

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
UNIVERSITÉ KASDI MERBAH OUARGLA

FACULTÉ DE MATHÉMATIQUES ET SCIENCES DE LA
MATIÈRE

DÉPARTEMENT DE MATHÉMATIQUES



Mémoire présenté En Vue De L'obtention Du

DIPLÔME DE MASTER

EN MATHÉMATIQUES

Option : Probabilité et Statistique

Par

MANSOURI Messaouda

Intitulé

Modélisation de la dépendance par les copules,
application en assurance

Membres du jury

AMARA Abdelkader	Prof.	UKMO	Président
BENBRAHIM Radhia	M. C. B	UKMO	Examineur
ARBIA Hanane	M. C. B	UKMO	Rapporteur

2024-2025

Table des matières

Table des matières	i
Dédicace	iii
Remerciements	iv
Table des figures	v
Notations et abriviations	vi
Introduction	1
1 Théorie des copules	3
1.1 Mesure d'association	3
1.1.1 Fonction de concordancer	3
1.1.2 Mesures de concordance	5
1.1.3 Mesure de dépendance	8
1.1.4 Dépendance de queue	10
1.2 Copules multivariées	10
1.2.1 Définition et théorème d'existence	11
1.2.2 Théorème de Sklar	12
1.2.3 Exemples de copules paramétriques multivariées	13
1.3 Copules bivariées	14
1.3.1 Définition d'une copule	14
1.3.2 Théorème de Sklar	15
1.3.3 Bornes de Fréchet Hoeffding	16
1.3.4 Densité de copule	17
1.4 Familles de copules	17
1.4.1 Copules elliptiques	17
1.4.2 Copule normale (Gaussienne)	18
1.4.3 Copule student	18

1.4.4	Copules Archimédiennes	19
1.4.5	Copule de valeur extrême	20
2	Estimation des paramètres d'une copule	21
2.1	Méthodes d'estimation	21
2.1.1	Méthode du maximum de vraisemblance exacte (MLE) . .	21
2.1.2	Méthode des fonctions d'inférence des marginales	23
2.1.3	Méthode de pseudo-maximum de vraisemblance	24
2.1.4	Méthode de moment	25
3	Modélisation par les copules sur des données réelles	28
3.1	Analyse descriptive des données	28
3.1.1	Coefficient de corrélation	29
3.2	Fonctions de distribution univariées	31
3.3	Fonctions copules	31
	Conclusion générale	32
	Bibliographie	33

Dédicace

J'ai le grand plaisir de dédier ce modeste travail à
mon père, que Dieu ait son âme,
ma chère mère, source inépuisable d'amour et de sagesse,
mon époux bien-aimé, Mourad, pour son soutien indéfectible,
mon fils précieux, Gharib Aladdin, lumière de ma vie,
mes frères et sœurs, piliers de mon existence,
Et à mon amie fidèle, Samira, dont l'amitié m'est si précieuse.
Avec tout mon amour et ma reconnaissance.

Remerciements

Je tiens à remercier "**Allah**" le tout puissant pour toute la volonté et le courage qu'il M'a donné pour l'achèvement de ce travail. Je tiens à exprimer ma sincère gratitude à ma directrice de mémoire, Prof **ARBIA Hanane**, pour son encadrement attentif, ses orientations précieuses et son soutien tout au long de l'élaboration de ce travail. Son expertise en probabilités et statistiques a grandement enrichi cette recherche.

Je remercie également le président du jury, Monsieur **AMARA Abdelkader**, ainsi que Prof **BENBRAHIM Radia**, examinateur du jury.

Je remercie également l'ensemble des enseignants de la Faculté des Sciences, Département de Mathématiques, spécialité Probabilités et Statistiques, pour la qualité de leur enseignement et leur implication dans notre formation. J'adresse particulièrement mes remerciements à Professeurs Mansoul Ibrahim et Benbrahim Radhia, ainsi qu'à Monsieur Saouli Mostefa Abdelouahab, pour leurs apports pédagogiques, leur disponibilité et leurs conseils tout au long de mon parcours.

Mes remerciements vont également à mes collègues et amis pour leur entraide, leur motivation et les échanges constructifs qui ont marqué cette période de recherche.

Enfin, je remercie du fond du cœur ma famille, en particulier mes parents, pour leur soutien inconditionnel, leur patience et leur encouragement constant tout au long de mes études.

Toutes et à tous, merci infiniment.

Table des figures

3.1	Nuage de points de la variable corporel en fonction de la variable matereil.	30
3.2	Nuage de points de la variable matereil en fonction de la variable corporel.	30

Notations et abriviations

v.a	variable aléatoire.
i.i.d	indépendantes identiquement distribuées.
ssi	si seulement si.
i.e.	c'est-à-dire.
$E(X)$	Espérance de X.
$Cov(X, Y)$	Covariance de X et Y.
$V(X)$	variance de X.
C, c	Copule et densité de copule.
X_1, \dots, X_n	Echantillon de taille n de la v.a X.
ϕ	Fonction de la loi normale standard.
C_ρ^{Ga}	Copule Gaussienne .
C_θ^A	Copule Archimédienne.
Π, M, W	Copules prouit, min et max.

$C_{\rho,v}^t$	Copule student.
Θ	Espace des paramètres.
Q	Fonction de concordance.
$\kappa_{X,Y}, \delta_{X,Y}$	Mesures de concordance et de dépendance.
τ, τ_{XY}, τ_c	Tau de Kendall.
ρ, ρ_{XY}, ρ_c	Rho de Spearman.
$\lambda, \lambda_L, \lambda_U$	Dépendance de queue.
H, F_i	Fonction de répartition jointe et marginales.
h, f_i	Densité jointe et marginales.
C_0	Famille de copules.
$\hat{\theta}_n^{MV}$	Estimateur de θ par la méthode de Maximum de vraisemblance.
$\hat{\theta}^{IFM}$	Estimateur de θ par la méthode de IFM.
$\hat{\theta}_n^{PMV}$	Estimateur de θ par la méthode de PMV.
$\hat{\theta}^{CM}$	Estimateur de θ par la méthode de moments.
M_k	Moments d'ordre k de la copule.

α, β	Fonctions strictement croissantes.
$L(\theta)$	Fonction log-vraisemblance de θ .
F^{-1}	Fonction quantile de F .
MLE	Méthode du maximum de vraisemblance exacte

Introduction

L'étude de la dépendance entre deux variables ou plus a suscité un grand intérêt de la part des statisticiens. Un riche ensemble de mesures de dépendance entre les v.a. a été proposé, comme le coefficient de corrélation de Pearson, le tau de Kendall, le rho de Spearman, etc. Bien que ces mesures soient simples à calculer et puissent être facilement interprétées, elles ne sont pas en mesure de détecter toutes les formes de dépendance. La fonction copule, qui représentent un outil innovant pour modéliser la structure de dépendance de plusieurs variables aléatoires.

Le terme copule (copula) vient du mot latin "copulae", qui signifie au sens figuré, liaison, lien, alliance ou union. Les copules constituent un sujet de recherche relativement moderne étant en plein essor depuis principalement les trois dernières décennies. Elles sont devenue en quelques années un outil important pour modéliser les risques multivariées (entre autres) surtout dans les domaines de la finance et l'assurance. Parmi les travaux les plus importants de copules en statistique, on peut citer ceux de Hoeffding, en 1940 qui a utilisé les copules pour étudier les mesures d'associations non paramétriques comme le rho de Spearman. En 1959 Sklar a proposé son fameux théorème concernant l'existence d'une relation entre la loi jointe et ses marginales. En assurance Tibiliti (1996) a introduit les copules pour modéliser la variations de la demande d'assurance.

Dans notre étude, nous avons tenté de montrer la dépendance entre les deux sous branches de l'assurance automobile, dommages matériels et corporels, à travers l'application de la théorie des copules.

Ce travail est constitué de trois chapitres :

Chapitre 1 : Théorie des copules

Dans ce chapitre, nous présentons l'aspect des mesures d'association, la mesure de concordance, la mesure de dépendance en forme de copule, telle que le rho de Spearman et le tau de Kendall. Nous introduisons les définitions de base des copules, leurs propriétés, ainsi que les principaux théorèmes liés à cette théorie, en particulier le théorème de Sklar. Nous nous intéressons dans un premier temps à l'étude du cas multivarié, puis nous passons au cas bivarié. De même quelque exemples de familles paramétriques de copules les plus utilisés.

