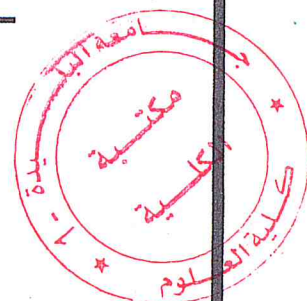


République Algérienne démocratique et populaire
Minister de l'enseignement supérieur
de la recherche Scientifique
Université SAAD DAHLAB, BLIDA 01

Faculté des Sciences
Département de mathématiques



Mémoire

Mémoire pour L'obtention du diplôme de master
En mathématiques

Option:

Modélisation Stochastique et Statistique

Thème :

**Modélisation de la dépendance par des copules, application sur la
sécheresse en algérie**

Présenté par:

- Elfertas Asmaa
- Mahdi Salma

Encadré par:

DR.A.Rassoul

Soutenue devant le jury composé de :

Nom et prénom	Grade	Établissement	En qualité de
Tami Omar	M.A.A	U.S.D.BLIDA	Président
Frihi.R	M.A.A	U.S.D.BLIDA	Examineur
Boukhari.M	M.A.A	U.S.D.BLIDA	Examineur

Année Universitaire 2017/2018

REMERCIEMENTS

Tout travail réussi dans la vie nécessite en premier lieu la bénédiction de Dieu, et ensuite l'aide et le support de plusieurs personnes. Je tiens donc à remercier et à adresser ma reconnaissance à toute personne qui m'a aidé de loin ou de près afin de réaliser l'un de mes rêves...

Je remercie tous les professeurs et les enseignants de mon cursus universitaire qui ont Contribué à ma formation.

Je remercie vivement et les membres du jury pour le soutien et l'intérêt qu'ils ont portés à ce travail.

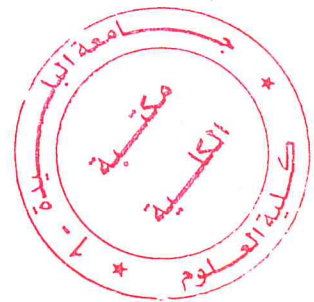
Je remercie fortement mon prometteur Abdelaziz RASSOUL

Je remercie fortement et de tout mon cœur mes vifs remerciements vont également à mon meilleur ami pour l'appui moral qu'ils m'ont témoigné,

Pour terminer, j'adresse mon grand amour à mes parents, mon père pour son sacrifice ma mère pour sa tendresse, mon frère, mes sœur, mon mari, C'est grâce à leur amour et leurs sacrifices que ce mémoire a été mené à bonne fin. Mon plus grand souhait dans cette vie, c'est de les voir toujours à côté de moi, en bonne santé, heureux et que la paix

soit avec eux.

Grand merci à vous tous



dédicace

A chaque fois qu'on achève une étape importante dans notre vie, on fait une pose pour regarder en arrière et se rappeler toutes ces personnes qui ont partagés avec nous tous les bons moments de notre existence, mais sur tout les mauvais. Ces personnes qui nous ont aidés sans le leur dire, soutenus sans réserve, aimé sans compter, ces personnes à qui notre bonheur devient directement le leur, à qui un malheur en nous, en eux se transforme en pleur.

Je dédie ce modeste travail à tous ceux que j'aime mais surtout :

A mes parents malika et nasseraddine qui ont toujours été les étoiles de mon ciel et ont illuminé mon chemin depuis ma naissance, je ne les remercierai jamais assez et a mon mari brahim aguenini .

Je le dédie également :

A mon promoteur MR RASSOUL

A mon meilleur ami elfertas asma

A mes sœurs: amina , nadjat , zola ,saida, et mes frères: mahmoud , omar , ali .et leurs enfants : donya zed ,nada, nihal, yasin, anas , inas, ayman , adam , akram , ishak , anfal , ahmad .

A mon amie d'enfance badiaa , amina .

A tous les enseignants de USDB qui m'ont accompagné durant mes études ;

A toute personne utilisant ce document pour un bon usage.

mahdi salma

dédicace

A chaque fois qu'on achève une étape importante dans notre vie, on fait une pose pour regarder en arrière et se rappeler toutes ces personnes qui ont partagés avec nous tous les bons moments de notre existence, mais sur tout les mauvais. Ces personnes qui nous ont aidés sans le leur dire, soutenus sans réserve, aimé sans compter, ces personnes à qui notre bonheur devient directement le leur, à qui un malheur en nous, en eux se transforme en
pleur.

Je dédie ce modeste travail à tous ceux que j'aime mais surtout :

A mes parents ferroum fatma et alaa qui ont toujours été les étoiles de mon ciel et ont illuminé mon chemin depuis ma naissance, je ne les remercierai jamais assez .

Je le dédie également :

A mon promoteur MR RASSOUL

A mon fiancé mahmoud

A mon meilleur ami mahdi salma et a moi mém.

A tout ma famille soualmi,ferroum,elfertas

A ma belle mère malika et mon père nasseraddine

A tout la famille mahdi mes sœurs et frères.

A tous les enseignants de USDB qui m'ont accompagné durant mes études ;

A toute personne utilisant ce document pour un bon usage.

elfertas asma

Table des matières

0.1	Résumé	6
0.2	Abstract	6
0.3	Introduction générale	7
1	Indépendance	9
1.1	De l'indépendance à la dépendance	9
1.2	Couples des variables aléatoires	10
1.2.1	Densités marginales	11
1.3	Variables indépendantes	11
1.4	Fonction génératrice	13
1.5	Indépendance	14
1.5.1	Distributions marginales	14
1.6	Fonction de répartition	15
1.6.1	Cas scalaire	15
1.6.2	Cas d'un couple aléatoire	16
2	DEPENDANCE ET COPULES	17
2.1	Dépendance ou corrélation ?	17
2.2	Qu'est-ce qu'une copule ?	18
2.2.1	Théorème de Sklar	18
2.2.2	Cas de deux variables	20
2.3	Bornes de Fréchet-Hoeffding	21
2.4	Mesures de concordance	22
2.5	Dépendance de queue	24
2.5.1	Coefficients de dépendance de queue	24
2.6	Familles des copules	25
2.6.1	Les Copules elliptiques	25
2.6.2	Copule normale	26
2.6.3	Copule des valeurs extrêmes	28
2.6.4	Copule archimédienne	30
2.7	Quelques copules bivariées	33
2.7.1	Copule produit	33

2.7.2	Copule de Farlie- Gumbel- Morgenstern (FGM)	35
2.7.3	Copules de survie	35
2.7.4	Copule HRT	36
2.7.5	Copule empirique	36
2.7.6	Quelques copules particulier et ses paramètres	37
2.8	Simulation de copules bivariées	37
2.8.1	Méthode des distributions	37
2.8.2	Méthode des distributions conditionnelles	38
2.8.3	Les méthodes dites analytiques	39
2.8.4	La méthode des quantiles empiriques	40
2.9	Conclusion	41
3	Testes statistiques pour les copules	42
3.1	Introduction	42
3.2	Méthode paramétrique d'estimation des copules	43
3.2.1	Fonction de répartition empirique	43
3.2.2	Copules empiriques	45
3.3	Méthodes paramétriques	46
3.3.1	Méthode des moments	46
3.3.2	Maximum de vraisemblance	46
3.4	Sélection et adéquation de la copule	47
3.4.1	Teste de Khi deux bidimensionnel	49
4	Application sur la modélisation de la sécheresse	51
4.1	Définition de la sécheresse	51
4.2	Type de la sécheresse	51
4.2.1	Sécheresses météorologiques	51
4.2.2	Sécheresses agricoles	52
4.2.3	Sécheresses hydrologiques	52
4.2.4	Sécheresse socio-économique	53
4.3	Prévision de la sécheresse	53
4.4	Analyse des caractéristiques de la sécheresse :	55
4.5	Indice de précipitation normalisé	56
4.5.1	Description de l'indice normalisé de précipitations (SPI)	56
4.5.2	Valeurs de l'indice de précipitations normalisé pour une période plus ou moins longue	58
4.6	Indices Hydrométriques (Stream flow Drought Index (SDI))	60
4.7	Ajustement des paramètres de la sécheresse par des lois de probabilités	61
4.7.1	SPI de 3 mois	61
4.7.2	Analyse statistiques de spi de 6 mois	62

4.8	Application de la copule tri-variée sur les caractéristiques de la sécheresse	63
4.8.1	Matrice de variance covariance	64
4.9	Modèles de copule bivariés	65
4.10	Modèles de copule trivariés	67
	Conclusion	67

Liste des tableaux

2.1	Les copules des valeurs extrêmes les plus utilisées	30
2.2	Les copules Archimédiennes les plus utilisées	31
2.3	Les paramètres de dépendance des copules Archimédiennes les plus utilisées	33
2.4	Quelques copules avec leurs paramètres et le codage sous R . . .	37
4.1	Valeurs de l'indice SPI.	58
4.2	Définition des états de sécheresse hydrologique à l'aide de SDI.	60
4.3	Paramètres statistiques des caractéristiques d'une sécheresse pour spi de trois mois	61
4.4	Ajustement de la durée de la sécheresse pour un spi de trois mois	61
4.5	Ajustement de l'intensité de la sécheresse pour un spi de trois mois	62
4.6	Ajustement de la sévérité de la sécheresse pour un spi de trois mois	62
4.7	Ajustement de la durée de la sécheresse pour un spi de six mois	62
4.8	Ajustement de l'intensité de la sécheresse pour un spi de six mois	63
4.9	Ajustement de la sévérité de la sécheresse pour spi de six mois	64
4.10	Corrélation des caractéristiques de la sécheresse pour les spi de 03 mois	64
4.11	Coefficients de dépendance entre les différentes caractéristiques des spi de 6 mois	64
4.12	Résultats d'ajustement par des copules bivariées pour le spi de 03 mois	67
4.13	Résultats d'ajustement par des copules trivariées	67

Table des figures

2.1	Schéma simplifié du théorème de Sklar avec F_x et F_y les fonctions marginales de la fonction de distribution bivariée F_{xy} . .	19
2.2	Densité du copule de Student bivariée	28
2.3	Densité des copules : Frank, Gumbel et Clayton	34
2.4	Lignes de niveaux des copules : Gumbel, Frank et Clayton . .	34
4.1	distributions d'ajustement des durrée, intensité et sévérité . .	63
4.2	Ajustement du durée, intensité et sévérité pour spi de six mois	65
4.3	Graphes de corrélations et ajustements des coefficients	66

0.1 Résumé

Ces dernières décennies, nous avons assisté à l'émergence du concept de copule en modélisation statistique. Cet essor est justifié par le fait que les copules permettent de faire une analyse séparée des marges et de la structure de dépendance induite par une distribution statistique. Cette séparation facilite l'incorporation de lois non gaussiennes et la prise en compte des dépendances non linéaires entre les variables aléatoires. L'applications des copules dans le domaines d'hydrologie, puisqu'il existe beaucoup de familles de copules bivariées, il sera toujours possible à l'utilisateur d'en choisir une qui lui convienne. Ce mémoire s'inscrit dans ce contexte. Nous proposons une modélisation des variables qui caractérisent la sécheresse par des copules multivariées avec des propriétés, elle permet de prendre en compte les différents degrés de dépendance entre les différentes paires de variables. Les méthodes classiques d'inférence nous permet d'estimer leurs paramètres. Nous abordons également l'estimation de copules dans le cas général, et exhibons les propriétés asymptotiques d'un estimateur des moindres carrés pondérés basé sur les coefficients de dépendance sans faire appel à des hypothèses sur les copules. Les modèles et méthodes proposés sont appliqués sur des données hydrologiques (précipitations).

0.2 Abstract

In recent decades, we have witnessed the emergence of the copula concept in statistical modeling. This development is justified by the fact that the copulas make it possible to make a separate analysis of the margins and the dependence structure induced by a statistical distribution. This separation facilitates the incorporation of non-Gaussian laws and the taking into account of nonlinear dependencies between the random variables. The application of copula in the field of hydrology, since there are many families of bivariate copulas, it will always be possible for the user to choose one that suits him. This memoir fits into this context. We propose a modeling of the variables that characterize the drought by multivariate copulas with properties, it allows to take into account the different degrees of dependence between the different pairs of variables. The classical methods of inference allow us to estimate their parameters. We also discuss copula estimation in the general case, and exhibit the asymptotic properties of a weighted least squares estimator based on dependence coefficients without resorting to copula assumptions. The proposed models and methods are applied to hydrological (precipitation) data.

0.3 Introduction générale

En statistique, la corrélation n'est qu'une mesure du lien linéaire entre les variables à l'étude, celle-ci ne repère pas les autres formes de liens. Pour récupérer la dépendance exacte, et ce, quel que soit le type de lien entre les variables, nous utilisons une fonction appelée copule.

Les copules, comme un modèle de la distribution multivariée, largement utilisé dans la finance, l'analyse actuarielle, a gagné la popularité de plus en plus dans l'étude hydrologique.

Le terme copule (copula) vient du mot latin "copeulae", qui signifie au sens figuré, liaison, lien, alliance ou union. Les copules représentent un outil innovant pour modéliser la structure de dépendance de plusieurs variables aléatoires. Elle est introduite par Sklar (1959) dans des études sur les espaces de probabilité métriques, ce dernier a élaboré un théorème fondamental dans la théorie des copules. Ce théorème permet la liaison entre la densité conjointe d'un vecteur aléatoire et ses densités marginales.

Le coefficient de corrélation linéaire de Pearson est un indicateur qui mesure la dépendance entre deux variables aléatoires X et Y qui prend toute valeur de l'intervalle $[-1, 1]$, cette mesure de dépendance facile à calculer est un indicateur performant lorsque la relation de dépendance est linéaire et l'univers considéré gaussien.

Il est très utile pour les familles de distributions elliptiques (car pour ces distributions la non corrélation implique l'indépendance). Cependant cette mesure de dépendance souvent utilisée par les praticiens possède plusieurs limites. Ainsi la non corrélation de deux variables non gaussiennes ne signifie pas une absence de dépendance, c'est à dire bien que le coefficient de corrélation soit nul, il y a une dépendance entre X et Y , donc ce dernier possède des insuffisances d'information.

En hydrologie, Favre et al. (2004) présentent notamment deux applications : la première concerne la modélisation des débits amont et aval d'un cours d'eau avec tributaire, et la seconde concerne la modélisation conjointe des débits maximaux annuels en fonction des volumes écoulés. De Michele et al. (2005) utilisent la copule de Gumbel pour modéliser la dépendance positive entre le débit de pointe et le volume de crue. Zhang & Singh (2006) utilisent les copules bivariées pour l'analyse fréquentielle des variables : débit de pointe et volume écoulé et volume de crue et sa durée.

Plus spécifiquement pour l'étude de la pluie, on peut mentionner les travaux suivants : Salvadori & De Michele (2004) ont utilisé la théorie des copules pour critiquer le calcul univarié des périodes de retour et ont montré que l'usage des copules permet de simplifier les calculs des isolignes des périodes de retour, dans le cas inconditionnel et conditionnel. Grimaldi et al. (2005)

ont étendu l'étude aux copules trivariées : ils ont analysé conjointement des données observées de hauteur critique, d'intensité maximale et de hauteur totale de pluie. Afin de déterminer la distribution conjointe trivariée, une analyse bivariée du couple intensité maximale–hauteur totale, conditionné par la hauteur critique est effectuée grâce à une fonction 3-copula. Salvadori & De Michele (2006) ont établi la dépendance entre l'intensité moyenne, la durée et l'épisode sec via une fonction 3-copula pour caractériser la structure temporelle des épisodes pluvieux.

L'analyse de la sécheresse, en tant que évènement extrême, consiste généralement à caractériser sa sévérité, sa durée et son intensité. Similaire à la plupart des problèmes hydrologiques, les variables telles que les caractéristiques de la sécheresse ne sont couramment pas indépendantes. Dans cette étude nous utilisons les fonctions copules pour analyser les caractéristiques de la sécheresse dans le bassin versant du Hodhna, Wilaya de M'sila sur la période 1979-2013 en se basant sur l'indice hydrométrique (SPI).

Dans ce mémoire on commence par présenter des mesures de dépendance entre les variables ; celles ci sont basées sur les rangs des observations. En outre, on expose des méthodes graphiques de prospection de la dépendance "moyenne" et asymptotique qui appuient les mesures. Ensuite, pour les commodités du lecteur, les copules archimédiennes sont détaillées. La section qui suit est consacrée à l'identification des copules qui peuvent modéliser la loi conjointe. L'ajustement de celles-ci est fait grâce à des comparaisons graphiques et des tests statistiques. Enfin, la dernière section utilise la méthodologie proposée pour l'étude des variables intensité–durée–sévérité d'une sécheresse.

L'organisation de ce mémoire est la suivante : au premier chapitre, on rappelle certains

concepts de dépendance et indépendance des variables aléatoires. Au deuxième chapitre, nous présentons la théorie générale des copules afin d'introduire le lecteur aux outils utilisés dans les chapitres suivants. Dans le troisième chapitre on présente l'inférence statistiques sur les copules, en particulier estimation des paramètres des copules et les testes statistiques sur les copules, les tests d'adéquation pour les copules qui s'ajustent le mieux aux données. Das le dernier chapitre nous présentons une modélisation basée sur les copules dans le cas où l'on utilise des données réelles en tant que données météorologiques. Ainsi, on présente une méthodologie pour déterminer la dépendance entre des variables qui caractérisent la sécheresse, en cas de dépendance, on l'ajuste par des copules convenables.

Chapitre 1

Indépendance

1.1 De l'indépendance à la dépendance

Il y a 250 ans, Thomas Bayes définissait l'indépendance de la manière suivante : « events are independent when the happening of any one of the does neither increase nor abate the probability of the rest ». était toutefois un peu plus précis lorsqu'il écrivait « two events are said independent when the probability of the happening of either of them is unaffected by our expectation of the occurrence of failure of the other ».

