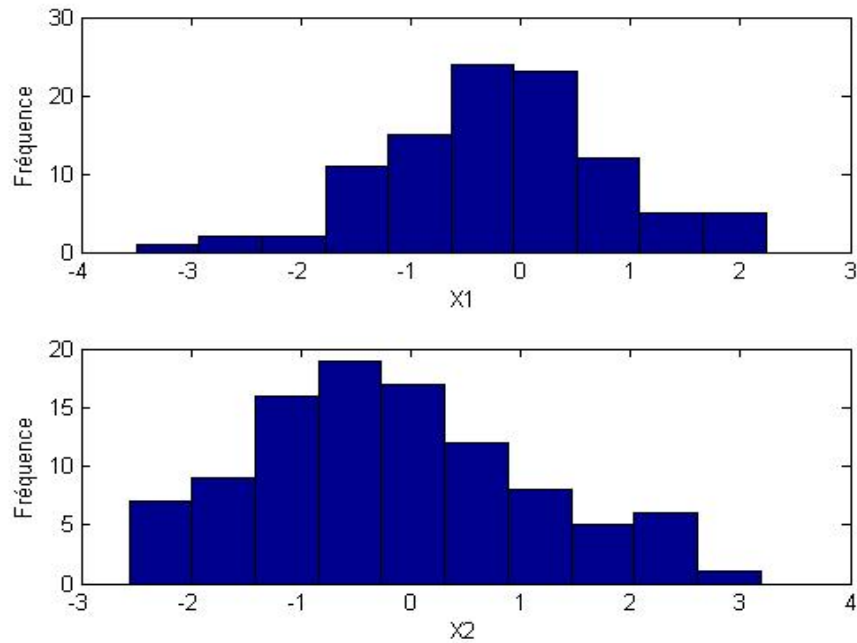
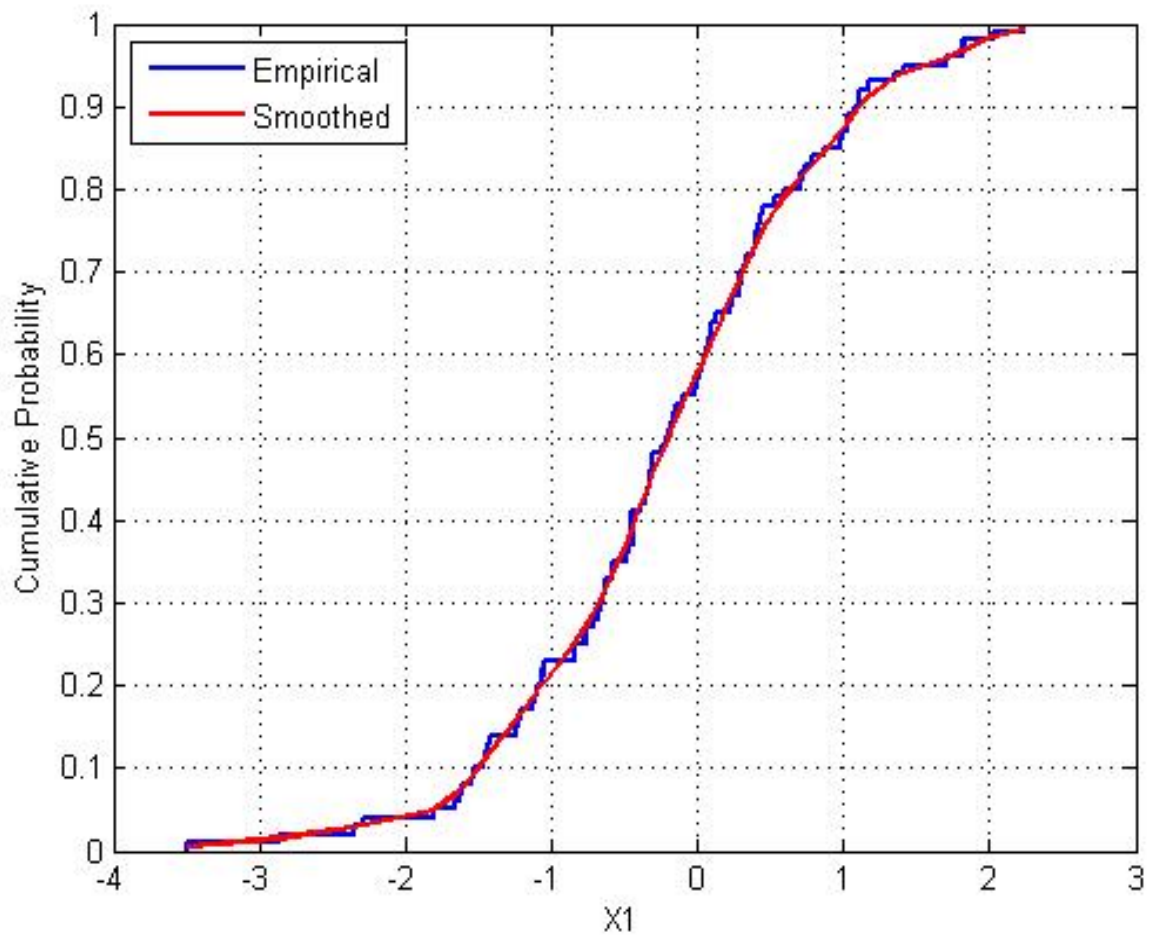