Chapitre 2 : Estimation de copules.

Ce chapitre est consacré aux méthodes d'estimation, où nous avons regroupé quelques méthodes d'estimations de copules paramétriques, telle que la méthode

du maximum de vraisemblance exacte, la méthode des fonction d'inférence des marginales, la méthode de pseudo-maximum de vraisemblance et la méthode de moment.

Chapitre 3 : Modélisation par les copules sur des données réelles

Dans ce chapitre, nous présentons une application des copules en assurance. Nous recherchons la copule qui modélise la dépendance entre deux sous-branches de l'assurance automobile : les dommages matériels et les dommages corporels dans la période de 2000 à 2024. L'implémentation sera réalisée à l'aide du logiciel statistique R.

Chapitre 1

Théorie des copules

Ce chapitre est consacré à la théorie des copules, nous définirons tout d'abord le mesure d'association ainsi que les copules multivariées. La section 3 traite le cas bivariées qui est considéré comme un cas particulier du copule multivarié. Ensuite, nous présentons dans les sections 4 et 5 la définition de la densité de couple et les familles de couples.

1.1 Mesure d'association

1.1.1 Fonction de concordance

Une mesure de dépendance permet de donner une idée sur la structure de dépendance entre deux variable, ceci est exprimé à l'aide d'un seul nombre.

Corrélation linéaire de Pearson

Soient X et Y deux v.a. La covariance de X et Y est le nombre

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y).$$

Soient X et Y deux v.a ayant des variances finies $V(X)$ et $V(Y)$ respectivement. Le coefficient de corrélation entre X et Y se calcule par la formule suivante :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}; \rho \in [-1, 1]$$

où $V(Z) = E(Z - E(Z))^2$ pour $Z = X$ et Y .

Remarque 1.1.1 *La dépendance et la corrélation sont des notions différentes. En effet, si X et Y sont des variables indépendantes elles sont non corrélées mais*

la réciproque est fautive sauf dans le cas où les variables sont gaussiennes car la dépendance est entièrement caractérisée par le coefficient de corrélation.

On peut donner le contre-exemple suivant :

Soient $X \rightsquigarrow N(0, 1)$ et $Y = X^2$. Les deux variables sont non-corrélées $Cov(X, Y) = E(X^3) = 0$ mais présentent un lien de dépendance évident.

Concept de concordance

Soit $\{(x_1, y_1), \dots, (x_n, y_n)\}$ un échantillon de n observations d'un couple (X, Y) . Il existe $C_n^2 = \frac{n!}{2!(n-2)!}$ paires de distributions distinctes de couples (x_i, y_i) et (x_j, y_j) qui sont dites concordantes ou discordantes si :

Concordantes :

$$(x_i - x_j)(y_i - y_j) > 0 \text{ i.e. } (x_i < x_j \text{ et } y_i < y_j) \text{ ou } (x_i > x_j \text{ et } y_i > y_j).$$

Discordantes :

$$(x_i - x_j)(y_i - y_j) < 0 \text{ i.e. } (x_i < x_j \text{ et } y_i > y_j) \text{ ou } (x_i > x_j \text{ et } y_i < y_j).$$

Définition 1.1.1 (*Fonction de concordance*) La fonction de concordance est la différence entre la probabilité de concordance et celle de discordance entre deux couples (X_1, Y_1) et (X_2, Y_2) sous la forme suivant :

$$Q = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]. \quad (1.1)$$

Comme les variables sont continues, donc

$$\begin{aligned} Q &= P[(X_1 - X_2)(Y_1 - Y_2) > 0] - (1 - P[(X_1 - X_2)(Y_1 - Y_2) > 0]) \\ &= 2P[(X_1 - X_2)(Y_1 - Y_2) > 0] - 1. \end{aligned}$$

Avec :

$$P[(X_1 - X_2)(Y_1 - Y_2) > 0] = P[X_1 > X_2, Y_1 > Y_2] + P[X_1 < X_2, Y_1 < Y_2].$$

Propriétés de la fonction de concordance

Le résultat suivant permet d'établir une relation entre la fonction de concordance Q et les copules associées aux deux couples.

Théorème 1.1.1 *Soient (X_1, Y_1) et (X_2, Y_2) deux couples de vas indépendantes de distributions jointes G_1 et G_2 avec des marges communes F et H respectivement (i.e. F pour X_1 et X_2 , H pour Y_1 et Y_2). Soient C_1 et C_2 les copules associées aux distributions G_1 et G_2 respectivement. On a*

$$Q = Q(C_1, C_2) = 4 \iint_{I^2} C_2(u, v) dC_1(u, v) - 1. \quad (1.2)$$

On peut montrer que les deux formules (1.1) et (1.2) sont équivalentes.

1.1.2 Mesures de concordance

Plus classiquement, les mesures d'association couramment utilisées d'après D'Agostino et Stephens [07], Joe [12] et Genest et Favre [09], pour la caractérisation de la dépendance sont : le coefficient de corrélation linéaire ρ de Pearson ; le coefficient de corrélation des rangs ρ de Spearman et le coefficient de corrélation des rangs τ de Kendall. Le coefficient de corrélation de Pearson mesure la dépendance linéaire entre deux vas. Pour l'étude d'une dépendance non linéaire, les deux autres coefficients sont plus adaptés. puisqu'ils ne dépendent que du rang de chaque observation par opposition au coefficient de corrélation de Pearson qui utilise les réalisations des variables. En outre, ρ et τ sont invariants par transformation strictement croissante des vas X et Y . Une mesure d'association entre deux vas X et Y de copule C est une mesure de concordance si elle vérifie les propriétés suivantes :

- κ est définie pour tout couple (X, Y) de va continues ;
- $-1 \leq \kappa_{X, -X} \leq \kappa_{X, Y} \leq \kappa_{X, X} = 1$;
- Si Y est presque sûrement une fonction croissante de X , alors $\kappa_{X, Y} = \kappa_M = 1$;
- Si Y est presque sûrement une fonction décroissante de X , alors $\kappa_{X, Y} = \kappa_W = -1$;
- Si X et Y sont indépendantes, alors $\kappa_{X, Y} = 0$;
- $\kappa_{X, Y} = \kappa_{Y, X}$;
- $\kappa_{-X, Y} = \kappa_{X, -Y} = -\kappa_{X, Y}$;
- Si α et β sont des fonctions strictement croissantes, alors $\kappa_{\alpha(X), \beta(Y)} = \kappa_{Y, X}$;
- Si C_1 et C_2 sont deux copules telles que $C_1 < C_2$ alors on a : $\kappa_{C_1} < \kappa_{C_2}$;
- Si $\{(X_n, Y_n)\}_{n \geq 1}$ est une suite de vas dont la copule est C_n , et si $\{C_n\}_{n \geq 1}$ converge vers C lorsque n tend vers l'infini alors κ_{C_n} converge vers κ_C .

Tau de Kendall

Le tau de Kendall et le rho de Spearman sont deux mesures de concordance bien connues en statistique. Elles offrent l'avantage de s'exprimer simplement en fonction de la copule associé au couple de vas. On peut exprimer le tau de Kendall de différentes manières, soit en fonction des observations, de la fonction de concordance et même en fonction de la copule associée. Le tau de Kendall de la série exprimé en fonction des observations s'écrit sous la forme suivante :

Définition 1.1.2 (*Version échantillon*) Soit une série de n observations $\{(x_i, y_i)\}_{1 \leq i \leq n}$ d'un couple (X, Y) . Le tau de Kendall est défini par

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

où

n_c : nombre de paires concordantes.

n_d : nombre de paires discordantes.

n : nombre total de paires.