Pour une définition formelle, Laplace affirmait que « si les évènements sont indépendants les uns des autres, la probabilité de l'existence de leur ensemble est le produit de leur probabilités particulières ». Autrement dit, deux évènements A et B sont indépendants si et seulement si :

$$P(A|B) = P(A \text{ et } B) = P(A, B) = P(A) \times P(B).$$

En utilisant la définition de l'espérance conditionnelle,

$$P(A \text{ et } B) = P(A|B) \times P(B),$$

l'indépendance se caractérise également par la relation :

$$P(A|B) = P(A).$$

On retrouve alors l'idée énoncée par Boole. Par opposition, on dira que deux évènements sont dépendants si

$$P(A \text{ et } B) = P(A, B) \neq P(A) \times P(B).$$

1.2 Couples des variables aléatoires

Donnons à présent les définitions précises relatives aux couples aléatoires, ces définitions généralisent celles concernant les variables aléatoires réelles.

Définition 1.1 *On appelle couple aléatoire – ou vecteur aléatoire de dimension deux – toute paire $Z = (X, Y)$ de variables aléatoires réelles. On a donc :*

$$Z : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}^2, \omega \mapsto Z(\omega) = (X(\omega), Y(\omega)).$$

Sur \mathbb{R} , nous avons considéré la plus petite tribu engendrée par les intervalles i.e. la tribu borélienne. Pour travailler dans \mathbb{R}^2 , on introduit la tribu borélienne de \mathbb{R}^2 , $\mathcal{B}(\mathbb{R}^2)$, qui est « la plus petite tribu » sur \mathbb{R}^2 contenant tous les pavés $I \times J$ où I et J sont deux intervalles de \mathbb{R} . Cette tribu contient tous les pavés du type $A \times B$ où A et B sont deux boréliens de \mathbb{R} mais elle contient également des ensembles beaucoup plus complexes qui ne sont pas des produits cartésiens.

Si Z est un couple aléatoire, on montre que pour tout borélien \mathcal{C} de \mathbb{R}^2 , l'image réciproque de \mathcal{C} par Z ,

$$Z^{-1}(\mathcal{C}) = \{\omega \in \Omega, Z(\omega) \in \mathcal{C}\},$$

est un élément de \mathcal{F} . Ceci donne un sens à la définition suivante.

Définition 1.2 *Soit $Z : \Omega \rightarrow \mathbb{R}^2$ un couple aléatoire. L'application \mathbb{P}_Z définie par*

$$\mathbb{P}_Z : \mathcal{B}(\mathbb{R}^2) \rightarrow [0, 1], \mathcal{C} \mapsto \mathbb{P}_Z(\mathcal{C}) = \mathbb{P}(Z^{-1}(\mathcal{C}))$$

est une mesure de probabilité appelée loi de Z .

Remarque 1.1 – *Si \mathcal{C} est un borélien de \mathbb{R}^2 et $Z = (X, Y)$ un couple aléatoire, l'ensemble $Z^{-1}(\mathcal{C})$, noté $\{Z \in \mathcal{C}\}$ par les probabilistes, n'est pas toujours très facile à déterminer.*

– *Néanmoins, si \mathcal{C} est le pavé $A \times B$ la situation est plus favorable, en effet, on a dans ce cas*

$$Z^{-1}(\mathcal{C}) = \{\omega \in \Omega, (X(\omega), Y(\omega)) \in A \times B\} = X^{-1}(\mathcal{A}) \times Y^{-1}(\mathcal{B}),$$

$$\text{soit encore } \{Z \in A \times B\} = \{X \in A, Y \in B\}.$$

La tribu des boréliens de \mathbb{R}^2 contient des parties complexes, toutefois, elle est engendrée par la classe des produits d'intervalles réels, ce qui conduit au résultat suivant :

Proposition 1.1 *Soit $Z = (X, Y)$ un couple aléatoire. La loi \mathbb{P}_Z est caractérisée par*

$$\mathbb{P}_Z(I \times J) = \mathbb{P}(X \in I, Y \in J),$$

pour tout couple (I, J) d'intervalles réels.

1.2.1 Densités marginales

Lois marginales du couple. Si $Z = (X, Y)$ est un couple aléatoire, les lois marginales de Z sont les lois des v.a.r. X et Y . Il est important de noter que si on connaît la loi du couple \mathbb{P}_Z on connaît les lois marginales. En effet, si A est un intervalle, une réunion d'intervalles ou un borélien de \mathbb{R} ,

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in \mathbb{R}) = \mathbb{P}_Z(A \times \mathbb{R}),$$

de même, pour la v.a.r. Y ,

$$\mathbb{P}_Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(X \in \mathbb{R}, Y \in B) = \mathbb{P}_Z(\mathbb{R} \times B).$$

1.3 Variables indépendantes

Définition 1.3 Soit X et Y deux variables aléatoires. On dit que X et Y sont indépendantes si :

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

Théorème 1.1 Soit X et Y deux variables aléatoires discrètes et indépendantes à valeurs dans E_1 et E_2 respectivement, soit $f : E_1 \rightarrow \mathbb{R}$, et $g : E_2 \rightarrow \mathbb{R}$, telles que $f(X)$ et $g(Y)$ sont intégrables alors $f(X)g(Y)$ est intégrable et on a :

$$E(f(X)g(Y)) = E(f(X))E(g(Y)). \quad (1.1)$$

Preuve: Posons $Z = (X, Y)$ et $h(z) = f(x)g(y)$, alors Z est à valeurs dans $E = E_1 \times E_2$, de plus d'après l'hypothèse d'indépendance, c'est à dire

$$P((X, Y) = (x, y)) = P(X = x)P(Y = y).$$

Alors

$$\begin{aligned} E(|h(Z)|) &= \sum_{k=1}^n |h(z_k)|P(Z = z_k) \\ &= \sum |f(x_k)||g(y_l)|P(X = x_k)P(Y = y_l) \\ &= \sum |f(x_k)|P(X = x_k) \sum |g(y_l)|P(Y = y_l) \end{aligned}$$

car $f(X)$ et $g(Y)$ sont intégrables, donc $E(|h(Z)|) < \infty$, par conséquent $h(Z)$ est intégrable. Par un raisonnement analogue à ci-dessus, mais sans la valeur absolue on obtient l'égalité (1.1). ■

Théorème 1.2 (Variance d'une somme de variables aléatoires indépendantes)
 Soit X_1 et X_2 deux variables aléatoires indépendantes, alors

$$V(X_1 + X_2) = V(X_1) + V(X_2). \quad (1.2)$$

Preuve: On sait que

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2cov(X_1, X_2),$$

où

$$cov(X_1, X_2) = E((X_1 - EX_1)(X_2 - EX_2))$$

est la covariance de X_1 et X_2 , mais d'après (1.1) on a

$$\begin{aligned} cov(X_1, X_2) &= E((X_1 - EX_1)(X_2 - EX_2)) \\ &= E((X_1 - EX_1))E((X_2 - EX_2)) = 0. \end{aligned}$$

■

Remarque 1.2 Si deux variables aléatoires X et Y sont indépendantes alors elles sont non corollées ($cov(X, Y) = 0$).

De plus, pour deux variables aléatoires de carré intégrable, il est possible d'obtenir une majoration de $cov(X, Y)$ à partir des variances de X et Y :

Proposition 1.2 (Inégalité de Cauchy Schwarz) Soient X et Y deux variables aléatoires de carré intégrable. On a

$$\begin{aligned} |E(XY)| &\leq E(|XY|) \leq \sqrt{E(X^2)E(Y^2)} \\ |cov(X, Y)| &\leq \sqrt{var(X)var(Y)}. \end{aligned}$$

Preuve: Le deuxième point s'obtient en appliquant le premier à $X - E(X)$ et $Y - E(Y)$.

Le premier point se montre de la même façon que dans le cas classique (produit scalaire de deux vecteurs) : on étudie le polynôme

$$R(\lambda) = E(X^2)\lambda^2 - 2E|XY|\lambda + E(Y^2).$$

Ce polynôme se factorise en

$$R(\lambda) = E[(\lambda|X| - |Y|)^2].$$

Il n'admet donc pas deux racines réelles distinctes, ce qui signifie que son discriminant (réduit) est négatif :

$$(E|XY|)^2 - E(X^2)E(Y^2) \leq 0.$$

On étudie de la même façon le cas d'égalité : le discriminant n'est nul que si le polynôme admet une racine réelle double. Dans ce cas, il existe donc λ_0 tel que

$$E[\lambda_0(|X| + |Y|)^2] = 0$$

l'espérance d'une variable aléatoire positive ne pouvant être nulle que si la variable aléatoire est (presque sûrement) nulle, on aura alors : $|Y| = -\lambda_0 |X|$, presque sûrement. ■

1.4 Fonction génératrice

Soit X une variable discrète à valeurs dans \mathbb{N} .

Définition 1.4 On appelle fonction génératrice de X , l'application $g_X : \mathbb{C} \rightarrow \mathbb{C}$ définie par :

$$g_X(s) = E(s^X) = \sum_{x \in X(\Omega)} s^x P(X = k), |s| \leq 1.$$

Remarque 1.3 La connaissance de la fonction génératrice g_X permet le calcul des moments de X , en effet on a :

$$EX = g'_X(1)$$

et

$$E(X^2) = g''_X(1) + g'_X(1), \quad (1.3)$$

où g'_X et g''_X sont les dérivées première et seconde de la fonction g_X .

Théorème 1.3 Soit X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{N} , alors :

$$g_{X+Y}(s) = g_X(s)g_Y(s), \quad (1.4)$$

Preuve:

$$g_{X+Y}(s) = E(s^{X+Y}) = E(s^X s^Y) = E(s^X)E(s^Y),$$

d'après. ■

Définition 1.5 La fonction génératrice d'un couple de variables aléatoires discrètes positives est la fonction définie sur $[0, 1]^2$ par

$$G_{X,Y}(s, t) = E(s^X t^Y) = \sum_{x \in X(\Omega), y \in Y(\Omega)} s^x t^y P(X = x, Y = y)$$

Proposition 1.3 La fonction génératrice détermine la loi du couple (X, Y) au sens où si deux couples de variables aléatoires positives ont la même fonction génératrice, alors ils suivent la même loi.

La fonction génératrice de (X, Y) permet de retrouver par exemple :

- la fonction génératrice de X : $G_X(s) = G_{(X,Y)}(s, 1)$ et de Y ,
- l'espérance de X si X est intégrable :

$$E(X) = \frac{\partial G_{(X,Y)}(1, 1)}{\partial s},$$

- l'espérance de X^2 si X est de carré intégrable :

$$E(X^2) = \frac{\partial^2 G_{(X,Y)}(1, 1)}{\partial s^2} + \frac{\partial G_{(X,Y)}(1, 1)}{\partial s},$$

- l'espérance de XY si X et Y sont de carré intégrable :

$$E(XY) = \frac{\partial^2 G_{(X,Y)}(1, 1)}{\partial s \partial t}.$$

1.5 Indépendance

1.5.1 Distributions marginales

Définition 1.6 Soient X et Y deux variables aléatoires à valeurs dans \mathbb{R} tels que (X, Y) admette la densité $f_{X,Y}(x, y)$. Alors on définit les densités marginales de X et de Y par :

$$f_X(x) = \int f_{X,Y}(x, y) dy \text{ et } f_Y(y) = \int f_{X,Y}(x, y) dx.$$

Théorème 1.4 Soit $X = (X_1, X_2)$ de densité f_1 et f_2 . Alors X_1 et X_2 sont des variables aléatoires indépendantes, si :

$$f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2), \quad (1.5)$$

où f_{X_i} est la densité marginale de la variable X_i . Réciproquement si f_X a la forme (1.5) où les f_{X_i} , sont des densités de probabilité, alors les X_i , sont indépendantes.

Théorème 1.5 Soit $X = (X_1, X_2)$ de densité f_X et de fonction caractéristique $\varphi_X(u)$. Pour que les variables aléatoires $X_i, 1 \leq i \leq 2$, soient indépendantes, il faut et il suffit

$$\varphi_X(u_1, u_2) = \varphi_{X_1}(u_1)\varphi_{X_2}(u_2), \quad (1.6)$$

où φ_{X_i} est la fonction caractéristique de la variable X_i .

Preuve: Supposons que les X_k sont indépendantes, alors $f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$, donc

$$\varphi_X(u) = \int e^{i(u_1x_1+u_2x_2)} f_X(x) dx = \int e^{iu_1x_1} e^{iu_2x_2} e^{iu_1x_1} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2,$$

donc d'après Fubini

$$\varphi_X(u) = \int_{-\infty}^{+\infty} e^{iu_1x_1} f_{X_1}(x_1) dx_1 \int_{-\infty}^{+\infty} e^{iu_2x_2} f_{X_2}(x_2) dx_2 = \varphi_{X_1}(u_1)\varphi_{X_2}(u_2),$$

Inversement, supposons que (1.5) est satisfaite, Soit $\tilde{X}_j, 1 \leq j \leq 2$, une suite de variables indépendantes telles que chaque \tilde{X}_j ait la même loi que X_j , et $\tilde{X}_j = (\tilde{X}_1, \tilde{X}_2)$. D'après la partie directe du théorème :

$$\varphi_{\tilde{X}}(u_1, u_2) = \varphi_{\tilde{X}_1}(u_1)\varphi_{\tilde{X}_2}(u_2),$$

et comme $\varphi_{\tilde{X}_i}(u_i) = \varphi_{X_i}(u_i)$, on a $\varphi_{\tilde{X}}(u) = \varphi_X(u)$, donc, par le théorème X et \tilde{X} ont la même loi ; d'après le théorème 9

$$f_X(x_1, x_2) = f_{\tilde{X}}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2).$$

■

1.6 Fonction de répartition

1.6.1 Cas scalaire

Définition 1.7 Soit X une variable aléatoire de densité f , la fonction de répartition de X est définie par la fonction F , telle que : $F : \mathbb{R} \rightarrow [0, 1]$ et

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Remarque 1.4 F est une fonction non décroissante telle que $F(-\infty) = 0$, $F(+\infty) = 1$, et $F'(x) = f(x)$.

1.6.2 Cas d'un couple aléatoire

Définition 1.8 Soit $X = (X_1, X_2)$ un couple aléatoire de densité $f(x)$, la fonction de répartition de X est définie par : $F : \mathbb{R}^2 \rightarrow [0, 1]$ et si $x = (x_1, x_2)$ alors :

$$F(x) = P(X \leq x) = P(X_1 \leq x_1, X_2 \leq x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(t_1, t_2) dt_1 dt_2.$$

On a aussi :

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}.$$

Chapitre 2

DEPENDANCE ET COPULES

Supposer l'indépendance entre les risques d'un portefeuille d'assurance est une hypothèse forte qui doit être testée. Les copules présentent de nombreux avantages pour modéliser la dépendance entre risques. D'une part, elles permettent de décrire le comportement individuel de chaque risque et « couplent » les lois marginales pour obtenir la loi jointe. D'autre part, elles offrent une représentation fonctionnelle de la dépendance qui donne une description très complète de la forme de cette dernière. Dans cette partie sont présentés les résultats fondamentaux de la théorie des copules.

2.1 Dépendance ou corrélation ?

Il est important de rappeler que la dépendance et la corrélation sont des notions différentes.

En effet, on a X et Y indépendantes $\implies X$ et Y non corrélées ou $\rho(X, Y) = 0$ mais la réciproque est fautive sauf dans le cas où les variables sont gaussiennes car la dépendance est alors entièrement caractérisée par le coefficient de corrélation. Le contre-exemple le plus connu dans la littérature est le suivant : soit $X \rightsquigarrow \mathcal{N}(0, 1)$ et $Y = X^2$, alors

$$\text{cov}(X, Y) = E(X^3) = 0.$$

Bien qu'il soit facile à calculer et fréquemment présent dans les travaux actuariels, en assurance comme en finance, le coefficient de corrélation doit être utilisé avec précaution car il n'est pertinent qu'en présence de distributions elliptiques (distribution multivariée Normale ou de Student) ou de dépendance linéaire. Les erreurs d'interprétation et limites liées à son utilisation sont discutées dans Embrechts et al.[1999].

2.2 Qu'est-ce qu'une copule ?

Le concept de copule a été introduit par Sklar en 1959 pour résoudre un problème de probabilité énoncé par Maurice Fréchet. A l'époque, Abe Sklar et Berthold Schweizer travaillent sur les travaux de Karl Menger concernant les espaces métriques aléatoires (Probabilistic Metric Space ou PMS), qui sont une généralisation de l'espace métrique usuel introduit par Fréchet en 1906. Même si les copules occupent une place importante dans l'oeuvre de Sklar et Schweizer, elles ne sont pas l'objet central de leurs recherches.

L'utilisation des copules par Sklar et Schweizer est assez originale : elles interviennent pour résoudre certains problèmes et ne font pas l'objet véritablement d'études appropriées.

Pendant de nombreuses années, les copules sont peu (ou pas) utilisées en statistiques. Il y a les travaux sur la dépendance de Kimeldorf et Sampson dans les années 1975 ou encore les recherches de Paul Deheuvels à la fin des années 1970, mais il faut attendre le milieu des années 1980 pour que celles-ci fassent l'objet d'une étude systématique par quelques statisticiens. Le point de départ est bien sûr l'article The joy of copulas de Genest et MacKey [1986] publié dans The American Statistician. Suivront de nombreux travaux de Christian Genest avec différents co-auteurs (MacKey, Louis-Paul Rivest,...). Maintenant, les copules sont un outil standard largement utilisé pour étudier la dépendance, les modèles de survie, etc.

Définition 2.1 (copule) *Une copule est une fonction de répartition multivariée C définie sur l'hypercube $[0, 1]^n$ et dont les marginales sont uniformes sur $[0, 1]$.*

En particulier, on donne ici la définition et les propriétés d'une copule bivariée ($n = 2$).

Définition 2.2 *On appelle copule bivariée toute fonction C définie de $[0, 1]^2 \rightarrow [0, 1]$ qui possède les propriétés suivantes :*

- $\forall u \in [0, 1], C(u, 0) = C(0, u) = 0;$
- $\forall u \in [0, 1], C(u, 1) = C(1, u) = u;$
- C est 2-croissante c'est-à-dire : $\forall (u_1, u_2), (v_1, v_2) \in [0, 1]^2$ avec $u_1 < u_2$ et $v_1 < v_2,$

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

2.2.1 Théorème de Sklar

Ce théorème est au coeur de la théorie des copules. Il constitue le fondement de la plupart des applications des copules. Le théorème de Sklar

élucide le rôle que jouent les copules dans l'identification de la relation, entre les fonctions de répartition à plusieurs variables et leurs marges univariées. Le théorème de Sklar s'énonce comme suit :

Théorème 2.1 (Théorème de Sklar [1959]) *Soit F une fonction de répartition n -dimensionnelle avec des marginales F_1, \dots, F_n , alors il existe une n -copule C telle que pour tout x de \mathbb{R}^n ,*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)).$$

Preuve: Voir Nelsen (2006). ■

Le théorème de Sklar appliqué à des variables aléatoires continues conduit au théorème suivant.