FIGURE 4.7 – Histogramme représente les fréquences des données

Notre objectif est de simuler des entrées (X_1, X_2) qui suivent les mêmes distributions que les données du tableau précédent à l'aide d'une copule. Pour simuler des données bivariées à l'aide d'une copule il est nécessaire de spécifier chacun des éléments suivant :

1. Les distributions marginales de chaque variable.
2. La famille de copules et tous les paramètres de forme.
3. Le coefficient de corrélation entre les deux variables.

Afin de déterminer les distributions marginales de chaque variables, nous allons tracer les fonctions de distribution cumulative (cdf) de chaque entrée et d'interpoler la cdf de chaque variable ou d'utiliser le lissage à noyau. La cdf des données de la première entrée et son interpolation sont données dans la représentation suivante.

FIGURE 4.8 – Comparaison entre la cdf empirique et l'estimation cumulative de cdf lissé par noyau pour X_1

On fait de même pour la cdf des données de la deuxième entrée.

Pour la simulation, on peut expérimenter différentes copules. Ici, nous utilisons une copule de student avec un paramètre de degré de liberté $\nu = 5$. Pour le paramètre de corrélation nous pouvons calculer le tau de kendall ou le paramètre de corrélation linéaire correspondant à la copule choisie, on obtient :

ν	τ	ρ
5	0.5180	0.7268

Ensuite, on génère les vecteurs aléatoires $U = (U_1, U_2)$ à partir de la copule de student et les paramètres précédents.

En utilisant les transformées inverses des cdf obtenues par interpolation ou lissage à noyau on obtient des entrées (X_1, X_2) . Nous donnons ci-dessous la représentation d'un échantillon de taille 1000 et la représentation sous forme d'un histogramme des réalisations de chaque entrée.