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

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

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

Définition 1.1.3 (*Version population*) Soient (X_1, Y_1) et (X_2, Y_2) deux couples de vecteurs continus iid de fonction de distribution jointe H . Le tau de Kendall est donné par

$$\tau_{XY} = \mathbf{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbf{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Son expression en terme de la copule associée est la suivante :

Définition 1.1.4 Soit (X, Y) un couple de v.a continues de copule C . Si les couples sont identiquement distribués i.e. $H_1 = H_2 = H$, le tau de Kendall a pour expression

$$\tau_c = Q(C, C) = 4 \int \int_{I^2} C(u, v) dC(u, v) - 1$$

Comme les couples sont uniformément distribués sur \mathbf{I} , alors

$$\iint_{\mathbf{I}^2} C(u, v) dC(u, v) = \mathbf{E}[C(u, v)].$$

Ainsi, on peut écrire :

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

Remarque 1.1.2 *Le tau de Kendall de X et Y d'une copule Archimédienne s'obtient simplement à l'aide du générateur de la copule, selon la formule ci-dessous*

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

Rho de Spearman

Le coefficient de corrélation rho de Spearman représente l'une des mesures les plus connues pour quantifier le degré d'association entre deux vas. La valeur de ce coefficient dénotée par ρ est équivalente au coefficient de corrélation de Pearson. Il a été développé par Spearman [21]. Le rho de Spearman de deux vas X et Y est égal au coefficient de corrélation entre les variables $F(X)$ et $G(Y)$ distribuées selon la loi uniforme sur l'intervalle \mathbf{I} tel que $\rho(X, Y) = \rho(F(X), G(Y))$.

Définition 1.1.5 *(Version échantillon) Soit une série de n observations $\{(x_i, y_i)\}_{1 \leq i \leq n}$ d'un couple (X, Y) . On définit le rho de Spearman par*

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

où $D_i = R_{X_i} - R_{Y_i}$, R_{X_i} et R_{Y_i} les rangs des observations de X et Y respectivement.

Définition 1.1.6 *(Version population) Soient (X_1, Y_1) , (X_2, Y_2) et (X_3, Y_3) trois vecteurs aléatoires indépendants de même loi H . La version population du rho de Spearman est définie comme étant proportionnelle à la différence de la probabilité de concordance et celle de discordance des couples aléatoires (X_1, Y_1) , (X_2, Y_3) . Elle s'écrit comme suit*

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

La distribution de (X_2, Y_3) étant \mathbf{II} (car les variables X_2 et Y_3 sont indépendantes) alors d'après ce qui précède on a le théorème suivant.

Théorème 1.1.2 (*Expression de rho de Spearman en terme de copule*)

$$\rho_c = 3Q(C, \mathbf{II}) = 3 \times \left(4 \iint_{I^2} C(u, v) d\mathbf{II}(u, v) - 1 \right) = 12 \iint_{I^2} C(u, v) d(uv) - 3.$$

Alors

$$\rho_c = 12 \iint_{I^2} C(u, v) dudv - 3; \rho \in [-1, 1].$$

1.1.3 Mesure de dépendance

Définition 1.1.7 *Une mesure numérique d'association δ entre deux variables aléatoires continues X, Y dont la copule est C est une mesure de dépendance si et seulement si elle satisfait les propriétés qui suivent (on la note $\delta_{X,Y}$) :*

1. $\delta_{X,Y}$ est définie pour chaque couple (X, Y) de variables aléatoires continues.
2. $0 \leq \delta_{X,Y} \leq 1$.
3. $\delta_{X,Y} = \delta_{Y,X}$.
4. $\delta_{X,Y} = 0$ si et seulement si X et Y sont indépendantes.
5. $\delta_{X,Y} = 1$ si et seulement si chacun de X, Y est une fonction strictement monotone de l'autre presque sûrement.
6. Si $\alpha(X)$ et $\beta(Y)$ sont des fonctions strictement monotones presque sûrement, alors

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

7. Si (X_n, Y_n) est une suite de variables aléatoires continues de copule C_n et C_n converge vers C , alors

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

Exemple de mesure de dépendance

Mesure de Schweizer et Wolffs : le rho de Spearman de deux v.a continue X, Y est définie par :

$$\rho_C = 12 \iint_{I^2} (C(u, v) - uv) dudv,$$

cet intégral représente la volume entre la copule C est la copule produit \mathbf{II} . Si on change la différence $(C(u, v) - uv)$ par $|C(u, v) - uv|$, alors on obtient une

mesure basée sur la distance L_1 entre le graphique de C et \mathbf{II} , cette distance représente la mesure σ de **Schweizer et Wolffs**, elle est définie par :

$$\sigma_c = \sigma_{X,Y} = 12 \iint_{I^2} |C(u, v) - uv| dudv. \quad (1.3)$$

Théorème 1.1.3 *Soit X, Y deux v.a continues d'une copule C . Alors la quantité σ_c définie dans (1.3) est une mesure de dépendance. Schweizer et Wolffs (1981) assure que toute distances entre les surfaces $z = C(u, v)$ et $z = uv$ représentent une mesure non paramétrique de dépendance [17], [23]. $\forall 1 \leq p < \infty$, la distance L_p entre C et \mathbf{II} définie par :*

$$L_p = \left(K_p \iint_{I^2} |C(u, v) - uv|^p dudv \right)^{\frac{1}{p}}. \quad (1.4)$$

telle que k_p est une constante.

A partir la quantité dans (1.4) L_p on définit les mesures de dépendances suivantes :

Mesure $\phi_{X,Y}$: si $p = 2$, alors nous avons

$$\phi_{X,Y} = \phi_c = \left(90 \iint_{I^2} |C(u, v) - uv|^2 dudv \right)^{\frac{1}{2}}.$$

Mesure $\Lambda_{X,Y}$: pour $p = \infty$, nous avons

$$\Lambda_{X,Y} = \Lambda_c = 4 \sup_{u,v \in I} |C(u, v) - uv|.$$

Remarque 1.1.3 *Après les définitions de mesures de dépendance σ_c , ϕ_c et Λ_c on trouve que ces dernières sont basées sur le coefficient de Rho de Spearman. Il existe d'autres mesures de dépendance basées sur un autre coefficient telle que le coefficient de Gini (voir[05]).*

1.1.4 Dépendance de queue

Le concept de dépendance de queue fournit une description de la dépendance au niveau des queues de distribution.

La dépendance de queue est une mesure locale contrairement au tau de Kendall et au rho de Spearman qui mesurent la dépendance sur l'ensemble de la distribution.

Définition 1.1.8 *Le coefficient de dépendance de queue inférieure "lower tail dependance coefficient" entre deux variables aléatoires X et Y , de fonctions de répartition respectives F_X et F_Y , est défini par la limite λ_L si elle existe :*

$$\lambda_L(X, Y) = \lim_{u \rightarrow 0^+} P(X < F_X^{-1}(u) | Y < F_Y^{-1}(u))$$

Le coefficient de dépendance de queue supérieure "upper tail dependance coefficient" entre deux variables aléatoires X et Y , de fonctions de répartition respectives F_X et F_Y , est défini par la limite λ_U si elle existe :

$$\lambda_U(X, Y) = \lim_{u \rightarrow 1^-} P(X > F_X^{-1}(u) | Y > F_Y^{-1}(u))$$

Si $\lambda \in]0, 1]$ on dit que X et Y sont asymptotiquement dépendantes au niveau supérieur de la queue de distribution

Si $\lambda_U = 0$ dit que X et Y sont asymptotiquement indépendantes au niveau supérieur de la queue de distribution.

Si $\lambda_L = 0$ alors, il n'ya pas de dépendance de queue au niveau inférieure de la distribution.

On peut exprimer λ_L et λ_U à l'aide de la copule C de couple (X, Y) tel que décrit dans la proposition suivante.

Proposition 1.1.1 *Etant donné deux variables aléatoires X et Y de copule $C(X, Y)$, on a :*

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

et

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

1.2 Copules multivariées

Cette section étend les résultats de la section précédente. Donc de la même manière, nous présentons les copules multivariées : définitions, théorème, propriétés et nous concluons cette section par quelques exemples de copules multivariées [08] et [18]. Nous notons $\mathbf{I}^d = [0, 1]^d$.

1.2.1 Définition et théorème d'existence

Soit le vecteur $u = (u_1, \dots, u_d) \in \mathbf{I}^d$. Soient A_1, \dots, A_d sous ensembles non vide de R^d et G une fonction définie sur $A_1 \times \dots \times A_d \rightarrow \mathbf{R}$.