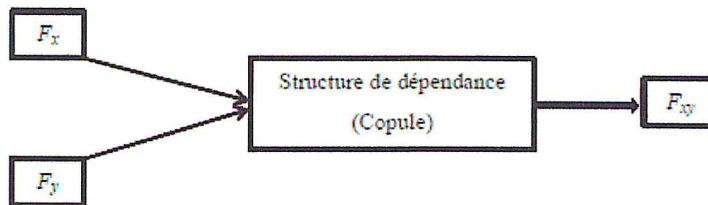


FIG. 2.1 – Schéma simplifié du théorème de Sklar avec F_x et F_y les fonctions marginales de la fonction de distribution bivariée F_{xy}

Corollaire 2.1 (inversion de Sklar) *Soient $C, F, F_1, F_2, \dots, F_n$ dans le théorème 2.1 et supposant que F_1, F_2, \dots, F_n sont continues si $F_1^{-1}, F_2^{-1}, \dots, F_n^{-1}$ les fonctions inverses de F_1, F_2, \dots, F_n :*

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)), \forall u_i \in I, i = 1, \dots, n.$$

Théorème 2.2 *Soit (X_1, \dots, X_n) un vecteur de variables aléatoires continues admettant F_1, \dots, F_n comme fonctions de répartition marginales et F comme fonction de répartition jointe, alors il existe une copule qui vérifie la relation du théorème 2.1. Si les marginales F_1, \dots, F_n sont continues, alors C est unique, autrement C est uniquement déterminée sur*

$$\text{Im}(F_1) \times \dots \times \text{Im}(F_n).$$

$\text{Im}(x)$ représentant l'ensemble des valeurs prises par X .

A partir du théorème 2.1, on voit qu'une copule permet d'exprimer une fonction de répartition multivariée selon ses marginales et que cette copule résume toute la structure de dépendance.

Une copule n'est autre qu'une fonction de répartition dont les lois marginales sont uniforme, ce qui conduit à une expression probabiliste simple de la copule :

$$C(u_1, \dots, u_n) = \text{pr}(U_1 \leq u_1, \dots, U_n \leq u_n).$$

Avec les copules, on ne travaille plus en nombre ou montant mais en rang.

Autrement dit, dans la pratique on transforme linéairement les réalisations x_1, \dots, x_n en uniformes empiriques u_1, \dots, u_n où

$$u_i = \frac{\text{Rang}(x_i)}{n + 1}$$

pour tout i de 1 à n . Cette manière de procéder se justifie par le fait que les rangs ne dépendent pas des lois marginales (Genest et Rémillard, 2004).

Définition 2.3 (densité d'une copule) *La densité d'une copule C , si elle existe, est définie comme suit*

$$c(u_1, \dots, u_n) = \frac{\partial^n C}{\partial u_1 \dots \partial u_n}(u_1, \dots, u_n)$$

2.2.2 Cas de deux variables

On considère un couple de v.a. X et Y de fonctions de répartition marginales :

$$F_X(x) = P(X \leq x), F_Y(y) = P(Y \leq y)$$

et de fonction de répartition conjointe :

$$F(x, y) = P(X \leq x, Y \leq y)$$

L'idée de base des copules est de travailler sur le couple $(F_X(x), F_Y(y))$ et non pas sur le couple (X, Y) . On travaille donc avec un vecteur aléatoire $(U = F_X(x), V = F_Y(y))$ dont les lois marginales suivent une loi uniforme sur $[0, 1]$. On appelle "copule" du couple (X, Y) la fonction de répartition C (unique si X et Y sont continues) telle que, pour tout $(x, y) \in \mathbb{R}^2$ (théorème de Sklar, 1959) :

$$F(x, y) = C(F_X(x), F_Y(y)) = (U, V)$$

La copule associée à (X, Y) est alors, pour tout $(u, v) \in [0, 1]^2$, donnée par :

$$C(U, V) = F(F_X^{-1}(u), F_Y^{-1}(v))$$

La densité f de F s'exprime alors par :

$$f(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y)$$

où f_X et f_Y sont les densités des variables X et Y et $c(u, v)$ est la densité de la copule définie par :

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$$

Les distributions conditionnelles $F_{Y|X}(y)$ et $F_{X|Y}(x)$ s'expriment en fonction des copules par :

$$F_{Y|X}(y) = C_1(F_X(x), F_Y(y)), F_{X|Y}(x) = C_2(F_X(x), F_Y(y))$$

où

$$C_1(F_X(x), F_Y(y)) = \frac{\partial C(u, v)}{\partial u}, C_2(F_X(x), F_Y(y)) = \frac{\partial C(u, v)}{\partial v}.$$

sont les copules conditionnelles.

2.3 Bornes de Fréchet-Hoeffding

Définissons tout d'abord trois copules importantes :

1. **La copule minimum** : qui a pour expression

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

avec $(u, v) \in [0, 1]^2$.

2. **La copule maximum** : qui est notée par

$$M = C^+(u, v) = \min(u, v)$$

avec $(u, v) \in [0, 1]^2$.

3. **La copule produit** : qui est notée par

$$\Pi = C(u, v) = uv$$

Les copules minimum C^- et maximum C^+ sont essentielles puisqu'elles définissent des copules extrémales de toute copule, que l'on appelle les bornes de Fréchet.

Proposition 2.1 *Pour toute copule C , nous avons :*

$$W = C^-(u, v) \leq C(u, v) \leq C^+(u, v) = M$$

2.4 Mesures de concordance

L'analyse des relations de dépendance entre variables aléatoires est l'un des sujets largement étudiés en probabilité et en statistique. La littérature évoque plusieurs mesures de dépendance et d'association, les plus connues sont le coefficient de Pearson, le Tau de Kendall et le Rho de Spearman. Les copules permettent, simultanément, de mesurer et de modéliser la dépendance. Ainsi, nous nous sommes intéressés dans cette section, à la présentation de la modélisation de la dépendance par les copules.

Soient (x, y) et (\tilde{x}, \tilde{y}) deux réalisations d'un vecteur aléatoire continu (X, Y) , alors (x, y) et (\tilde{x}, \tilde{y}) sont dites

- **Concordantes** si $(x - \tilde{x}) - (y - \tilde{y}) > 0$
- **Discordantes** si $(x - \tilde{x}) - (y - \tilde{y}) < 0$.

Théorème 2.3 *Soit K une mesure de concordance pour des variables aléatoires continues X et Y .*

1. *si Y est une fonction croissante de X , alors $K \langle X, Y \rangle = 1$;*
2. *si Y est une fonction décroissante de X , alors $K \langle X, Y \rangle = -1$;*
3. *si α et β sont des fonctions strictement croissantes, alors*

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

Il est facile de construire des exemples où le coefficient de corrélation linéaire de Pearson n'est pas invariant, par transformation strictement croissante. Par conséquent, la corrélation linéaire n'est pas une mesure de concordance.

En revanche, le Tau de Kendall et le Rho de Spearman sont deux mesures de concordance bien connues en statistique. Elles donnent une mesure de la corrélation entre les rangs des observations, à la différence du coefficient de corrélation linéaire qui lui apprécie la corrélation entre les valeurs des observations.

Elles offrent par ailleurs l'avantage de s'exprimer simplement en fonction de la copule associée au couple de variables aléatoires.

Définition 2.4 (Tau de Kendall) *Soient (X, Y) un couple de vecteurs aléatoires et (X', Y') une copie de (X, Y) , c'est-à-dire un couple de vecteurs en tout point identique à (X, Y) . Le tau de Kendall noté τ est défini par :*

$$\tau(X, Y) = \text{pr}\{(X - X')(Y - Y') > 0\} - \text{pr}\{(X - X')(Y - Y') < 0\}.$$

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

Son expression en terme de copule est la suivante :

Théorème 2.4 Soit (X, Y) un couple de variables aléatoires continues de copule C , alors

$$\tau(X, Y) = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1 = 1 - 4 \iint_{[0,1]^2} \partial_u C(u, v) \partial_v C(u, v) dudv.$$

ce qui peut s'écrire encore

$$\tau(X, Y) = 4E(C(U, V)) - 1, \text{ avec, } U, V \rightsquigarrow \mathcal{U}([0, 1]).$$

Un estimateur du tau de Kendall se construit à partir d'un échantillon $\{(x_1, y_1), \dots, (x_T, y_T)\}$ de (X, Y) de la façon suivant :

$$\hat{\tau}(X, Y) = \frac{2}{T(T-1)} \sum_{j=2}^T \sum_{i=1}^{j-1} \text{sign}\{(x_j - x_i)(y_j - y_i)\} \text{ où } \text{sign}(z) = \begin{cases} 1, & \text{si } z \geq 1 \\ -1, & \text{si } z < 1 \end{cases}$$

Définition 2.5 (rho de Spearman) Soient (\tilde{X}, \tilde{Y}) et (X', Y') deux couples de vecteurs aléatoires copies d'un vecteur aléatoire (X, Y) , alors le **rho de Spearman** est égal à :

$$\rho_S(X, Y) = 3[\text{pr}\{(X - \tilde{X}) - (Y - Y') > 0\} - \text{pr}\{(X - \tilde{X}) - (Y - Y') < 0\}]$$

Le rho de Spearman s'écrit aussi en fonction du coefficient ρ de corrélation linéaire de Pearson :

$$\rho_S(X, Y) = \rho(F_X(X), F_Y(Y))$$

où F_X et F_Y sont les fonctions de répartition respectives de X et Y .

Soit (X, Y) un couple de variables aléatoires continues de copule C , alors

$$\rho_S(X, Y) = 12 \iint_{[0,1]^2} uv dC(u, v) - 3 = 12 \iint_{[0,1]^2} C(u, v) dudv - 3$$

Un estimateur du rho de Spearman se construit à partir d'un échantillon $\{(x_1, y_1) \dots (x_T, y_T)\}$ de (X, Y) de la façon suivante :

$$\rho_S(X, Y) = \frac{\sum_{i=1}^T (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^T (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^T (S_i - \bar{S})^2}}$$

où R_i est le rang de x_i , S_i celui de y_i et

$$\bar{Z} = \frac{1}{T} \sum_{i=1}^T Z_i.$$

2.5 Dépendance de queue

Le concept de dépendance de queue fournit une description de la dépendance au niveau des queues de distribution, très intéressante pour étudier la survenance simultanée de valeurs extrêmes. C'est une mesure locale contrairement au tau de Kendall et au rho de Spearman qui mesurent la dépendance sur l'ensemble de la distribution.

2.5.1 Coefficients de dépendance de queue

Définition 2.6 *Le coefficient de dépendance de queue inférieure ou (lower tail dependence coefficient) de deux variables aléatoires X et Y , de fonctions de répartition respectives F_X et F_Y , est défini par*

$$\lambda_L(X, Y) = \lim_{\alpha \rightarrow 0^+} \Pr[X \leq F_X^{-1}(\alpha) | Y \leq F_Y^{-1}(\alpha)] \quad (2.1)$$

$$= \lim_{\alpha \rightarrow 0^+} \frac{C(\alpha, \alpha)}{\alpha} \quad (2.2)$$

(si cette limite existe).

Si $\lambda_L = 0$, alors la dépendance de queue inférieure est nulle.

Les variables X et Y n'ont pas de dépendance de queue à gauche.

Si $\lambda_L \in (0, 1]$, cela signifie qu'il y a une probabilité non nulle que la variable Y dépasse

un seuil u sachant que la variable X est déjà supérieure à ce seuil α .

Autrement dit, λ_L est la probabilité d'avoir un extrême en Y sachant que X est extrême.

De façon similaire, on définit le coefficient de dépendance de queue supérieure.

Définition 2.7 *Le coefficient de dépendance de queue supérieure ou (upper tail dependence coefficient) de deux variables aléatoires X et Y de fonctions de répartition respectives F_X et F_Y , est défini par*

$$\lambda_U(X, Y) = \lim_{\alpha \rightarrow 1^-} \Pr[Y > F_Y^{-1}(\alpha) | X > F_X^{-1}(\alpha)] \quad (2.3)$$

$$= \lim_{\alpha \rightarrow 1^-} \frac{1 - 2\alpha + C(\alpha, \alpha)}{1 - \alpha}$$

$$= \lim_{\alpha \rightarrow 1^-} \frac{\bar{C}(\alpha, \alpha)}{1 - \alpha}. \quad (2.4)$$

(si cette limite existe).

On dit que X et Y sont asymptotiquement dépendantes au niveau supérieur de la queue de distribution si $\lambda_U \in (0, 1]$ et asymptotiquement indépendantes au niveau supérieur de la queue de distribution si $\lambda_U = 0$.

Lorsque $\lambda_U > 0$, une interprétation simple peut être donnée lorsqu'on étudie la sinistralité extrême concomitante sur deux branches d'assurance : sachant qu'un sinistre extrême est survenu dans une branche, il existe une probabilité non nulle qu'un sinistre d'une intensité relative comparable survienne concomitamment dans l'autre branche.

La notion de copule de survie est très utile pour l'étude de la dépendance de queue. La définition en est la suivante.

2.6 Familles des copules

Il existe plusieurs familles de copules dont les plus connues sont les copules elliptiques, les copules archimédiennes et les copules de valeurs extrêmes.

2.6.1 Les Copules elliptiques

Les copules elliptiques sont définies à partir des lois de distribution elliptique.

Dans ce qui suit, nous donnons quelques définitions de la distribution elliptique ainsi que deux exemples classiques de cette famille de copules qui sont la copule Gaussienne et la copule de Student.

Définition 2.8 *Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est de distribution elliptique s'il admet la représentation suivante :*

$$X = \mu + RAU$$

où,

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

Définition 2.9 *La fonction de densité d'une distribution elliptique (si elle existe) est donnée par :*

$$f(x) = |\Sigma|^{-1/2} g((x - \mu)^T \Sigma^{-1} (x - \mu)), x \in \mathbb{R}^n$$

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

Définition 2.10 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de distribution elliptique. La fonction caractéristique $\varphi_X(t)$, $t \in \mathbb{R}^n$ est donnée par :

$$\begin{aligned}\varphi_X(t) &= E(\exp(it^T X)) \\ &= E(\exp(it^T \mu + RAU)) \\ &= \exp(it^T \mu)g(t^T \Sigma t).\end{aligned}$$

Définition 2.11 On appelle copule elliptique toute copule de la forme

$$\begin{aligned}C_\rho(u_1, u_2) &= \int_{-\infty}^{\phi_g^{-1}(u_1)} \int_{-\infty}^{\phi_g^{-1}(u_2)} \frac{1}{\sqrt{1-\rho^2}} g\left(\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{\sqrt{1-\rho^2}}\right) dx_1 dx_2 \\ &= H_\rho(\phi_{g_1}^{-1}(u_1), \phi_{g_2}^{-1}(u_2))\end{aligned}$$

avec H_ρ est la distribution conjointe des variables X_1 et X_2 et $\phi_{g_1}^{-1}$ et $\phi_{g_2}^{-1}$ les fonctions quantiles respectives et ρ leur coefficient de corrélation.

Les copules elliptiques sont les extensions multivariées des distributions elliptiques dont deux des plus célèbres sont la distribution normale et la distribution de Student. Elles sont symétriques et modélisent de la même manière la dépendance des queues gauches et des queues droites des distributions marginales.

Copule Gaussienne

Définition 2.12 La copule Gaussienne bivariée est définie de la façon suivante :

$$C(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$$

- ρ est une matrice symétrique définie positive avec $\text{diag}(\rho) = 1$
- Φ_ρ la distribution normale bivariée standard de matrice de corrélation

2.6.2 Copule normale

Soit H_ρ une distribution normale bivariée de coefficient de corrélation ρ telle que $\forall x_1, x_2 \in \mathbb{R}$ on a

$$H_\rho(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^x \int_{-\infty}^y \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{t_1 - m_1}{\sigma_1}\right)^2 - \frac{2\rho(t_1 - m_1)(t_2 - m_2)}{\sigma_1\sigma_2} + \left(\frac{t_2 - m_2}{\sigma_2}\right)^2\right]\right\} dt_1 dt_2$$

En particulier si $X_1 \rightsquigarrow \mathcal{N}(0, 1)$ et $X_2 \rightsquigarrow \mathcal{N}(0, 1)$, on a

$$\begin{aligned} C_\rho(u_1, u_2) &= \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{2\rho x_1 x_2 - x_1^2 - x_2^2}{2(1-\rho^2)}\right) dx_1 dx_2 \end{aligned}$$

ou de manière équivalent et d'après Roncalli (2004)

$$C_\rho(u_1, u_2) = \int_0^{u_1} \Phi\left(\frac{\Phi^{-1}(u_2) - \rho\Phi^{-1}(x_1)}{\sqrt{1-\rho^2}}\right) dx_1$$

avec $\rho \in]-1, 1[$, avec $\beta_1 = \Phi^{-1}(x_1)$, $\beta_2 = \Phi^{-1}(x_2)$.

L'une des observations qui ressort dans l'étude de Brechmann et Schepsmeier (2013) est que la copule normale modélise passablement une dépendance non linéaire ou en présence d'évènements extrêmes (sauf pour une corrélation parfaite). Néanmoins, il est très facile de simuler des variables aléatoires avec une distribution conjointe Normale (ou Gaussienne).

Copule de Student

la copule de student (ou la t-copule) est la fonction de dépendance associée à la fonction de répartition de student multidimensionnelle.

Définition 2.13 La copule Student bivariée est définie de la façon suivante :

$$C(u, v; \rho, \nu) = t_{\rho, \nu}(t_\nu^{-1}(u), t_\nu^{-1}(v))$$

Avec ρ la matrice de corrélation symétrique définie positive, $t_{\rho, \nu}^{-1}$, la distribution de Student bivariée standard de matrice de corrélation et de degré de liberté.