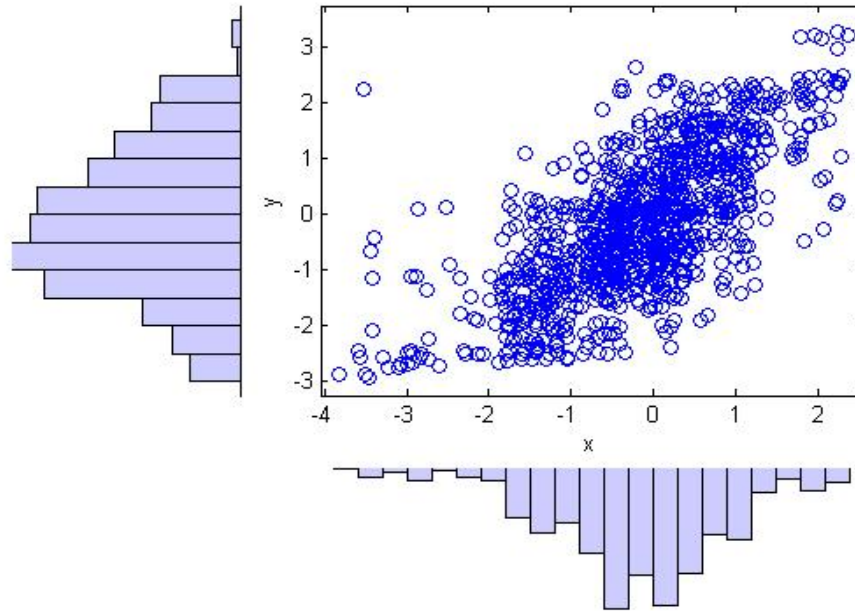


FIGURE 4.9 – histogramme représente les réalisations de chaque entrée.

Les histogrammes marginaux des données simulées sont une version lissée des histogrammes des données d'origine (4.7) .

Remarque

Cet exemple illustre l'utilité des copules dans la simulation des données dépendantes et que les copules s'intègrent naturellement avec d'autres fonctions d'adaptation de distribution.

Conclusion

Dans ce travail, on a d'abord présenté les copules, leurs propriétés principales ainsi que leur structure de dépendance. Ensuite, on a décrit les différentes méthodes d'estimation qui permettent d'évaluer la valeur approximative du paramètre de la copule, on a aussi donné les différentes méthodes de simulation et on a simulé quelques copules par ces méthodes. Puis on a estimé les paramètres de quelques copules. A la fin on a donné un exemple d'application sur des données réelles pour simuler des entées dépendantes (X_1, X_2) .

Bibliographie

- [AG05] Belkacem Abdous and Kilani Ghoudi. Non-parametric estimators of multivariate extreme dependence functions. *Nonparametric Statistics*, 17(8) :915–935, 2005.
- [CFG00] Philippe Capéraà, Anne-Laure Fougères, and Christian Genest. Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis*, 72(1) :30–49, 2000.
- [CJ81] R Dennis Cook and Mark E Johnson. A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society : Series B (Methodological)*, 43(2) :210–218, 1981.
- [CRRL00] Arnaud Costinot, Gaël Riboulet, Thierry Roncalli, and Crédit Lyonnais. Stress testing et théorie des valeurs extrêmes : une vision quantifiée du risque extrême. *Groupe de Recherche Opérationnelle, Crédit Lyonnais, France*, 2000.
- [Dal56] Giorgio Dall’Aglia. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 10(1-2) :35–74, 1956.
- [Deh79] Paul Deheuvels. La fonction de dépendance empirique et ses propriétés. un test non paramétrique d’indépendance. *Bulletins de l’Académie Royale de Belgique*, 65(1) :274–292, 1979.
- [Dev86] Luc Devroye. Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pages 260–265, 1986.
- [DMK01] Dominique Drouet Mari and Samuel Kotz. *Correlation and dependence*. World Scientific, 2001.
- [Fér56] R Féron. Sur les tableaux de corrélation dont les marges sont données. cas de l’espace a trois dimensions. In *Annales de l’ISUP*, number 1, pages 3–12, 1956.