Définition 1.2.1 Soit a_i les plus petits éléments de $A_i, i = 1, \dots, n$. La fonction G est dite attachée si elle est nulle pour tout $v \in A_1 \times \dots \times A_d$ pour au moins un indice k telle que $v_k = a_k$

$$G(v_1, v_2, \dots, v_{k-1}, a_k, a_{k+1}, \dots, a_n) = 0.$$

Définition 1.2.2 Soient S_1, \dots, S_d des parties mesurables non vides de $\bar{\mathbf{R}}$. Soit $B = [a, b]$ un d -pavé dont les sommets sont dans $\text{Dom}G$. Le G volume de B est alors défini par

$$V_G(B) = \sum \text{sgn}(c)G(c).$$

où la somme s'effectue sur tous les sommets c de B et le $\text{sgn}(c)$ est donné par

$$\text{sgn}(c) = \begin{cases} 1 & \text{si } c_k = a_k \text{ pour un nombre pair de } k, \\ -1 & \text{si } c_k = a_k \text{ pour un nombre impair de } k \end{cases}$$

Définition 1.2.3 (la copule)

La fonction G est dite décroissante si $V_G(B) \geq 0$ pour tout B dont les sommets sont dans $\text{Dom}G$. Une copule d -dimensionnelle est une fonction C de \mathbf{I}^d dans \mathbf{I} ayant les propriétés suivantes :

1. $\forall u \in \mathbf{I}^d$

$$C(u) = 0 \text{ pour au moins une coordonné de } u = 0.$$

2. pour toutes les coordonnés égale 1 sauf u_i , alors

$$C(u) = u_i$$

3. $\forall u, v \in \mathbf{I}^d$ telle que $u \leq v$

$$V_C([u, v]) \geq 0.$$

1.2.2 Théorème de Sklar

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

$$H(x) = H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (1.5)$$

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

Inversement, si C est une copule et si F_1, \dots, F_d sont des fonctions de répartition univariées, alors la fonction H définie dans (1.5) est une fonction de répartition d-dimensionnelle de marginales F_1, \dots, F_d . Pour la preuve de ce théorème voir [16].

Corollaire 1.2.1 *Soit H une fonction de répartition d-dimensionnelle de fonctions de répartition marginales F_1, \dots, F_d et soient $F_1^{-1}, \dots, F_d^{-1}$ les inverses généralisés F_1, \dots, F_d respectivement. Alors, la copule C associée à H est donnée par :*

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

Propriétés de copules multivariées

Théorème 1.2.1 (Continuité uniforme)

Une copule C est uniformément continue sur son domaine. En particulier, pour tout u, v dans \mathbf{I}^d , nous avons :

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

Théorème 1.2.2 (Invariance)

Soient (X_1, \dots, X_d) un vecteur de variables aléatoires continues, de fonction de répartition F associée à une copule C et $(\alpha_1, \alpha_2, \dots, \alpha_d)$ une suite de fonctions strictement croissantes. Alors, la fonction de répartition jointe du vecteur aléatoire $(\alpha_1(X_1), \dots, \alpha_d(X_d))$ est aussi associée à la même copule C [10].

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

Théorème 1.2.3 (Dérivées partielles)

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

$$0 \leq \frac{\partial C(u)}{\partial u_i} \leq 1,$$

de plus, les fonctions $u \rightarrow \frac{\partial C(u)}{\partial u_i}$ sont non décroissantes.

Densité d'une copule multivariée

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

$$\begin{aligned} c(F_1(x_1), \dots, F_d(x_d)) &= \frac{\partial^d C(F_1(x_1), \dots, F_d(x_d))}{\partial F_1(x_1), \dots, \partial F_d(x_d)} \\ &= \frac{h(x_1, \dots, x_d)}{f_1(x_1), \dots, f_d(x_d)} \\ &= \frac{h(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{\prod_{i=1}^d f_i(F_i^{-1}(u_i))} \end{aligned} \quad (1.6)$$

telle que h est la densité de H et f_i est la densité de F_i et $x_i = F_i^{-1}(u_i)$.

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

$$h(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i) \quad (1.7)$$

1.2.3 Exemples de copules paramétriques multivariées

1. Copule d'indépendance :

Les variables aléatoires X_1, \dots, X_d sont indépendantes si et seulement si

$$F(x_1, \dots, x_d) = F_1(x_1) \dots F_d(x_d).$$

Nous définissons donc la copule d'indépendance multivariée par

$$\prod^d(u) = u_1 u_2 \dots u_d.$$

telle que $u_i = F_i(x_i)$ pour $i = 1, \dots, d$.

2. Copule Gaussienne :

Soit X_1, \dots, X_d des v.a Gaussiennes, de moyenne μ , de matrice de covariance Σ et de matrice de corrélation R . Soit Φ une distribution normale standard multivariées. Alors la copule Gaussienne C^{Ga} est définie par :

$$C^{Ga}(u) = \Phi(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)).$$

où Φ^{-1} est l'inverse de la fonction de répartition standard normale univarié : [08]et [14].

D'après (1.7) nous avons :

$$h(x_1, \dots, x_d) = c(\Phi(x_1), \dots, \Phi(x_d)) \prod_{i=1}^d \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \right)$$

et

$$h(x_1, \dots, x_d) = \frac{1}{(2\pi)^{\frac{d}{2}} |R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}x'R^{-1}x\right)$$

alors la densité de la couple Gaussiennes est donne par

$$\begin{aligned} c(\Phi(x_1), \dots, \Phi(x_d)) &= \frac{h(x_1, \dots, x_d)}{\prod_{i=1}^d \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \right)} \\ &= \frac{\frac{1}{(2\pi)^{\frac{d}{2}} |R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}x'R^{-1}x\right)}{\prod_{i=1}^d \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \right)} \end{aligned}$$

3. Copule Archimédienne :

Les copules Archimédiennes multivariées sont définies par un générateur $\varphi(t)$ (voir [04] et [17]). La forme générale de cette famille de copules est présenté comme suit :

$$C^A(u) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_d)),$$

et dont la densité est :

$$C^A(u) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_d)) \dot{\varphi}(u_1) \dot{\varphi}(u_2) \dots \dot{\varphi}(u_d)$$

1.3 Copules bivariées

1.3.1 Définition d'une copule

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

Soient U et V deux variables aléatoires uniformes sur I dont la fonction de répartition :

$$P(U \leq u) = \begin{cases} 0 & \text{si } u < 0, \\ u & \text{si } 0 \leq u < 1 \\ 1 & \text{si } u \geq 1 \end{cases}$$

Alors $\forall (u, v) \in I^2$ on a

$$C(u, v) = P(U \leq u, V \leq v)$$

cette définition assure donc que la copule est une fonction de répartition avec des marges uniformes.

Définition 1.3.1 *La copule bivariée C est une fonction de $\mathbf{I}^2 = [0, 1]^2$ dans \mathbf{I} définie par les caractéristiques suivantes :*

i) La copule C est attachée (grounded), c'est à dire

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

ii) Les marges sont uniformes, c'est à dire

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

iii) La copule C est 2-croissantes : $\forall u_1, v_1, u_2, v_2 \in \mathbf{I}$ telle que $u_1 \leq u_2$ et $v_1 \leq v_2$

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

1.3.2 Théorème de Sklar

Ce théorème dans le cas bivarié est donné par :

Théorème 1.3.1 *Soient X, Y deux v.a de fonction de répartition jointe H et des marginales F et G . Alors il existe une copule C telle que pour tout $(x, y) \in \mathbb{R}^2$.*

$$H(x, y) = C(F(x), G(y)). \quad (1.8)$$

Si F et G sont continues, alors C est unique.

Inversement, si C une copule et F et G sont des fonctions de répartition, alors la fonction H définie dans (1.8) est une fonction de répartition jointe dont les marginales sont F et G [09], [21].

Définition 1.3.2 *Soit F une fonction de répartition univarié. Le quantile de F (l'inverse généralisé) est une fonction, notée F^{-1} de domaine de définition \mathbf{I} . telle que.*

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

Proposition 1.3.1 Soit X une v.a de fonction de répartition F , alors

1. Si U est uniforme dans $[0; 1]$ alors

$$F^{-1}(U) \underset{d}{\sim} F.$$

2. Si F est continue, alors

$$F(X) \underset{d}{\sim} U_{[0,1]}$$

Définition 1.3.3 Une copule C bivariable définie sur I^2 est une fonction de répartition jointe et dont les marginales F et G sont uniformes sur I .

Corollaire 1.3.1 (Inverse de théorème de Sklar) Soit H une fonction de répartition 2-dimensionnelle de fonctions de répartition marginales F et G . Alors la copule C associée à H est donnée par :

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)) \text{ pour tout } (u, v) \in I^2.$$

1.3.3 Bornes de Fréchet Hoeffding

Toute copule bivariable C sont bornées par deux copules. Ces bornes sont déterminées dans le théorèmes suivant :

Théorème 1.3.2 Soit C une copule, alors

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

W et M représentent les bornes inférieure et supérieure de Fréchet-Hoeffding (respectivement).