La fonction de student bivariée est donnée par :

$$t_\nu(x) = \int_{-\infty}^x \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\beta^2}{\nu}\right)^{-\frac{\nu+1}{2}} d\beta.$$

Nous en déduisons que sa densité s'écrit :

$$t_{\rho, \nu}(x) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(1 + \frac{\beta_1^2 + \beta_2^2 - 2\rho\beta_1\beta_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} d\beta_1 d\beta_2.$$

La copule de student bivariée est donnée par :

$$C_{\rho, \nu}(u, v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(1 + \frac{\beta_1^2 + \beta_2^2 - 2\rho\beta_1\beta_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+2}{2}} d\beta_1 d\beta_2,$$

avec $0 \leq \rho < 1$.

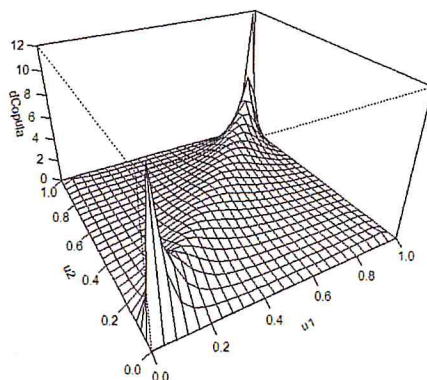


FIG. 2.2 – Densité du copule de Student bivariée

Remarque 2.1 Les copules Normale et celle de Student sont des copules symétriques et relativement simples à utiliser parce que les distributions associées sont bien connues. Elles sont souvent appelées les copules implicites car elles n'ont pas de forme analytique explicite. Ces copules s'expriment, par conséquent, en fonction des distributions bivariées auxquelles elles sont associées (Théorème de Sklar).

2.6.3 Copule des valeurs extrêmes

Définition 2.14 Une copule de valeurs extrêmes vérifie la relation suivante :

$$C(u_1^t, \dots, u_n^t) = C^t(u_1, \dots, u_n)$$

pour tout réel t positif.

Pour $n = 2$, Joe (1997) montre qu'une copule de valeurs extrêmes est vérifie la relation

$$C_A(u, v) = \exp \left\{ \ln(uv) A \left(\frac{\ln(v)}{\ln(u) + \ln(v)} \right) \right\}.$$

où $A : [0, 1] \rightarrow [0.5, 1]$ est une fonction convexe est appelée fonction de dépendance de Pickands, qui doit vérifier la condition

$$\max(t, 1 - t) \leq A(t) \leq 1$$

en tout $t \in [0, 1]$.

Les copules de valeurs extrêmes ne permettent de modéliser qu'une dépendance positive car pour tous $u, v \in [0, 1]$, elles vérifient la condition

$$uv \leq C_A(u, v) \leq \min(u, v).$$

Lorsque $A(t) = 1$ en tout $t \in [0, 1]$, on retrouve la copule d'indépendance. Ainsi, la copule $\Pi(u, v) = uv$ est un modèle extrême.

- Les copules dites de **Gumbel–Hougaard** appartiennent aussi à cette classe. Ces copules dépendent d'un paramètre $\theta \geq 1$ et sont engendrées par la fonction de Pickands donnée en tout $t \in (0, 1)$ par la formule

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

- A titre d'exemple, Une copule qui s'écrit de la manière suivante

$$C(u_1, u_2, a) = u_1 u_2 \exp\{[(-\ln u_1)^{-a} + (-\ln u_2)^{-a}]^{-\frac{1}{a}}\}$$

avec $a \in]0, \infty[$, est une copule de valeurs extrêmes

$$\begin{aligned} C(u_1^t, u_2^t) &= u_1^t u_2^t \exp\{[(-\ln u_1)^{-a} + (-\ln u_2)^{-a}]^{-\frac{1}{a}}\} \\ &= (u_1 u_2)^t \exp\{[t^{-a}(-\ln u_1)^{-a} + t^{-a}(-\ln u_2)^{-a}]^{-\frac{1}{a}}\} \\ &= (u_1 u_2)^t \exp\{[t(-\ln u_1)^{-a} + (-\ln u_2)^{-a}]^{-\frac{1}{a}}\} \\ &= \{u_1 u_2 \exp\{(-\ln u_1)^{-a} + (-\ln u_2)^{-a}\}^{-\frac{1}{a}}\} \\ &= C^t(u_1, u_2) \end{aligned}$$

Alors que la copule

$$\begin{aligned} C(u_1^t, u_2^t) &= u_1^t u_2^t + (1 + a(1 - u_1^t)(1 - u_2^t)) \\ &= u_1^t u_2^t (1 + a - a u_1^t - a u_2^t + a u_1^t u_2^t)^t \\ &\neq u_1^t u_2^t (1 + a - a u_1 - a u_2 + a u_1 u_2)^t \\ &\neq C^t(u_1, u_2, a) \end{aligned}$$

Remarque 2.2 Les copules de valeurs extrêmes sont max-stable c'est à dire si $(X_1, Y_n), \dots, (X_n, Y_n)$ sont des couples de variables iid dont la structure de dépendance peut être modélisée par une copule de valeurs extrêmes C , alors la structure de dépendance de

$$M_n = \max(X_1, \dots, X_n)$$

et

$$N_n = \min(Y_1, \dots, Y_n)$$

sont modélisée par la même copule C .

Le tableau suivant contient les copules de valeurs extrêmes les plus connues avec $\tilde{u} = -\ln u$, $\vartheta(u_1, u_2; \theta) = \Phi(1/\theta + 0.5\theta \ln(\ln u_1 / \ln u_2))$ et $\xi(w; \theta) = \xi(w, 1 - w; \theta)$.

TAB. 2.1 – Les copules des valeurs extrêmes les plus utilisées

Copule	θ	$C(u_1, u_2)$	$A(w)$
C^\perp		$u_1 u_2$	1
Gumbel A	$[1, +\infty)$	$\exp \left[-(\tilde{u}_1^\theta + \tilde{u}_2^\theta)^{1/\theta} \right]$	$[w^\theta + (1-w)^\theta]^{1/\theta}$
Gumbel B	$[0, 1]$	$u_1 u_2 \exp \left[\theta \frac{\tilde{u}_1 \tilde{u}_2}{\tilde{u}_1 + \tilde{u}_2} \right]$	$\theta w^2 - \theta w + 1$
Galambos	$[0, +\infty)$	$u_1 u_2 \exp \left[(\tilde{u}_1^{-\theta} + \tilde{u}_2^{-\theta})^{-1/\theta} \right]$	$1 - [w^{-\theta} + (1-w)^{-\theta}]^{-1/\theta}$
Hüsler-Reiss	$[0, +\infty)$	$\exp \left[\begin{array}{l} -\tilde{u}_1 \vartheta(u_1, u_2; \theta) \\ -\tilde{u}_2 \vartheta(u_1, u_2; \theta) \end{array} \right]$	$w\xi(w; \theta) + (1-w)\xi(1-w; \theta)$
Marshall-Olkin	$[0, 1]^2$	$u_1^{1-\theta_1} u_2^{1-\theta_2} \min(u_1^{\theta_1}, u_2^{\theta_2})$	$\max(1 - \theta_1 w, 1 - \theta_2(1-w))$

2.6.4 Copule archimédienne

Les copules Archimédiennes sont indiscutablement associées au statisticien canadien Christian Genest. Ce n'est pas lui qui les a inventées, mais il est le premier à avoir adopté une analyse statistique de ces fonctions copules. Et ses nombreuses publications ont largement contribué à les faire connaître. En fait, ces copules archimédiennes dérivent des t-normes archimédiennes de Ling [1965]. Comme il existe un lien très fort entre copules et t-normes, la construction de ces copules est immédiate (Schweizer et Sklar [1983]).

Cette méthode de construction des copules peut être facilement étendue en dimension n . Elle permet de construire une grande variété des copules.

Définition 2.15 Soit φ une fonction continue, strictement décroissante et définie de $[0, 1]$ dans $[0, +\infty[$ telle que $\varphi(1) = 0$. Le pseudo-inverse (inverse généralisé) de φ est la fonction $\varphi^{[-1]} : [0, +\infty[\rightarrow [0, 1]$ donnée par

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

Notons que $\varphi^{[-1]}$ est continue et décroissante sur $[0, +\infty]$, et strictement décroissante sur $[0, \varphi(0)]$. En outre, $\varphi^{[-1]}(\varphi(u)) = u$ dans $[0, 1]$, et

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

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

Définition 2.16 Soit φ une fonction continue, strictement décroissante de $[0, 1]$ dans $[0, \infty[$ telle que $\varphi(1) = 0$. $\varphi^{[-1]}$ est le pseudo-inverse de φ . Soit C une fonction de $[0, 1]^2$ dans $[0, 1]$ donnée par

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2)).$$

alors C est une copule si et seulement si φ est convexe. Une copules C de la forme

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$$

est dite copule archimédienne avec un générateur φ . Elle est symétrique et associative.

Le grand avantage de cette famille c'est qu'elle permet de décrire des structures de dépendance très diverses dont notamment les dépendances asymétriques où les coefficients de queue inférieure et de queue supérieure sont différents. Genest et MacKay (1986) ont présenté les propriétés de cette classe des copules. Le choix du générateur φ détermine la famille des copules.

A titre d'exemple, prenons $\varphi(t) = t^{-a} - 1$ avec $0 < a < \infty$. On peut vérifier que $\varphi(t)$ est continue et $\varphi(1) = 0$. $\varphi' = -at^{-a-1}$ donc φ est strictement décroissante de $[0, 1]$ dans $[0, \infty]$. $\varphi''(t) \geq 0$ dans $[0, 1]$, φ est convexe.

En plus, $\varphi(0) = \infty$ et φ est un générateur strict. On a

$$\begin{aligned} C(u_1, u_2) &= \varphi^{-1}(\varphi(u_1) + \varphi(u_2)) \\ &= [(u_1^{-a} - 1) + (u_2^{-a} - 1) + 1]^{-\frac{1}{a}} \\ &= (u_1^{-a} + u_2^{-a} - 1)^{-\frac{1}{a}} \end{aligned}$$

Définition 2.17 (Copule de Clayton) La copule de la forme

$$C(u_1, u_2) = (u_1^{-a} + u_2^{-a} - 1)^{-\frac{1}{a}}$$

est une copule archimédienne de générateur $\varphi(t) = t^{-a} - 1$. Elle est appelée copule de Clayton.

TAB. 2.2 – Les copules Archimédiennes les plus utilisées

Copule	$\varphi(u)$	$C(u, v)$
Gumbel	$(-\ln u)^a, a \geq 1$	$\exp(-[(-\ln u)^a + (-\ln v)^a]^{\frac{1}{a}})$
Franck	$\ln\left(\frac{\exp^{-au} - 1}{\exp^{-a} - 1}\right), a \neq 0$	$\frac{1}{a} \ln\left[1 + \frac{(\exp^{-au} - 1)(\exp^{-av} - 1)}{(\exp^{-a} - 1)}\right]$
Clayton	$\frac{(u^{-a} - 1)}{a}, a > 0$	$(u^{-a} + v^{-a} - 1)^{-\frac{1}{a}}$
Joe	$-\ln(1 - (1 - u)^\alpha)$	$1 - (\bar{u}^\alpha + \bar{v}^\alpha + \bar{u}^\alpha \bar{v}^\alpha)^{1/\alpha}$
C^\perp	$-\ln u$	uv

Calcul du tau de Kendall pour les copules archimédiennes

Pour les copules archimédiennes, la détermination du tau de Kendall et des coefficients de dépendance des queues par les expressions présentées précédemment est un peu compliquée. Une expression permettant de calculer ce coefficient plus facilement consiste à l'exprimer en fonction du générateur $\varphi(t)$ de la copule de la manière suivante

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt$$

Pour cette copule C , le tau de Kendall est

$$\begin{aligned} \tau &= 1 + 4 \int_0^1 \frac{t^{-a} - 1}{-at^{-a-1}} dt = 1 - \frac{4}{a} \int_0^1 t - t^{a+1} dt \\ &= 1 - \frac{4}{a} \left[\frac{1}{2} t^2 - \frac{1}{a+2} t^{a+2} \right]_0^1 = \frac{a}{a+2}. \end{aligned}$$

Calcul des coefficients de dépendance des queues pour les copules archimédiennes

On peut exprimer le coefficient de dépendance de queue supérieure en fonction du générateur de la copule C comme suit

$$\begin{aligned} \lambda_U &= \lim_{u \rightarrow 1} \frac{C(u, u)}{(1-u)} = \lim_{u \rightarrow 1} \frac{[1 - 2u + \varphi^{-1}(2\varphi(u))]}{(1-u)} \\ &= \lim_{u \rightarrow 1} 2 + \frac{\varphi^{-1}(2\varphi(u)) - 1}{1-u} \end{aligned}$$

On pose $s = \varphi(u)$, on a

$$\begin{aligned} \lambda_U &= \lim_{s \rightarrow 1} 2 + \frac{\varphi^{-1}(2s) - 1}{1 - \varphi^{-1}(s)} \\ &= \lim_{s \rightarrow 0} 2 - 2 \frac{[\varphi^{-1}(2s) - 1] / 2s}{[\varphi^{-1}(s) - 1] / s} = 2 - 2 \lim_{s \rightarrow 0} \frac{\varphi^{-1}'(2s)}{\varphi^{-1}'(s)} \end{aligned}$$

Si $\varphi^{-1}'(0) \in (-\infty, 0)$, C n'admet pas de dépendance de queue supérieure. Prenons l'exemple du même copule de Clayton, on a

$$\lambda_U = 2 \lim_{s \rightarrow 0} \left[\frac{(2s+1)^{-1/a-1}}{(s+1)^{-1/a-1}} \right] = 2 \cdot 2^{-1/a-1} = 2^{-1/a}$$

Dans le tableau suivant, nous donnons quelques exemples de copules Archimédiennes (nous utilisons les notations de Joe [1997] pour être plus concis : $\bar{u} = 1 - u$ et $u = -\ln u$) :

TAB. 2.3 – Les paramètres de dépendance des copules Archimédiennes les plus utilisées

Copule	λ_L	λ_U	τ_a
Gumbel	0	$2 - 2^{-\frac{1}{a}}$	$1 - \frac{1}{a}$
Franck	0	0	$1 - \frac{4(1-D_1(a))}{a}$
Clayton	$2^{-\frac{1}{a}}$	0	$\frac{a}{a+2}$
Joe			
C^\perp			

- La copule de Gumbel n’appréhende que des dépendances positives et possède la caractéristique de pouvoir représenter des risques dont la structure de dépendance est plus accentuée sur la queue supérieure. Elle est à ce titre particulièrement adaptée en assurance et en finance pour étudier l’impact de la survenance d’événements de forte intensité sur la dépendance entre branches d’assurance ou actifs financiers.
- La copule de Franck permet de modéliser les dépendances aussi bien positives que négatives. On note qu’il n’existe pas de dépendance de queue pour cette copule. Le tau de Kendall s’exprime en fonction de a grâce à la fonction Debye définie par

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{\exp^t - 1} dt.$$

- Comme la copule de Gumbel, la copule de Clayton ne permet de modéliser que les dépendances positives. A l’inverse de la copule de Gumbel, elle vise à rendre compte d’une dépendance sur les événements de faible intensité.

2.7 Quelques copules bivariées

Il existe une multitude de copules permettant de caractériser l’ensemble des structures de dépendance possibles. Cet ensemble est constitué de nombreuses familles qui offrent une diversité suffisante pour modéliser toutes les structures de dépendance les plus courantes.

2.7.1 Copule produit

C’est la copule la plus simple et elle a la forme suivante

$$C(u_1, u_2) = C^\perp = u_1 u_2$$

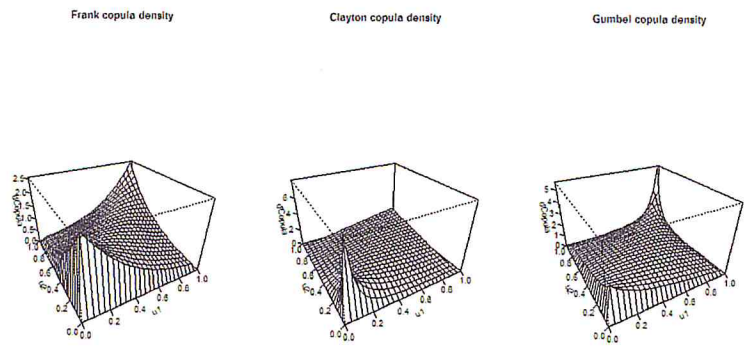


FIG. 2.3 – Densité des copules : Frank, Gumbel et Clayton

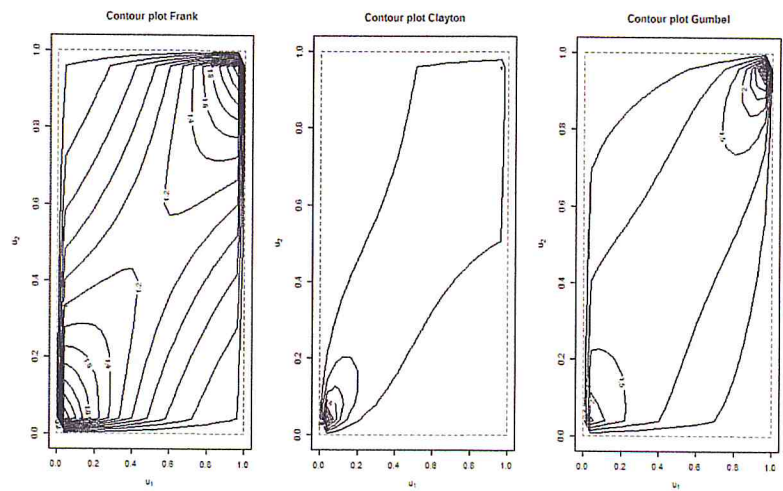


FIG. 2.4 – Lignes de niveaux des copules : Gumbel, Frank et Clayton

elle est très importante comme "Benchmark" puisqu'elle correspond à l'indépendance entre les deux variables. elle est appelée également copule d'indépendance. Pour deux variables indépendantes, on a

$$C(u_1, u_2) = u_1 u_2$$

et

$$C(u_1, u_2) = \frac{\partial C(u_1, u_2)}{\partial(u_1)} = u_2$$

On peut vérifier que la distribution conditionnelle de $Y/X = x$ n'est autre que la distribution marginale de Y .

$$F_{Y/X}(y) = \frac{\text{pr}(Y < y) \text{pr}(X < x)}{\text{pr}(X < x)} = C_1(F_X(x), F_Y(y)) = F_Y(y) = \text{pr}(Y < y)$$

2.7.2 Copule de Farlie- Gumbel- Morgenstern (FGM)

La copule de Farlie-Gumbel-Morgenstern a été proposée par Morgenstern (1956) et elle a la forme suivante

$$C(u_1, u_2; a) = u_1 u_2 (1 + a(1 - u_1)(1 - u_2))$$

Cette copule est simple mais restrictive dans la mesure où elle ne peut être utilisée que pour étudier une dépendance de faible amplitude entre les variables. Le rho de Spearman prend des valeurs dans un intervalle très restreint $[-1/3, 1/3]$ et c'est pour cette raison qu'on ne peut pas l'utiliser pour modéliser les variables aléatoires qui présentent une forte dépendance. Si le paramètre a est nul, cette copule correspond à une copule produit.