- [FM53] Robert Fortet and Edith Mourier. Convergence de la répartition empirique vers la répartition théorique. In *Annales scientifiques de l'École Normale Supérieure*, volume 70, pages 267–285, 1953.
- [Fré51] Maurice Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, 3^e e serie, Sciences, Sect. A*, 14 :53–77, 1951.
- [Gal75] Janos Galambos. Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association*, 70(351a) :674–680, 1975.
- [Gen87] Christian Genest. Frank's family of bivariate distributions. *Biometrika*, 74(3) :549–555, 1987.
- [GM86a] Christian Genest and R Jock MacKay. Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian journal of statistics*, 14(2) :145–159, 1986.
- [GM86b] Christian Genest and R Jock MacKay. Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian journal of statistics*, 14(2) :145–159, 1986.
- [GR89] Christian Genest and Louis-Paul Rivest. A characterization of gumbel's family of extreme value distributions. *Statistics & probability letters*, 8(3) :207–211, 1989.
- [Gum60] Emil Julius Gumbel. Distributions des valeurs extremes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris*, 9 :171–173, 1960.
- [HF06] Laurens Haan and Ana Ferreira. *Extreme value theory : an introduction*, volume 3. Springer, 2006.
- [Hut90] Timothy P Hutchinson. Continuous bivariate distributions emphasising applications. Technical report, 1990.
- [JX96] Harry Joe and James Jianmeng Xu. The estimation method of inference functions for margins for multivariate models. 1996.
- [KM97] Claudia Klüppelberg and Thomas Mikosch. *Modelling extremal events for insurance and finance*. Springer, 1997.
- [KS75] George Kimaldorf and Allan Sampson. One-parameter families of bivariate distributions with fixed marginals. *Communications in Statistics-Theory and Methods*, 4(3) :293–301, 1975.
- [Lin65] C-H Ling. Representation of associative functions. *Publ. Math. Debrecen*, 12 :189–212, 1965.
- [Mau85] Sylvette Maury. Influence de la question dans une épreuve relative à la notion d'indépendance (influence of the question in a task on the concept of independence). *Educational studies in mathematics*, pages 283–301, 1985.
- [MFM08] Paul Meylan, Anne-Catherine Favre, and André Musy. *Hydrologie fréquentielle : une science prédictive*. PPUR presses polytechniques, 2008.

- [MO67] Albert W Marshall and Ingram Olkin. A generalized bivariate exponential distribution. *Journal of applied probability*, 4(2) :291–302, 1967.
- [MS06] Yannick Malevergne and Didier Sornette. *Extreme financial risks : From dependence to risk management*. Springer Science & Business Media, 2006.
- [Nel91] Roger B Nelsen. Copulas and association. *Advances in probability distributions with given marginals : beyond the copulas*, pages 51–74, 1991.
- [Nel07] Roger B Nelsen. *An introduction to copulas*. Springer science & business media, 2007.
- [PBG05] Christian Partrat, Jean-Luc Besson, and Fabrice Genest. Assurance non-vie : modélisation, simulation. (*No Title*), 2005.
- [PLA17] Frédéric PLANCHET. Modèles financiers en assurance et analyses dynamiques. *Support de cours ISFA*, 2017.
- [Sch61] B Schweizer. Associative functions and statistical triangle inequalities. *Publicationes Mathematicae, Debrecen*, 8 :169–186, 1961.
- [Skl59] M Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8 :229–231, 1959.
- [Taw88] Jonathan A Tawn. Bivariate extreme value theory : models and estimation. *Biometrika*, 75(3) :397–415, 1988.
- [Vos08] David Vose. *Risk analysis : a quantitative guide*. John Wiley & Sons, 2008.

RÉSUMÉ

Une copule est un outil mathématique venant de la théorie des probabilités permettant de modéliser la dépendance entre les variables aléatoires, de fournir des expressions relativement simples des structures de dépendance de ces variables aléatoires et de distinguer le comportement de leurs distributions jointes. L'objectif du travail consiste à introduire les différentes familles des copules et leurs propriétés, ainsi, de présenter les différentes méthodes d'estimation. Fournir ensuite des tests d'adéquation pour le choix de la bonne copule. Puis donner les différentes méthodes de simulation .

Mots clés : Copule, distribution marginale, tau de kendall, dépendance.

ABSTRACT

A copula is a mathematical tool coming from the probabilities theory to model dependence between the random variables, to provide expressions relatively simple dependency structures of these random variables and to distinguish the behavior of their joint distributions. The objective of the work is to introduce the different families of copulas and their properties, as well, to present the different estimate methods . Then provide testing of adequacy for the choice of the good copula. Then give the different simulation methods .

Key words : Copula, marginal distribution, Kendall's tau, dépendence