Comme une conséquence de théorème de Sklar, si X et Y sont des variables aléatoires dont la fonction de répartition jointe est H et des marginales F et G , respectivement, alors pour tout $x, y \in \mathbb{R}$

$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) \leq \min(F(x), G(y)).$$

Car M et W sont des copules, les bornes ci-dessus sont des fonctions de répartition jointes et sont appelés les bornes de Fréchet-Hoeffding pour une fonction de répartition jointe H et des marginales F et G .

1.3.4 Densité de copule

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

Définition 1.3.4 La densité $c(F(x), G(y))$ associée à la copule $C(F(x), G(y))$ est définie par :

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

D'après le théorème de Sklar, on donne la représentation canonique suivante :

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

1.4 Familles de copules

Il existe nombreuses familles de copules qui diffèrent dans le détail de la dépendance qu'elles représentent. Une famille en général a plusieurs paramètres qui se rapportent à la force et la forme de la dépendance. Certaines familles de copules sont décrites ci-dessous.

1.4.1 Copules elliptiques

Les copules elliptiques sont définies à partir des familles lois elliptiques. En cas particulier on a la copule gaussienne et la copule de student.

Définition 1.4.1 On appelle copule elliptique toute copule qui s'écrit de la forme suivante :

$$\begin{aligned} C_\rho(u, v) &= F_\rho(\Phi_{g,1}^{-1}(u), \Phi_{g,2}^{-1}(v)) \\ &= \frac{1}{\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi_{g,1}^{-1}(u)} \int_{-\infty}^{\Phi_{g,2}^{-1}(v)} g\left(\frac{x^2 - 2\rho xy + y^2}{\sqrt{1-\rho^2}}\right) dx dy \end{aligned}$$

où F_ρ est la distribution jointe des v.a X et Y , $\Phi_{g,1}^{-1}(u)$ et $\Phi_{g,2}^{-1}(v)$ sont leurs fonctions quantiles respectives et leurs coefficients de corrélation.

1.4.2 Copule normale (Gaussienne)

La copule gaussienne bivariée est donnée par :

$$C_{\rho}^{Ga}(u, v) = \Psi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v))$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left[-\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}\right] dx dy$$

telle que $\rho \in]-1, 1]$ et $\rho = \frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$

Ψ_{ρ} : distribution normale bivariée centrée de matrice de covariance $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$

Φ : fonction de répartition de loi normale $N(0, 1)$:

$$\Phi(x) = P(X_i \leq x) = \frac{1}{2\pi} \int_{-\infty}^x \exp\left[-\frac{t^2}{2}\right] dt.$$

- Si $\rho = 0$ alors $C_{\rho}(u, v) = \Pi(u, v)$.
- Si $\rho = -1$ alors $C_{\rho}(u, v) = W(u, v)$.
- Si $\rho = 1$ alors $C_{\rho}(u, v) = M(u, v)$.

1.4.3 Copule student

Comparativement à la copule gaussienne, la copule de Student permet grâce à son degré de liberté, de mieux tenir compte des queues de distribution épaisses. En outre, lorsque le degré de liberté tend vers l'infini, la copule de Student est égale à la copule Gaussienne. La copule de Student est définie de la façon suivante :

$$C_{\rho,v}^t(u, v) = t_{\rho,v}(t_v^{-1}(u), t_v^{-1}(v))$$

$$= \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 + t^2 - 2\rho st}{v(1-\rho^2)}\right)^{-\frac{v+1}{2}} ds dt.$$

avec une matrice diagonale définie positive, $\text{diag } \rho = 1$ et t_v^{-1} est la fonction inverse de la distribution standard de Student à degrés de liberté où $v > 0$.

Remarque 1.4.1 *Les copules elliptiques sont des copules symétriques et relativement simples d'utilisation du fait que l'on connaît bien les distributions auxquelles elles sont associées. Elles sont souvent appelées copules implicites car n'ayant pas de forme analytique explicite et s'expriment par conséquent en fonction de leurs distributions bivariées.*

1.4.4 Copules Archimédiennes

Les copules archimédiennes offrent une très grande flexibilité au niveau de la structure de dépendance. Parmi les copules appartenant à cette famille notons les copules de Clayton, de Gumbel et de Frank. Elles sont construites à l'aide d'une fonction φ qui est appelée la fonction génératrice de la copule.

Définition 1.4.2 Une fonction $\varphi : [0; 1] \rightarrow \mathbb{R}^{*+}$ est dite génératrice si elle est continue, décroissante et convexe telle que $\varphi(1) = 0$. Le pseudo-inverse de φ est définie par :

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

Remarquons que si $\varphi(0) = \infty$, alors φ est strictement décroissante et au moins deux fois continuellement dérivable telle que :

$$\varphi'(u) < 0 \text{ et } \varphi''(u) > 0, \forall 0 \leq u \leq 1$$

On peut remplacer le pseudo-inverse $\varphi^{[-1]}$ par l'inverse ordinaire.

Définition 1.4.3 Une copule est dite Archimédienne si elle s'écrit sous la forme :

$$C_{\theta}^A(u, v) = \varphi_{\theta}^{-1}(\varphi_{\theta}(u) + \varphi_{\theta}(v))$$

Avec φ est un générateur.

Exemples de copule Archimédienne

Le tableau suivant présente quelques familles Archimédiennes classiques :

copule	$\varphi_{\theta}(t)$	$C_{\theta}(u, v)$	λ_L	λ_U	τ
Frank	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$-\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$; $\theta \in \mathbb{R} \setminus \{0\}$	0	0	$1 - \frac{4}{\theta} + \frac{4}{\theta^2} \int_0^{\theta} \frac{t}{e^t-1} dt$
Clayton	$(t^{-\theta} - 1)/\theta$	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$; $\theta \in [-1, \infty[\setminus \{0\}$	$2^{-\frac{1}{\theta}}$	0	$\frac{\theta}{\theta+2}$
Gumbel	$(-\ln t)^{\theta}$	$\exp\left\{-[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta}\right\}$; $\theta \geq 1$	0	$2 - 2\frac{1}{\theta}$	$1 - 1/\theta$

TAB. 1.1 – Générateurs des principales copules Archimédiennes et leurs dépendance du queue et tau de Kendall.

1.4.5 Copule de valeur extrême

Comme leur nom l'indique, les copules de valeurs extrêmes sont issues de la théorie de valeurs extrêmes. Dans cette dernière, on considère les variables aléatoires de maximum suivantes :

$$M_{(n,X)} = \max_{1 \leq i \leq n} (X_i) \text{ et } M_{(n,Y)} = \max_{1 \leq j \leq n} (Y_j).$$

On étudie alors la distribution de probabilité de la variable :

$$M_n = (M_{(n,X)}, M_{(n,Y)}).$$

et les cas pour lesquels cette variable, qu'on a renormalisée, ne suit pas une loi dégénérée.

Définition 1.4.4 *Soit n une constante réelle positive. Une copule de valeurs extrême est une copule qui vérifie la relation suivante :*

$$C(u^n, v^n) = C^n(u, v), \forall u, v \in I^2$$

ou :

$$C^{\frac{1}{n}}(u^n, v^n) = C(u, v), \forall u, v \in I^2, \forall n > 0.$$

Chapitre 2

Estimation des paramètres d'une copule

L'estimation en statistiques est une opération qui permet de calculer la valeur inconnue d'un paramètre du modèle statistique à partir d'un échantillon observé (X_1, \dots, X_n) . Il existe plusieurs méthodes d'estimation paramétriques telle que la méthode du maximum de vraisemblance (méthode classique), la méthode des moments...etc, et non paramétrique tel que la copule empirique et comme les copules sont des fonctions de répartition paramétrique comme on a vu dans le premier chapitre, alors on a besoin d'estimer ces paramètres.

2.1 Méthodes d'estimation

Dans cette section, on va expliquer les méthodes d'estimation telle que la méthode de maximum de vraisemblance exacte, méthode des fonctions d'inférence des marginales qui sont des méthodes paramétriques, et deux méthodes sont semi-paramétriques, telle que la méthode de pseudo-maximum de vraisemblance et la méthode de moment, on termine par la méthode de la distance minimale.