2.7.3 Copules de survie

Définition 2.18 Soit $\bar{C}(u_1, \dots, u_n)$ la fonction définie par

$$\bar{C}(u_1, \dots, u_n) = C(1 - u_1, \dots, 1 - u_n)$$

où

$$\bar{C}(u_1, \dots, u_n) = \text{Pr}[U_1 > u_1, \dots, U_n > u_n]$$

Alors $\bar{C}(u_1, \dots, u_n)$ est appelée copule de survie (survival copula) de la copule C .

Définition 2.19 Soient X et Y deux variables aléatoires de copule C , alors on a

$$\lambda_L(X, Y) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}$$

et

$$\lambda_U(X, Y) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}.$$

On montre facilement que le coefficient de dépendance de queue inférieure (resp. supérieure) de C est le coefficient de dépendance de queue supérieure (resp. inférieure) de \bar{C} . Autrement dit, à partir d'une copule donnée, il est possible de créer une autre copule présentant une structure de dépendance de queue inversée.

2.7.4 Copule HRT

La copule HRT n'appartient pas à la famille des copules archimédiennes. Elle a été introduite par Venter [2001] pour modéliser la dépendance sur des événements extrêmes de forte intensité. Elle a été construite comme étant la copule de survie de la copule de Clayton et présente donc une structure de dépendance inversée.

$$\begin{aligned} C(u, v) &= u + v - 1 + ((1 - u)^{-a} + (1 - v)^{-a} - 1)^{\frac{1}{a}}, \\ \lambda_L &= 0, \lambda_U = 2^{\frac{-1}{a}}, \tau_a = \frac{a}{a + 2}. \end{aligned}$$

2.7.5 Copule empirique

Deheuvels [1979] a introduit la notion de copule empirique Soit $\{\vartheta_1^t, \dots, \vartheta_n^t\}$ la statistique de rang associée à l'échantillon multivarié $\{x_1^t, \dots, x_n^t\}$ dont chaque vecteur possède J observations ($\forall i \in [1, n]$, ϑ_i^t est le rang de x_i^t parmi $(x_i^t)_{t=1 \dots J}$) alors toute copule \hat{C} définie sur le treillis

$$\left\{ \left(\frac{t_1}{T}, \dots, \frac{t_n}{T} \right) \mid 1 \leq k \leq n, 0 \leq t_k \leq T \right\}$$

par

$$\hat{C} \left(\frac{t_1}{T}, \dots, \frac{t_n}{T} \right) = \frac{1}{T} \sum_{t=1}^T \prod_{k=1}^n 1_{\{\vartheta_k^t \leq t_k\}}$$

est une copule empirique.

2.7.6 Quelques copules particulier et ses paramètres

Le tableau 2.4 représente quelques copule avec sont code sous le logiciel R ainsi les paramètres correspondants

TAB. 2.4 – Quelques copules avec leurs paramètres et le codage sous R

famille de cpule	famille	par	par1
Gaussian	1	$[-1, 1]$	*
t-Student	2	$[-1, 1]$	$(2, \infty)$
Clayton(survie)	3, 13	$(0, \infty)$	-
Clayton Rotate 90 et 270	23, 33	$(-\infty, 0)$	-
Gumbel(Survi)	4,14	$[1, +\infty)$	-
Gumbel Rotate 90 et 270	24, 34	$(-\infty, -1]$	-
Frank	5	R^*	-
(Survival) Joe	6, 16	$(1, +\infty)$	-
Rotated Joe (90 and 270 degrees)	26, 36	$(-\infty, -1)$	-
(Survival) Clayton-Gumbel (BB1)	7, 17	$(0, \infty)$	$[1, +\infty)$
Rotated Clayton-Gumbel (90 and 270 degrees)	27, 37	$(-\infty, 0)$	$(-\infty, -1]$

2.8 Simulation de copules bivariées

2.8.1 Méthode des distributions

Nous utilisons la transformation

$$X = (F_1^{-1}(U_1), \dots, F_n^{-1}(U_n)).$$

Nous avons donc

$$F(U_1, \dots, U_n) = C(F_1^{-1}(U_1), \dots, F_n^{-1}(U_n)).$$

Pour simuler $U = (U_1, \dots, U_n)$, nous pouvons simuler le vecteur aléatoire $X = (X_1, \dots, X_n)$ de fonction de distribution F et appliquer la transformation

$$U = (F_1(X_1), \dots, F_n(X_n)).$$

Théorème 2.5 Soit C une copule et soit (U, V) un vecteur aléatoire de loi C . On définit pour tout $(u, v) \in (0, 1)^2$

$$L_v(u) = Pr(U \leq u | V = v) = \frac{\partial}{\partial v} C(u, v).$$

Soient maintenant T_1 et T_2 deux variables aléatoires indépendantes telle que $T_1 \rightsquigarrow \mathcal{U}([0, 1])$ et $T_2 \rightsquigarrow \mathcal{U}([0, 1])$. Si $U = L_{T_2}^{-1}(T_1)$ et $V = T_2$, alors la paire (U, V) est de loi C .

On peut donc établir un algorithme de simulation de copule comme suit :

- Algorithm 2.6**
1. On génère des variables aléatoires et indépendantes V, T qui sont $\mathcal{U}([0, 1])$.
 2. On pose $U = L_V^{-1}(T)$.
 3. Le couple désiré est (U, V) .

2.8.2 Méthode des distributions conditionnelles

Considérons le cas bivarié. Soit $U = (U_1, U_2)$ un vecteur aléatoire de fonction de distribution C . Nous savons que $P[U_1 \leq u_1] = u_1$ et

$$P[U_2 \leq u_2 | U_1 = u_1] = C_{2|1}(u_1, u_2)$$

Comme $C(U_1, 1)$ et $C_{2|1}(u_1, u_2)$ sont deux variables aléatoires uniformes, nous obtenons l'algorithme suivant :

1. Simuler deux variables aléatoires uniformes v_1 et v_2 .
2. Prendre u_1 égal à v_1 .
3. Soit $C(u_2, u_1) = C_{2|1}(u_1, u_2)$.
4. Prendre u_2 égal à $C^{-1}(v_2, u_1)$.

Cet algorithme est suggéré par Genest et MacKay (1986). Il est utilisé par Genest (1987) pour simuler la copule Frank. Nous rappelons que la fonction de cette copule est

$$C(u_1, u_2; \theta) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right)$$

Nous en déduisons que

$$C_{2|1}(u_1, u_2; \theta) = \frac{e^{-\theta u_1}(e^{-\theta u_2} - 1)}{(e^{-\theta} - 1) + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}$$

Nous obtenons finalement

$$\begin{aligned} C^{-1}(u_1, u_2) &= \{u_2, C_{2|1}(u_1, u_2; \theta) = u\} \\ &= -\frac{1}{\theta} \ln \left(1 + \frac{u(e^{-\theta} - 1)}{u + (1 - u)e^{-\theta u_1}} \right) \end{aligned}$$

La méthode bivariée s'étend sans difficultés (sur le plan mathématique) au cas multivarié. Prenons par exemple le cas trivarié. Nous avons l'algorithme suivant :

1. Simuler trois variables aléatoires uniformes v_1, v_2 et v_3 .
2. Prendre u_1 égal à v_1
3. Soit $C(u_2, u_1) = C_{2|1}(u_1, u_2, 1)$. Prendre u_2 égal à $C^{-1}(v_2, u_1)$.
4. Soit $C(u_3; u_1, u_2) = C_{3|1,2}(u_1, u_2, u_3)$. Prendre u_3 égal à $C^{-1}(v_3; u_1, u_2)$.

Il existe plusieurs algorithmes pour simuler les copules Archimédiennes. Genest et MacKay (1986) prônent l'idée de simuler la distribution conjointe du vecteur aléatoire X en procédant par des simulations récursives des distributions conditionnelles de X_j sachant X_i .

1. Générer un nombre aléatoire uniformément distribué U_1 .
2. Poser $X_1 = F_1^{-1}(U_1)$.
3. Pour $j = 2, \dots, n$, on calcule récursivement :

$$U_j = F_j(X_j|x_1, \dots, x_{j-1}) = \frac{\varphi^{-1(j-1)} [c_{j-1} + \varphi(F_j(x_j))]}{\varphi^{-1(j-1)}(c_{j-1})}$$

où $c_j = \varphi F_1(x_1) + \dots + \varphi F_j(x_j)$ et $\varphi^{-1}(j)$ est la $j^{\text{ième}}$ dérivée de la fonction inverse de la fonction φ .

4. Poser $X_j = F_j^{-1}(U_j)$.

2.8.3 Les méthodes dites analytiques

Dans ce paragraphe, la simulation est spécifique à chaque copule ou à un type de copule.

Pour simuler la copule Clayton (pour $\theta > 0$), nous pouvons employer l'algorithme donné par Devroye (1986) :

1. Simuler deux variables aléatoires exponentielles standards x_1 et x_2 .
2. Simuler une variable aléatoire x de fonction de distribution $\Gamma(1, \theta)$.
3. Prendre $u_1 = (1 + x_1/x)^{-\theta}$ et $u_2 = (1 + x_2/x)^{-\theta}$. Pour simuler des vecteurs aléatoires X ayant une copule Gaussienne de dimension n , on propose l'algorithme suivant :
4. Générer Z , une réalisation de la loi normale multivariée de moyenne $0_{n \times 1}$ et de variance $\Sigma_{n \times n}$ avec, $\Sigma_{ii} = 1$ et $\Sigma_{ij} = r$.
5. Utiliser la décomposition $\Sigma = A^t A$.

6. Générer Y_1, \dots, Y_n une suite de n variables aléatoires i.i.d de loi normale standard.
7. Calculer $X = (X_1, \dots, X_n)^t = AY$ de $Y = (Y_1, \dots, Y_n)^t$.
8. Poser $U_i = \Phi(X_i), i = 1, \dots, n$ avec Φ est la fonction de distribution normale standard.

Afin de simuler des vecteurs aléatoires X ayant une copule de Student de dimension n à v degrés de liberté, on propose l'algorithme suivant :

1. Générer Z , une réalisation de la loi normale multivariée de moyenne $0_{n \times 1}$ et de variance $\Sigma_{n \times n}$ avec, $\Sigma_{ii} = 1$ et $\Sigma_{ij} = r$.
2. Générer $V \rightsquigarrow \chi_v^2$.
3. Dédurre $W = v/V$.
4. Poser $T = Z\sqrt{W}$.
5. Soit t_v la distribution de Student multivariée avec v degrés de liberté, de fonction de répartition marginale F_T . Soit T_j le $j^{\text{ième}}$ élément de T . On calcule la variable aléatoire $U_j : U_j = F_T(T_j)$.
6. Calculer $X_j = F_j^{-1}(U_j)$, où F_j est la fonction de répartition de X_j .

2.8.4 La méthode des quantiles empiriques

Le problème de cette section n'est plus la simulation du vecteur U dont la distribution est une copule C , mais concerne la simulation du vecteur X dont la copule est C et les marginales pas forcément uniformes. Dans les sections précédentes, X est simulé à partir de la transformation suivante :

$$(F_1^{-1}(U_1), \dots, F_n^{-1}(U_n))^t$$

Cela implique la connaissance des distributions F_1, \dots, F_n . Ce qui n'est pas toujours le cas. Néanmoins, s'il est possible de simuler les marginales F_1, \dots, F_n , alors, nous pouvons simuler la distribution multidimensionnelle F grâce à la méthode des quantiles empiriques. Soit $\mathbb{F}_{i,m}$ le processus de distribution empirique (non normalisé).

Nous avons le résultat suivant :

$$\sup_x \|\mathbb{F}_{i,m} - \mathbb{F}_i(x)\|, a.s \rightarrow 0, \text{ lorsque } m \rightarrow 0.$$

Soient U_m et F_m les processus de distribution empirique correspondants aux distributions $C(u_1, \dots, u_n)$ et $F(x_1, \dots, x_n)$. En utilisant un argument de type Glivenko-Cantelli, nous avons

$$\sup_{u_1, \dots, u_n} |U_p(F_{1,m(u_1)}, \dots, F_{n,m(u_n)})|, a.s \rightarrow 0, \text{ lorsque } m \wedge p \rightarrow 0.$$

avec \wedge indique le maximum.

2.9 Conclusion

Ce chapitre nous a permis d'explorer les bases de la théorie des copules, en particulier ses performances en terme de modélisation et de caractérisation des relations de dépendance. Les applications des copules sont multiples ; elles varient de l'hydrologie et de la météorologie, à la finance et aux assurances. Cependant, leur utilisation en économie reste très restreinte. Ainsi, nous contribuons à ce domaine par une application en économie, en présentant dans le prochain chapitre l'utilité des copules pour modéliser la dépendance des variables macroéconomiques et boursières. Cette étude consistera à comparer les prévisions fournies par les copules à celles déduites des modèles classiques (modèles de séries chronologiques).

Chapitre 3

Testes statistiques pour les copules

3.1 Introduction

La grande propriété des modèles à base de copules est que différentes formes fonctionnelles de copule pourraient être utilisées afin de simuler toute structure de dépendance entre des variables aléatoires auxquelles les scientifiques ont pensé. D'un autre côté, la grande variété de formes fonctionnelles plausibles est l'inconvénient majeur de l'usage plus général de Copula dans la pratique, car il n'existe généralement aucune indication claire de la forme de copule paramétrique à utiliser.

De plus, il n'ya qu'un consensus sur le choix de la forme fonctionnelle appropriée, indiquant que ce processus appartient aux éléments les plus méconnus de la modélisation des copules. Le fait de savoir si une famille de copules choisie modélise de manière appropriée la structure de dépendance d'un ensemble de données observé constitue donc un problème fondamental.

Lorsque le jeu de données doit être analysé, le problème à résoudre se compose généralement de trois parties. En utilisant d'abord l'expérience des modèles de type similaire, nous suggérons plusieurs formes fonctionnelles de copule qui sont supposées correspondre aux données fournies.

Dans un deuxième temps, nous estimons les paramètres de toutes les fonctions proposées. Enfin, à l'aide de tests de qualité de l'ajustement, nous analysons laquelle des copules en particulier semble être la plus appropriée. Si l'analyse est effectuée correctement, le résultat final est une fonction de copule bien ajustée qui, au moyen d'une simulation de Monte-Carlo, permet en outre d'analyser en détail le comportement probable des variables examinées.

Nous présentons deux méthodes d'estimation des paramètres de copule.

Selon le traitement avec les marges de copule, nous différons sur les modèles dits paramétriques et semi-paramétriques. Les modèles paramétriques spécifient et estiment les formes paramétriques des marges ainsi que de la fonction de copule. Au contraire, les modèles semi-paramétriques ne supposent aucune hypothèse sur les paramètres des marges et se substituent aux distributions marginales par les fonctions de distribution cumulative empiriques univariées. Nous décrivons les deux techniques en détail dans les deux sections suivantes, puis fermons en introduisant plusieurs tests de qualité de l'ajustement.

3.2 Méthode paramétrique d'estimation des copules

3.2.1 Fonction de répartition empirique

En premier lieu, on va expliquer comment estimer les marges F et G , ainsi que la loi conjointe H . Soit donc X_1, \dots, X_n , un échantillon aléatoire tiré d'une loi dont la fonction de répartition est F . Si F est inconnue, on peut l'estimer par la fonction de répartition empirique

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}, \quad (3.1)$$

Pour $x \in \mathbb{R}$ fixé, $F_n(x)$ correspond au nombre d'observations qui sont inférieures à x . Nous allons nous intéresser à présent au comportement asymptotique de F_n .

Tout d'abord, il faut noter que la loi des grands nombres assure que, pour x fixé, $F_n(x)$ converge en probabilité vers $F(x)$.

Le Théorème de Glivenko Cantelli étend largement ce résultat en démontrant la convergence uniforme de F_n vers F pour $x \in \mathbb{R}$. Ce résultat est énoncé dans la suite.

Théorème 3.1 *Soit $(X_n)_{n \geq 1}$, une suite de variables aléatoires de même fonction de répartition F . Alors*

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

converge en probabilité vers zéro.

Considérons maintenant le processus empirique

$$\mathbb{F}_n(x) = \sqrt{n} \{F_n(x) - F(x)\}.$$

Une application du Théorème central limite assure que pour un $x \in \mathbb{R}$ fixé, $F_n(x)$ converge vers une variable aléatoire Normale de moyenne nulle et de variance

$$\sigma_x^2 = F(x)(1 - F(x)).$$

Le Théorème de Donsker va beaucoup plus loin en obtenant le comportement limite en loi de $F_n(x)$ en tant que fonction aléatoire définie pour tout $x \in \mathbb{R}$. Avant, on va définir ce qu'est un pont Brownien.

Définition 3.1 *Un pont brownien \mathbb{B} est une fonction aléatoire définie sur $[0, 1]$ telle que*

$$\text{cov}\{\mathbb{B}(s), \mathbb{B}(t)\} = \min(s, t) - st, \forall (s, t) \in [0, 1]^2$$

et $\mathbb{B}(s)$ est une variable aléatoire suivant une loi normale de moyenne nulle et de variance $s(1 - s)$.