2.1.1 Méthode du maximum de vraisemblance exacte (MLE)

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

$$H_0 : C \in C_0, \tag{2.1}$$

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

$$\dot{H}_0 = F_1 \in \mathcal{F}_1, \dots, F_d \in \mathcal{F}_d$$

telle que F_j pour $j = 1, \dots, d$ sont les lois marginales de X_1, \dots, X_d et $\mathcal{F}_j = \{\mathcal{F}_{\gamma_j} : \gamma_j \in \Gamma_j\}$, où Γ_j est un sous ensemble de \mathbb{R} et D'après le théorème de Sklar la fonction de répartition de X est donnée par :

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

et sa densité par

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

en maximisant la fonction log-vraisemblance $L(\theta)$ définie par

$$L(\theta) = \sum_{i=1}^n \ln h(x_1, \dots, x_d) \quad (2.2)$$

telle que $h(x_1, \dots, x_d)$ est la fonction de densité jointe de x_1, \dots, x_d . Comme on peut voir la fonction $L(\theta)$ peut s'écrire

$$\begin{aligned} L(\theta) &= \sum_{i=1}^n \ln \left(c_\theta(F_1(x_1^i), \dots, F_d(x_d^i)) \prod_{j=1}^d f_j(x_j^i) \right) \\ &= \sum_{i=1}^n \ln c_\theta(F_1(x_1^i), \dots, F_d(x_d^i)) + \sum_{i=1}^n \sum_{j=1}^d \ln f_j(x_j^i), \end{aligned}$$

où c_θ représente la densité de la copule C_θ alors l'estimateur de θ , noté $\hat{\theta}_n^{MV}$ est

$$\hat{\theta}_n^{MV} = \arg \max L(\theta).$$

Exemple 2.1.1 (*La copule Gaussienne multivariée*) Soit la fonction de distribution de la copule Gaussienne multivariée C_{G_a} définie comme suit :

$$C^{G_a}(u_1, \dots, u_n; R) = \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

où R est sa matrice de corrélation, sa fonction de densité est donnée par :

$$h(x_1, \dots, x_d) = \frac{1}{(2\pi)^{\frac{n}{2}} |R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}x'R^{-1}x\right)$$

$$h(x_1, \dots, x_n) = c(\Phi(x_1), \dots, \Phi(x_n)) \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \right) \quad (2.3)$$

On a donc :

$$c(\Phi(x_1), \dots, \Phi(x_n)) = \frac{\frac{1}{(2\pi)^{\frac{n}{2}} |R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}x'R^{-1}x\right)}{\prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \right)} \quad (2.4)$$

Soit $u_i = \Phi(x_i) \Leftrightarrow x_i = \Phi^{-1}(u_i)$, on peut alors récrire la densité de la copule comme suit :

$$c(u_1, \dots, u_n) = \frac{1}{|R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}S'(R^{-1} - I)S\right)$$

avec $S = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))'$.

Soit l'échantillon $\{x_1^t, \dots, x_n^t\}_{t=1}^T$, la fonction log-vraisemblance est donnée par :

$$L(\theta) = -\frac{T}{2} \ln |R| - \frac{1}{2} \sum_{t=1}^T S'(R^{-1} - I)S_t$$

où θ est l'ensemble de tous les paramètres et $S_t = \Phi(\Phi^{-1}(u_1^t), \dots, \Phi^{-1}(u_n^t))$.

L'estimateur du maximum de vraisemblance de R est donnée par (voir Magnus end Neudecker [20])

$$\hat{R}^{MV} = \frac{1}{T} \sum_{t=1}^T S_t' S_t.$$

2.1.2 Méthode des fonctions d'inférence des marginales

La méthode du maximum de vraisemblance peut engendrer des temps de calcul très longs pour une copule multidimensionnelle car l'estimation des paramètres des lois marginales et des paramètres de la copule se fait d'une façon simultanée.

Joe et Xu (1996) notent cependant que la représentation par la copule permet de séparer les paramètres spécifiques aux distributions marginales $(\beta_1, \dots, \beta_n)$ et les paramètres communs de la structure de dépendance α . Il suggèrent une estimation à deux étapes (voir Shih et Louis (1995)) :

1. Nous estimons les paramètres β_1, \dots, β_n des lois marginales dans un premier temps.

2. Etant donné ces estimations, nous estimons ensuite le paramètre de la copule.

Nous pouvons donc écrire $\theta = (\alpha, \beta_1, \dots, \beta_n)$, par conséquent

$$H(x_1, \dots, x_n; \theta) = C(F_1(x_1; \beta_1), \dots, F_n(x_n; \beta_n); \alpha; \beta_1, \dots, \beta_n),$$

Notons $c(u_1, \dots, u_n; \alpha)$ la contrepartie paramétrique de la densité de la copule. L'approche par maximum de vraisemblance réduit la maximisation à la quantité suivante :

$$L(x; \alpha, \beta_1, \dots, \beta_n) = \sum_{t=1}^T \ln c(F_1(x_1^t; \beta_1), \dots, F_n(x_n^t; \beta_n); \alpha) + \sum_{t=1}^T \ln f_1(x_1^t; \beta_1) + \dots + \sum_{t=1}^T \ln f_n(x_n^t; \beta_n).$$

Ainsi, au lieu de chercher le maximum global

$$(\hat{\alpha}^{MV}, \hat{\beta}_1^{MV}, \dots, \hat{\beta}_n^{MV}) = \arg \max_{\alpha, \beta_1, \dots, \beta_n} L(x; \alpha, \beta_1, \dots, \beta_n)$$

on peut effectuer en deux étapes la procédure d'estimation des paramètres :

- On estime la distribution marginale univariée

$$\hat{\beta}_i = \arg \max_{\beta_i} \sum_{t=1}^T \ln f_i(x_i^t; \beta_i),$$

- puis on estime α en tenant compte des estimateurs obtenus ci-dessus

$$\hat{\alpha} = \arg \max_{\alpha} \sum_{t=1}^T \ln c(F_1(x_1^t; \hat{\beta}_1), \dots, F_n(x_n^t; \hat{\beta}_n); \alpha)$$

L'estimateur IFM des paramètres du modèle est alors défini de la façon suivante :

$$\hat{\theta}^{IFM} = (\hat{\beta}_1, \dots, \hat{\beta}_n; \hat{\alpha})'$$

2.1.3 Méthode de pseudo-maximum de vraisemblance

Cette méthode a été proposée dans le cas où les marges F_1, \dots, F_d associées aux X_1, \dots, X_d sont inconnus, elle contient deux étapes :

- On remplace les marges F_1, \dots, F_d par leurs estimations naturelles (estimateur empirique), elles sont définies par :

$$\hat{F}_{j,n}(x_{ij}) = \frac{1}{n} \sum_{i=1}^n 1(X_{ij} \leq x). \quad (2.5)$$

– En maximisant la pseudo log-vraisemblance pour estimer θ , telle que

$$L(\theta) = \sum_{i=1}^n \ln c_{\theta} \left\{ \hat{F}_{1,n}(x_{i1}), \hat{F}_{2,n}(x_{i2}), \dots, \hat{F}_{d,n}(x_{id}) \right\} \quad (2.6)$$

alors l'estimateur $\hat{\theta}_n^{PMV}$ est

$$\hat{\theta}_n^{PMV} = \arg \max L(\theta) \quad (2.7)$$

Remarque 2.1.1 Nous avons pour tout $i \in \{1, 2, \dots, n\}$ et $j \in \{1, 2, \dots, d\}$ $R_{ij} = n\hat{F}_{j,n}(x_{ij})$, où R_{ij} est le rang de X_{ij} dans l'échantillon univariés X_{1j}, \dots, X_{nj} , donc l'estimateur tiré de cette méthode est basé sur les rangs des observations.