Le Théorème de Donsker peut maintenant être formellement énoncé.

Théorème 3.2 *Soit $\mathcal{D}(\mathbb{R})$, l'ensemble des fonctions définies sur \mathbb{R} qui sont continues à droite et qui possèdent une limite à gauche. Alors le processus empirique \mathbb{F}_n converge dans l'espace $\mathcal{D}(\mathbb{R})$ vers un processus gaussien \mathbb{F} tel que $\mathbb{F}(x) = \mathbb{B}\{F(x)\}$. Ainsi, \mathbb{F} est de moyenne nulle et possède la fonction de covariance*

$$\text{cov}\{\mathbb{F}(x), \mathbb{F}(y)\} = F\{\min(x, y)\} - F(x)F(y).$$

Soit maintenant $(X_1, Y_1), \dots, (X_n, Y_n)$, un échantillon aléatoire bivarié tiré d'une loi dont la fonction de répartition conjointe est H . Pour l'estimer, on définit la fonction de répartition empirique bivariée

$$H_n(x, y) = \frac{1}{n} \sum_{i=1}^n 1_{(X_i \leq x, Y_i \leq y)}. \quad (3.2)$$

Soit maintenant le processus empirique

$$\mathbb{H}_n(x, y) = \sqrt{n} \{H_n(x, y) - H(x, y)\}.$$

Une application du Théorème central limite assure que lorsque $(x, y) \in \mathbb{R}^2$ est fixé, $\mathbb{H}_n(x, y)$ converge vers une variable Normale de moyenne nulle et de variance

$$\sigma_{x,y}^2 = H(x, y)(1 - H(x, y)).$$

Le résultat qui suit est beaucoup plus fort dans la mesure où la convergence du processus empirique $\mathbb{H}_n(x, y)$ s'applique pour tout $(x, y) \in \mathbb{R}^2$.

Théorème 3.3 *Le processus empirique \mathbb{H}_n converge dans l'espace $\mathcal{D}(\mathbb{R}^2)$ vers un processus gaussien centré \mathbb{H} dont la fonction de covariance est*

$$\text{cov}\{\mathbb{H}(x, y), \mathbb{H}(x', y')\} = H\{\min(x, x'), \min(y, y')\} - H(x, y)H(x', y').$$

Avant de clore cette section, il est à noter que \mathbb{F}_n et \mathbb{H}_n sont reliés via

$$\mathbb{F}_n(x) = \lim_{y \rightarrow \infty} \mathbb{H}_n(x, y),$$

On peut ainsi déduire le comportement limite de \mathbb{F}_n à partir de celui de \mathbb{H}_n .

3.2.2 Copules empiriques

Dans cette section, nous allons expliquer comment estimer une copule C avec la copule empirique C_n . D'abord, on rappelle que

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

Un estimateur naturel de C consiste à remplacer H par H_n , F par F_n et G par G_n . On a alors

$$\tilde{C}_n(u, v) = H_n\{F_n^{-1}(u), G_n^{-1}(v)\}.$$

De la définition de H_n , on a donc

$$\tilde{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq F_n^{-1}(u), Y_i \leq G_n^{-1}(v)\}}.$$

À noter que l'inégalité $X_i \leq F_n^{-1}(u)$ est à peu près équivalente à $F_n(X_i) \leq u$, et similairement pour $Y_i \leq G_n^{-1}(v)$. On peut donc écrire que

$$\tilde{C}_n(u, v) = C_n(u, v) \frac{1}{n} \sum_{i=1}^n 1_{\{F_n(X_i) \leq u, G_n(Y_i) \leq v\}}.$$

La fonction C_n ci-dessus est la définition généralement admise de la copule empirique.

En notant R_i le rang de X_i parmi X_1, \dots, X_n , et S_i le rang de Y_i parmi Y_1, \dots, Y_n , on a que

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1_{\{\frac{R_i}{n} \leq u, \frac{S_i}{n} \leq v\}}$$

Le résultat suivant concerne le comportement asymptotique du processus de copule empirique défini par

$$\mathbb{C}_n(u, v) = \sqrt{n} \{C_n(u, v) - C(u, v)\}.$$

Théorème 3.4 *Le processus empirique C_n converge faiblement dans l'espace $\ell^\infty([0, 1]^2)$ des fonctions réelles bornées sur $[0, 1]$ vers un processus gaussien de représentation*

$$\mathbb{C}(u, v) = \mathbb{D}_C(u, v) - C_{10}(u, v)\mathbb{D}_C(u, 1) - C_{01}(u, v)\mathbb{D}_C(1, v),$$

où $C_{10}(u, v) = \partial C(u, v)/\partial u$ et $C_{01}(u, v) = \partial C(u, v)/\partial v$ sont les dérivées partielles de C et \mathbb{D}_C est un processus gaussien sur $[0, 1]^2$ de fonction de covariance

$$\text{cov}\{\mathbb{C}(u, v), \mathbb{C}(u', v')\} = C\{\min(u, u'), \min(v, v')\} - C(u, v)C(u', v').$$

On constate que \mathbb{C} possède une structure de covariance qui dépend de la copule et de ses dérivées partielles, ces dernières sont inconnues. Comme l'estimation de C_{10} et C_{01} est difficile, une approximation de cette distribution limite est requise en utilisant des procédures bootstrap ; voir Fermanian, Radulović & Wegkamp (2004) pour plus de détails. Toutefois, l'usage du bootstrap n'est pas approprié, et ce dans plusieurs situations énumérées dans Bickel, Götze & van Zwet (1997). À cet effet, on présente dans ce qui suit une procédure de ré-échantionnage alternative, à savoir la méthode du multiplicateur, voir Quessy (2014), Scaillet (2005) et Rémillard & Scaillet (2009).

3.3 Méthodes paramétriques

On se place ici dans le cas où la distribution conjointe dépend d'un paramètre, que l'on cherche à estimer.

3.3.1 Méthode des moments

Cette méthode est notamment utilisée pour les mesures de dépendance, l'estimateur des moments de la mesure de dépendance considérée est alors simplement obtenu en égalant l'expression paramétrique (analytique) de la mesure avec un estimateur non paramétrique de cette même mesure.

3.3.2 Maximum de vraisemblance

Considérons un vecteur aléatoire X , absolument continu, telle que la densité jointe s'écrive

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \cdot \prod_{i=1}^d f_i(x_i)$$

où $c : [0, 1]^d \rightarrow \mathbb{R}^+$ est la densité de la copule associée à X , et où f_i est la densité de la variable X_i .

La log-vraisemblance $\log \mathcal{L}$ associée à un échantillon X_1, \dots, X_n i.i.d. s'écrit

$$\log \mathcal{L}_n = \sum_{k=1}^n \log c(F_1(x_1^k), \dots, F_d(x_d^k)) + \sum_{i=1}^d \sum_{k=1}^n \log f_i(x_i^k)$$

et peut se décomposer en deux termes : celui de gauche est associé à la structure de dépendance, et le second aux lois marginales. Notons que le second terme est le seul qui apparaît si l'on suppose que les composantes du vecteur X sont indépendantes.

On supposera que la copule C appartient à une famille paramétrique $C = \{C_\theta, \theta \in \Theta\}$ et que les lois marginales sont également dans des familles paramétriques,

$$F_i \in \mathcal{F}_i = \{F_{\alpha_i}, \alpha_i \in A_i\}.$$

Sous les conditions usuelles de régularités, l'estimateur du maximum de vraisemblance $(\hat{\theta}, \hat{\alpha})$, solution de

$$(\hat{\theta}, \hat{\alpha}) = \arg \max \{\log \mathcal{L}_n(\theta, \alpha)\},$$

est consistant et asymptotiquement Gaussien, au sens où

$$\sqrt{n} \left\{ (\hat{\theta}, \hat{\alpha}) - (\theta, \alpha) \right\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

avec

$$\Sigma = - \left(\lim_{n \rightarrow \infty} \frac{\partial^2 \log \mathcal{L}_n(\theta, \alpha)}{\partial(\theta, \alpha) \partial(\theta, \alpha)'} \Big|_{(\theta, \alpha)} \right)^{-1}.$$

3.4 Sélection et adéquation de la copule

Comme pour les fonctions de distribution à une variable, une famille de copules est choisie pour ses particularités à représenter un ou plusieurs aspects des données. Ensuite, l'ajustement doit être réalisé pour retenir la copule qui modélise au mieux les observations.

A cet effet, Venter (2002, 2003) et Belguisse (2001) proposent d'utiliser des fonctions descriptives K , J , M , L et R pour tester l'adéquation d'une copule ou pour sélectionner une copule ayant des caractéristiques spécifiques par exemple au niveau des queues de distribution. Il s'agit de comparer les valeurs empiriques et les valeurs estimées pour les différentes copules. Cette comparaison peut être faite par un test de Kolmogorov-Smirnov ou du χ^2

au seuil α . Cette procédure a été proposée pour la fonction K par Genest & Rivest (1993) et Genest et al. (2006); ils utilisent les tests de Kolmogorov-Smirnov et Cramér-von Mises pour l'adéquation des copules.

Pour un ajustement complémentaire des copules, Hürlimann (2004) propose le test du χ^2 bidimensionnel.

- **La fonction $K(z)$** : Il s'agit de la fonction de répartition de la variable aléatoire $C(U, V)$. Pour une copule archimédienne, $K(z)$ (Genest & Rivest, 1993) est définie par :

$$K(z) = z - \frac{\varphi'(z)}{\varphi(z)} \quad (3.3)$$

La version empirique de la fonction $K(z)$ est exprimée par :

$$K(z) = \frac{1}{n} \sum_{i=1}^n 1_{\{Z_i \leq z\}}. \quad (3.4)$$

- **La fonction $J(z)$ ou de τ cumulatif** : Le coefficient de corrélation des rangs τ de Kendall est lié à la copule par l'expression suivante :

$$\tau = \frac{-1 + 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv}{-1 + 4E(C(\mathbf{u}, \mathbf{v}))} \quad (3.5)$$

$$-1 + 4E(C(\mathbf{u}, \mathbf{v})) \quad (3.6)$$

où $E(\cdot)$ est l'opérateur de l'espérance mathématique.

La fonction $J(z)$ s'écrit :

$$J(z) = -1 + 4 \int_0^z \int_0^z C(u, v) c(u, v) du dv / C(z, z)^2.$$

La double intégrale représente la moyenne pondérée de $C(u, v)$, le poids étant $C(z, z)$; le quotient représente la moyenne de $C(u, v)$ qui croît en fonction de z ; le second $C(z, z)$ compare la moyenne de C à sa valeur maximale. On a $J(1) = \tau$.

La version empirique de la fonction $J(z)$ est exprimée par :

$$J(z) = -1 + 4I(z)/C(z, z)^2$$

où $I(z)$ est définie par

$$I(z) = \frac{1}{n} \sum_{i=1}^n z_i 1_{\{u_i \leq z, v_i \leq z\}}.$$

avec 1 est la fonction indicatrice et $u_i = \frac{\text{rang}(x_i)}{n+1}$ et $v_i = \frac{\text{rang}(y_i)}{n+1}$.

- **La fonction $M(\mathbf{z})$** : est la moyenne conditionnelle cumulative, elle est définie par :

$$M(z) = E(V|U \leq z) = \left(\int_{u=0}^z \int_{v=0}^1 v \cdot c(u, v) dudv \right) / z \quad (3.7)$$

et vérifiant $M(1) = 1/2$. En définissant $\mathbf{D}(\mathbf{z})$ et $\mathbf{N}(\mathbf{z})$ comme suit et ,

$$\mathbf{D}(\mathbf{z}) = \sum_{i=1}^n 1_{\{u_i \leq z\}}, \mathbf{N}(\mathbf{z}) = \sum_{i=1}^n v_i 1_{\{u_i \leq z\}}$$

la version empirique de $M(z)$ est donnée par :

$$M(z) = \mathbf{N}(\mathbf{z})/\mathbf{D}(\mathbf{z}),$$

avec $D(1) = n$ et $N(1) = n/2$.

- **Les fonctions $\mathbf{L}(\mathbf{z})$ et $\mathbf{R}(\mathbf{z})$** : Ces deux fonctions montrent les concentrations des queues gauche (Left) et droite (Right) de la distribution. Les événements extrêmes sont mis en évidence particulièrement aux limites qui, si elles existent, représentent la dépendance asymptotique définie par Joe (1997) : λ_U équation (2.3) pour les petites valeurs extrêmes (queue gauche), et λ_L équation (2.1) pour les grandes valeurs extrêmes (queue droite).

Les deux fonctions $\mathbf{L}(\mathbf{z})$ et $\mathbf{R}(\mathbf{z})$ sont alors :

$$\mathbf{L}(\mathbf{z}) = \mathbf{P} \quad (U \leq z, V \leq z) / z^2 = C(z, z) / z^2 \quad (3.8)$$

$$\mathbf{R}(\mathbf{z}) = \mathbf{P} \quad (U > z \text{ ou } V > z) / (1 - z)^2 \quad (3.9)$$

$$= (1 - 2z + c(z, z)) / (1 - z)^2. \quad (3.10)$$

On s'intéresse à la fonction \mathbf{L} pour tout $z \in [0, 0.5]$ et à la fonction \mathbf{R} pour tout $z \in [0.5, 1]$, ces deux fonctions ont leurs versions empiriques.

3.4.1 Teste de Khi deux bidimensionnel

Il s'agit ici de mettre en place le test du χ^2 bidimensionnel pour l'adéquation de la distribution bidimensionnelle $F(x, y)$ obtenue grâce à la copule. La méthodologie adoptée reprend celle proposée par Hürlimann (2004). L'espace est subdivisé en pavés $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$. Les fréquences théoriques $f_{i,j}$ de chacun des pavés sont calculées par :

$$f_{i,j} = C(F_X(x_i), F_Y(y_j)) - C(F_X(x_{i-1}), F_Y(y_j)) - C(F_X(x_i), F_Y(y_{j-1})) + C(F_X(x_{i-1}), F_Y(y_{j-1}))$$

Ces pavés sont regroupés en k classes, telles que chacune d'elles contient au moins 5% des observations afin de pouvoir utiliser la statistique suivante du χ^2

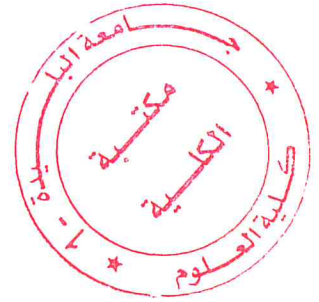
$$\chi_{obs}^2 = \sum_{k=1}^m \frac{(O_k - E_k)^2}{E_k}$$

où m est le nombre de classes, O_k et E_k sont les effectifs respectivement observé et théorique de la classe k , $k = 1, \dots, m$. Cette statistique suit une loi du chi-deux à $(m - r - 1)$ degrés de liberté avec r le nombre de paramètres estimés de la copule. On se donne ensuite un seuil critique α pour la zone de rejet et on calcule la p-valeur $\hat{\alpha} = P(\chi_{m-r-1}^2 > \chi_{obs}^2)$. On rejette l'adéquation de la copule paramétrique si $\hat{\alpha} < \alpha$.

Il faut préciser pour les praticiens que la p-valeur exacte $\hat{\alpha}$ se situe en fait dans l'intervalle suivant :

$$P(\chi_{m-1}^2 > \chi_{obs}^2) \leq \hat{\alpha} \leq P(\chi_{m-r-1}^2 > \chi_{obs}^2)$$

Cet encadrement étant de faible amplitude lorsque r est petit vis à vis de n , le seuil critique α sera le plus souvent en dehors de cet encadrement et on conclura sans ambiguïté. Parmi les copules acceptées, on sélectionne alors celle qui maximise la p-valeur.



Chapitre 4

Application sur la modélisation de la sécheresse

4.1 Définition de la sécheresse

Plusieurs définitions existent, la meilleure définition est celle de Périer, l'état de la sécheresse peut cependant être caractérisé comme un déficit hydrique marqué dans une ou plusieurs composante(s) du cycle hydrologique. Ce manque d'eau est généralement dû à de trop faibles précipitations (Alley, 1984 ; Chang et Cleopatra, 1991) sur une période donnée, par rapport à la moyenne des apports observés sur cette période et a un impact direct sur l'alimentation des différents compartiments du bassin versant (surface, sol et nappes). Si les précipitations sont trop faibles ou inexistantes sur une période prolongée, l'apport d'eau à la surface du sol et dans les couches de sol plus profondes est par conséquent amoindri et l'eau disponible dans les cours d'eau et/ou pour la végétation peut alors elle aussi être déficitaire. Pour cette raison, trois grandes catégories de sécheresses ont été définies dans un premier temps par Dracup et al. (1980) et reprises par Wilhite et Glantz (1985) et sont aujourd'hui couramment utilisées par les climatologues et les hydrologues pour l'étude et le suivi des sécheresses.

4.2 Type de la sécheresse

4.2.1 Sécheresses météorologiques

Les sécheresses météorologiques sont caractérisées par un déficit des précipitations, solides et liquides (Palmer, 1965 ; Boken, 2005 ; Keyantash & Dracup, 2002). Ainsi, il s'agit d'une période, qui peut varier du mois à l'année, voire

dans des cas extrêmes, a plusieurs années, durant laquelle les précipitations sont inférieures à la normale. Les sécheresses météorologiques sont souvent déclenchées par des anomalies persistantes de grande échelle des températures de surface de la mer (Bjerknes, 1969 ; Rasmusson et Wallace, 1983 ; Folland et al. 1986 ; Lamb et Pepler, 1992 ; Ting et Wang, 1997 ; Trenberth et Shea, 2005).

4.2.2 Sécheresses agricoles

Les sécheresses agricoles (ou édaphiques) sont caractérisées par un déficit lié à la réserve en eau du sol. Il s'agit d'une période durant laquelle l'humidité du sol est inférieure à sa valeur moyenne, ce qui a des conséquences directes sur la végétation, qu'elle soit naturelle ou cultivée (Palmer, 1965 ; Rosenberg, 1978 ; Wilhelmi, 2002).

Ces sécheresses sont généralement provoquées par un cumul des précipitations inférieure à la normale (Narasimhan et Srinivasan, 2005), ou par une distribution temporelle plus irrégulière (c.a.d. des précipitations moins fréquentes mais plus intenses, phénomènes souvent observés sur le pourtour méditerranéen), mais peuvent parfois être engendrées par des taux d'évapotranspiration plus élevés (Klocke et Hergert, 1990 ; Rind et al., 1990 ; Hanson, 1991 ; Vicente-Serrano et al., 2010) ou des processus de ruissellement plus intenses, en comparaison à la normale saisonnière. Les sécheresses agronomiques ont souvent de lourdes conséquences sur la production agricole (Panu et Sharma, 2002).

4.2.3 Sécheresses hydrologiques

Les sécheresses hydrologiques peuvent à la fois définir le débit d'un cours d'eau comme trop faible mais peuvent aussi représenter le fait qu'un réservoir du sol ou du sous-sol n'est pas suffisamment réalimenté. Ainsi, les sécheresses hydrologiques dépendent du degré d'approvisionnement en eaux de surface et en eaux souterraines des lacs, réservoirs, aquifères et cours d'eau. L'impact d'une sécheresse hydrologique est important sur les activités humaines, puisqu'elle va avoir de fortes conséquences sur l'irrigation, les activités touristiques, la production d'énergie hydroélectrique, les transports (dans certains pays), l'alimentation en eau domestique et la gestion/protection de l'environnement.

4.2.4 Sécheresse socio-économique

La sécheresse socio-économique englobe les effets économiques et sociaux des formes précédentes (exode rurale, famine, baisses des rendements agricoles, perte de cheptel, augmentation des prix des produits de première nécessité, problème de santé humaine et animale, etc). “La sécheresse socio-économique est donc une forme de choc interne de l’offre, c’est-à-dire, une grave perturbation provoquée par des événements hors du contrôle du pays, et qui a des effets importants sur les variables économiques intérieurs”

4.3 Préviation de la sécheresse

Contrairement à toutes les autres catastrophes, la sécheresse est le phénomène naturel le plus lent à se développer. Cela permet généralement d’avoir plus du temps d’atténuer ses effets et mettre en oeuvre des plans de suivi appropriés.

La préviation de la sècheresse permettra ainsi de prévoir les conditions de sécheresse à l’avance par soit quelques mois, saisons ou quelques années. Par ailleurs, divers outils et méthodes pour la préviation des sécheresses ont été proposés et testés dans différentes régions au cours des dernières décennies (Dastorani & Afkhami, 2011) en se basant sur deux approches fondamentales, à savoir :

- La première est la prédiction des conditions hydrologiques
- La seconde est la prédiction des indices de sécheresse.

La préviation des conditions hydrologiques comporte habituellement la préviation du climat et la préviation d’écoulement dans les rivières. Généralement la préviation du climat est basée sur l’analyse statistique des données observées et les méthodes dynamiques qui résolvent numériquement les équations physiques qui régissent le système climatique. Actuellement la détection et la surveillance des conditions de sécheresse sont essentiellement basées sur certains indices. Ces indices permettent de déterminer d’une façon scientifique le seuil indiquant la sécheresse à différentes échelles de temps et de définir des classes d’appartenance à cet événement en fonction de sa sévérité et de sa position.

La communauté scientifique a développé et étudié un certain nombre d’indicateurs de sècheresse au cours des cinquante dernières années afin d’anticiper, de suivre et de caractériser les différents épisodes extrêmes qui touchent diverses régions du globe. La plupart des indices de sécheresse ont été élaborés aux Etats-Unis, mais sont régulièrement employés sur les autres continents. Heim (2002) a répertorié tous les indices développés au cours du XXIème

siècle aux États-Unis.

Foley (1957) fut le premier à suggérer le calcul d'un indice donnant des informations sur la sévérité d'une sécheresse. Ce n'est que plus tard que des indicateurs de sécheresse plus pertinents furent définis. Très utilisé aux États-Unis (Heim, 2002), le Palmer Drought Severity Index (PDSI) (Palmer, 1965), par exemple, est un indicateur calculé à partir d'un bilan d'eau simplifié. Ainsi, en fonction de la valeur du PDSI on peut connaître l'intensité d'une sécheresse en un lieu donné au fil du temps. Le PDSI a été calibré en fixant arbitrairement les sécheresses extrêmes à une valeur correspondant aux sécheresses extrêmes historiques. Comme nous l'avons vu dans la partie précédente, la sécheresse en un point donné est un état relatif par rapport à la moyenne habituellement observée en ce point et un étalonnage réalisé pour une région donnée n'a donc pas forcément de sens en d'autres régions du globe.

Un indicateur pertinent pour certaines zones géographiques peut ainsi manquer de cohérences ailleurs. Pour pallier l'arbitraire de cet étalonnage, un scPDSI (self calibrated PDSI) a alors été considéré (Briffa et al., 2009). Ces indicateurs ne proposent cependant pas d'échelle de temps explicite permettant de détecter à la fois les sécheresses courtes et les sécheresses longues (Alley, 1984; Gutmann et al., 1992; Lloyd-Hughes et Saunders, 2002).

Un autre type d'indicateur, recommandé par l'OMM (Organisation Météorologique Mondiale) en 2009 pour surveiller les sécheresses et gérer les risques liés au climat, est le Standardized Index (Gibbs, 1975). Très largement utilisé par la communauté scientifique (McKee et al., 1993; Lloyd-Hughes et Saunders, 2002; Labedzki, 2007; Bordi et al., 2009; Vidal et al., 2010, etc.), cet indicateur peut être calculé à partir de différentes variables, telles que les précipitations, l'humidité du sol, les débits ou encore l'évapotranspiration potentielle. Le plus connu et le plus utilisé est le Standardized Precipitation Index (SPI, McKee et al., 1993). Il est basé, comme son nom l'indique sur les données de précipitation et permet de quantifier les sécheresses météorologiques. Il a été développé en vue de caractériser les déficits de précipitation pour une période donnée et correspond à la transformation de la série temporelle des précipitations en une distribution normale standardisée de moyenne nulle et d'écart-type unitaire (distribution gaussienne). Cela revient donc à ajuster une fonction de densité de probabilité à la distribution de la fréquence des précipitations sommées ou moyennes, sur une échelle temporelle définie (généralement, 3, 6, 9, 12 et 24 mois). L'ajustement se fait séparément pour chacun des mois de l'année afin de conserver la saisonnalité. Chaque densité de probabilité cumulée est ensuite transformée en une distribution normale standardisée (Gaussienne), de moyenne nulle et d'écart type unitaire. Chaque valeur du SPI est alors reliée à un degré de sévérité et une probabilité d'oc-

currence de la sécheresse. Il a été montré à plusieurs reprises que le SPI était plus adapté que le PDSI pour suivre les sécheresses.

Actuellement, les variables autres que les précipitations sont aussi étudiées, notamment l'évapotranspiration (Vicente-Serrano et al.2010), afin de définir l'indicateur le plus pertinent à utiliser en fonction de la région que l'on considère. En effet, si les précipitations semblent être aujourd'hui l'outil le plus performant pour le suivi des sécheresses, car elles conditionnent rapidement la quantité d'eau dans tous les réservoirs du cycle hydrologique, il a par ailleurs été montré que dans le cadre du réchauffement global la demande évaporatoire va augmenter (Wang et al.2011 ; Stefanon et al., 2012), ce qui aura probablement un poids dans la genèse des sécheresses. D'autres indicateurs basés sur l'étude des déciles (Gibbs et Maher, 1967 ; Meko, 1985) sont utilisés par la communauté scientifique et permettent notamment, par exemple, de mettre en évidence les sécheresses hydrologiques.

L'inconvénient de tous ces indices est qu'ils sont pertinents seulement lorsqu'ils sont calculés sur des séries longues de données (au moins 30 ans). Ainsi, il paraît difficile d'anticiper les sécheresses en tout point du globe, à partir de données observées. Si plusieurs types de jeux de données observées, notamment les données satellitaires et les données in situ, permettent aujourd'hui d'étudier et de suivre précisément, voire d'anticiper les épisodes de sécheresses sur certaines régions du globe, il s'avère intéressant d'exploiter au maximum aussi les données simulées grâce à des modèles numériques, qui permettent généralement de produire des données sur de plus longues périodes à l'échelle globale. Enfin, l'impact des sécheresses sur l'agriculture doit être évalué à une échelle temporelle plus détaillée que le mois (par exemple par périodes de 10 jours) et des observations plus directement liées au fonctionnement de la végétation peuvent avoir une valeur ajoutée importante.

4.4 Analyse des caractéristiques de la sécheresse :

L'identification du début et la fin d'un épisode de sécheresse reste toujours difficile. Cependant, l'utilisation des indices comme celle de SPI permet non seulement le contrôle et la prédiction de la sécheresse mais aussi le suivi de ses caractéristiques. Les caractéristiques de sécheresse font principalement référence à la durée, la sévérité et l'intensité.

- **La durée** : est la longueur de la période dans laquelle les valeurs de l'indice de la sécheresse sont inférieures à certain seuil. Ce seuil est généralement déterminé par la communauté scientifique suivant l'indice

utilisé.

- **La sévérité** : est la cumulation des valeurs de l'indice sur la période du temps de la sécheresse
- **L'intensité** : parfois appelé magnitude, est définie comme le rapport entre la sévérité et la durée de sécheresse (Dingman 1994 ; Shiao 2006).
- Une autre caractéristique importante de la sécheresse est le temps inter-arrival, qui est définie comme le temps entre le début d'une sécheresse et le début de la suivante (Shiao 2006).

4.5 Indice de précipitation normalisé

L'indice SPI (McKee et al., 1993, 1995) est un indice à la fois puissant, souple d'utilisation et simple à calculer. Les données sur les précipitations constituent en fait le seul paramètre requis. En outre, l'indice SPI se révèle tout aussi efficace pour analyser les périodes ou cycles humides que les périodes ou cycles secs.

Pour calculer l'indice SPI, il faut disposer idéalement de relevés mensuels s'étalant sur au moins 20 à 30 ans, mais de préférence sur 50 à 60 ans.

4.5.1 Description de l'indice normalisé de précipitations (SPI)

- L'indice SPI est fondé sur la probabilité de précipitations sur un laps de temps donné. La probabilité des précipitations observées est transformée en un indice qui sert aussi bien à l'expérimentation qu'à l'exploitation dans plus de 70 pays.
- **Destinataires** : Les responsables de la planification en matière de sécheresse sont nombreux à apprécier la souplesse d'utilisation de l'indice SPI. L'indice est utilisé aussi dans divers instituts de recherche, des universités et des Services météorologiques et hydrologiques nationaux du monde entier, dans le cadre d'activités de suivi de la sécheresse et d'alerte précoce dans le domaine.
- **Avantages** : Les précipitations représentent le seul paramètre dont il faut disposer. Il est possible de calculer l'indice pour diverses échelles de temps, celui-ci permettant de détecter rapidement les situations de sécheresse et d'en évaluer la gravité.
- **Inconvénients** : Il ne permet de quantifier que le déficit de précipitations ; les valeurs basées sur des données préliminaires peuvent changer et les valeurs varient si la durée de la période des relevés s'allonge.

- L'indice SPI a été conçu pour quantifier le déficit de précipitations à de multiples échelles de temps. Ces échelles de temps traduisent les incidences de la sécheresse sur la disponibilité des différents types de ressources en eau. L'humidité du sol réagit relativement vite aux anomalies de précipitations, tandis que les eaux souterraines, le débit des cours d'eau et les volumes stockés dans les réservoirs sont sensibles aux anomalies de précipitations à plus long terme. C'est pourquoi McKee et al. (1993) ont initialement calculé l'indice SPI pour des laps de temps de 3, 6, 12, 24 et 48 mois.
- Le calcul de l'indice SPI en un lieu, quel qu'il soit, est basé sur un historique de précipitations sur une longue durée correspondant au laps de temps étudié.
- On ajuste une distribution de probabilité à cette longue série de relevés, puis on la transforme en une distribution normale pour que l'indice SPI moyen, du lieu considéré et pour le laps de temps étudié, soit égal à zéro (Edwards et McKee, 1997).
- Les valeurs positives de l'indice SPI indiquent des précipitations supérieures à la médiane et les valeurs négatives, des précipitations inférieures à la médiane.
- Étant donné que l'indice est normalisé, il est possible de représenter de la même manière les climats humides et les climats arides; ainsi est-il possible aussi, grâce à l'indice SPI, d'assurer une surveillance des périodes humides. McKee et al. (1993) ont utilisé le système de classification présenté dans le tableau des valeurs de l'indice SPI figurant ci-après (tableau 4.1) pour définir l'intensité des épisodes de sécheresse en fonction de la valeur de l'indice.
- Ils ont aussi défini les critères d'un épisode de sécheresse pour une échelle de temps quelle qu'elle soit. Une sécheresse sévit quand l'indice présente de façon continue une valeur négative de -1,0 ou moins et se termine lorsque l'indice devient positif.
- Par conséquent, on distingue pour chaque épisode de sécheresse une durée, avec un commencement et une fin, et une intensité pour chaque mois durant lequel l'épisode se poursuit. Pour obtenir ce que l'on peut appeler la «magnitude» de la sécheresse, il suffit d'ajouter toutes les valeurs de l'indice pour l'ensemble des mois sur lesquels porte

TAB. 4.1 – Valeurs de l'indice SPI.

l'épisode.2.0 et plus	Extrêmement humide
De 1.5 à 1.99	Très humide
De 1.0 à 1.49	Modérément humide
de -0.99 à 0.99	Proche de la normale
de -1.0 à -1.49	Modérément sec
de -1.99 à -1.5	Très sec
<-2	Extrêmement sec

4.5.2 Valeurs de l'indice de précipitations normalisé pour une période plus ou moins longue

Indice SPI sur 1 mois

Une carte de l'indice SPI calculé sur 1 mois ressemble beaucoup à une carte représentant le pourcentage de la normale des précipitations pour une période de 30 jours. L'indice dérivé offre en fait une représentation plus exacte des précipitations mensuelles, car la distribution est normalisée.

L'indice SPI sur 1 mois peut fournir une approximation des conditions représentées par l'indice d'humidité des cultures, qui fait partie de la série d'indices que regroupe l'indice de sécheresse de Palmer.

Il est possible de mal interpréter l'indice SPI sur 1 mois si l'on ne tient pas compte des caractéristiques climatiques du lieu. Dans les régions où il est normal que les pluies soient faibles au cours d'un mois donné, il est possible d'obtenir des valeurs négatives ou positives importantes alors que l'écart par rapport à la moyenne est relativement faible. L'indice SPI sur 1 mois peut aussi induire en erreur lorsque les valeurs des précipitations sont inférieures à la normale dans des régions où, pour le mois considéré, le total normal des précipitations est faible. Tout comme les cartes du pourcentage de la normale des précipitations, celles de l'indice SPI sur 1 mois fournissent des informations utiles, mais il faut faire preuve de prudence quand on les analyse.

Indice SPI sur 3 mois

L'indice SPI sur 3 mois établit la comparaison entre le total des précipitations sur la période de trois mois examinée et les totaux des précipitations pour cette même période de trois mois de toutes les années pour lesquelles on dispose de relevés. En d'autres termes, l'indice SPI sur 3 mois se terminant fin février permet de comparer le cumul des précipitations des mois de

décembre, janvier et février de l'année examinée aux cumuls de précipitations de décembre à février de toutes les années figurant dans l'historique des observations exécutées à la station étudiée. Chaque nouvelle année, les données d'observation viennent s'ajouter à la série des relevés dont la durée gagne ainsi une année et les valeurs de toutes les années sont une nouvelle fois prises en compte. Les valeurs peuvent alors évoluer, ce qui sera le cas quand on comparera, d'un point de vue historique et statistique, l'année en cours avec toutes les années précédentes composant la série chronologique des observations.

Indice SPI sur 6 mois

L'indice SPI sur 6 mois établit la comparaison entre les précipitations sur la période de six mois examinée et les totaux des précipitations pour cette même période de six mois de toutes les années pour lesquelles on dispose de relevés. À titre d'exemple, un indice SPI sur 6 mois se terminant fin septembre permet de comparer le cumul des précipitations d'avril à septembre de l'année examinée aux cumuls de précipitations de cette même période pour toutes les années écoulées.

L'indice SPI sur 6 mois fournit une indication sur les tendances des précipitations sur une saison et jusqu'à moyenne échéance; on considère que pour cette échelle de temps, il présente encore davantage de sensibilité aux conditions que l'indice Palmer.

Indice SPI sur 9 mois

L'indice SPI sur 9 mois fournit une indication sur les régimes de précipitations inter-saisonniers, à moyenne échéance. Il faut en général au moins une saison pour que des conditions de sécheresse s'installent. Un indice SPI établi sur 9 mois et présentant des valeurs inférieures à -1,5 est habituellement un bon indicateur de conditions sèches aux conséquences importantes pour l'agriculture et pouvant affecter d'autres secteurs également. Pour certaines régions, on s'apercevra que la représentation cartographique de l'indice Palmer correspond assez bien à celle de l'indice SPI sur 9 mois. Pour d'autres régions, le rapprochement se fera davantage entre l'indice Palmer et l'indice SPI sur 12 mois. C'est à partir de 9 mois qu'on commence à établir le lien entre une sécheresse saisonnière de courte durée et une sécheresse à plus long terme pouvant se transformer en sécheresse hydrologique ou en sécheresse s'étalant sur plusieurs années.

4.6 Indices Hydrométriques (Stream flow Drought Index (SDI))

Comme suggéré par Mckee et al. (1993), la procédure de calcul SPI peut également être appliquée à d'autres variables d'eau, telle que l'humidité du sol, l'accumulation de neige, écoulement dans les rivières, les nappes et les eaux souterraines.

Le SDI développé par Nalbantis et Tsakiris (2009) et l'indice de ruissellement normalisé (SRI) développé par (Shukla et Wood 2008), ont des procédures de calcul très similaires à celle de SPI. Les différentes classes de indices SDI et leur signification sont identiques à celle utilisés dans pour l'indice de sécheresse SPI et RDI (tableau 4.2).