Cas bivarié

Soit X, Y deux variables aléatoires dont les marges sont F et G respectivement. Soit C la copule associée à X, Y de paramètre θ . La pseudo log-vraisemblance est définie par :

$$L(\theta) = \sum_{i=1}^n \ln c_{\theta} \{F_n(x_i), G_n(y_i)\}, \quad (2.8)$$

telle que

$$F_n(x_i) = \frac{1}{n} \sum_{i=1}^n 1(X_i \leq x) \text{ et } G_n(y_i) = \frac{1}{n} \sum_{i=1}^n 1(Y_i \leq y). \quad (2.9)$$

comme $nF_n = R_i$ et $nG_n = S_i$, où R_i et S_i sont les rangs de X_i et Y_i dans leurs échantillons univariés respectivement, donc $L(\theta)$ devient

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

2.1.4 Méthode de moment

Cette méthode est proposé par Brahimi et al. (2012) [27] pour estimer plus d'un paramètre d'une copule C_{θ} . Cette méthode est similaire à la méthode de moments dans le cas univarié. Elle consiste à calculer les moments classique d'ordre k de la copule C_{θ} . D'où nous obtenons ces formules en fonction de paramètres de dépendance. On résume dans ce qui suit l'algorithme correspond à cette méthode

où nous supposons que nous avons une copule paramétrique C_θ de paramètre $\theta = (\theta_1, \dots, \theta_r)$:

1. Calculer les moments d'ordre k de la variable $C(U)$ par :

$$M_k(C) = E [(C(U))^k] = \int_{[0,1]^d} (C(u))^k dC(u), \quad k = 1, 2, \dots \quad (2.10)$$

pour $k = 1$, M_k correspond à l'espérance de copule donné comme suit :

$$M_1(C) = \mathbf{E}[(C(U))] = \frac{(2^{d-1} - 1)\tau + 1}{2d}$$

et car la copule C est paramétrique, donc $M_k(C)$ devient $M_k(\theta)$ et (1.10) Réécrire :

$$M_k(\theta) = \int_{I^2} (C_\theta(u))^k dC_\theta(u), \quad k = 1, 2, \dots \quad (2.11)$$

2. Calculer la version empirique de Moments \hat{M}_k selon les étapes suivants :

(a) Calculer le vecteur \hat{U}_{ji} où

$$\hat{U}_{ji} = F_{jn}(X_{ji}), \quad j = 1, \dots, d$$

telle que $F_{jn}(x_j) = n^{-1} \sum_{i=1}^n \mathbf{1}(X_{ji} \leq x_j)$ et construire le vecteur $\hat{U}_i = (\hat{U}_{1i}, \dots, \hat{U}_{di})$, pour $i = 1, \dots, n$.

(b) Calculer les \hat{M}_k :

$$\hat{M}_k = n^{-1} \sum_{i=1}^n \left(C_n(\hat{U}_i) \right)^k$$

telle que : $C_n(u) = F_n(F_{1n}^{-1}(u_1), \dots, F_{1n}^{-1}(u_d))$ pour $u \in \mathbf{I}^d$ et $F_{jn}^{-1}(u_j)$ est la fonction de quantil empirique correspond de la fonction F_{jn} .

3. Résoudre le système d'équation suivant :

$$\begin{cases} M_1(\theta_1, \dots, \theta_r) = M_1 \\ M_2(\theta_1, \dots, \theta_r) = M_2 \\ \vdots \\ M_r(\theta_1, \dots, \theta_r) = M_r \end{cases} \quad (2.12)$$

La solution de ce système nous donne l'estimateur de θ , noté $\hat{\theta}^{CM} = (\hat{\theta}_1, \dots, \hat{\theta}_r)$.

Exemple 2.1.2 Comme un exemple illustratif de cette méthode, la copule choisie dans [9] est la copule de Gumbel pour deux paramètres qui est définie par :

$$C_{\alpha, \beta(u)} = \left(\left(\sum_{j=1}^d (u_j^{-\alpha} - 1)^\beta \right)^{1/\beta} + 1 \right)^{-1/\alpha} \quad \text{pour } u = (u_1, \dots, u_d) \in \mathbf{I}^d, \quad (2.13)$$

avec un generateur :

$$\varphi_{\alpha, \beta}(t) = (\theta^{-\alpha} - 1)^\beta.$$

En substituant (1.13) dans (1.11), alors les moments M_k sont donnés par :

$$M_k(\alpha, \beta) = \frac{(k+1)\beta + \alpha\beta - k}{(k+1)^2\beta + (k+1)\alpha\beta}$$

En particulier, les premiers moments sont les suivants :

$$M_1(\alpha, \beta) = \frac{2\beta + \alpha\beta - 1}{4\beta + 2\alpha\beta}$$

$$M_2(\alpha, \beta) = \frac{3\beta + \alpha\beta - 2}{9\beta + 3\alpha\beta}$$

En résolvant le système (1.12), donc les estimateurs $\hat{\alpha}, \hat{\beta}$ de et β deviennent :

$$\hat{\alpha} = \frac{8\hat{M}_1 - 9\hat{M}_2 - 1}{1 - 4\hat{M}_1 + 3\hat{M}_2},$$

$$\hat{\beta} = \frac{1 - 4\hat{M}_1 + 3\hat{M}_2}{(1 - 2\hat{M}_1)(1 - 3\hat{M}_2)}$$

Chapitre 3

Modélisation par les copules sur des données réelles

Ce chapitre explore le potentiel d'applications de la théorie des copules à l'assurance. En particulier, nous allons mesurer la dépendance entre la branche automobile à savoir la responsabilité civile matérielle et la responsabilité civile corporelle.

L'étude a été menée sur un échantillon de l'agence d'assurance ouargla (saa). Les données portent sur les montants de règlements sinistres pour la responsabilité civile matérielle et la responsabilité civile corporelle. Elles sont présentées dans des triangles de règlements par année de survenance et par année de développement sur la période allant de 2000 à 2024.

3.1 Analyse descriptive des données

Le tableau 1 présente les caractéristiques statistiques des variables la responsabilité civile matérielle et la responsabilité civile corporelle. D'après le tableau 1, la kurtosis est positif pour les deux variables corporel et matereil cela indique qu'on observe dans les données plus de valeurs extrêmes.

	Mediane	Moyenne	Ecart type	Kurtosis
CORPOREL	2856149	3495806	3280929	1.727639
MATEREIL	10350307	7924808	6136148	1.542289

TAB. 3.1 – Caractéristiques statistiques des variables la responsabilité civile matérielle et la responsabilité civile corporelle.

3.1.1 Coefficient de corrélation

La relation entre les variables la responsabilité civile matérielle et la responsabilité civile corporelle est très bonne (tableau 2), elle reflète une forte corrélation entre ces deux variables. Le coefficient de dépendance de rang de Spearman est 0.710. Le coefficient de corrélation linéaire Pearson est 0.709. Le coefficient de corrélation de rang de Kendall est 0.536. Alors L'analyse multivariée était donc préférable à l'analyse univariée. Ce qui suggère l'importance d'utiliser des copules.

Coefficient de corrélation	Valeur
Spearman	0.710
Pearson	0.709
Kendall	0.536

TAB. 3.2 – Corrélation entre les variables la responsabilité civile matérielle et la responsabilité civile corporelle.

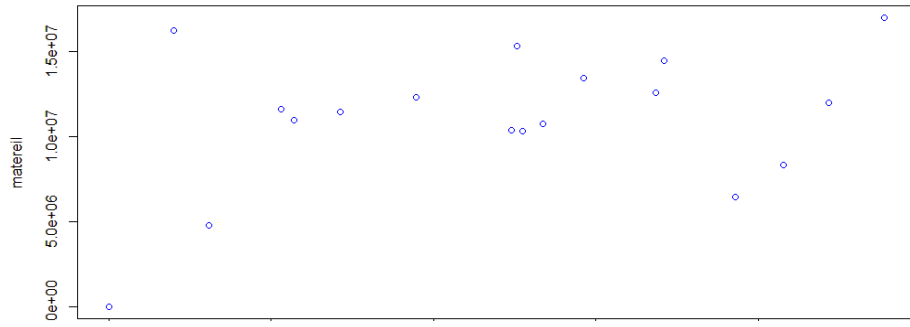


FIG. 3.1 – Nuage de points de la variable corporel en fonction de la variable materiel.

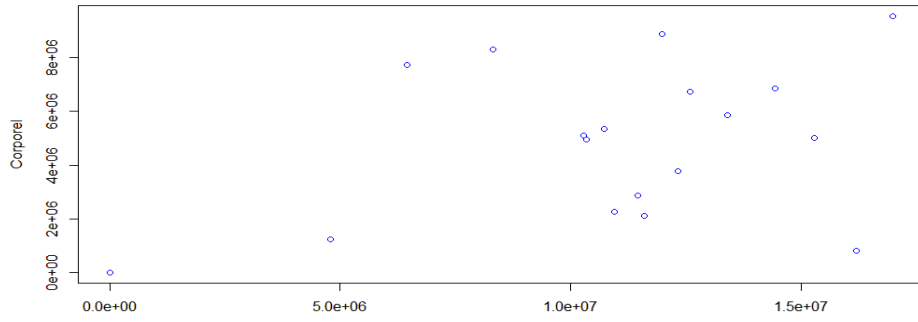


FIG. 3.2 – Nuage de points de la variable materiel en fonction de la variable corporel.

Les figures 1 et 2, montrent que les variables matériels et corporels sont fortement corrélés, ce qui affirme l'existence de la relation de dépendance.

3.2 Fonctions de distribution univariées

L'erreur quadratique moyenne (RMSE) est utilisée pour sélectionner la meilleure fonction de distribution, les distributions Gamma et Exponentielle est choisie car les RMSE sont les plus basses.

Variables	Distributions	paramètres	RMSE
CORPOREL	Gamma	2.19	1.4
MATEREIL	Exponentielle	3.35	1.6

TAB. 3.3 – Les distributions marginales.

3.3 Fonctions copules

Les valeurs des paramètres estimés de chacune des fonctions copules à l'aide de la méthode MLE et du logarithme MLE sont données dans le tableau 4. D'après les valeurs de logMLE, on peut conclure que la copule de Frank serait la meilleure pour modéliser les distributions conjointes de corporel et materiel.

Copules	paramètre	logMLE
Frank	18.5	30.02
Clayton	5.6	27.8
Gumbel	4.9	28.4
Student t	1.2, 3	25
Normal	1.1	26.05

TAB. 3.4 – Les fonctions copules et logarithme MLE.

la distribution conjointe est

$$C_{\theta}(u, v) = -\frac{1}{18.5} \ln \left\{ 1 + \frac{(e^{-18.5u} - 1)(e^{-18.5v} - 1)}{e^{-18.5} - 1} \right\}.$$

Conclusion

La copule est un concept permettant de modéliser la structure de dépendance entre des variables. Contrairement aux différentes mesures classiques de dépendance, la copule permet de mieux décrire la relation entre les variables. Autrement dit, au lieu de résumer cette structure par un seul indicateur, comme le coefficient de corrélation, on utilise la fonction copule qui offre une description plus complète de la dépendance.

Dans notre étude, nous avons tenté de mettre en évidence la dépendance entre deux sous-branches de l'assurance automobile : les dommages matériels et les dommages corporels, à travers l'application de la théorie des copules. Nous avons constaté que cette dépendance est significative, et que ne pas en tenir compte pourrait entraîner une sous-estimation importante des provisions techniques, ce qui pourrait impacter la solvabilité de la compagnie.

La théorie des copules permet ainsi une meilleure maîtrise du risque lié à l'évaluation des provisions techniques. Elle contribue également à une gestion plus rigoureuse de la solvabilité de l'assureur, en prenant en compte la dépendance à différents niveaux, notamment la dépendance de queue, ce qui constitue à la fois sa spécificité et son intérêt majeur.

Bibliographie

- [1] Abderrahmane, N., Zeghmar, R., (2016). Modélisation De La Dépendance Par Les Copules. Mémoire de master (Université Boumerdes).
- [2] Benelmir. I., (2018). Modélisation de la Dépendance par les Copules, Thèse de doctorat (université BISKRA).
- [3] Brahimi, B., Necir, A., (2012). Semiparametric estimation of copula models based on the method of moments, *Statistical Methodology*, 9, 467-477.
- [4] Breyman, W., Dias, A., and Embrechts, P., (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative finance*, 3, 1-14.
- [5] Cherubini, U., Luciano, E., and Vecchiato, W., (2004). Copula methods in finance. Wiley Finance Series.
- [6] Chine, A., (2011). About copula statistics. Thèse de doctorat (université BISKRA).
- [7] D'Agostino, R.B., Stephens, M.A, (1986). Goodness-of-Fit Techniques, New York : Marcel Dekker Inc.
- [8] Embrechts, P., Alexander, j., Neil, Mc., and Frey, F., (2005). Quantitative risk management : concepts, techniques and tools. Princeton university press.
- [9] Genest, C., Favre ; A.C., (2007). Everything you always wanted to know about copula modelling but were afraid to ask, *J, Hydrol, Eng*, 12 (4), 347-368.
- [10] Gregor, N. F., (2009). Copula parameter estimation by Maximum-Likelihood and Minimum-Distance estimators , A simulation study.
- [11] Guessouri, A., (2019). Sur les copules et applications, Mémoire de master (université BISKRA).
- [12] Joe, H., (1997). Multivariate models and dependence concepts, Chapman & Hall, London, 73.
- [13] Joe, H., et Xu, J., (1996). The estimation method of inference functions for margins for multivariate models, Technical Report, 166, Department of Statistics, University of British Columbia.
- [14] Klemelä, J., (2009). Smoothing of Multivariate Data : Density Estimation and Visualization, Willey finance.

- [15] Lounas, F., (2011). Modélisation de la dépendance par les copules et applications. Mémoire de magistère (université Tizi-Ouzou).
- [16] Magnus, J.R., et Neudecker. H. Matrix Differential Calculus with Applications in Statistics and Econometrics. John Wiley et Sons, New York (1988).
- [17] Nelsen, R.B, (2006). An introduction to copulas, second edition, Springer, New York.
- [18] Rüschendorf, L., (2009). On the distributional transform, Sklar's theorem, and the empirical copula process. *J. Statist. Plann. Inference*, 139(11), 3921-3927.
- [19] Schweizer, B., Wolf, E., (1981). On non-parametric measures of dependence for random variables, *Ann, Statist*, 9, 879–885.
- [20] Shih, J.H., & Louis, T.A., (1985). Inferences on the association parameter in copula models for bivariate survival data, *Biometrics*, 51, 1384-1399.
- [21] Spearman, C., (1904). The proof and measurement of association between two things, *Am J Psychol*, 15, 72–101.
- [22] Vose, D., (2008). Risk analysis : a quantitative guide, John Wiley and sons, England.
- [23] Wolff, E.F, (1977). Measures of Dependence Derived from Copulas, Ph.D, thesis. University of Massachusetts, Amherst.

Résumé

Domages corporel et materiel sont deux principales variables utilisé dans l'analyse de l'assurance automobile. L'objet de cet étude est la mise en évidence de la structure de dépendance de deux variables d'assurance de Saa ouargla (2000- 2024). Après avoir réalisé l'analyse bivariée, La copule de Frank a été sélectionnée comme étant la meilleure fonction de distribution bivariée.

Mots clés: Modélisation , la copule , la dépendance.

Abstract

Corporal et material damages are two main variables used in car insurance analysis. The purpose of this study underlines the dependence structure of two insurance variables of Saa Ouargla (2000-2024). After conducting bivariate analysis, Frank copula was selected to be the best bivariate distribution function.

Keywords: Modeling, copula, dependence.

ملخص

الأضرار الجسدية المادية متغيران رئيسيان مستخدمان في تحليل تأمين السيارات. تهدف هذه الدراسة إلى تسليط الضوء على بنية الارتباط لمتغيرين تأمينيين في منطقة Saa ورقلة (2000-2024). بعد إجراء التحليل ثنائي المتغيرات، اختيرت كوبيلا فرانك كأفضل دالة توزيع ثنائية المتغيرات.

الكلمات المفتاحية: النمذجة، كوبيلا، الارتباط.

Résumé

Dommages corporel et materiel sont deux principales variables utilisé dans l'analyse de l'assurance automobile. L'objet de cet étude est la mise en évidence de la structure de dépendance de deux variables d'assurance de Saa ouargla (2000- 2024). Après avoir réalisé l'analyse bivariée, La copule de Frank a été sélectionnée comme étant la meilleure fonction de distribution bivariée.

Mots clés: Modélisation , la copule , la dépendance.

Abstract

Corporal et material damages are two main variables used in car insurance analysis. The purpose of this study underlines the dependence structure of two insurance variables of Saa Ouargla (2000-2024). After conducting bivariate analysis, Frank copula was selected to be the best bivariate distribution function.

Keywords: Modeling, copula, dependence.

ملخص

الأضرار الجسدية المادية متغيران رئيسيان مستخدمان في تحليل تأمين السيارات. تهدف هذه الدراسة إلى تسليط الضوء على بنية الارتباط لمتغيرين تأمينيين في منطقة Saa ورقلة (2000-2024). بعد إجراء التحليل ثنائي المتغيرات، اختيرت كوبيلا فرانك كأفضل دالة توزيع ثنائية المتغيرات.

الكلمات المفتاحية: النمذجة، كوبيلا، الارتباط.