TAB. 4.2 – Définition des états de sécheresse hydrologique à l'aide de SDI.

state	Description	Critère
0	Non-drought	$SDI=0$
1	Mild drought	$-1 < SDI < 0.0$
2	Moderate drought	$-1.5 < SDI < -1.0$
3	Severe drought	$-2 < SDI < -1.5$
4	Extreme drought	$SDI < -2.0$

Bien que Mckee ait suggéré que la distribution gamma peut être bien ajusté les données d'écoulements dans les rivières, alors que Nalbantis 2008 ; Shukla et Wood 2008 ont constaté que la distribution log-normale est la loi le mieux adéquat pour l'ajustement des débits.

Dans notre travail, nous nous ajustons la série des débits à la fois à la loi gamma et à la distribution log-normale et nous déterminons la qualité de l'ajustement pour les cours d'eau Algériens en fonction de l'échelle des temps. Donc, pour l'application de la copule, les données de débits cumulés sont ajustées sur les deux distributions pour les différentes échelles des temps afin de sélectionner la meilleure distribution pour le cas des cours d'eau Algériens.

4.7 Ajustement des paramètres de la sécheresse par des lois de probabilités

4.7.1 SPI de 3 mois

Les différents paramètres des caractéristiques de trois mois sont présentées dans le tableau suivant :

TAB. 4.3 – Paramètres statistiques des caractéristiques d’une sécheresse pour spi de trois mois

state	Description	Critère
0	Non-drought	SDI=0
1	Mild drought	$-1 < \text{SDI} < 0.0$
2	Moderate drought	$-1.5 < \text{SDI} < -1.0$
3	Severe drought	$-2 < \text{SDI} < -1.5$
4	Extreme drought	$\text{SDI} < -2.0$

Les résultats d’ajustement des spi de trois mois par des différentes distributions de probabilités sont présentées dans les trois tableaux suivants :

La durée :

TAB. 4.4 – Ajustement de la durée de la sécheresse pour un spi de trois mois

Lois	paramètre1	error1	paramètre2	error2	ks	p-value
Weibull	forme :2.0654	0.3029	échelle : 6.9916	0.6634	0.0998	0.9349
exponentielle	rate :0.1620	0.0301	-	-	0.2768	0.0235
gamma	forme :3.4314	0.8609	rate : 0.5559	0.1502	0.7171	2.237e-13
log-normale	mean :1.6674	0.1078	sd :0.5805	0.0762	0.1366	0.6516

L’intensité :

La sévérité :

TAB. 4.5 – Ajustement de l'intensité de la sécheresse pour un spi de trois mois

Lois	paramètre1	error1	paramètre2	error2	ks	p-value
Weibull	forme : 1.3351	0.1799	échelle : 0.6375	0.0941	0.1860	0.2685
exponentielle	rate : 1.7221	0.3198	-	-	0.2101	0.1543
gamma	forme : 1.8754	0.4555	rate : 3.2296	0.8984	0.8428	2.2e-16
log-normale	mean : -0.8332	0.1436	sd : 0.7734	0.1016	0.1290	0.7201

TAB. 4.6 – Ajustement de la sévérité de la sécheresse pour un spi de trois mois

Lois	paramètre1	error1	paramètre2	error2	ks	p-value
Weibull	forme : 2.0168	0.2876	échelle : 0.1094	0.0106	0.1563	0.478
exponentielle	rate :10.3571	1.9233	-	-	0.3008	0.0105
gamma	forme :3.5108	0.8817	rate : 36.3619	9.8171	1	2.2e-16
log-normale	mean : -2.4868	0.1055	sd :0.5684	0.0746	0.1310	0.702

4.7.2 Analyse statistiques de spi de 6 mois

La durée :

TAB. 4.7 – Ajustement de la durée de la sécheresse pour un spi de six mois

Lois	paramètre1	error1	paramètre2	error2	ks	p-value
Weibull	forme :1.8690	0.2791	échelle : 7.3243	0.7676	0.1385	0.6343
exponentielle	rate :0.1543	0.0286	-	-	0.2670	0.032
gamma	forme :2.7929	0.6941	rate : 0.4308	0.1173	0.8729	2.2e-16
log-normale	mean :1.6793	0.1218	sd :0.6558	0.0861	0.1429	0.5954

L'intensité :

La sévérité :

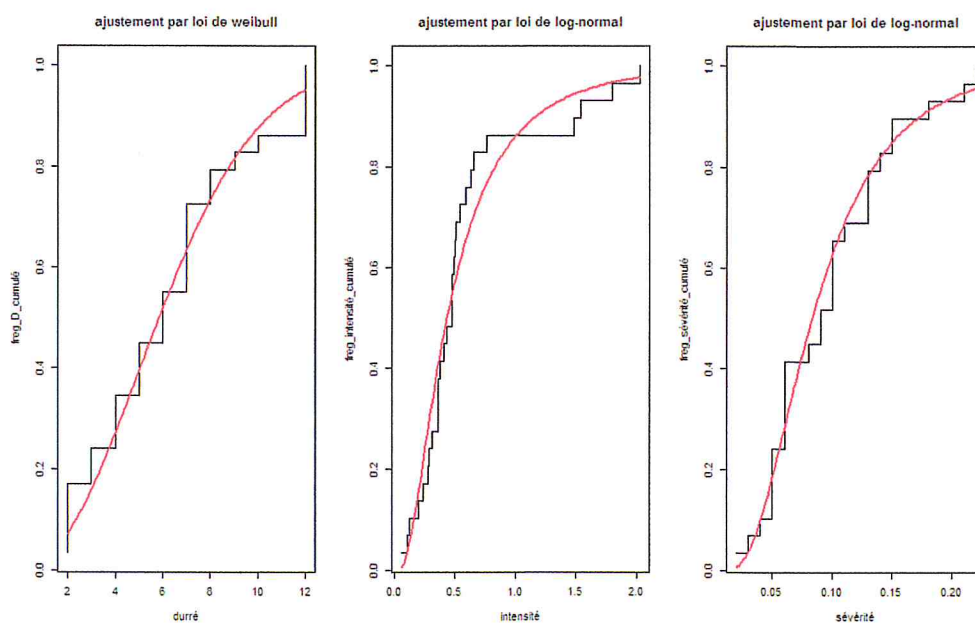


FIG. 4.1 – distributions d’ajustement des durrée, intensité et sévérité

TAB. 4.8 – Ajustement de l’intensité de la sécheresse pour un spi de six mois

Lois	paramètre1	error1	paramètre2	error2	ks	p-value
Weibull	forme : 1.0517	0.1455	échelle : 0.5184	0.0968	0.1423	0.5997
exponentielle	rate : 1.9714	0.3661	-	-	0.1597	0.4505
gamma	forme : 1.1579	0.2711	rate : 2.2827	0.6640	0.6352	1.378e-10
log-normale	mean : -1.1691	0.2004	sd : 1.0791	0.1417	0.1931	0.2295

4.8 Application de la copule tri-variée sur les caractéristiques de la sécheresse

Le coefficient de corrélation de Pearson, le coefficient de rho de Spearman et la corrélation de tau de Kendall ont été appliqués aux deux variables de la sécheresse (à savoir la durée, sévérité et l’intensité) et leurs fonctions de distribution cumulative (CDF) associés.

Le tableau IV.9 résume les valeurs de coefficient de corrélation de Pearson, le Tau de

Kendall et rho de Spearman entre toutes les paires possibles des trois caractéristiques de la sécheresse. Les résultats montrent que la corrélation entre la durée et la sévérité à la plus forte valeur du coefficient de corrélation, aussi

TAB. 4.9 – Ajustement de la sévérité de la sécheresse pour spi de six mois

Lois	paramètre1	error1	paramètre2	error2	ks	p-value
Weibull	forme : 1.1591	0.1517	échelle : 0.0787	0.0134	0.1421	0.602
exponentielle	rate :13.4884	2.5047	-	-	0.1675	0.3901
gamma	forme :1.5204	0.3638	rate : 20.5072	5.7986	0.9982	2.2e-16
log-normale	mean : -2.9654	0.1542	sd : 0.8306	0.1091	0.0849	0.9849

bien à l'échelle de 3 mois qu'à l'échelle 6 mois, avec les trois méthodes utilisée (rho de Spearmann, tau de Kendall et coefficient de corrélation Pearson).

TAB. 4.10 – Corrélation des caractéristiques de la sécheresse pour les spi de 03 mois

Corrélation	Pearsen		tau de Kendell		rho de Spearmann	
	cor	p-value	tau	p-value	rho	p-value
Durée-intensité	0.6134	0.0004	0.5388	8.64e-05	0.7013	2.25e-05
Durée-sévérité	0.01	0.9588	-0.0642	0.6472	-0.0556	0.7742
sévérité-intensité	0.6855	4.01e-05	0.6859	4.01e-05	0.5562	0.0017

TAB. 4.11 – Coefficients de dépendance entre les différentes caractéristique des spi de 6 mois

Corrélation	Pearsen		tau de Kendell		rho de Spearmann	
	cor	p-value	tau	p-value	rho	p-value
Durée-intensité	0.3688	0.0489	0.3774	0.0062	0.5276	0.0033
Durée-sévérité	0.0492	0.7998	0.2094	0.1361	0.3095	0.1024
sévérité-intensité	0.7744	8e-07	0.2562	0.0588	0.3290	0.0814

4.8.1 Matrice de variance covariance

Pour le spi de trois mois, on obtient la matrice

$$\text{Si } gma_{3m} \begin{pmatrix} 10.5049 & 0.9845 & 0.0017 \\ 0.9845 & 0.2452 & 0.0176 \\ 0.0017 & 0.0176 & 0.0027 \end{pmatrix}$$

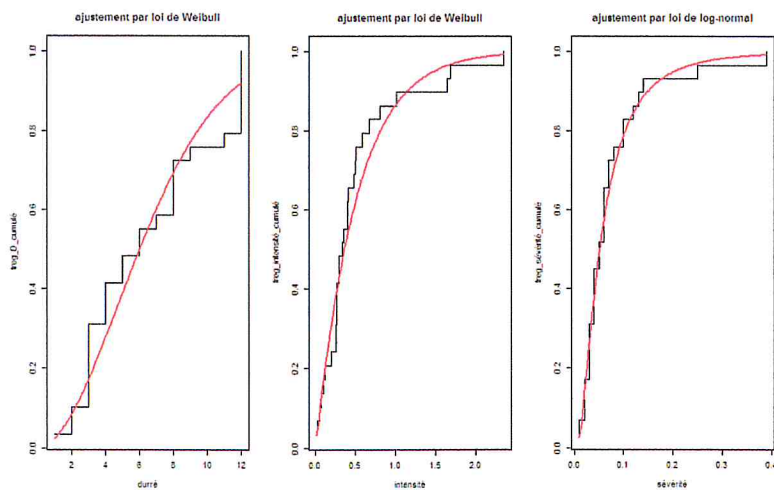


FIG. 4.2 – Ajustement du durée, intensité et sévérité pour spi de six mois

Pour le spi de six mois, on obtient la matrice

$$\text{Si } gma_{6m} = \begin{pmatrix} 13.7586 & 0.9439 & 0.01436 \\ 0.9439 & 0.2915 & 0.0354 \\ 0.01436 & 0.0354 & 0.0062 \end{pmatrix}$$

4.9 Modèles de copule bivariés

Tout d'abord, des modèles de copule bivariés sont développés pour représenter la dépendance conjointe de paires de variables de la sécheresse telles que la durée, l'intensité et la sévérité.

Quatre modèles de copule, à savoir les copules de Clayton, de Gumbel, de Frank, et le copule de Student sont ajustés et leurs résultats correspondants sont présentés dans le tableau 4.12. Les paramètres des fonctions de copule sont estimés à l'aide de l'approche de pseudo-vraisemblance maximale.

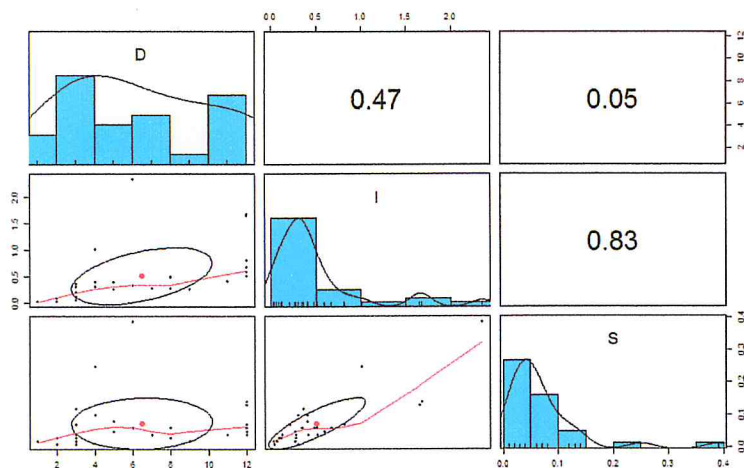


FIG. 4.3 – Graphes de corrélations et ajustements des coefficients

LogLik : est la valeur de maximum de vraisemblance,

AIC : Critère d'information Akaike

$$AIC = 2 \times \log(\text{LogLik}) + 2 \times m,$$

où m est le nombre des paramètres du modèle de copule.

Pour la copule t-Student, comme $\sigma_{12} = \sigma_{21}$, la matrice de corrélation

$$\Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{21} & 1 \end{pmatrix}.$$

Les paramètres des copule sont estimés à l'aide du package `copula` de logiciel statistique R.

Le tableau 4.12 montre que la copule t-Student a entraîné des valeurs de log-Likelihood et des valeurs AIC minimales plus élevées pour deux combinaisons de paires DuréeIntensité, par contre pour le paire Durée-Sévérité est mieux ajusté par la copule de Clayton. Ces modèles bivariés sont utilisés pour calculer les périodes de retour des variables bivariées et caractéristiques de la sécheresse trivariées.

TAB. 4.12 – Résultats d’ajustement par des copules bivaries pour le spi de 03 mois

Liaison	Copule	paramètres	tau	LogLik	AIC
durée-intensité	Clayton	1.68 -	0.46	10.41	-16.82
	Frank	5.65 -	0.5	9.14	-16.28
	Gumbel	2.05 -	0.51	8.90	-15.80
	t-Student	$\nu=0.76$ $\sigma=30$	0.55	10.60	-19.19
durée-sévérité	Clayton	-0.15 -	-0.07	0.48	3.05
	Frank	-0.39 -	-0.04	00	2.0
	Gumbel	-1.09 -	-0.08	0.05	1.90
	t-Student	-0.18 2.07	-0.11	00	2.0
intensité-sévérité	Clayton	1.48	0.42	8.26	-12.53
	Frank	4.51	0.42	5.72	-12.71
	Gumbel	1.88	0.47	7.26	-12.52
	t-Student	0.6 2	0.41	7.46	-12.93

4.10 Modèles de copule trivariés

La forme complètement imbriquée des copules de Clayton, Gumbel et Frank, ainsi que la copule d’une famille elliptique sont choisies pour modéliser les caractéristiques de la sécheresse. Les fonctions de copule sont équipées d’un estimateur de pseudo maximum-vraisemblance .

Les paramètres de copule estimés et les valeurs correspondantes de la fonction de loglikelihood sont présentés dans le tableau 4.13, qui affiche la plus grande valeur log-vraisemblable pour la copule de Clayton. Les performances des familles de copules sont comparées à l’aide de statistiques LogLik, AIK.

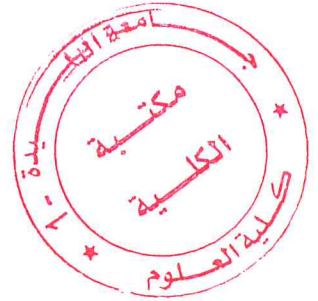
TAB. 4.13 – Résultats d’ajustement par des copules trivariées

Copule	paramètres	tau	LogLik	AIC
Clayton	0.64 -	0.24	3.35	-2.71
Frank	2.08 -	0.22	1.31	-0.62
Gumbel	1.35 -	0.26	1.76	0.48
t-Student	$\nu=0.31$ $\sigma=2$	0.2	1.70	-1.40

4.11 Conclusion

Dans ce travail, on a modélisé la dépendance dans des données échangeables à l’aide de la théorie des copules. Cette modélisation fait intervenir

un paramètre de dépendance pour une copule dont la dimension varie selon la taille de la série statistique. Dans un premier temps, on a réalisé une analyse statistique pour chaque variables qui caractérisent la sécheresse, estimation des paramètres et l'ajustement des données réelles par des densités de probabilités usuelles. Dans la deuxième partie, on a étudié la dépendance statistiques entre les variables de la sécheresse deux à deux, et l'ajustement bivariée des ces variables par des copules bivariés, et après la dépendance entre les trois variables à la fois.



Bibliographie

- [1] Genest, C. Frank's family of bivariate distributions. *Biometrika*, 74, 549-555, (1987).
- [2] Genest, C. et J. MacKay. Bivariate distributions with uniform marginals, The joy of copulas *American Statistician*, 40, 280-283, (1986).
- [3] Genest, C. et Rivest, L. Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association*, 88, 1034-1043, (1993).
- [4] Genest, C., Quessy, J. F. & Rémillard, B. (2006) Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.* 33(2), 337-366.
- [5] Joe, H. Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. Probab. Lett.* 9(1), 75-81, (1990).
- [6] Joe, H. Multivariate extreme-value distributions with applications to environmental data. *Canad. J. Statist.* 22(1), 47-64, (1994).
- [7] Joe, H. (1997) *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability, Chapman & Hall, London, UK.
- [8] Hürlimann, W. (2004) Fitting bivariate cumulative returns with copulas *Comp. Statis. Data Anal.* 45, 355-372.
- [9] McKee, T.B., N.J. Doesken et J. Kleist, 1993 : The relationship of drought frequency and duration to time scale. In : *Proceedings of the Eighth Conference on Applied Climatology*, Anaheim, California, du 17 au 22 janvier 1993. Boston, American Meteorological Society, 179-184.
- [10] Nelson, R. B. : *An Introduction to Copulas*. Springer Series in Statistics, Springer Science+Business Media, Inc., New York, 2nd édition, 2006.
- [11] Palmer, W. C. (1965) : *Meteorological Drought*, Weather Bureau, Research Paper, 45, U.S. Dept. of Commerce, Washington, DC, 58 pp.
- [12] Sklar, A. (1959). *Fonctions de répartition à dimensions et leurs marges*, *Publ. Inst. Statist, Univ. Paris*, 8, 229-231.